Journal of Medical Internet Research

Impact Factor (2018): 4.945 - ranked #1 medical informatics journal by Impact Factor
Volume 23 (2021), Issue 10 ISSN: 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Original Papers

Viewpoints

Tutorial

Corrigenda and Addendas

Review

Comparative Effectiveness of Enhanced Patient Instructions for Bowel Preparation Before Colonoscopy: Network Meta-analysis of 23 Randomized Controlled Trials

Xu Tian[1*], BSc, MSN, PhD; Li-Juan Yi[2*], BSc, MSN; Yang Han[3*], BSc; Hui Chen[3], BSc; Xiao-Ling Liu[3], BSc; Wei-Qing Chen[3*], MD; Maria F Jiménez-Herrera[1], PhD

[1]Nursing Department, Universitat Rovira i Virgili, Tarragona, Spain

[2]Department of Nursing, Hunan Traditional Chinese Medical College, Zhuzhou, China

[3]Chongqing University Cancer Hospital, Chongqing, China

[*]these authors contributed equally

Corresponding Author:
Maria F Jiménez-Herrera, PhD
Nursing Department
Universitat Rovira i Virgili
Avinguda Catalunya, 35
Tarragona, 43002
Spain
Phone: 34 977299426
Email: maria.jimenez@urv.cat

Abstract

Background: Various enhanced patient instructions (EPIs) have been used for bowel preparation (BP) and our previous meta-analysis also demonstrated the efficacy of EPIs in increasing the colonic polyp and adenoma detection rates; however, the optimal method for adequate BP has not yet been developed.

Objective: We performed a network meta-analysis to determine the optimal instructions.

Methods: We searched for randomized controlled trials (RCTs) comparing the effectiveness of EPIs with each other or standard patient instructions (SPIs) for BP. We performed direct and Bayesian network meta-analyses for all instructions and used the GRADE (Grading of Recommendations Assessment, Development and Evaluation) criteria to appraise the quality of evidence.

Results: We included 23 RCTs (7969 patients) comparing 10 different instructions. In direct meta-analyses, most of the EPIs, except visual aids and mobile apps, increased the adequate preparation rate (APR). Network meta-analyses showed that additional explanations were superior to visual aids (odds ratio [OR] 0.35, 95% CI 0.19-0.59), telephone calls (OR 0.62, 95% CI 0.37-0.99), educational videos (OR 0.79, 95% CI 0.5-0.77), and mobile apps (OR 0.33, 95% CI 0.14-0.68) with low-to-high-quality evidence; newly designed booklets (OR 3.28, 95% CI 1.59-6.16), SMS text messaging (OR 2.33, 95% CI 1.28-3.91), telephone calls (OR 1.86, 95% CI 1.03-1.78), educational videos (OR 2.33, 95% CI 1.40-3.65), and social media applications (OR 2.42, 95% CI 1.4-3.93) were superior to visual aids and mobile apps with low-to-high-quality evidence. SMS text messaging, telephone calls, and social media applications increase adherence to and satisfaction with the BP regime. Social media applications reduce the risk of adverse events (AEs). Telephone calls and social media applications increase the polyp detection rate (PDR).

Conclusions: Newly designed booklets, telephone calls, educational videos, and social media applications can improve the quality of BP. Telephone calls and social media applications improve adherence to and satisfaction with the BP regime, reduce the risk of AEs, and increase the PDR.

Trial Registration: INPLASY (International Platform of Registered Systematic Review and Meta-analysis Protocols) INPLASY2020120103; https://inplasy.com/inplasy-2020-12-0103/

(J Med Internet Res 2021;23(10):e19915) doi:10.2196/19915

KEYWORDS

colonoscopy; bowel preparation; patient instruction; systematic review; network meta-analysis

Introduction

Colorectal cancer ranks third among the most common cancers and second in terms of the incidence and cause of cancer-related mortality worldwide, with 1.8 million new cases and 0.88 million deaths in 2018 [1]. Colonoscopy has been regarded as the criterion standard approach for early detection and safe removal of colorectal lesions [2-4]. In particular, screening colonoscopy has been associated with decreased colorectal cancer incidence and mortality [2,5]. The quality of bowel preparation (BP) is an important contributor toward successful and safe colonoscopy [6]. However, approximately 18% to 30.5% inadequate BP has been reported in previous studies [7-9]. It is discouraging that inadequate BP is associated with decreased polyp detection rates (PDRs), increased risk of adverse events (AEs), prolonged working time, and increased medical expenditure [10,11].

Previous studies have determined several factors that can influence the quality of BP, such as appropriate dietary restrictions and proper administration of preparation solutions [12]. Of all the factors, adequate comprehension of the BP and colonoscopy details is a critical contributor to adequate BP [13]. Patients usually receive written booklets or verbal instructions from professionals before colonoscopy for details regarding BP and dietary restrictions, which are defined as standard patient instructions (SPIs) [14]. However, the effect of SPIs on improving the quality of BP is not enough [15]. Therefore, researchers and practitioners have been developing most of the enhanced patient instructions (EPIs) by including cartoon pictures, SMS text messaging, telephone calls, mobile apps, and social media applications to improve the quality of BP prior to colonoscopy [14].

Thus far, several traditional pairwise meta-analyses investigating the comparative efficacy between EPIs and SPIs for the quality of BP have been published, and they have demonstrated improved BP [13,14,16-18]. Moreover, our previous meta-analysis also demonstrated the efficacy of EPIs in increasing the PDR and adenomas detection rate (ADR) [19]. However, only 2 3-arm randomized controlled trials (RCTs) investigated the comparative efficacy of telephone calls or WeChat versus SMS text messaging for BP in patients receiving outpatient colonoscopy. It is still unclear which EPIs should preferably be used by decision makers for BP before colonoscopy. To address those issues that could not be addressed by traditional pairwise meta-analysis, network meta-analysis, which can simultaneously assess the comparative efficacy of multiple interventions, has been developed [20,21]. Therefore, we performed direct pairwise and Bayesian network meta-analyses combining direct and indirect evidence comparing the relative efficacy of all patient instructions to determine the optimal educational instructions for BP before colonoscopy. We also used the GRADE (Grading of Recommendations Assessment, Development and Evaluation) criteria to appraise the quality of evidence.

Methods

We conducted this systematic review and network meta-analysis in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [22] and reported all the outcomes according to the International Society for Pharmacoeconomics and Outcomes Research Task Force on Indirect Treatment Comparisons Good Research Practices [23]. No formal protocol for the present study has been published. We registered our systematic review on the INSPLAY (International Platform of Registered Systematic Review and Meta-analysis Protocols) platform, and the trial registration number is INPLASY2020120103.

Search Strategy

We constructed the search strategy with the assistance of an experienced medical librarian using full-text words and MeSH (Medical Subject Headings). We also refined the search strategy according to the specific requirements of each database. All potential RCTs comparing EPIs to each other or SPIs for BP were captured in PubMed, the Cochrane Central Register of Controlled Trials, and Embase until December 2019. The last search was updated in February 2020. Details of the search strategies used for the 3 targeted databases are presented in Multimedia Appendix 1.

Study Selection

We first excluded duplicate records through running the Finding Duplicate function embedded in EndNote (version X9, Clarivate Analytics). Then, we checked the titles and abstracts of the articles to exclude irrelevant articles. Next, we screened the full texts to further check the eligibility of all the remaining studies. The inclusion criteria were as follows: (1) patients: adult patients who were assigned to receive selective outpatient colonoscopy; (2) interventions: all EPIs or SPIs for BP; (3) outcomes: the quality of BP assessed with the adequate preparation rate (APR), adherence to instruction (AI), satisfaction with the BP solution, willingness to repeat the same BP solution, PDR, and AEs including abdominal discomfort, nausea and vomiting, and sleep disturbance; (4) study design: RCTs. Language restrictions were not imposed. The exclusion criteria included (1) animal studies and (2) conference abstracts without sufficient data or unpublished studies.

The eligibility was checked by 2 investigators (XT and HC) independently, and any divergences were resolved through the consensus principle. When no agreement could be reached, a third investigator (WQC) was consulted for determining the eligibility.

Data Extraction

Essential data including the leading author, year of publication, study design, country where the study was conducted, age and sex of the patients, details of the BP regime, details of the instructions used, and outcomes of interest were extracted by 2 independent investigators (LJY and XT) using a data extraction sheet designed in advance. EPIs were classified as additional explanations, visual aids, new visual aids, newly designed booklets, SMS text messages, telephone calls, mobile apps, social media applications, and educational videos. The

classification and comparison of EPIs are documented in Multimedia Appendix 2.

The primary outcome of the present meta-analysis was the comparative efficacy of EPIs for improving the quality of BP before colonoscopy, which was assessed with respect to the APR. The secondary outcome was the comparative efficacy of EPIs with respect to the AI, satisfaction with the BP solution, willingness to repeat the same BP solution, PDR, and AEs.

Quality Assessment

We assigned 2 investigators to independently assess the risk of bias of each eligible study with the Cochrane risk of bias assessment tool [24]. We labeled each study as having low, unclear, or high risk of bias according to the match between the actual information and the following assessment criteria: random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other biases. A third investigator (WQC) was consulted to solve any discrepancies.

Statistical Analysis

In traditional pairwise meta-analysis, we calculated the pooled odds ratio (OR) with 95% CI to express the dichotomous data [25]. We performed Cochran Q tests to qualitatively assess the heterogeneity and used the I^2 statistic to quantitatively estimate the level of heterogeneity [26]. All pairwise meta-analyses were performed based on the random-effect model because this model simultaneously incorporates within- and between-study heterogeneities. Publication bias is assessed using a funnel plot if the accumulated number of eligible studies for individual outcomes was more than 10 [27], and an asymmetry suggests the presence of publication bias [28]. Direct meta-analysis was conducted using Review Manager 5.3 (Cochrane Collaboration).

After completing direct meta-analysis, we conducted random-effect network meta-analyses to estimate all the relative effects using Markov chain Monte Carlo methods in OpenBUGS 3.2.3 (MRC Biostatistics Unit) following the methods described by Lu and Ades [29,30]. We used the initial value that was automatically generated from the software to fit the model [31]. To achieve convergence, we performed each Markov chain Monte Carlo method with 50,000 iterations and 20,000 burn-ins. We drew the comparison-adjusted funnel plot to assess the small-study effects when the number of studies included in one comparison pair exceeded 10 [32].

We assessed the probability that each instruction was the most efficacious one for improving quality of BP, the second best, the third best, and so on by calculating the OR for each instruction compared with an arbitrary common control group and counting the number of iterations of the Markov chain in which each instruction had the highest OR, the second highest OR, and so on [33].

Sensitivity Analysis

We designed several sensitivity analyses to evaluate the robustness of the summarized findings according to the following principles: (1) BP assessment scale (excluding studies in which uncommon scales were used except for the Boston Bowel Preparation Scale [BBPS], Ottawa Bowel Preparation Scale [OBPS], and Aronchick Bowel Preparation Scale [ABPS]); (2) risk of bias (excluding studies with high risk); (3) study design (excluding studies with multicenter design).

Quality of Evidence

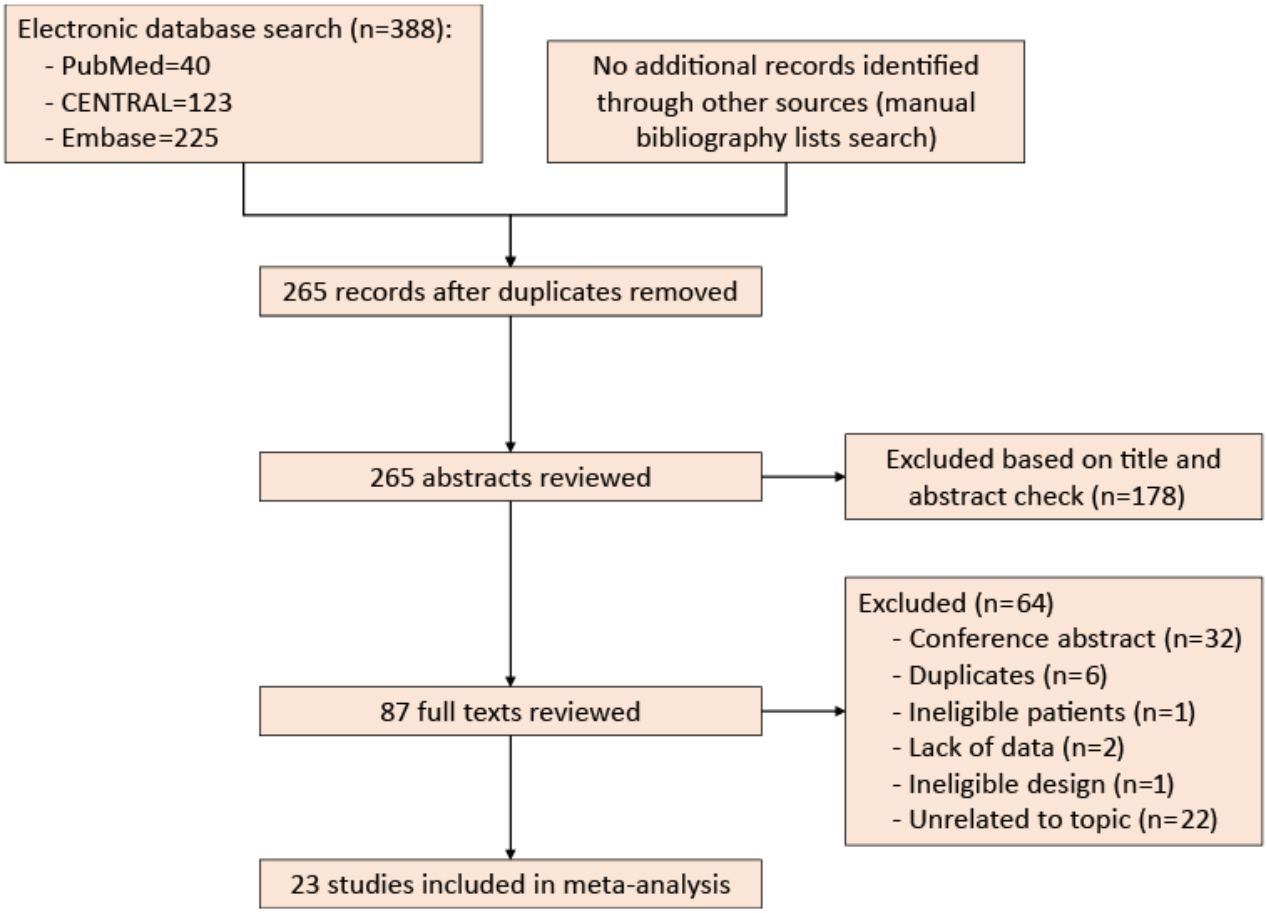
We rated the quality of evidence of the primary outcomes with the GRADE working group approach [34,35]. In this approach, the quality of direct evidence based on RCTs would be first rated as high and the level could be reduced to moderate, low, or very low according to 5 domains, including risk of bias, indirectness, imprecision, inconsistency, and publication bias [35]. The quality of indirect evidence was consistent with the level of the lowest rating of the 2 pairwise estimates that contribute as first-order loops to the indirect estimates and imprecision or intransitivity can further reduce the level [35]. If the assumption of coherence between direct and indirect estimates was confirmed, then the higher one of their levels would be assigned to the results from network meta-analysis [35].

Results

Study Selection

Figure 1 presents the schematic flowchart of the study selection process. From a total of 388 unique studies identified using the search strategy, we included 23 RCTs in this network meta-analysis [36-58]. Among these, 1 RCT involving a newly designed educational booklet was excluded (because of inpatient enrollment) [59]. Further, 1 RCT comparing modified BP protocols (multimedia education) to standard BP protocols was excluded owing to the design implementation using historical control data [60]. Moreover, 2 RCTs comparing educational videos or additional explanations with SPIs were excluded owing to the lack of essential data [61,62].

Figure 1. Flow diagram showing retrieval and selection of literature. CENTRAL: Cochrane Central Register of Controlled Trials.



Study Characteristics

The basic characteristics of the patients who were included in all eligible studies are summarized in Table 1. We have also documented the characteristics of the eligible studies in Multimedia Appendix 3. The studies were published between 2009 and 2019 [36-56,58]. The sample sizes of the individual studies ranged from 92 to 1018 with a total of 7969 participants. Among the 23 studies included, 2 were designed as 3-arm trials [41,55]; among these, Lee et al [41] compared the efficacy between SMS text messaging, telephone calls, and SPIs, and Wang et al compared the efficacy between SMS text messaging, social media applications and SPIs [55]. All other studies had 2-arm designs [36-40,42-54,56,58]. Further, 2 studies had multicenter designs [49,54]. All the studies reported the quality of BP assessed using the BBPS (9 trials) [36,41,45,50,53-57], OBPS (8 trials) [39,40,42,43,47-49,52], ABPS (4 trials) [37,38,51,58], Harefield Cleansing Scale (1 trial) [44], and Universal Preparation Assessment Scale (1 trial) [46].

Table 1. Basic characteristics of the patients included in studies on educational instructions for bowel preparation before colonoscopy.

Study	Country	Comparison	Sample size (male vs female participants)	Age (years, male vs female participants), mean (SD)	Male vs female participants (%)	Outcomes
Back et al 2018 [36]	Korea	Educational videos vs SPIs[a] (verbal instructions and instructional leaflets)	283 (139 vs 144)	55.4 (12.8) vs 57.6 (13.1)	53.4 vs 56.2	APR[b], AI[c], and SE[d]
Calderwood et al 2011 [57]	United States	Visual aids vs SPIs (written information)	969 (477 vs 492)	57.3 (8.0) vs 57.1 (7.3)	41.5 vs 41.7	APR, CIT[e], WT[f], PDR[g], and AEs[h]
Cho et al 2015 [58]	Korea	Educational videos vs SPIs (verbal education)	101 (51 vs 50)	n.r.[i]	52.9 vs 52	APR
Elvas et al 2017 [37]	Portugal	Additional explanations vs SPIs (oral and written information)	229 (116 vs 113)	60.0 (13.0) vs 59.0 (11.0)	50.9 vs 58.4	APR and WRBP[j]
Garg et al 2016 [38]	United States	Educational videos vs SPIs (verbal education)	94 (48 vs 46)	59.3 (18.1) vs 57.3 (19.4)	43.8 vs 45.7	APR, PDR, ADR[k], CIT, and WT
Jeon et al 2018 [36]	Korea	Educational videos vs SPIs (written information)	281 (140 vs 141)	46.7 (9.9) vs 49.9 (9.6)	57.1 vs 57.4	APR, AI, PDR, ADR, CIT, WT AEs, and SDT[l]
Kang et al 2016 [40]	China	Social media applications vs SPIs (verbal and written instructions)	770 (387 vs 383)	45.5 (13.0) vs 44.4 (13.2)	52.2 vs 49.9	APR, AI, WRBP, ADR, CIT, WT, ICIBP[m], AEs, and SDT
Lee et al 2015 [41]	South Korea	SMS text messaging vs telephone calls vs SPIs (written information)	390 (127 vs 126 vs 137)	45.7 (12.4) vs 46.0 (12.2) vs 47.1 (11.8)	59.8 vs 62.7 vs 53.3	APR, CIR[n], AI, SE, WRBP, PDR, ADR, CIT, WT, ICIBP, AEs, and SDT
Liu et al 2018 [42]	China	Educational videos vs SPIs (written information)	281 (239 vs 237)	55.1 (6.3) vs 54.4 (8.6)	61.5 vs 64.9	APR, PDR, CIT, WT, and ICIBP
Liu et al 2014 [43]	China	Telephone calls vs SPIs (verbal and written instructions)	605 (300 vs 305)	44.8 (12.5) vs 45.7 (12.6)	53.3 vs 48.2	APR, AI, WRBP, PDR, CIR, CIT, WT, AEs, and SDT
Lorenzo et al 2015 [44]	Spain	Mobile app vs SPIs (written information)	260 (108 vs 152)	48.3 (13.5) vs 52.5 (14.0)	44.4 vs 40.1	APR, AI, SE, and WRBP
Meng 2015 [45]	China	Additional explanations vs SPIs (verbal education)	618 (318 vs 300)	59.0 (15.7)	n.r.	APR and AI
Modi et al 2009 [46]	United States	Additional explanations vs SPIs (written and verbal instructions)	164 (84 vs 80)	57.9 (9.1) vs 57.3 (9.1)	45.2 vs 33.8	APR, CIT, and WT
Park et al 2016 [47]	South Korea	Educational videos vs SPIs (oral and written information)	502 (250 vs 252)	49.2 (8.6) vs 47.3 (9.2)	62.8 vs 66.3	APR, PDR, CIT, and WT
Pillai et al 2018 [48]	United States	Educational videos vs SPIs (verbal and written instructions)	104 (56 vs 48)	n.r.	44.6 vs 50	APR

Study	Country	Comparison	Sample size (male vs female participants)	Age (years, male vs female participants), mean (SD)	Male vs female participants (%)	Outcomes
Prakash et al 2013 [49]	United States	Educational videos vs SPIs (written information)	133 (67 vs 66)	n.r.	52.2 vs 36.4	APR and SE
Sharara et al 2017 [51]	Lebanon	Mobile app vs SPIs (written information)	160 (80 vs 80)	52.0 (13.0) vs 55.0 (12.0)	60 vs 46.3	APR and AI
Spiegel et al 2011 [52]	United States	Newly designed booklet vs SPIs (written information)	436 (216 vs 220)	60.0 (10.7) vs 60.0 (12.3)	95.8 vs 97.7	APR
Tae et al 2012 [53]	Korea	New visual aids vs SPIs (oral and written information)	205 (102 vs 103)	48.6 (8.8) vs 47.6 (9.2)	71.6 vs 68.9	APR, PDR, CIT, CST, and WT
Walter et al 2019 [54]	Germany	SMS text messaging vs SPIs (oral and written information)	495 (248 vs 247)	47.5 (13.6) vs 47.2 (14.8)	50.8 vs 46.9	APR, PDR, ADR, CST, and WT
Wang et al 2019 [55]	China	SMS text messaging vs social media applications vs SPIs (oral and written information)	384 (129 vs 128 vs 127)	52.6 (12.7) vs 48.9 (13.0) vs 51.5 (12.1)	54.3 vs 61.7 vs 53.5	APR, AI, SE, WRBP, PDR, ADR, CIR, CIT, WT, AEs, and SDT
Zhang et al 2018 [56]	China	Social media applications vs SPIs (oral and written information)	1018 (511 vs 507)	51.2 (18.5) vs 50.7 (17.9)	52.6 vs 50.1	APR, AI, WRBP, ADR, CIR, CIT, WT, AEs, and SDT
Rice et al 2016 [50]	United States	Educational videos vs SPIs (oral and written information)	92 (42 vs 50)	60.1 (8.8) vs 61.0 (7.9)	61.9 vs 50.4	APR

[a]SPIs: standard patient instructions.
[b]APR: adequate preparation rate.
[c]AI: adherence to instruction.
[d]SE: satisfaction with the education.
[e]CIT: cecal intubation time.
[f]WT: withdrawal time.
[g]PDR: polyp detection rate.
[h]AEs: adverse events.
[i]n.r.: not reported.
[j]WRBP: willingness to repeat the same BP solution.
[k]ADR: adenomas detection rate.
[l]SDT: sleep disturbance.
[m]ICIBP: incomplete colonoscopy due to inadequate bowel preparation.
[n]CIR: cecal intubation rate.

Quality Assessment

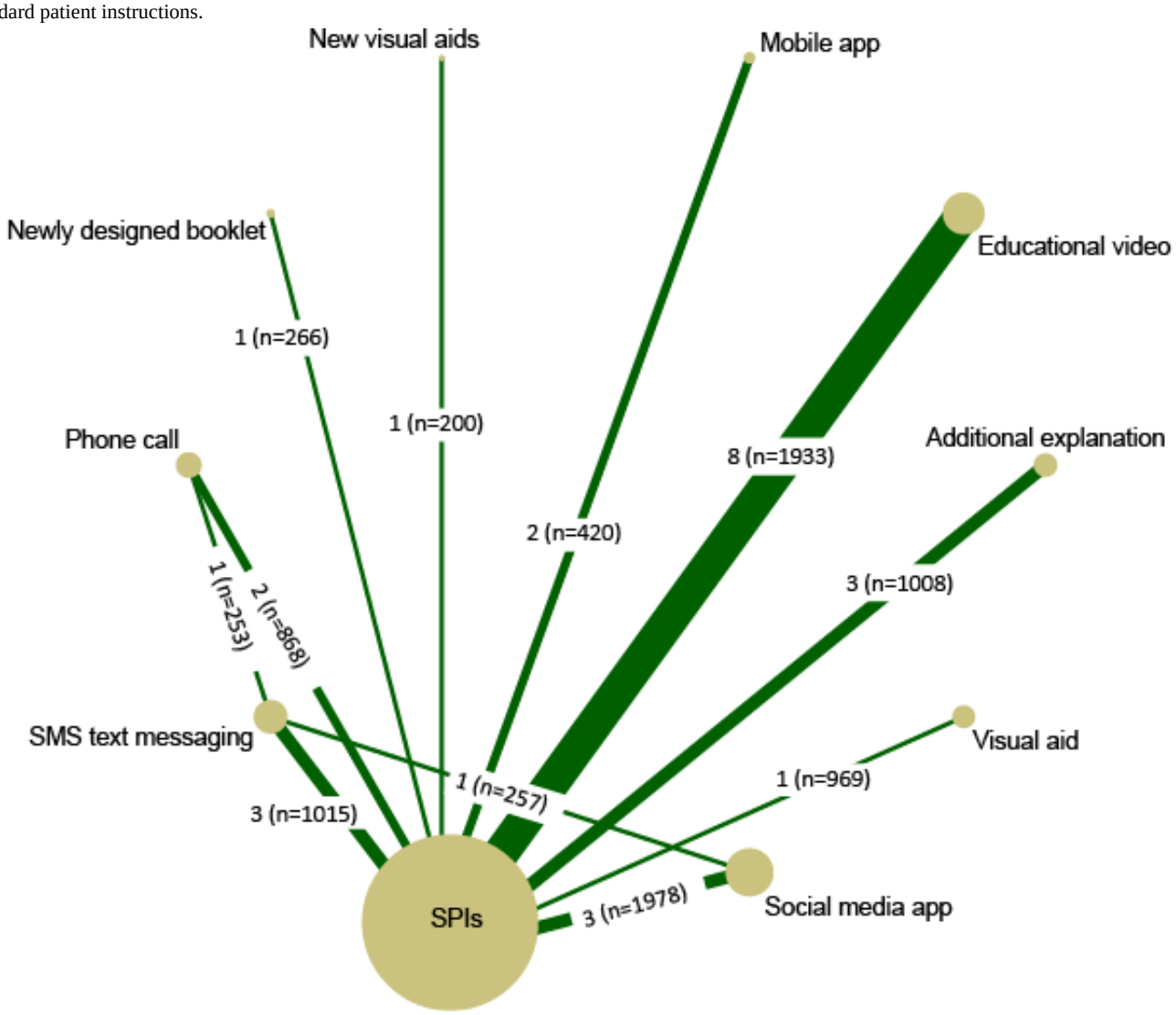
The overall and study-level quality assessments are outlined in Figure S1 of Multimedia Appendix 4. Overall, most of the studies (18/23, 78.3%) appeared to have been at low-to-moderate risk of bias, with 4 (17.4%) not reporting the details of generating the random sequence [45,47,49,56] and 8 (34.8%) not reporting the details for allocation concealment [36,38,45-48,53,57]. Further, 10 (43.5%) did not report the details of blinding the outcome assessor [39-41,43-45,47,51,53,54]. More importantly, 4 studies (17.4%) did not blind the assessors and were rated to be at high risk of bias [38,46,48,49]. In addition, 4 (17.4%) were rated to be at high risk of attrition bias [38,46,49,53]. Other bias sources were not detected in all studies.

Direct Treatment Effects

Primary Outcomes

Figure 2 demonstrates the available direct comparisons and network meta-analyses. Compared to SPIs, additional explanations (3 RCTs; OR 3.34, 95% CI 1.45-7.69), newly designed booklets (1 RCT; OR 3.63, 95% CI 2.15-6.12), new visual aids (1 RCT; OR 3.05, 95% CI 1.21-7.68), SMS text messaging (3 RCTs; OR 2.77, 95% CI 1.86-4.14), telephone calls (2 RCTs; OR 2.64, 95% CI 1.03-6.74), educational videos (8 RCTs; OR 2.82, 95% CI 1.83-4.35), and social media applications (3 RCTs; OR 2.70, 95% CI 1.75-4.18) adequately increased the BP rate, but not visual aids (1 RCT; OR 1.18, 95% CI 0.78-1.8) and mobile apps (2 RCTs; OR 1.85, 95% CI 0.15-23.24). Head-to-head meta-analysis showed that telephone calls (1 RCT; OR 1.25, 95% CI 0.33-4.77) or social media applications (1 RCT; OR 1.32, 95% CI 0.63-2.78) were not superior to SMS messaging. Significant heterogeneity was observed in trials comparing additional explanations (I^2=79%), telephone calls (I^2=60%), educational videos (I^2=62%), and mobile apps (I^2=68%). All the pooled results are delineated in Figure S2 in Multimedia Appendix 4.

Figure 2. Evidence network for adequate preparation rate. The numbers outside and inside parentheses refer to the number of trials and the total number of participants in these trials, respectively, and the thickness of the connecting lines corresponds to the number of trials between comparators. SPIs: standard patient instructions.



Secondary Outcomes

Compared to SPIs, additional explanations (1 RCT; OR 8.38, 95% CI 3.73-18.84), SMS text messaging (2 RCTs; OR 7.04, 95% CI 2.79-17.81), telephone calls (2 RCTs; OR 4.54, 95% CI 2.92-7.05), and social media applications (3 RCTs; OR 3.88, 95% CI 2.05-7.35) are associated with improved AI, but not educational videos (1 RCT; OR 3.16, 95% CI 0.99-10.06) or mobile apps (2 RCTs; OR 1.26, 95% CI 0.73-2.18). Head-to head meta-analysis showed that telephone calls (1 RCT; OR 1.25, 95% CI 0.33-4.77) and social media applications (1 RCT; OR 0.86, 95% CI 0.3-2.45) were not superior to SMS text messaging. Significant heterogeneity was observed in trials comparing social media applications (I^2=63%) to SPIs. All the pooled results are delineated in Figure S3 in Multimedia Appendix 4.

Compared to SPIs, social media applications (1 RCT; OR 2.56, 95% CI 1.35-4.87) and telephone calls (1 RCT; OR 1.85, 95% CI 1.01-3.41) indicated improved satisfaction with the

instructions but not SMS text messaging (2 RCTs; OR 1.92, 95% CI 0.94-4.05) or educational videos (1 RCT; OR 1.47, 95% CI 0.69-3.12). Head-to-head meta-analysis showed that telephone calls (1 RCT; OR 1.22, 95% CI 0.73-2.06) and social media applications (1 RCT; OR 0.95, 95% CI 0.52-1.72) were not superior to SMS text messaging. All the pooled results are delineated in Figure S4 in Multimedia Appendix 4.

Compared to SPIs, social media applications (3 RCTs; OR 2.16, 95% CI 1.51-3.09) and mobile apps (1 RCT; OR 2.48; 95% CI 1.22-5.04) were associated with improved willingness to repeat the same BP regime, but not additional explanations (1 RCT; OR 2.04, 95% CI 0.37-11.35), SMS text messaging (2 RCTs; OR 1.24, 95% CI 0.57-2.71), or telephone calls (2 RCTs; OR 1.58, 95% CI 0.67-3.69). Head-to-head meta-analysis showed that telephone calls (1 RCT; OR 1.44, 95% CI 0.61-3.37) and social media applications (1 RCT; OR 0.99, 95% CI 0.44-2.23) were not superior to SMS text messaging. Significant heterogeneity was observed in trials comparing telephone calls (I^2=69%) to SPIs. All the pooled results are delineated in Figure S5 in Multimedia Appendix 4.

Compared to SPIs, telephone calls (2 RCTs; OR 1.85, 95% CI 1.36-2.5) were associated with an increased PDR, but not visual aids (1 RCT; OR 0.99, 95% CI 0.76 -1.28), new visual aids (1 RCT; OR 0.99, 95% CI 0.57-1.73), SMS text messaging (3 RCTs; OR 1.07, 95% CI 0.81-1.41), educational videos (5 RCTs; OR 1.05, 95% CI 0.72-1.54), or social media applications (1 RCT; OR 1.25, 95% CI 0.65-2.41). Head-to-head meta-analysis showed that telephone calls (1 RCT; OR 1.61, 95% CI 0.97-2.68) were not superior to SMS text messaging. Significant heterogeneity was observed in the trials comparing educational videos (I^2=52%) to SPIs. All the pooled results are delineated in Figure S6 in Multimedia Appendix 4.

Compared to SPIs, social media applications were associated with reduced risk of abdominal discomfort (3 RCTs; OR 0.67, 95% CI 0.5-0.9), but not visual aids (1 RCT; OR 0.9, 95% CI 0.63-1.28), SMS text messaging (2 RCTs; OR 0.67, 95% CI 0.4-1.12), telephone calls (2 RCTs; OR 0.94, 95% CI 0.65-1.35), or educational videos (1 RCT; OR 1.01, 95% CI 0.37-2.76). Head-to-head meta-analysis showed that telephone calls (1 RCT; OR 0.83, 95% CI 0.47-1.48) or social media applications (1 RCT; OR 1.22, 95% CI 0.61-2.45) were not superior to SMS text messaging. All the pooled results are delineated in Figure S7 in Multimedia Appendix 4.

Compared to SPIs, social media applications indicated reduced risk of nausea and vomiting (3 RCTs; OR 0.7, 95% CI 0.55-0.88), but not visual aids (1 RCT; OR 1.27, 95% CI 0.67-2.41), SMS text messaging (2 RCTs; OR 0.64, 95% CI 0.27-1.53), telephone calls (2 RCTs; OR 0.8, 95% CI 0.56-1.13), or educational videos (1 RCT; OR 0.49, 95% CI 0.14-1.66). Head-to-head meta-analysis showed that telephone calls (1 RCT; OR 0.55, 95% CI 0.33-0.91) were superior to SMS text messaging, but not social media applications (1 RCT; OR 1.19, 95% CI 0.7-2.03). Significant heterogeneity was observed in the trials comparing SMS text messaging (I^2=74%) to SPIs. All the pooled results are delineated in Figure S8 in Multimedia Appendix 4.

Compared to SPIs, SMS text messaging (2 RCTs; OR 1.1, 95% CI 0.71-1.71), telephone calls (2 RCTs; OR 0.69, 95% CI 0.41-1.16), educational videos (1 RCT; OR 2.32, 95% CI 0.78-6.86), and social media applications (3 RCTs; OR 0.62, 95% CI 0.35-1.10) were not associated with increased risk of sleep disturbance. Head-to-head meta-analysis showed that telephone calls (1 RCT; OR 1.08, 95% CI 0.65-1.8) and social media applications (1 RCT; OR 1.08, 95% CI 0.66-1.77) were not superior to SMS text messaging. Significant heterogeneity was observed in the trials comparing telephone calls (I^2=55%) and social media applications (I^2=85%) to SPIs. All the pooled results are delineated in Figure S9 in Multimedia Appendix 4.

Network Meta-Analysis and Quality Assessment

Primary Outcomes

For the primary outcome of APR, the results of direct and indirect comparisons were largely similar with overlapping CIs, although differences were observed in the effect size and statistical significance in some cases, as shown in Table 2. In the network meta-analysis, we calculated the mixed-effect estimate as the weighted average of the direct (where available) and indirect treatment effects. In this analysis, when compared with SPIs, additional explanations (OR 3.56,95% CI 2.46-5; moderate quality of evidence), newly designed booklets (OR 3.81, 95% CI 2.19-6.3; very low quality of evidence), new visual aids (OR 3.61, 95% CI 1.29-8.63; very low quality of evidence), SMS text messaging (OR 2.7, 95% CI 1.85-3.86; high quality of evidence), telephone calls (OR 2.15, 95% CI 1.49-3.02; moderate quality of evidence), educational videos (OR 2.7, 95% CI 2.12-3.41; low quality of evidence), and social media applications (OR 2.81, 95% CI 2.07-3.73; high quality of evidence) increased the APR in patients undergoing colonoscopy, but not visual aids (OR 1.21, 95% CI 0.78-1.81; moderate quality of evidence) and mobile apps (OR 1.15, 95% CI 0.53-2.19; very low quality of evidence), as observed in Table 2.

Network meta-analysis demonstrated that visual aids (OR 0.35, 95% CI 0.19-0.59; high quality of evidence), telephone calls (OR 0.62, 95% CI 0.37-0.99; high quality of evidence), educational videos (OR 0.79, 95% CI 0.5-0.77; moderate quality of evidence), and mobile apps (OR 0.33, 95% CI 0.14-0.68; low quality of evidence) were inferior to additional explanations; newly designed booklets (OR 3.28, 95% CI 1.59-6.16; low quality of evidence), SMS text messaging (OR 2.33, 95% CI 1.28-3.91; high quality of evidence), telephone calls (OR 1.86, 95% CI 1.03-1.78; high quality of evidence), educational videos (OR 2.33, 95% CI 1.4-3.65; moderate quality of evidence), and social media applications (OR 2.42, 95% CI 1.4-3.93; high quality of evidence) were superior to visual aids; mobile apps were inferior to newly designed booklets (OR 0.32, 95% CI 0.12-0.7; low quality of evidence) or SMS text messaging (OR 0.44, 95% CI 0.18-0.9; low quality of evidence); social media applications (OR 0.43, 95% CI 0.19-0.85; low quality of evidence) were inferior to educational videos, and mobile apps (OR 0.42, 95% CI 0.18-0.83; low quality of evidence) were inferior to social media applications in increasing the APR.

Table 2. Pooled summary estimates and quality of evidence derived from direct and indirect estimates and network meta-analysis informing on comparative efficacy of educational interventions for improving quality of bowel preparation before colonoscopy[a].

Educational instruction	Direct estimate		Indirect estimate		Network meta-analysis	
	OR[b] (95% CI)	Quality of evidence	OR (95% CI)	Quality of evidence	OR (95% CI)	Quality of evidence
Compared with standard patient instructions						
Additional explanations	*3.34 (1.45-7.69)*	Moderate[c]	n.e.[d]	N/A[e]	*3.56 (2.46-5)*	Moderate
Visual aids	*1.18 (0.78-1.8)*	Moderate[f]	n.e.	N/A	1.21 (0.78-1.81)	Moderate
Newly designed booklets	*3.63 (2.15-6.12)*	Very low[g,h]	n.e.	N/A	*3.81 (2.19-6.3)*	Very low
New visual aids	*3.05 (1.21-7.68)*	Very low[c,g,i]	n.e.	N/A	*3.61 (1.29-8.63)*	Very low
SMS text messaging	*2.77 (1.86-4.14)*	High	2.12 (0.96-4.03)	Very low	*2.7 (1.85-3.86)*	High
Telephone calls	*2.64 (1.03-6.74)*	Moderate[j]	2.22 (0.55-8.93)	Very low	*2.15 (1.49-3.02)*	Moderate
Educational videos	*2.82 (1.83-4.35)*	Low[k]	n.e.	N/A	*2.7 (2.12-3.41)*	Low
Social media applications	*2.7 (1.75-4.18)*	High	1.03 (0.57-1.85)	Low	*2.81 (2.07-3.73)*	High
Mobile apps	1.85 (0.15-23.24)	Very low[j,h]	n.e.	N/A	1.15 (0.53-2.19)	Very low
Compared with additional explanations						
Visual aids	N/A	N/A	*0.34 (0.14-0.87)*	Moderate	*0.35 (0.19-0.59)*	High
Newly designed booklets	N/A	N/A	1.05 (0.39-2.83)	Very low	1.11 (0.56-2)	Very low
New visual aids	N/A	N/A	0.89 (0.26-3.08)	Very low	1.05 (0.34-2.61)	Very low
SMS text messaging	N/A	N/A	0.81 (0.32-2.03)	Moderate	0.79 (0.45-1.26)	Moderate
Telephone calls	N/A	N/A	0.77 (0.22-2.7)	Moderate	*0.62 (0.37-0.99)*	High
Educational videos	N/A	N/A	0.82 (0.32-2.1)	Low	*0.79 (0.5-0.77)*	Moderate
Social media applications	N/A	N/A	0.79 (0.31-2.01)	Moderate	0.81 (0.5-1.26)	Moderate
Mobile apps	N/A	N/A	0.54 (0.04-7.66)	Very low	*0.33 (0.14-0.68)*	Low
Compared with visual aids						
Newly designed booklets	N/A	N/A	*3.08 (1.58-6.01)*	Very low	*3.28 (1.59-6.16)*	Low
New visual aids	N/A	N/A	2.59 (0.94-7.13)	Very low	3.12 (0.98-7.96)	Very low
SMS text messaging	N/A	N/A	*2.35 (1.32-4.19)*	Moderate	*2.33 (1.28-3.91)*	High
Telephone calls	N/A	N/A	2.24 (0.8-6.26)	Moderate	*1.86 (1.03-1.78)*	High
Educational videos	N/A	N/A	*2.39 (1.31-4.36)*	Low	*2.33 (1.4-3.65)*	Moderate
Social media applications	N/A	N/A	*2.29 (1.25-4.18)*	Moderate	*2.42 (1.4-3.93)*	High
Mobile apps	N/A	N/A	1.57 (0.86-2.87)	Very low	0.99 (0.4-2.05)	Very low
Compared with newly designed booklets						
New visual aids	N/A	N/A	0.84 (0.29-2.43)	Very low	1.02 (0.3-2.68)	Very low
SMS text messaging	N/A	N/A	0.76 (0.4-1.47)	Very low	0.76 (0.38-1.37)	Very low
Telephone calls	N/A	N/A	0.73 (0.25-2.13)	Very low	0.61 (0.3-1.09)	Very low
Educational videos	N/A	N/A	0.78 (0.39-1.53)	Very low	0.76 (0.41-1.3)	Very low
Social media applications	N/A	N/A	0.74 (0.38-1.47)	Very low	0.79 (0.41-1.37)	Very low
Mobile apps	N/A	N/A	0.51 (0.04-6.69)	Very low	*0.32 (0.12-0.7)*	Low
Compared with new visual aids						
SMS text messaging	N/A	N/A	0.91 (0.33-2.49)	Very low	0.94 (0.29-2.24)	Very low
Telephone calls	N/A	N/A	0.87 (0.23-3.23)	Very low	0.75 (0.23-1.76)	Very low

Educational instruction	Direct estimate		Indirect estimate		Network meta-analysis	
	OR[b] (95% CI)	Quality of evidence	OR (95% CI)	Quality of evidence	OR (95% CI)	Quality of evidence
Educational videos	N/A	N/A	0.93 (0.33-2.57)	Very low	0.95 (0.3-2.15)	Very low
Social media applications	N/A	N/A	0.89 (0.32-2.46)	Very low	0.98 (0.31-2.27)	Very low
Mobile apps	N/A	N/A	0.61 (0.04-8.9)	Very low	0.4 (0.1-1.05)	Very low
Compared with SMS text messaging						
Telephone calls	1.25 (0.33-4.77)	Very low[f,i]	0.95 (0.34-2.65)	Moderate	0.82 (0.48-1.31)	Moderate
Educational videos	N/A	N/A	1.02 (0.57-1.84)	Low	1.04 (0.65-1.56)	Low
Social media applications	1.32 (0.63-2.78)	Low[l]	0.98 (0.54-1.76)	High	1.07 (0.67-1.62)	High
Mobile apps	N/A	N/A	0.67 (0.05-8.58)	Very low	*0.44 (0.18-0.9)*	Low
Compared with telephone calls						
Educational videos	N/A	N/A	1.07 (0.38-3.01)	Low	1.3 (0.83-1.95)	Low
Social media applications	N/A	N/A	1.02 (0.36-2.88)	Moderate	1.35 (0.82-2.07)	Moderate
Mobile apps	N/A	N/A	0.70 (0.05-10.33)	Very low	0.55 (0.23-1.13)	Very low
Compared with educational videos						
Social media applications	N/A	N/A	0.96 (0.52-1.77)	Low	1.05 (0.71-1.51)	Low
Mobile apps	N/A	N/A	0.66 (0.05-8.47)	Very low	*0.43 (0.19-0.85)*	Low
Compared with social media applications						
Mobile apps	N/A	N/A	0.69 (0.05-8.85)	Very low	*0.42 (0.18-0.83)*	Low

[a]The italicized values indicate statistically significant differences.

[b]OR: odds ratio.

[c]One was rated with high risk.

[d]n.e.: not estimable.

[e]N/A: not applicable.

[f]Only one was captured.

[g]Point estimates between two studies were conflicting.

[h]One with only 266 participants was included.

[i]One with only 200 participants was included.

[j]A wide 95% CI was generated.

[k]Two were rated with high risk.

[l]One with only 257 participants was included.

For the primary outcome of APR, newly designed booklets had the highest probability of being ranked the best (85.8%), followed by additional explanations (80.6%), new visual aids (71.6%), social media applications (64.6%), educational videos (61.0%), SMS text messaging (60.4%), telephone calls (40.9%), visual aids (15.9), mobile apps (12%), and SPIs (7.2%), as shown in Figure S10 in Multimedia Appendix 4.

Secondary Outcomes

Network meta-analysis showed that when compared to SPIs, additional explanations (OR 9.84, 95% CI 4.13-21.78), SMS text messaging (OR 6.99, 95% CI 3.57-12.92), telephone calls (OR 8.48, 95% CI 3.08-7.42), educational videos (OR 4.24, 95% CI 1.12-12.61), and social media applications (OR 3.76, 95% CI 2.7-5.13) increased adherence to the preparation regime, but not mobile apps (OR 1.33, 95% CI 0.74-2.2), as shown in Table 3. Additional explanations significantly increased adherence to the preparation regime when compared to social media applications (OR 0.46, 95% CI 0.16-0.96) or mobile apps (OR 0.16, 95% CI 0.05-0.37). Moreover, SMS messaging (OR 0.21, 95% CI 0.08-0.44), telephone calls (OR 0.29, 95% CI 0.13-0.54), and social media applications (OR 0.36, 95% CI 0.18-0.65) significantly increased adherence to the preparation regime compared to mobile apps.

Table 3. Pooled relative risks of secondary outcomes based on combined direct and indirect evidence from Bayesian network meta-analysis with different educational instructions in patients undergoing colonoscopy[a].

Education instruction	Adherence to instruction	Satisfaction with BP_b	Willingness to repeat	Abdominal discomfort	Nausea and vomiting	Sleep disturbance	Polyp detection rate
Compared with standard patient instructions							
Additional explanations	*9.84 (4.13-21.78)*	N/A[c]	1.02 (0.19-3.24)	N/A	N/A	N/A	N/A
Visual aids	N/A	N/A	N/A	0.91 (0.62-1.27)	1.35 (0.67-2.46)	N/A	1 (0.76-1.28)
New visual aids	N/A	N/A	N/A	N/A	N/A	N/A	1.03 (0.56-1.74)
SMS text messaging	*6.99 (3.57-12.92)*	*1.79 (1.22-2.55)*	1.44 (0.86-2.32)	0.71 (0.46-1.03)	*0.66 (0.46-0.9)*	0.8 (0.58-1.1)	1.1 (0.85-1.41)
Telephone calls	*4.84 (3.08-7.42)*	*1.97 (1.19-3.12)*	1.36 (0.95-1.89)	0.95 (0.67-1.31)	0.92 (0.67-1.25)	*0.69 (0.51-0.91)*	*1.86 (1.4-2.44)*
Educational videos	*4.24 (1.12-12.61)*	1.6 (0.7-3.18)	N/A	1.18 (0.36-2.93)	2.77 (0.65-8.51)	2.88 (0.84-7.93)	0.81 (0.6-1.07)
Social media applications	*3.76 (2.7-5.13)*	*2.03 (1.17-3.34)*	*2.22 (1.77-2.75)*	*0.69 (0.52-0.9)*	*0.69 (0.54-0.85)*	*0.53 (0.44-0.65)*	1.37 (0.87-2.01)
Mobile apps	1.33 (0.74-2.2)	N/A	*2.73 (1.27-5.34)*	N/A	N/A	N/A	N/A
Compared with additional explanations							
SMS text messaging	0.85 (0.25-2.07)	N/A	2.39 (0.4-8.14)	N/A	N/A	N/A	N/A
Telephone calls	0.59 (0.2-1.29)	N/A	2.25 (0.4-7.44)	N/A	N/A	N/A	N/A
Educational videos	0.52 (0.09-1.74)	N/A	N/A	N/A	N/A	N/A	N/A
Social media applications	*0.46 (0.16-0.96)*	N/A	3.68 (0.66-12.1)	N/A	N/A	N/A	N/A
Mobile apps	*0.16 (0.05-0.37)*	N/A	4.53 (0.67-16.48)	N/A	N/A	N/A	N/A
Compared with visual aids							
New visual aids	N/A	N/A	N/A	N/A	N/A	N/A	1.05 (0.54-1.87)
SMS text messaging	N/A	N/A	N/A	0.8 (0.46-1.32)	0.54 (0.24-1.05)	N/A	1.13 (0.77-1.59)
Telephone calls	N/A	N/A	N/A	1.08 (0.64-1.7)	0.76 (0.35-1.45)	N/A	*1.90 (1.28-2.72)*
Educational videos	N/A	N/A	N/A	1.34 (0.37-3.47)	2.28 (0.43-7.59)	N/A	0.82 (0.55-1.19)
Social media applications	N/A	N/A	N/A	0.79 (0.49-1.2)	0.57 (0.27-1.05)	N/A	1.39 (0.82-2.2)
Compared with new visual aids							
SMS text messaging	N/A	N/A	N/A	N/A	N/A	N/A	1.16 (0.6-2.04)

Education instruction	Adherence to instruction	Satisfaction with BP_b	Willingness to repeat	Abdominal discomfort	Nausea and vomiting	Sleep disturbance	Polyp detection rate
Telephone calls	N/A	N/A	N/A	N/A	N/A	N/A	1.96 (1-3.46)
Educational videos	N/A	N/A	N/A	N/A	N/A	N/A	0.85 (0.43-1.52)
Social media applications	N/A	N/A	N/A	N/A	N/A	N/A	1.44 (0.67-2.71)
Compared with SMS text messaging							
Telephone calls	0.77 (0.33-1.49)	1.12 (0.66-1.79)	0.99 (0.55-1.65)	1.39 (0.86-2.14)	1.44 (0.95-2.1)	0.88 (0.59-1.27)	*1.71 (1.2-2.36)*
Educational videos	0.67 (0.14-2.19)	0.93 (0.36-1.95)	N/A	1.74 (0.48-4.55)	4.34 (0.95-13.54)	*3.68 (1.01-10.24)*	0.74 (0.49-1.07)
Social media applications	0.59 (0.28-1.09)	1.16 (0.65-1.92)	1.64 (0.93-2.65)	1.02 (0.63-1.57)	1.07 (0.73-1.52)	*0.68 (0.47-0.95)*	1.25 (0.79-1.88)
Mobile apps	*0.21 (0.08-0.44)*	N/A	2.02 (0.77-4.42)	N/A	N/A	N/A	N/A
Compared with telephone calls							
Educational videos	0.92 (0.22-2.87)	0.86 (0.32-1.89)	N/A	1.28 (0.37-3.32)	3.07 (0.69-9.65)	*4.24 (1.17-11.87)*	*0.44 (0.29-0.65)*
Social media applications	0.82 (0.45-1.34)	1.09 (0.52-2.01)	*1.69 (1.1-2.48)*	0.75 (0.48-1.12)	0.76 (0.51-1.09)	0.79 (0.55-1.1)	0.75 (0.44-1.17)
Mobile apps	*0.29 (0.13-0.54)*	N/A	2.08 (0.88-4.29)	N/A	N/A	N/A	N/A
Compared with educational videos							
Social media applications	1.31 (0.28-3.5)	1.48 (0.53-3.3)	N/A	0.78 (0.48-1.97)	0.38 (0.08-1.08)	*0.26 (0.07-0.64)*	*1.73 (1.01-2.77)*
Mobile apps	0.46 (0.09-1.32)	N/A	N/A	N/A	N/A	N/A	N/A
Compared with social media applications							
Mobile apps	*0.36 (0.18-0.65)*	N/A	1.25 (0.56-2.5)	N/A	N/A	N/A	N/A

[a]The italicized values indicate statistically significant differences.

[b]BP: bowel preparation.

[c]N/A: not applicable.

The analysis also demonstrated that besides educational videos, SMS text messaging (OR 1.79, 95% CI 1.22-2.55), telephone calls (OR 1.97, 95% CI 1.19-3.12), and social media applications (OR 2.03, 95% CI 1.17-3.34) indicated satisfaction with the BP regime when compared to SPIs.

Network meta-analysis also showed that when compared to SPIs, social media applications (OR 2.22, 95% CI 1.77-2.75) or mobile apps (OR 2.73, 95% CI 1.27-5.34) were associated with increased willingness to repeat the same BP regime. Social media applications (OR 1.69, 95% CI 1.1-2.48) significantly increased the willingness to repeat the same BP regime when compared to telephone calls.

Furthermore, telephone calls significantly increased the PDR when compared to SPIs (OR 1.86, 95% CI 1.4-2.44) and SMS messaging (OR 1.71, 95% CI 1.2-2.36). Social media applications (OR 1.73, 95% CI 1.01-2.77) were associated with an increased PDR compared to educational videos.

The analysis also showed that besides social media applications (OR 0.69, 95% CI 0.54-0.85), no other instructions were associated with decreased abdominal discomfort, when compared to SPIs. SMS text messaging (OR 0.66, 95% CI 0.46-0.9) or social media applications (OR 0.69, 95% CI 0.54-0.85) were superior to SPIs in reducing the risk of nausea and vomiting. Telephone calls or social media applications were associated with reduced risk of sleep disturbance but not SMS

text messaging or educational videos, when compared to SPIs. Moreover, SMS text messaging (OR 3.68, 95% CI 1.01-10.24) or telephone calls (OR 4.24, 95% CI 1.17-11.87) were also superior to educational videos in improving sleep disturbance. Social media applications were also superior to SMS text messaging (OR 0.68, 95% CI 0.47-0.95) or educational videos (OR 0.26, 95% CI 0.07-0.64) in reducing the risk of sleep disturbance

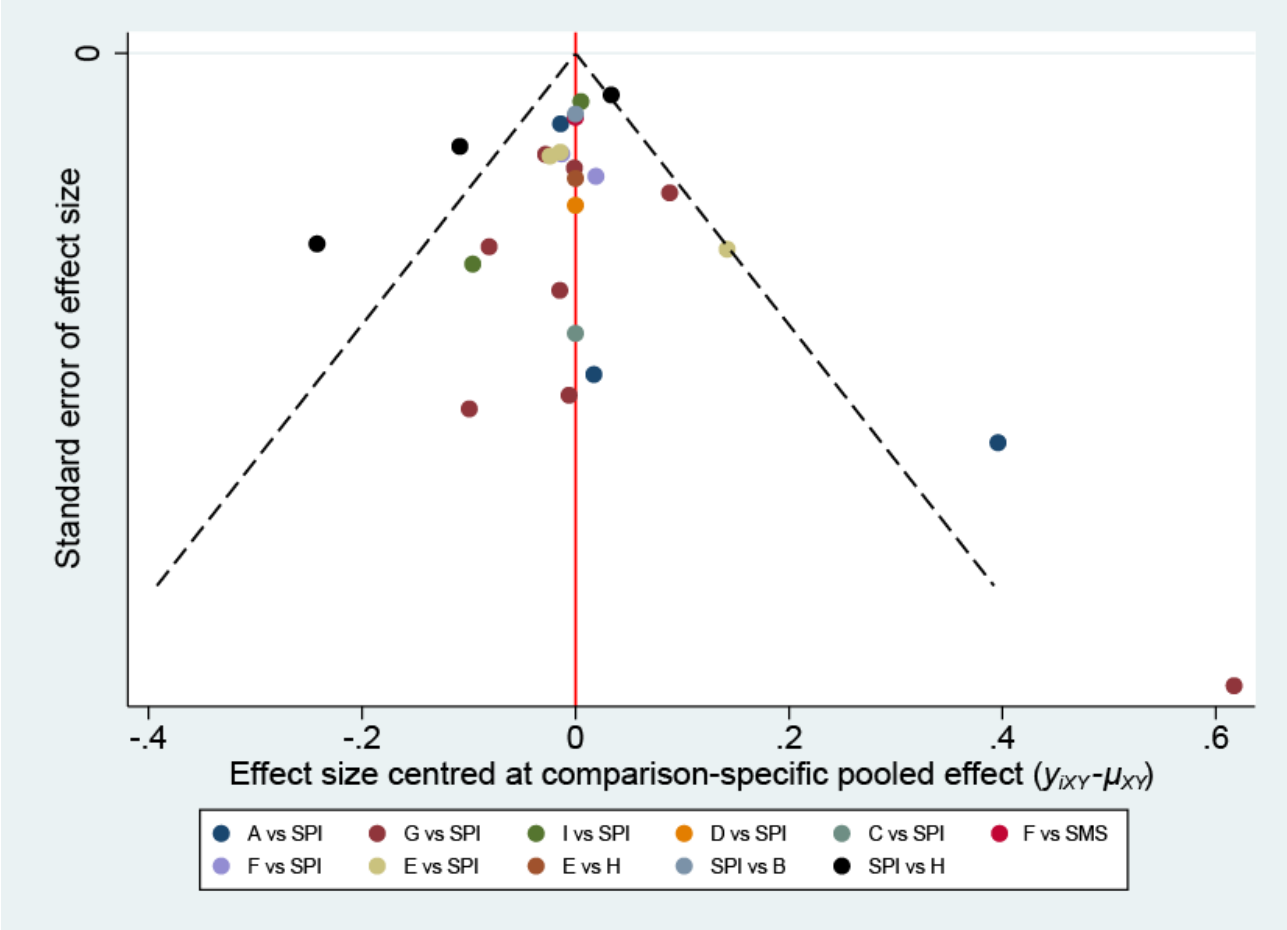
Sensitivity Analysis

Results from multiple prespecified sensitivity analyses are presented in Multimedia Appendix 5. Overall, for the primary outcome, the results were largely similar to the main analysis in the sensitivity analyses based on the (1) BP assessment scale (excluding studies in which uncommon scales were used except for BBPS, OBPS, and ABPS), (2) risk of bias (excluding studies with high risk), and (3) study design (excluding studies with multicenter designs). After excluding the studies using uncommon BP assessment scales, new visual aids (OR 3.16, 95% CI 1.02-8.01) were statistically superior to visual aids or mobile apps (OR 0.27, 95% CI 0.06-0.75) in increasing the APR; the difference (OR 0.49, 95% CI 0.2-1.03) between mobile apps and social media applications was not significant in increasing adherence to the BP regime. Excluding the studies with high risk also showed that telephone calls (OR 0.63, 95% CI 0.37-1.01) and educational videos (OR 0.79, 95% CI 0.5-1.19) were not inferior to additional explanations in increasing the APR. Excluding studies using uncommon preparation assessment scales showed that social media applications were not superior to mobile apps (OR 0.49, 95% CI 0.2-1.03) in increasing adherence to the BP regime. Furthermore, excluding the studies with high risk revealed that the difference (OR 1.59, 95% CI 0.88-2.66) between mobile apps and social media applications in increasing the PDR was not significant.

Publication Bias and Network Coherence

We did not find evidence of publication bias based on the funnel plot asymmetry in Figure 3, although the number of studies included in each comparison was very small, thereby making the available methods for evaluating publication bias somewhat unreliable. There were no significant differences between direct and indirect estimates where both were available, and the 2 methods had overlapping CIs for all interventions, as observed in Table 2.

Figure 3. Comparison-adjusted funnel plot for adequate preparation rate. The vertical axis represents the standard error of the effect size. The horizontal axis indicates the effect size centered at the comparison-specific pooled effect. The symmetrical funnel plot indicates the absence of publication bias. PDR: polyp detection rate; EPI: enhanced patient instruction; SPI: standard patient instruction; APR: adequate preparation rate; A: additional explanation; B: visual aid; C: newly designed booklet; D: new visual aids; E: SMS text messaging; F: telephone call; G: educational video; H: social media application; I: mobile app.



Discussion

Several meta-analyses [13,14,16-18] examined the quality of BP, colonic outcomes, and AEs including abdominal discomfort, nausea and vomiting, and sleep disturbance by comparing EPIs and SPIs. They concluded that EPIs are more effective and safer techniques for improving the quality of BP before colonoscopy. However, they did not completely explain which instruction is superior among the various EPIs because their analyses were based solely on direct comparisons between 2 given techniques. Our network meta-analysis of 23 RCTs involving 7969 patients is the first study that comprehensively analyzes direct and indirect evidence in EPIs for BP prior to colonoscopy.

In this systematic review and network meta-analysis, we made several key observations: (1) SMS text messaging and social media applications increase the APR, with high confidence in the estimates; additional explanations and telephone calls improve the APR with moderate confidence in the estimates; educational videos, newly designed booklets, and new visual aids also increase the APR but with low-to-very-low confidence in the estimates. (2) Based on high-to-moderate-quality evidence, additional explanations offer significant advantages over visual aids, telephone calls, and educational videos, but not SMS text messaging and social media applications; additional explanations are also superior to mobile apps but not to newly designed booklets and new visual aids, although the quality of evidence ranges from low to very low. (3) SMS text messaging, telephone calls, social media applications, and educational videos offer significant advantages over visual aids for increasing the APR, based on moderate-to-high-quality evidence; newly designed booklets also increase the APR with low-quality evidence. (4) According to low-quality evidence, newly designed booklets, SMS text messaging, educational videos, and social media applications significantly increase the APR but not new visual aids and telephone calls when compared to mobile apps. Overall, we observed that newly designed booklets had the highest probability of being ranked the best (for increasing APR), followed by additional explanations, new visual aids, social media applications, educational videos, SMS text messaging, telephone calls, and visual aids. (5) SMS text messaging, telephone calls, and social media applications increase adherence to and satisfaction with the BP regime; social media applications are associated with decreased risk of AEs; telephone calls or social media applications increase the PDR. It should be noted that we analyzed an inadequate number of eligible studies, and additional studies are warranted.

The quality of BP is the metric to determine if colonoscopy is successful and safe [63]. In this regard, 2 studies investigated the comparative efficacy of telephone calls and social media applications (WeChat) compared to SMS text messaging for BP and colonic outcomes. Lee and colleagues found no significant difference in the quality of BP between the telephone call and SMS text messaging groups, which is consistent with our finding (OR 0.82, 95% CI, 0.48-1.31; moderate quality evidence). However, the finding of Wang et al conflicts with the finding in the present study (OR 1.07, 95% CI 0.67-1.62; high quality of evidence), wherein social media applications are superior to SMS text messaging in increasing the APR. As a critically important quality metric of colonoscopy [63,64], the PDR was also assessed in these 2 studies, where no differences between the telephone call or social media application, and SMS text messaging groups were detected. However, our study demonstrates that telephone calls are associated with an increased PDR when compared to SMS text messaging (OR 1.71, 95% CI 1.2-2.36). Moreover, only 1 study with a small sample size was performed to compare telephone calls or social media applications to SMS text messaging. However, more studies performing indirect comparisons were included in our study to calculate the combined estimates. Further, this finding was supported by moderate- or high-quality evidence in our study.

The strengths of our analyses include the comprehensive and simultaneous assessment of the relative efficacy of all EPIs for BP prior to colonoscopy. Given the limited comparative effectiveness studies, it remains difficult for patients and physicians to make informed decisions about which instructions are most effective for improving the quality of BP. We used the GRADE methodology [34,35] to assess the quality of evidence for this network meta-analysis, which can be directly applied in guideline development.

However, there are certain limitations associated with direct comparative effectiveness related to network analyses and individual studies, which merit further discussion. First, there is a paucity of studies. Second, network meta-analysis may be vulnerable to misinterpretation. The biggest threat to the validity of network meta-analysis is conceptual heterogeneity, wherein there are considerable differences among the participants, interventions, background treatments, and outcome assessments, thus limiting the comparability of trials. It assumes that patients enrolled in all the included studies could have been sampled from the same theoretical population [20,35]. Although there were subtle differences in the patients (proportion of patients having undergone colonoscopy previously and those undergoing screening colonoscopy, diagnostic colonoscopy, or surveillance colonoscopy), BP solutions (such as polyethylene glycol [PEG] alone, PEG plus prokinetic agent, and low or standard doses), solution administration methods (split or single dose), and dietary restrictions (low residue, clear liquid, or low fiber), we tried to minimize this conceptual heterogeneity by performing multiple sensitivity analyses, including excluding trials using uncommon preparation assessment scales with high risk or multicenter design; the overall findings were unchanged, suggesting the robustness of these analyses. Third, the ranking probabilities may be challenging to interpret and do not always imply a clinically important difference. However, instead of focusing only on the summaries of the effect estimates, we used GRADE to rate the overall quality of evidence considering the risk of bias, imprecision, indirectness, inconsistency, and other biases for rating the confidence in the estimates [35].

There were similar limitations in the individual studies, which also undermine the strength of the meta-analysis. Most studies focused on the quality of BP, with a limited number of studies on colonic outcomes. BP-related AEs were poorly reported, limiting the assessments regarding the benefits of instructions; hence, a thorough assessment of risk-benefit profiles could not be performed. Studies were also at risk of detection bias with

suboptimal reporting of the blinding to assessor outcomes. Various dietary restrictions were imposed in different eligible studies; however, previous meta-analyses have demonstrated no difference between low residues and clear liquids [12,65], and thus, sensitivity analysis was not designed according to these aspects. Although BP solutions with different doses were used in the included studies, we did not design the subgroup or sensitivity analysis according to the dose of the BP solution because no significant difference between low and traditional doses was confirmed in our meta-analysis [6]. Split doses proved beneficial compared to single doses for improving BP [66] and increasing the PDR and ADR [67]; however, the insufficient number of eligible studies poses limitations for designing further sensitivity analyses. Moreover, 5 BP assessment scales were used in all the eligible studies; thus, we performed a sensitivity analysis to examine the robustness of the pooled results through excluding studies in which uncommon scales were used and found that most of the results were unchanged.

Despite these limitations, our network meta-analysis provides a better understanding regarding the comparative efficacy of EPIs for BP prior to colonoscopy. Newly designed booklets, telephone calls, educational videos, and social media applications can increase the quality of BP. Telephone calls or social media applications may be associated with improved adherence to and satisfaction with the BP regime, decreased risk of AEs, or an increased PDR.

Acknowledgments

This study was supported by the Medical Research Project which was jointly approved by the Chongqing Science and Technology Bureau and Health Commission of Chongqing Municipal City (approval number: 2020MSXM064). We also sincerely appreciate all the authors who performed the studies included in the network meta-analysis performed in this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full search strategy.
[DOCX File , 18 KB - jmir_v23i10e19915_app1.docx]

Multimedia Appendix 2
Classification and comparison of all types of enhanced patient instructions included in this study.
[DOCX File , 16 KB - jmir_v23i10e19915_app2.docx]

Multimedia Appendix 3
Characteristics of eligible studies.
[DOCX File , 17 KB - jmir_v23i10e19915_app3.docx]

Multimedia Appendix 4
All supplementary figures supporting our findings.
[DOCX File , 3862 KB - jmir_v23i10e19915_app4.docx]

Multimedia Appendix 5
All sensitivity analyses according to predesigned criteria.
[DOCX File , 20 KB - jmir_v23i10e19915_app5.docx]

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Sep:394-424 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
2. Imperiale TF, Glowinski EA, Lin-Cooper C, Larkin GN, Rogge JD, Ransohoff DF. Five-year risk of colorectal neoplasia after negative screening colonoscopy. N Engl J Med 2008 Sep;359(12):1218-1224. [doi: 10.1056/nejmoa0803597]
3. Winawer SJ, Zauber AG, Ho MN, O'Brien MJ, Gottlieb LS, Sternberg SS, et al. Prevention of colorectal cancer by colonoscopic polypectomy. N Engl J Med 1993 Dec;329(27):1977-1981. [doi: 10.1056/nejm199312303292701]
4. Zauber AG, Winawer SJ, O'Brien MJ, Lansdorp-Vogelaar I, van Ballegooijen M, Hankey BF, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. N Engl J Med 2012 Feb;366(8):687-696. [doi: 10.1056/nejmoa1100370]
5. Zarchy TM, Ershoff D. Risk of colorectal cancer in families of patients with adenomatous polyps. N Engl J Med 1996 May;334(20):1339-1340. [doi: 10.1056/nejm199605163342016]

6. Yi L, Tian X, Shi B, Chen H, Liu X, Pi Y, et al. Low-volume polyethylene glycol improved patient attendance in bowel preparation before colonoscopy: a meta-analysis with trial sequential analysis. Front Med 2019 May;6:92. [doi: 10.3389/fmed.2019.00092]
7. Sherer EA, Imler TD, Imperiale TF. The effect of colonoscopy preparation quality on adenoma detection rates. Gastrointest Endosc 2012 Mar;75(3):545-553. [doi: 10.1016/j.gie.2011.09.022] [Medline: 22138085]
8. Tian X, Shi B, Chen H, Liu X, Tang R, Pi Y, et al. Comparative efficacy of 2 L polyethylene glycol alone or with ascorbic acid vs. 4 L polyethylene glycol for colonoscopy: a systematic review and network meta-analysis of 12 randomized controlled trials. Front Med (Lausanne) 2019 Aug;6:182 [FREE Full text] [doi: 10.3389/fmed.2019.00182] [Medline: 31497604]
9. Tian X, Shi B, Liu X, Chen H, Chen W. A randomized trial of split dose 3 L polyethylene glycol lavage solution, 2 L polyethylene glycol lavage combined with castor oil, and 1 L of polyethylene glycol lavage solution combined with castor Oil and ascorbic acid for preparation for colonoscopy. Front Med (Lausanne) 2019 Jul;6:158 [FREE Full text] [doi: 10.3389/fmed.2019.00158] [Medline: 31334239]
10. Rex D, Imperiale TF, Latinovich DR, Bratcher LL. Impact of bowel preparation on efficiency and cost of colonoscopy. Am J Gastroenterol 2002 Jul;97(7):1696-1700. [doi: 10.1111/j.1572-0241.2002.05827.x] [Medline: 12135020]
11. Harewood GC, Sharma VK, de Garmo P. Impact of colonoscopy preparation quality on detection of suspected colonic neoplasia. Gastrointest Endosc 2003 Jul;58(1):76-79. [doi: 10.1067/mge.2003.294] [Medline: 12838225]
12. Song G, Tian X, Ma L, Yi L, Shuai T, Zeng Z, et al. Regime for bowel preparation in patients scheduled to colonoscopy: low-residue diet or clear liquid diet? Evidence From Systematic Review With Power Analysis. Medicine (Baltimore) 2016 Jan;95(1):e2432 [FREE Full text] [doi: 10.1097/MD.0000000000002432] [Medline: 26735547]
13. Kurlander JE, Sondhi AR, Waljee AK, Menees SB, Connell CM, Schoenfeld PS, et al. How efficacious are patient education interventions to improve bowel preparation for colonoscopy? a systematic review. PLoS One 2016 Oct;11(10):e0164442 [FREE Full text] [doi: 10.1371/journal.pone.0164442] [Medline: 27741260]
14. Guo X, Yang Z, Zhao L, Leung F, Luo H, Kang X, et al. Enhanced instructions improve the quality of bowel preparation for colonoscopy: a meta-analysis of randomized controlled trials. Gastrointest Endosc 2017 Jan;85(1):90-97.e6. [doi: 10.1016/j.gie.2016.05.012] [Medline: 27189659]
15. Ness RM, Manam R, Hoen H, Chalasani N. Predictors of inadequate bowel preparation for colonoscopy. Am J Gastroenterol 2001 Jun;96(6):1797-1802. [doi: 10.1111/j.1572-0241.2001.03874.x] [Medline: 11419832]
16. Chang C, Shih S, Wang H, Chu C, Wang T, Hung C, et al. Meta-analysis: the effect of patient education on bowel preparation for colonoscopy. Endosc Int Open 2015 Dec;3(6):E646-E652 [FREE Full text] [doi: 10.1055/s-0034-1392365] [Medline: 26716129]
17. Desai M, Nutalapati V, Bansal A, Buckles D, Bonino J, Olyaee M, et al. Use of smartphone applications to improve quality of bowel preparation for colonoscopy: a systematic review and meta-analysis. Endosc Int Open 2019 Feb;7(2):E216-E224 [FREE Full text] [doi: 10.1055/a-0796-6423] [Medline: 30705956]
18. Gkolfakis P, Tziatzios G, Papanikolaou IS, Triantafyllou K. Strategies to improve inpatients' quality of bowel preparation for colonoscopy: a systematic review and meta-analysis. Gastroenterol Res Pract 2019;2019:5147208 [FREE Full text] [doi: 10.1155/2019/5147208] [Medline: 31191646]
19. Tian X, Xu L, Liu X, Chen W. Enhanced patient education for colonic polyp and adenoma detection: meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2020 Jun;8(6):e17372 [FREE Full text] [doi: 10.2196/17372] [Medline: 32347798]
20. Cipriani A, Higgins JP, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. Ann Intern Med 2013 Jul;159(2):130-137. [doi: 10.7326/0003-4819-159-2-201307160-00008] [Medline: 23856683]
21. Mills EJ, Ioannidis JPA, Thorlund K, Schünemann HJ, Puhan MA, Guyatt GH. How to use an article reporting a multiple treatment comparison meta-analysis. JAMA 2012 Sep;308(12):1246-1253. [doi: 10.1001/2012.jama.11228] [Medline: 23011714]
22. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug;151(4):264-269 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
23. Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health 2011 Jun;14(4):417-428 [FREE Full text] [doi: 10.1016/j.jval.2011.04.002] [Medline: 21669366]
24. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct;343:d5928-d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
25. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986 Sep;7(3):177-188. [Medline: 3802833]
26. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
27. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions version 6.2. United Kingdom: Cochrane Training; Feb 2021.

28. Page MJ, McKenzie JE, Higgins JPT. Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review. BMJ Open 2018 Mar;8(3):e019703 [FREE Full text] [doi: 10.1136/bmjopen-2017-019703] [Medline: 29540417]
29. Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2. Med Decis Making 2012 Oct;33(5):607-617. [doi: 10.1177/0272989x12458724]
30. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Statist Med 2004 Oct;23(20):3105-3124. [doi: 10.1002/sim.1875]
31. Sutton A, Ades AE, Cooper N, Abrams K. Use of indirect and mixed treatment comparisons for technology assessment. Pharmacoeconomics 2008;26(9):753-767. [doi: 10.2165/00019053-200826090-00006] [Medline: 18767896]
32. Song G, Liu X, Bian W, Wu J, Deng Y, Zhang H, et al. Systematic review with network meta-analysis: comparative efficacy of different enteral immunonutrition formulas in patients underwent gastrectomy. Oncotarget 2017 Apr;8(14):23376-23388 [FREE Full text] [doi: 10.18632/oncotarget.15580] [Medline: 28423579]
33. Singh S, Murad MH, Chandar AK, Bongiorno CM, Singal AK, Atkinson SR, et al. Comparative effectiveness of pharmacological interventions for severe alcoholic hepatitis: a systematic review and network meta-analysis. Gastroenterology 2015 Oct;149(4):958-970.e12. [doi: 10.1053/j.gastro.2015.06.006] [Medline: 26091937]
34. Guyatt G, Oxman AD, Sultan S, Brozek J, Glasziou P, Alonso-Coello P, et al. GRADE guidelines: 11. making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. J Clin Epidemiol 2013 Feb;66(2):151-157. [doi: 10.1016/j.jclinepi.2012.01.006] [Medline: 22542023]
35. Puhan MA, Schünemann HJ, Murad MH, Li T, Brignardello-Petersen R, Singh JA, GRADE Working Group. A GRADE working group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ 2014 Sep;349:g5630. [doi: 10.1136/bmj.g5630] [Medline: 25252733]
36. Back SY, Kim HG, Ahn EM, Park S, Jeon SR, Im HH, et al. Impact of patient audiovisual re-education via a smartphone on the quality of bowel preparation before colonoscopy: a single-blinded randomized study. Gastrointest Endosc 2018 Mar;87(3):789-799.e4. [doi: 10.1016/j.gie.2017.09.007] [Medline: 28939500]
37. Elvas L, Brito D, Areia M, Carvalho R, Alves S, Saraiva S, et al. Impact of personalised patient education on bowel preparation for colonoscopy: prospective randomised controlled trial. GE Port J Gastroenterol 2017 Jan;24(1):22-30 [FREE Full text] [doi: 10.1159/000450594] [Medline: 28848777]
38. Garg S, Girotra M, Chandra L, Verma V, Kaur S, Allawy A, et al. Improved bowel preparation with multimedia education in a predominantly African-American population: a randomized study. Diagn Ther Endosc 2016;2016:2072401 [FREE Full text] [doi: 10.1155/2016/2072401] [Medline: 27006590]
39. Jeon SC, Kim JH, Kim SJ, Kwon HJ, Choi YJ, Jung K, et al. Effect of sending educational video clips via smartphone mobile messenger on bowel preparation before colonoscopy. Clin Endosc 2019 Jan;52(1):53-58 [FREE Full text] [doi: 10.5946/ce.2018.072] [Medline: 30153724]
40. Kang X, Zhao L, Leung F, Luo H, Wang L, Wu J, et al. Delivery of instructions via mobile social media app increases quality of bowel preparation. Clin Gastroenterol Hepatol 2016 Mar;14(3):429-435.e3. [doi: 10.1016/j.cgh.2015.09.038] [Medline: 26492848]
41. Lee Y, Kim E, Choi J, Lee K, Park K, Cho K, et al. Impact of reinforced education by telephone and short message service on the quality of bowel preparation: a randomized controlled study. Endoscopy 2015 Nov;47(11):1018-1027. [doi: 10.1055/s-0034-1392406] [Medline: 26182387]
42. Liu C, Song X, Hao H. Educational video followed by retelling bowel preparation process to improve colonoscopy bowel preparation quality: a prospective nursing intervention study. Med Sci Monit 2018 Aug;24:6029-6037. [doi: 10.12659/msm.909572]
43. Liu X, Luo H, Zhang L, Leung FW, Liu Z, Wang X, et al. Telephone-based re-education on the day before colonoscopy improves the quality of bowel preparation and the polyp detection rate: a prospective, colonoscopist-blinded, randomised, controlled study. Gut 2014 Jan;63(1):125-130. [doi: 10.1136/gutjnl-2012-304292] [Medline: 23503044]
44. Lorenzo-Zúñiga V, Moreno de Vega V, Marín I, Barberá M, Boix J. Improving the quality of colonoscopy bowel preparation using a smart phone application: a randomized trial. Dig Endosc 2015 Jul;27(5):590-595. [doi: 10.1111/den.12467] [Medline: 25708251]
45. Meng X. Individualized health education improves bowel preparation before colonoscopy. WCJD 2015;23(14):2323. [doi: 10.11569/wcjd.v23.i14.2323]
46. Modi C, Depasquale JR, Digiacomo WS, Malinowski JE, Engelhardt K, Shaikh SN, et al. Impact of patient education on quality of bowel preparation in outpatient colonoscopies. Qual Prim Care 2009;17(6):397-404. [Medline: 20051190]
47. Park J, Kim MS, Kim H, Kim SI, Shin CH, Lee HJ, et al. A randomized controlled trial of an educational video to improve quality of bowel preparation for colonoscopy. BMC Gastroenterol 2016 Jun;16(1):64 [FREE Full text] [doi: 10.1186/s12876-016-0476-6] [Medline: 27317249]
48. Pillai A, Menon R, Oustecky D, Ahmad A. Educational colonoscopy video enhances bowel preparation quality and comprehension in an inner city population. J Clin Gastroenterol 2018 Jul;52(6):515-518. [doi: 10.1097/MCG.0000000000000893] [Medline: 28742732]

49. Prakash SR, Verma S, McGowan J, Smith BE, Shroff A, Gibson GH, et al. Improving the quality of colonoscopy bowel preparation using an educational video. Can J Gastroenterol 2013 Dec;27(12):696-700 [FREE Full text] [doi: 10.1155/2013/292636] [Medline: 24340313]
50. Rice SC, Higginbotham T, Dean MJ, Slaughter JC, Yachimski PS, Obstein KL. Video on diet before outpatient colonoscopy does not improve quality of bowel preparation: a prospective, randomized, controlled trial. Am J Gastroenterol 2016 Nov;111(11):1564-1571. [doi: 10.1038/ajg.2016.450] [Medline: 27753434]
51. Sharara AI, Chalhoub JM, Beydoun M, Shayto RH, Chehab H, Harb AH, et al. A customized mobile application in colonoscopy preparation: a randomized controlled trial. Clin Transl Gastroenterol 2017 Jan;8(1):e211 [FREE Full text] [doi: 10.1038/ctg.2016.65] [Medline: 28055031]
52. Spiegel BMR, Talley J, Shekelle P, Agarwal N, Snyder B, Bolus R, et al. Development and validation of a novel patient educational booklet to enhance colonoscopy preparation. Am J Gastroenterol 2011 May;106(5):875-883. [doi: 10.1038/ajg.2011.75] [Medline: 21483463]
53. Tae JW, Lee JC, Hong SJ, Han JP, Lee YH, Chung JH, et al. Impact of patient education with cartoon visual aids on the quality of bowel preparation for colonoscopy. Gastrointest Endosc 2012 Oct;76(4):804-811. [doi: 10.1016/j.gie.2012.05.026] [Medline: 22840295]
54. Walter B, Klare P, Strehle K, Aschenbeck J, Ludwig L, Dikopoulos N, et al. Improving the quality and acceptance of colonoscopy preparation by reinforced patient education with short message service: results from a randomized, multicenter study (PERICLES-II). Gastrointest Endosc 2019 Mar;89(3):506-513.e4. [doi: 10.1016/j.gie.2018.08.014] [Medline: 30138612]
55. Wang S, Wang Q, Yao J, Zhao S, Wang L, Li Z, et al. Effect of WeChat and short message service on bowel preparation. Eur J Gastroenterol Hepatol 2019;31(2):170-177. [doi: 10.1097/meg.0000000000001303]
56. Zhang Q, Li J, Zhang Q, Li Y, Lei C, Shang B, et al. Effect of education by messaging software on the quality of bowel preparation for colonoscopy. Chin Med J (Engl) 2018 Jul;131(14):1750-1752 [FREE Full text] [doi: 10.4103/0366-6999.235881] [Medline: 29998899]
57. Calderwood AH, Lai EJ, Fix OK, Jacobson BC. An endoscopist-blinded, randomized, controlled trial of a simple visual aid to improve bowel preparation for screening colonoscopy. Gastrointest Endosc 2011 Feb;73(2):307-314. [doi: 10.1016/j.gie.2010.10.013] [Medline: 21168840]
58. Cho YY, Kim HO. Effects of a patient educational video program on bowel preparation prior to colonoscopy. J Korean Acad Nurs 2015 Oct;45(5):704-712. [doi: 10.4040/jkan.2015.45.5.704] [Medline: 26582115]
59. Ergen WF, Pasricha T, Hubbard FJ, Higginbotham T, Givens T, Slaughter JC, et al. Providing hospitalized patients with an educational booklet increases the quality of colonoscopy bowel preparation. Clin Gastroenterol Hepatol 2016 Jun;14(6):858-864. [doi: 10.1016/j.cgh.2015.11.015] [Medline: 26681487]
60. Hsu W, Liang C, Lin C, Lee T, Chung C. A modified bowel preparation protocol improves the quality of bowel cleansing for colonoscopy. Advances in Digestive Medicine 2016 Mar;3(3):144-147. [doi: 10.1016/j.aidm.2015.12.001]
61. Kakkar A, Jacobson BC. Failure of an internet-based health care intervention for colonoscopy preparation: a caveat for investigators. JAMA Intern Med 2013 Jul;173(14):1374-1376. [doi: 10.1001/jamainternmed.2013.6477] [Medline: 23877080]
62. Rosenfeld G, Krygier D, Enns RA, Singham J, Wiesinger H, Bressler B. The impact of patient education on the quality of inpatient bowel preparation for colonoscopy. Can J Gastroenterol 2010 Sep;24(9):543-546 [FREE Full text] [doi: 10.1155/2010/718628] [Medline: 21152458]
63. Kaminski MF, Regula J, Kraszewska E, Polkowski M, Wojciechowska U, Didkowska J, et al. Quality indicators for colonoscopy and the risk of interval cancer. N Engl J Med 2010 May;362(19):1795-1803. [doi: 10.1056/nejmoa0907667]
64. Corley DA, Jensen CD, Marks AR, Zhao WK, Lee JK, Doubeni CA, et al. Adenoma detection rate and risk of colorectal cancer and death. N Engl J Med 2014 Apr;370(14):1298-1306 [FREE Full text] [doi: 10.1056/NEJMoa1309086] [Medline: 24693890]
65. Nguyen DL, Jamal MM, Nguyen ET, Puli SR, Bechtold ML. Low-residue versus clear liquid diet before colonoscopy: a meta-analysis of randomized, controlled trials. Gastrointest Endosc 2016 Mar;83(3):499-507.e1. [doi: 10.1016/j.gie.2015.09.045] [Medline: 26460222]
66. Martel M, Barkun AN, Menard C, Restellini S, Kherad O, Vanasse A. Split-dose preparations are superior to day-before bowel cleansing regimens: a meta-analysis. Gastroenterology 2015 Jul;149(1):79-88. [doi: 10.1053/j.gastro.2015.04.004] [Medline: 25863216]
67. Zawaly K, Rumbolt C, Abou-Setta AM, Neilson C, Rabbani R, Zarychanski R, et al. The efficacy of split-dose bowel preparations for polyp detection: a systematic review and meta-analysis. Am J Gastroenterol 2019 Jun;114(6):884-892. [doi: 10.14309/ajg.0000000000000155] [Medline: 30865011]

Abbreviations

ABPS: Aronchick Bowel Preparation Scale
AEs: adverse events
AI: adherence to instruction

APR: adequate preparation rate
BBPS: Boston Bowel Preparation Scale
CENTRAL: Cochrane Central Register of Controlled Trials
EPIs: enhanced patient instructions
GRADE: Grading of Recommendations Assessment, Development and Evaluation
OBPS: Ottawa Bowel Preparation Scale
OR: odds ratio
PDR: polyp detection rate
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SPIs: standard patient instructions

Edited by R Kukafka; submitted 07.05.20; peer-reviewed by K Mears, P Copanitsanou; comments to author 29.09.20; revised version received 13.10.20; accepted 22.03.21; published 25.10.21.

Please cite as:
Tian X, Yi LJ, Han Y, Chen H, Liu XL, Chen WQ, Jiménez-Herrera MF
Comparative Effectiveness of Enhanced Patient Instructions for Bowel Preparation Before Colonoscopy: Network Meta-analysis of 23 Randomized Controlled Trials
J Med Internet Res 2021;23(10):e19915
URL: https://www.jmir.org/2021/10/e19915
doi:10.2196/19915
PMID:34694227

Review

Supportive Care Interventions for People With Cancer Assisted by Digital Technology: Systematic Review

Michael Marthick[1], PhD; Deborah McGregor[2], MHS; Jennifer Alison[3,4], PhD; Birinder Cheema[5], PhD; Haryana Dhillon[6], PhD; Tim Shaw[1], PhD

[1]Research in Implementation Science and eHealth Group, Faculty of Medicine and Health, University of Sydney, Camperdown, Australia

[2]Faculty of Science, Medicine and Health, University of Wollongong, Wollongong, Australia

[3]Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[4]Sydney Local Health District, Sydney, Australia

[5]School of Health Sciences, Western Sydney University, Penrith, Australia

[6]Centre for Medical Psychology & Evidence-based Decision-making, School of Psychology, University of Sydney, Sydney, Australia

Corresponding Author:
Michael Marthick, PhD
Research in Implementation Science and eHealth Group
Faculty of Medicine and Health
University of Sydney
Charles Perkins Centre
John Hopkins Dr
Camperdown, 2006
Australia
Phone: 61 (02) 9351 2222
Email: mmar7320@uni.sydney.edu.au

Abstract

Background: Although relatively new, digital health interventions are demonstrating rapid growth because of their ability to facilitate access and overcome issues of location, time, health status, and most recently, the impact of a major pandemic. With the increased uptake of digital technologies, digital health has the potential to improve the provision of supportive cancer care.

Objective: This systematic review aims to evaluate digital health interventions for supportive cancer care.

Methods: Published literature between 2000 and 2020 was systematically searched in MEDLINE, PubMed, Embase, PsycINFO, Cochrane Central Register of Controlled Trials, and Scopus. Eligible publications were randomized controlled trials of clinician-led digital health interventions to support adult cancer patients. The interventions included were determined by applying a digital health conceptual model. Studies were appraised for quality using the revised Cochrane risk of bias tool.

Results: Twenty randomized controlled trials met the inclusion criteria for the analysis. Interventions varied by duration, frequency, degree of technology use, and applied outcome measures. Interventions targeting a single tumor stream, predominantly breast cancer, and studies involving the implementation of remote symptom monitoring have dominated the results. In most studies, digital intervention resulted in significant positive outcomes in patient-reported symptoms, levels of fatigue and pain, health-related quality of life, functional capacity, and depression levels compared with the control.

Conclusions: Digital health interventions are helpful and effective for supportive care of patients with cancer. There is a need for high-quality research. Future endeavors could focus on the use of valid, standardized outcome measures, maintenance of methodological rigor, and strategies to improve patient and health professional engagement in the design and delivery of supportive digital health interventions.

Trial Registration: PROSPERO CRD42020149730; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=149730

(J Med Internet Res 2021;23(10):e24722) doi:10.2196/24722

KEYWORDS

digital health; telehealth; eHealth; neoplasm; supportive care; systematic review; mobile phone

Introduction

Background

Approximately 18.1 million new cancer cases and 9.6 million cancer-related deaths occurred globally in 2018 [1]. The rising tide of cancer diagnoses in many developed countries has been attributed to both population aging and the increasing prevalence of primary risk factors, including physical inactivity, obesity, and metabolic disease [1,2]. As the cancer population continues to grow, there is an urgent need to improve supportive care services [2].

Supportive care focuses on assisting people with cancer and their families to cope with the disease and its treatment [3]. The management of cancer treatment-related symptoms and side effects and the maintenance of health-related quality of life from early diagnosis to end-of-life are key aims of supportive cancer care [3,4]. Supportive care interventions vary and may involve multidisciplinary team support, including doctors, nurses, pharmacists, and allied health professionals [5]. Recently, a shared follow-up approach between primary and secondary providers has been promoted to successfully meet increasing demands for survivorship care [6,7], as innovative methods for long-term cancer care are constantly needed [2].

There have been ongoing attempts to improve access to supportive care cancer services through the use of digital health technology [8]. Digital health interventions, with telemedicine as its oldest form dating back to the 1920s, have been increasing dramatically in recent years [9]. The terms *digital health* and *eHealth* are frequently used interchangeably, with numerous varied definitions. Eysenbach [10] defined eHealth as *an emerging field in the intersection of medical informatics, public health, and business, referring to health services and information delivered or enhanced through the internet and related technologies*. Elbert [11] and McLean et al [12] assert 3 key elements of digital health: (1) data obtained from the patient, (2) electronic transfer of data over a distance, and (3) patient-tailored feedback from a health care professional. Furthermore, a recent conceptual model proposed by Shaw et al [13] acknowledges the role of telehealth consultations, web-based forums, mobile devices and apps, and social media, in enabling real-time communication between health professionals and consumers.

Recent systematic reviews evaluating the impact of digital health interventions on health and health care costs provide promising evidence of effectiveness and cost-effectiveness [11,14]. Digital health has demonstrated potential in engaging people in their care [15], including as a tool for the treatment and self-management training of chronically ill patients [16,17]. Digital health interventions have been shown to be effective for managing cancer-related fatigue [18], may improve physical activity among cancer survivors [19], and can lead to positive effects addressing the supportive cancer care needs of individuals with different preferences and priorities [20]. Cancer survivors have been found to have a positive attitude toward digital health [21], suggesting that digital health interventions have the potential to overcome common challenges associated with access to supportive care in this population. Health professional-led, digital health–enabled, supportive care interventions may prove particularly useful in increasing accessibility of services to those with limited access because of location, health, time, and public health emergencies [22-30]. Despite the abundance of recent digital health literature, there remains an acknowledged lack of quality evidence regarding the effectiveness of supportive digital health care interventions for people with cancer [19,20,31-33].

Objective

Over the past several decades, studies have investigated the implementation and effects of digital health interventions in people with cancer. In previous systematic reviews evaluated in 2013, the design features of supportive digital health interventions for patients with cancer [20]; in 2014, the use of technology in cancer follow-up [31]; in 2015 and 2017, the effect of telehealth interventions in cancer survivors' general quality of life [32,34], and in 2020, the benefits and limits of digital health for optimal supportive care in oncology [35]. The last review by Aapro et al [35] conducted an article search up to November 2018 and focused on the technical features of digital technologies. This is a rapidly growing area of health care because of advances in information technology and the uptake of digital technologies by both health professionals and patients. Therefore, this systematic review aims to explore the effect of supportive care interventions assisted by digital technologies on the outcomes of patients with cancer.

Methods

Search Strategy

This systematic review was registered in the PROSPERO (International Prospective Register of Systematic Reviews) and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [36]. A search was performed in August 2020, using the following web-based databases: MEDLINE (OvidSP, Wolters Kluwer), PubMed (National Center for Biotechnology Information, US National Library of Medicine), Embase (OvidSP, Wolters Kluwer), PsycINFO (American Psychological Association), Cochrane Central Register of Controlled Trials (John Wiley & Sons), and Scopus (SciVerse, Elsevier). Three main keywords were searched: *supportive care*, *digital health*, and *cancer patients*. Additional search terms were included based on synonyms of these keywords and medical subject headings. Multimedia Appendix 1 indicates the search strategy used.

Inclusion Criteria

The inclusion criteria were as follows: (1) studies in English describing a randomized controlled trial (RCT), published between January 2000 and August 2020; (2) intervention recipients were adults with a diagnosis of cancer; (3) involved clinician-led digital health interventions; and (4) interventions implemented to provide supportive cancer care.

The determination of digital health interventions was on the basis of the conceptual model of Shaw et al [13], which consists of 3 core domains:

1. Health in our hands: Using digital technologies to monitor, track, and inform health, for example, smartphones, tablets, clinical devices, mobile sensors and wearables, apps, social media, and web-based information.
2. Interacting for health: Using digital technologies to enable health communication among practitioners and between health professionals and clients or patients, for example, traditionally dominated by teleconferencing and videoconferencing, this domain increasingly includes a range of synchronous and asynchronous tools, such as SMS and push notifications from mobile apps, dedicated portals, social media platforms, and virtual or simulated therapy tools.
3. Data enabling health: Collecting, managing, and using digital health data, for example, technologies that provide expanded knowledge and insights about the health of an individual, community, or population.

To be included in the review, it was essential that the intervention satisfied the 2 domains *health in our hands* and *interacting for health*. The third domain, *data enabling health* was deemed nonessential because of the known inconsistent reporting of these criteria. The essential criterion *health in our hands* was captured in the *Intervention* column and outlines the nature of digital health experience. *Interacting for health* was captured in the *Interactions* column of review data and outlines the individuals involved in any form of communicative exchange that supports the health and well-being of the patient and caregiver.

Studies with interventions involving automated systems, such as interactive voice response and similar web-based systems, to monitor symptoms were included if the intervention featured an internet or web-based component and triggered health professional or researcher involvement when a threshold, such as a pain score, was reached.

Exclusion Criteria

Studies were excluded if they were (1) not RCTs; (2) reported only self-managed interventions, patient-to-patient interventions, prevention tools, or alternative treatments, or (3) focused solely on interventions involving telephone delivery that replicated a clinical service.

Data Extraction and Synthesis

Two reviewers (MM and DM) independently reviewed the titles and abstracts, followed by a full-text review of all publications. In cases of disagreement, a consensus was sought through discussion. Disagreement persisted for 4 studies; therefore, a third reviewer (TS) was consulted to adjudicate.

Endnote software (Clarivate Plc) [37] was used to manage references, and Covidence software (Veritas Health Innovation Ltd) was used to import and extract studies [38]. Two reviewers (MM and DM) independently applied the revised Cochrane risk of bias tool (The Cochrane Collaboration) [39] to establish the quality of the included studies. A matrix was developed by the authors and applied in the collection and analysis of structured data. Matrix criteria also included whether a study adhered to the CONSORT (Consolidated Standards of Reporting Trials) eHealth Checklist [40], a tool developed to improve the standard of reporting in digital health trials.

Quality Assessment

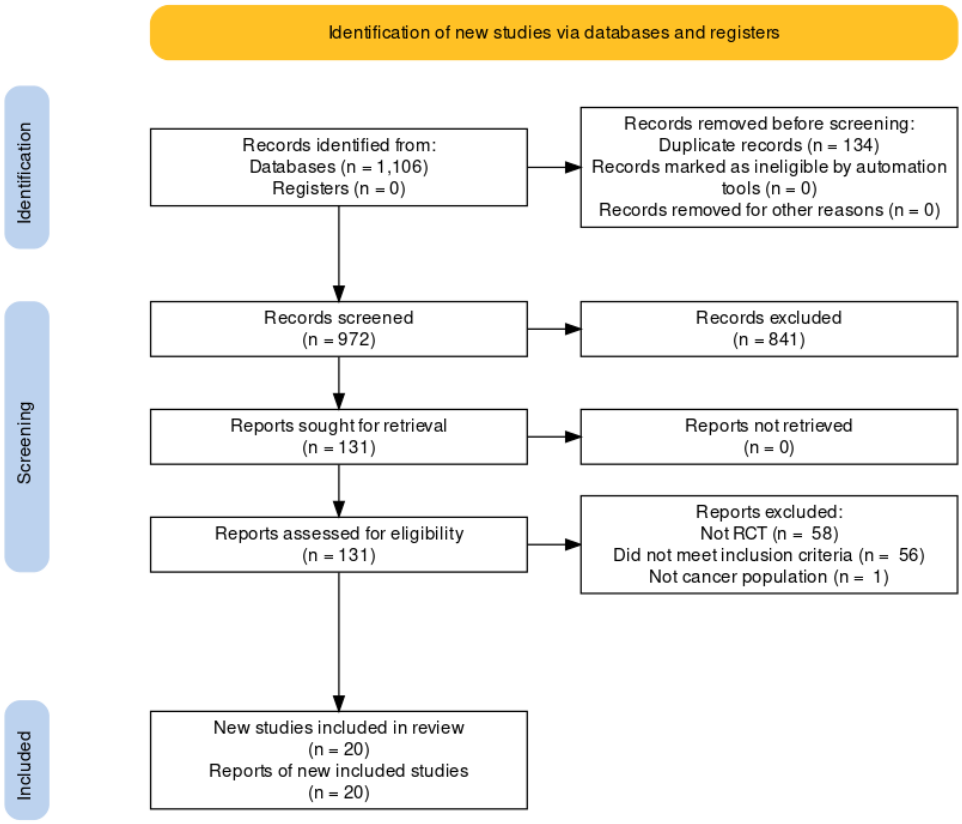
Using the revised Cochrane Collaboration's tool for assessing risk of bias (RoB) in randomized trials (Risk of Bias 2.0) [39], each study's methodological quality was assessed in 5 major domains: randomization process, deviation from intentional interventions, missing outcomes, measurement of outcomes, and selection of reported results. The RoB for each domain was rated as *some concerns*, *low*, or *high*. The overall RoB for each study was rated as *some concerns* or *high*.

Results

Study Selection

The initial search generated 972 records. After title and abstract review, 135 publications were retained. The full text for each of these 135 publications was reviewed for eligibility, resulting in the identification of 17 publications. Three additional studies were found using a reference search, generating a total of 20 digital health–enabled supportive cancer care interventions for inclusion in the review (Figure 1).

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



Study Characteristics

Population

Details of the study design and characteristics are given in Multimedia Appendix 2 [41-60]. Although the search was published in 2000, the earliest result identified was published in 2007, and the remaining included studies were published after 2009. A total of 20 studies were conducted across 8 countries [41-48], with 11 from the United States. The sample sizes ranged from 52 to 516 participants, with a median of 118.

Of the 20 included studies, 12 (60%) were designed for participants within a single tumor stream. Interventions for breast cancer were dominant, with a total of 9 studies [43,46-53]. A total of 2 studies identified were for lung cancer [54,55]. Of those targeting multiple tumor types, 1 recruited participants with breast or prostate cancer [45], and 1 recruited participants with lung, breast, or colorectal cancer [44]. The remaining 8 studies included several tumor types, such as participants living with any type or stage of cancer [42,56-58] or those with solid tumors attending ambulatory oncology clinics for chemotherapy [59,60].

Intervention Design and Features

The duration of interventions ranged from 4 weeks to 12 months, with a variable frequency of clinician-patient interactions ranging from biweekly to every 3 months. There were 5 studies involving the use of a web-based portal or web-based experience [45,46,48,55,57]; 8 studies included the use of a telephone or smartphone [43,44,47,52-54,59,60]; 2 studies used a combination of web-based and telephone interactions [51,58]; 2 studies used social media networks or social networks [51,55]; and 4 studies used wearable activity trackers [42,51-53]. Multidisciplinary care was identified in 30% (6/20) of the publications [44,46,47,55,56,58]. The study by Børøsund et al [46] was nurse-led, with referrals to either physicians or social workers. Uni-disciplinary interactions dominated, with 5 nurse-led [45,48,50,56,60], 1 social worker-led [49], and 1 led with a medical specialist experienced in mindfulness program delivery [42]. Bruggeman-Everts et al [41] involved a psychologist or physiotherapist assigned to participants in different arms of the study. Steel et al [57] outlined a collaborative care intervention, whereby a care coordinator provided information to the patient's medical team, as well as patients and caregivers. The interventions included digital health tutoring, psychotherapy, nursing support, remote exercise or rehabilitation program delivery, and digital mindfulness.

Digital supportive care interventions included interactive voice response, tele and video counseling, internet-based patient-provider communication, exercise based on the internet, support systems, symptom monitoring, and self-management, mobile phone-based remote monitoring, and activity monitoring with tracking devices. The programs included digital health tutoring, psychotherapy, nursing support, remote exercise, rehabilitation program delivery, and digital mindfulness interventions. All varied in terms of design, features, and use of multimedia components. Only 2 publications referred to the CONSORT Digital Health Checklist [46,55].

RoB Assessment

With regard to the overall RoB, no included studies were rated as *low* overall RoB; instead, 8 had *some concerns*, and 12 were *high risk*. A summary of the RoB assessment can be found in Multimedia Appendix 3 [41-60], describing the methodological quality of each domain according to the Cochrane tool for assessing RoB in randomized trials (RoB 2.0). Figures 2 and 3 describe a graphical representation of the RoB assessments. Figure 2 includes studies in which an intention-to-treat analysis was performed, whereas Figure 3 contains studies with a per-protocol analysis.

Figure 2. Risk of bias in studies with intention-to-treat analysis.

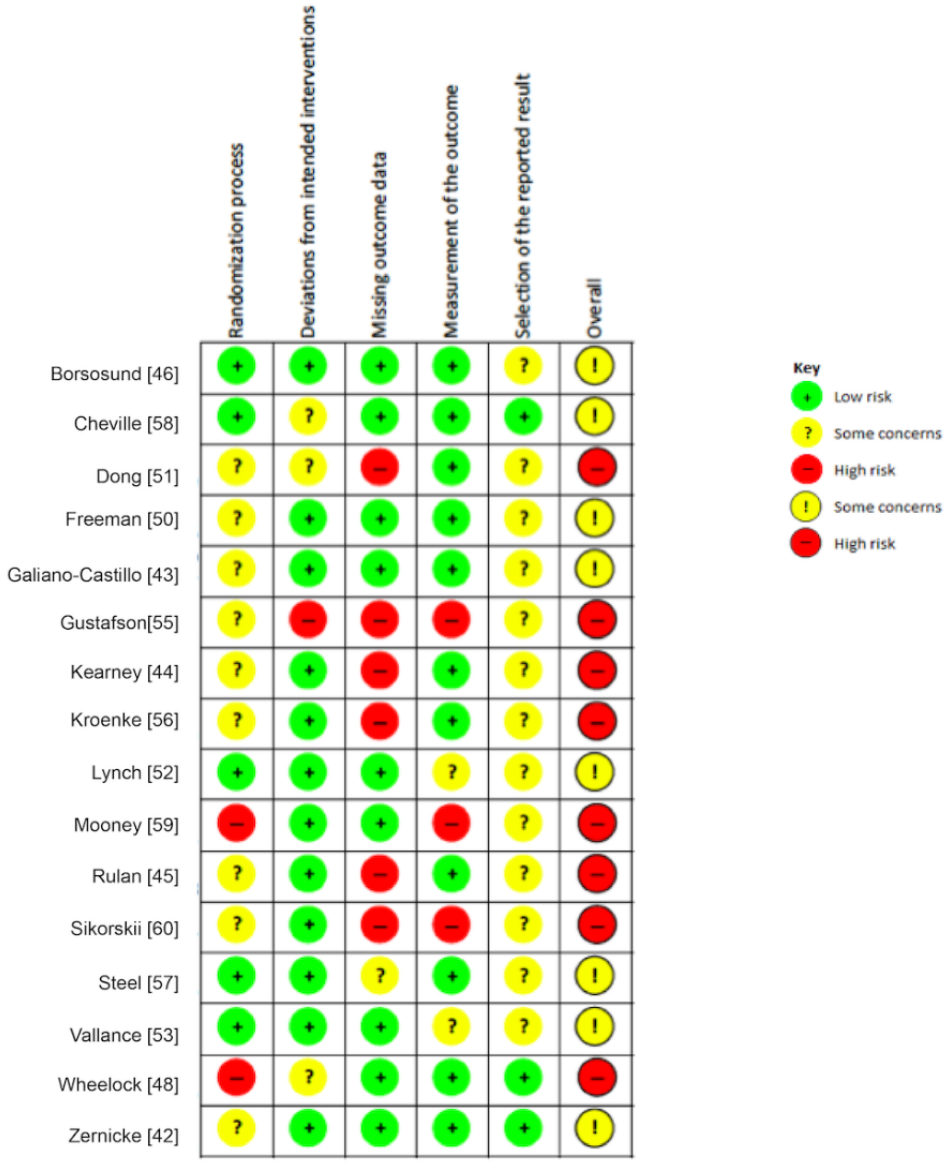
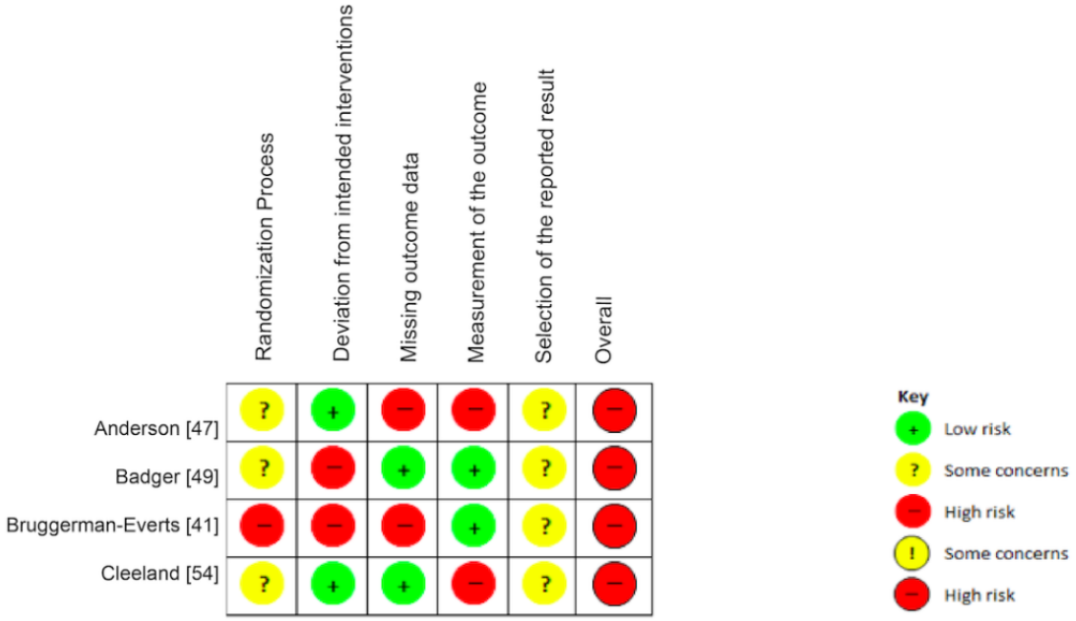


Figure 3. Risk of bias in studies with per-protocol analysis.



Intervention Outcomes

Overview

Interventions were analyzed according to the before-after test design, with most of the interventions including repeated measurement points. Although 20 studies were included, the interventions and outcomes were heterogeneous and did not enable meta-analyses. Outcomes were synthetized using a model for quality of life among cancer survivors developed by Ferrell and Dow [61,62]. The model encompasses dimensions of physical, psychological, social, and spiritual well-being, specifying the content for each dimension in the context of cancer survivors. Details of the intervention outcomes and statistically significant results are presented in Table 1.

Table 1. Intervention outcomes.

Author	Primary and secondary outcomes (grouped)	Measures	Results
Anderson et al [47]	• Pain • Sleep • Fatigue	• MDASI[a]	• Decreased pain severity from baseline to time point 1 (0.6 vs 2.3; *P*=.03; 95% CI 0.13 to 3.3) and from baseline to time point 2 (1.2 vs 3.5; *P*=.02; 95% CI 0.47 to 4.2) in the intervention group. • Improved reported sleep.
Badger et al [49]	• Depression • Symptom distress • Social well-being; spiritual well-being	• CES-D[b]; GSDS[c]; Social and spiritual; Well-being scales	• Depression, symptoms, and spiritual well-being improved in intervention groups (*P*=.01). No between-group differences. Social well-being improved for tele and video groups.
Børøsund et al [46]	• Symptom distress • Anxiety • Depression • Self-efficacy	• MSAS[d] • HADS[e] • CBI[f]	• WebChoice lower symptom distress (−0.16, 95% CI −.25 to −0.06; *P*=.001), anxiety (−0.79, 95% CI −1.49 to −0.09; *P*=.03), and depression (−0.79, 95% CI 1.18 to −0.05; *P*=.03) compared with control. • Internet-based communication group lower depression (−0.69, 95% CI −1.32 to −0.05; *P*=.03) compared with usual care. No change in symptom distress or anxiety.
Bruggeman-Everts et al [41]	• Fatigue severity • Mental health • Distress	• CIS-FS[g] • The positive and negative affect schedule • HADS	• Clinically changes in fatigue severity in 66% (41/62) of patients in home-based physiotherapist guided protocol (AAF[h]), 49% (27/55) of patients in web-based psychologist-guided intervention (eMBCT[i]), and 12% (6/50) of patients in psycho-education e-mails only.
Cheville et al [58]	• Functional capacity • Pain • HRQoL[j]	• AM-PAC[k] • BPI[l] • 5-item EQ-5D-3L	• Telerehabilitation by physical therapist-physician team (intervention group 1) improved function (difference, 1.3; 95% CI 0.08 to 2.35; *P*=.03) and quality of life (difference, 0.04; 95% CI 0.004 to 0.071; *P*=.01) compared with control. • Intervention groups 1 and 2 showed reduced pain interference and average intensity (intervention group 1, −0.4; 95% CI −0.78 to −0.07; *P*=.02; and intervention group 2, −0.5; 95% CI −0.84 to −0.11; *P*=.006).
Cleeland et al [54]	• Symptom threshold events • Cumulative distribution of symptom threshold events • Symptom severity	• MDASI	• Both groups had decreased symptom threshold events, control group reported more events at the end of the study period. For both groups together, the effect size of reduction in symptom severity was 0.72, effect size of 0.68 in the control group and 0.75 in the intervention group.

Author	Primary and secondary outcomes (grouped)	Measures	Results
Dong et al [51]	• HRQoL • Muscle strength • Cardiorespiratory capacity	• SF-36[m] • Stand-up or sit-down chair test and arm lifting test (30 seconds) • Modified Bruce treadmill protocol	• CEIBISMS[n] intervention showed improvements after 12 weeks in role-physical (P=.009), general health (P=.02), mental health (P=.01), vitality (P=.01) and health transition (P=.007). • In comparison with control group, differences in vitality (P=.009), mental health (P=.001), and health transition (P=.048).
Freeman et al [50]	• HRQoL • Functional capacity • Fatigue, sleep • Spiritual well-being	• SF-36 • FACT-B[o] • FACIT-F[p] and cog • FACT-Sp[q]	• Less fatigue (P=.002), cognitive dysfunction (P=.001), and sleep disturbance (P<.001) for both intervention groups compared with control. • No differences between live delivery and telemedicine delivery of therapy. • No group effect on overall quality of life; however, there was a time effect.
Galiano-Castillo et al [43]	• Functional capacity • Cognitive function	• 6-minute walk test • Trail making test • ACT[r]	• After intervention, the telerehabilitation group had significantly improved distances as well as percentage of predicted 6-minute walk test compared with the control group (P<.001).
Gustafson et al [55]	• Caregiver surveys reporting patient symptom distress	• Modified ESAS[s]	• Caregivers in the CHESS[t] arm consistently reported lower patient physical symptom distress than caregivers in the internet arm at 4 months (P=.03); and at 6 months (P=.004)
Kearney et al [44]	• 6 chemotherapy-related symptoms	• Common toxicity criteria adverse events chemotherapy symptom assessment scale	• Difference between groups in fatigue, higher in the control group (OR[u] 2.29, 95% CI 1.04 to 5.05; P=.04) and in hand-foot syndrome lower in control group (OR control or intervention 0.39, 95% CI 0.17 to 0.92; P=.03)
Kroenke et al [56]	• Depression • Pain severity	• HSCL-20[v] • BPI	• Improvements for the intervention group: >30% decrease in pain index (P<.001) and >50% decrease in the depression scale (P<.001). Effect size between-group differences at 3 months was 0.67 (95% CI 0.33 to 1.02) for pain and 0.42 (95% CI 0.16 to 0.69) for depression. • Intervention group had better outcomes for several HrQoL domains, including mental health, vitality, anxiety, and physical symptom burden.
Lynch [52][w]	• Vigorous physical activity (MVPA[x])	• Actigraph and activPAL accelerometers	• Between-group difference in MVPA at T2 (69 min/week; 95% CI 22 to 116); decreased total time of sitting 37 min/day (95% CI −72 to −2) and prolonged bouts of sitting 42 min/day (95% CI −83 to −2), favoring the intervention group
Vallance [53][w]			

Author	Primary and secondary outcomes (grouped)	Measures	Results
	• HRQoL • Fatigue	• FACT-B • FACIT-F	• Intervention group improvement in fatigue at T2 4.6 (95% CI 1.3 to 7.8). Within groups: intervention group, increase in fatigue at T2 5.1 (95% CI 2.0 to 8.2) and at T-3 3.3 (95% CI 0.1 to 6.41). No effects on HRQoL.
Mooney et al [59]	• Symptom severity • Distress	• Single item scale	• No significant difference between symptom severity or distress scores between groups.
Ruland et al [45]	• Symptom distress • Depression • Self-efficacy • HRQoL	• MSAS-SF[y] • Centre for Epidemiological Cancer Behavior Inventory Studies-Depression Scale 15 d	• Decreased distress on one subscale of MSAS. • Group differences on symptom distress were significant for the MSAS-SF (slope estimate, −0.052, 95% CI −0.101 to −0.004; t244=4.42; *P*=.04). There were no significant within- or between-group differences on the other MSAS-SF subscales.
Sikorskii et al [60]	• Symptom severity	• MDASI	• Decreased symptom severity across both intervention groups after 10 weeks. No between-group differences. Effect sizes were similar for NASM[z] (0.56) and ATSM[aa] (0.59)
Steel et al [57]	• Depression • Pain • Serum cytokine levels natural killer cell numbers	• CES-D • BPI • Functional assessment of cancer therapy–anemia, and hepatobiliary	• Reductions in pain (Cohen d=0.62), fatigue (Cohen d=0.26), depression (Cohen d=0.71), and significant changes in HRQoL with an effect size of Cohen d=0.99 at 6 months follow-up (*P*=.05) when compared with those in the enhanced usual care arm at 6 months.
Wheelock et al [48]	• Time between symptoms • Health care use	• Clinic visits • Health service use	• Did not meet primary objective, no difference in health service use
Zernicke et al [42]	• Feasibility • Mood • Stress • Posttraumatic growth inventory	• Monitoring interest, eligibility, and participation • Profile of mood states • CSOSI[ab]	• Significant improvements and moderate effect sizes in the web-based MBCR[ac] group relative to controls for mood disturbance (Cohen d=0.44; *P*=.049), stress (Cohen d=0.49; *P*=.02), spirituality (Cohen d=0.37; *P*=.04), and mindfully acting with awareness (Cohen d=0.50; *P*=.03).

[a]MDASI: MD Anderson symptom inventory.

[b]CES-D: Center for Epidemiological Studies-Depression.

[c]GSDS: General Sleep Disturbance Scale.

[d]MSAS: Memorial Symptom Assessment Scale.

[e]HADS: Hospital Anxiety And Depression Scale.

[f]CBI: Cancer Behavioral Inventory.

[g]CIS-FS: Checklist Individual Strength-Fatigue Severity.

[h]AAF: Ambulant Activity Feedback

[i]eMBCT: Web-based Mindfulness-Based Cognitive Therapy.

[j]HRQoL: health-related quality of life.

[k]AM-PAC: Activity Measure for Postacute Care.

[l]BPI: Brief Pain Inventory.

[m]SF-36: 36-item Short Form Health Survey.

[n]CEIBISMS: combined exercise intervention based on internet and social media software.

[o]FACT-B: Functional Assessment of Cancer Therapy–Breast.

[p]FACIT-F: Functional Assessment of Chronic Illness Therapy–Fatigue.

[q]FACT-Sp: Functional Assessment of Chronic Illness Therapy-Spiritual Well-Being.

[r]ACT: acceptance and commitment therapy.

[s]ESAS: Edmonton Symptom Assessment Score.

[t]CHESS: Comprehensive Health Enhancement Support System.

[u]OR: odds ratio.

[v]HSCL-20: Hopkins Symptom Checklist Depression Scale.

[w]Lynch [52] and Vallance [53] are 2 publications with different outcomes of the same randomized controlled trial.

[x]MVPA: moderate-to-vigorous intensity physical activity.

[y]MSAS-SF: Memorial Symptom Assessment Scale – Short Form

[z]NASM: nurse-assisted symptom management

[aa]ATSM: automated telephone symptom management

[ab]CSOSI: Calgary Symptoms of Stress Inventory.

[ac]MBCR: mindfulness-based cancer recovery.

Physical Well-being

Despite measuring similar outcomes, heterogeneous self-reported instruments were used across the studies. Control or reduction of symptoms and maintenance of function and independence comprise this domain. A total of 16 studies reported statistically significant changes in outcomes within the physical well-being domain [41,43-47,49,50,52-58,60]. Furthermore, 4 studies showed a significant reduction in pain [47,56-58], of which 3 used the Brief Pain Inventory [56-58]. Decreased fatigue was reported in 4 studies, favoring the intervention groups [41,44,50,53]. The most common outcome measure reported was clustered symptom changes, as referenced in 6 studies [45,46,49,54,55,60]. Finally, functional capacity-related outcomes were significant in 3 studies [43,52,58].

Health-Related Quality of Life

Health-related quality of life was the second most common domain, with significant improvements for the intervention group referenced in 4 studies [51,56-58].

Psychological Well-being

The most common outcome in this domain reporting significant improvements in the 4 interventions was depression [46,49,56,57]. Anxiety was reduced compared with the control group in one study [46], and significant improvements in mood disturbance were observed in one study [42].

Social and Spiritual Well-being

A single study reported a significant improvement in social well-being in the intervention group [49]. Badger et al [49] was also the only study that evaluated spiritual well-being changes, resulting in statistically significant improvements for the intervention group but no between-group differences.

Discussion

Principal Findings

Overview

In the studies integrated in this review, the interventions included digital health education, psychotherapy, nursing support, remote exercise, rehabilitation program delivery, and digital mindfulness interventions. All interventions satisfied the domains *health in our hands* and *interacting for health* of the digital health model by Shaw et al [13]. Digital supportive care interventions have been shown to improve cancer-related symptoms [45,46,49,54,55,60], pain [47,56-58], fatigue [41,44,50,53], health-related quality of life [51,56-58], functional capacity [43,52,58], and depression [46,49,56,57]. Only 2 RCTs included in this review did not report significant changes in one or more outcomes [48,59].

Digital Supportive Cancer Care Interventions

The digital interventions reviewed have been shown to be beneficial and independent of disease and demographic factors. This is similar to findings reported in other reviews [20,31,32,35,47,63] and meta-analyses [34]. In addition, the use of technology for cancer follow-up appears to be acceptable to patients, is clinically safe [31], and improves health knowledge and self-management practices [64]. However, such interventions vary in design and features, most lack or fail to report theoretical frameworks, and they use outcome measures making pooling or comparison between studies difficult. Another issue when comparing digital supportive care interventions is that it may be possible that interventions vary in their efficacy across different populations and technologies used for delivery. Furthermore, past studies have reported potential challenges impacting the implementation of digital health care, such as technical problems, lack of technology knowledge, and data security [65], which need to be considered when planning future studies. Recent systematic reviews concluded that a range of strategies should be implemented in digital supportive care [20,35,63] and general digital health interventions [66]. The study by O'Connor et al [66] recommends increasing public awareness of different

technologies and understanding of how they work, personalization of care, clinical accreditation of interventions, improving focus on health literacy, and safeguarding privacy of personal information as key areas for investigation. Key areas for digital supportive care design and implementation noted in previous cancer-related reviews that should be further explored are mechanisms for participant feedback to drive the co-design of digital interventions [20,67], the efficiency of delivering relevant and tailored health care information [65], and ways to integrate supportive care services at all stages of the cancer treatment pathway [35].

A supportive digital care intervention model should be underpinned by a theoretical framework that anticipates not only the outcomes and the tools to measure these, but also the process of achieving the outcomes from a particular intervention [20,63,68]. However, many studies do not address the validity of patient-reported health outcomes, and most of them use self-reported measures in pre-post test design, which leaves them at a RoB. This might be in part because one of the main challenges in the development of an evidence-based digital supportive cancer care intervention is the velocity of technology development in comparison with the often-long process of conducting and evaluating clinical trials.

Digital Health for Chronic Disease Care

Evidence from the research and implementation of digital health interventions across other disease groups may facilitate the transferability of digitally enabled supportive cancer care. Cancer is a chronic disease for many people [69]. Applications of research findings in other chronic diseases, such as cardiovascular disease [70], hypertension [12], and diabetes [71], which have a larger evidence base in digital health–enabled interventions with positive effects, should be used where possible. A recent systematic review [72] focused on the broader application of these symptom-reporting systems within multiple patient groups and concluded that although further research needs to be completed, most studies reported positive health outcomes. For example, in the case of diabetes, Greenwood et al [71] found that the most effective digital interventions incorporated all components of a technology-assisted self-management feedback loop, connected people with diabetes and their health care team using two-way communication, analyzed patient-generated health data, tailored education programs, and individualized feedback. A 2018 systematic review focused on using remote monitoring in people with a history of type 2 diabetes [73] significantly improved glycated hemoglobin and self-management. Evidence from diabetes research seems more cohesive, in part, because the outcomes of lowering glucose levels and glycated hemoglobin allow homogeneous measurement across studies.

Several successful digital health interventions focus on both behavior change and increasing patient engagement [74,75]. Barello [75] concluded that most studies failed to account for the complexity of patient engagement and that a more holistic approach might help maximize the potential of digital health technology [75]. Another recent review focused on mobile health apps for chronic disease management and found that regular symptom assessments, automated reminders, and feedback loops were common features, with most studies reporting significant improvement in health outcomes [73].

Current Challenges and Need for Quality Information

This review compiles evidence regarding the potential of digital health interventions for supportive cancer care in different settings, including remote areas and emergency situations [33]. However, the challenges facing public health systems worldwide in terms of emergencies, such as the COVID-19 pandemic, have rapidly increased the use of digital health interventions [29]. Health care systems, including cancer care, are adapting in response to the need for social distancing, lockdowns, and other public health initiatives. Cancer clinics have reduced clinical appointments, administration encounters, and postponed elective cancer surgeries [28], which has increased the need for follow-up and management without visiting hospitals [76]. This situation has advanced the use of digital health and telehealth apps and programs worldwide [30]. Digital supportive cancer care has been implemented out of necessity and is becoming a common delivery model [77]. Two of the suggested strategies to enable supportive cancer care during COVID-19 are (1) empowering patients and caregivers through the use of digital communication and (2) increasing the use of existing digital health platforms [28].

There is an urgent need to agree on relevant outcomes, methods of assessment, and there is a need for improved quality of primary studies and RCTs, as shown by our quality assessment. The lack of high-quality randomized trials identified in this review reflects the ongoing problem of low-quality research. Moreover, only 2 of the included publications referenced the CONSORT Digital Health Checklist, which was published in 2011. Only 3 included studies were published before 2011, making it disappointing that most later publications failed to reference this standard. Comments of Dickinson in BMC Cancer [31] remain relevant in 2020:

> *Nevertheless, there are surprisingly few randomized trials given the explosion in technological innovation in recent years. It could be that technology is evolving so fast that potential innovative technological interventions become outdated before they can mature sufficiently to be subjected to randomized trials.*

This insight has been recurrent in different systematic reviews [32,74].

Strengths and Limitations

Although the review was a rigorous evaluation of RCTs, there were a small number of included studies that indicated there may be significant literature in the phase I or II feasibility spectrum. Although 20 studies were included, the interventions and outcomes were heterogeneous and did not enable meta-analyses. Owing to the nature of this review, there was also a heterogeneous population, variable outcome measures, variable study quality, and methodological limitations. These characteristics have been found to contribute to a lack of evidence regarding the benefits of digital health [19]. It is difficult to draw conclusions and synthesize studies with inconsistent outcome measures, and a systematic approach to using standardized measures is required. The CONSORT Digital

Health Checklist should be used routinely. The review was limited to studies written in English. Therefore, it is possible that research papers published in other parts of the world were missed. As with many studies in oncology, this review found that breast cancer survivorship dominated. Thus, the development of supportive care interventions across other tumor streams is required. As most of the included studies were conducted in the United States and Europe, it is unclear whether the findings from these studies can be generalized to other countries and populations, particularly in developing nations.

Future Directions

Living well with cancer has gained greater relevance as the survival rates of many cancer types are increasing. The future of digital health in oncology supportive care brings a range of new and exciting possibilities. There is a need to evaluate the efficacy and efficiency of digital interventions in real-world conditions and standardize a core set of outcomes included in all studies to facilitate comparisons between interventions and digital technologies.

Conclusions

Digital health–enabled supportive cancer care is capable of improving health-related quality of life, symptom burden including self-report of pain and fatigue, depression, and, to a lesser extent, functional capacity. Supportive digital interventions in the field of cancer are being used and have been reported to be helpful for patients, independent of other factors. However, there is a need for higher quality research and clearer reporting than is evident in the current RCTs. Future research should focus on using valid, standardized outcome measures, increasing the methodological rigor of studies undertaken, and the development and evaluation of strategies to improve both patient and health professional engagement in the design and delivery of supportive digital health interventions.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 52 KB - jmir_v23i10e24722_app1.docx]

Multimedia Appendix 2
Population and intervention characteristics.
[DOCX File , 35 KB - jmir_v23i10e24722_app2.docx]

Multimedia Appendix 3
Summary of risk of bias assessment.
[DOCX File , 96 KB - jmir_v23i10e24722_app3.docx]

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394-424 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
2. Bluethmann SM, Mariotto AB, Rowland JH. Anticipating the "Silver Tsunami": prevalence trajectories and comorbidity burden among older cancer survivors in the United States. Cancer Epidemiol Biomarkers Prev 2016 Jul;25(7):1029-1036 [FREE Full text] [doi: 10.1158/1055-9965.EPI-16-0133] [Medline: 27371756]
3. Hui D, De La Cruz M, Mori M, Parsons HA, Kwon JH, Torres-Vigil I, et al. Concepts and definitions for "supportive care," "best supportive care," "palliative care," and "hospice care" in the published literature, dictionaries, and textbooks. Support Care Cancer 2013 Mar;21(3):659-685 [FREE Full text] [doi: 10.1007/s00520-012-1564-y] [Medline: 22936493]
4. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. J Am Med Assoc 2017 Jul 11;318(2):197-198 [FREE Full text] [doi: 10.1001/jama.2017.7156] [Medline: 28586821]
5. Olver IN. The importance of supportive care for patients with cancer. Med J Aust 2016 Jun 20;204(11):401-402. [doi: 10.5694/mja16.00279] [Medline: 27318394]
6. Rubin G, Berendsen A, Crawford SM, Dommett R, Earle C, Emery J, et al. The expanding role of primary care in cancer control. Lancet Oncol 2015 Sep;16(12):1231-1272. [doi: 10.1016/S1470-2045(15)00205-3] [Medline: 26431866]
7. Lisy K, Kent J, Piper A, Jefford M. Facilitators and barriers to shared primary and specialist cancer care: a systematic review. Support Care Cancer 2021 Jan;29(1):85-96. [doi: 10.1007/s00520-020-05624-5] [Medline: 32803729]
8. Nasi G, Cucciniello M, Guerrazzi C. The performance of mHealth in cancer supportive care: a research agenda. J Med Internet Res 2015 Feb 13;17(1):e9 [FREE Full text] [doi: 10.2196/jmir.3764] [Medline: 25720295]
9. Rooij TV, Marsh S. eHealth: past and future perspectives. Per Med 2016 Jan;13(1):57-70. [doi: 10.2217/pme.15.40] [Medline: 29749870]

10. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
11. Elbert NJ, van Os-Medendorp H, van Renselaar W, Ekeland AG, van Roijen LH, Raat H, et al. Effectiveness and cost-effectiveness of ehealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014 Apr 16;16(4):e110 [FREE Full text] [doi: 10.2196/jmir.2790] [Medline: 24739471]
12. McLean G, Band R, Saunderson K, Hanlon P, Murray E, Little P, DIPSS co-investigators. Digital interventions to promote self-management in adults with hypertension systematic review and meta-analysis. J Hypertens 2016 Apr;34(4):600-612 [FREE Full text] [doi: 10.1097/HJH.0000000000000859] [Medline: 26845284]
13. Shaw T, McGregor D, Brunner M, Keep M, Janssen A, Barnet S. What is eHealth (6)? Development of a conceptual model for eHealth: qualitative study with key informants. J Med Internet Res 2017 Oct 24;19(10):e324 [FREE Full text] [doi: 10.2196/jmir.8106] [Medline: 29066429]
14. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. Int J Med Inform 2010 Nov;79(11):736-771. [doi: 10.1016/j.ijmedinf.2010.08.006] [Medline: 20884286]
15. Zanaboni P, Ngangue P, Mbemba GI, Schopf TR, Bergmo TS, Gagnon M. Methods to evaluate the effects of internet-based digital health interventions for citizens: systematic review of reviews. J Med Internet Res 2018 Jun 07;20(6):e10202 [FREE Full text] [doi: 10.2196/10202] [Medline: 29880470]
16. Eland-de Kok P, van Os-Medendorp H, Vergouwe-Meijer A, Bruijnzeel-Koomen C, Ros W. A systematic review of the effects of e-health on chronically ill patients. J Clin Nurs 2011 Nov;20(21-22):2997-3010. [doi: 10.1111/j.1365-2702.2011.03743.x] [Medline: 21707807]
17. Kuijpers W, Groen WG, Aaronson NK, van Harten WH. A systematic review of web-based interventions for patient empowerment and physical activity in chronic diseases: relevance for cancer survivors. J Med Internet Res 2013 Feb 20;15(2):e37 [FREE Full text] [doi: 10.2196/jmir.2281] [Medline: 23425685]
18. Seiler A, Klaas V, Tröster G, Fagundes CP. eHealth and mHealth interventions in the treatment of fatigued cancer survivors: a systematic review and meta-analysis. Psychooncology 2017 Sep;26(9):1239-1253. [doi: 10.1002/pon.4489] [Medline: 28665554]
19. Roberts AL, Fisher A, Smith L, Heinrich M, Potts HW. Digital health behaviour change interventions targeting physical activity and diet in cancer survivors: a systematic review and meta-analysis. J Cancer Surviv 2017 Dec;11(6):704-719 [FREE Full text] [doi: 10.1007/s11764-017-0632-1] [Medline: 28779220]
20. Ventura F, Ohlén J, Koinberg I. An integrative review of supportive e-health programs in cancer care. Eur J Oncol Nurs 2013 Aug;17(4):498-507. [doi: 10.1016/j.ejon.2012.10.007] [Medline: 23158437]
21. Jansen F, van Uden-Kraan CF, van Zwieten V, Witte BI, Verdonck-de Leeuw IM. Cancer survivors' perceived need for supportive care and their attitude towards self-management and eHealth. Support Care Cancer 2015 Jun;23(6):1679-1688. [doi: 10.1007/s00520-014-2514-7] [Medline: 25424520]
22. Butow PN, Phillips F, Schweder J, White K, Underhill C, Goldstein D, Clinical Oncological Society of Australia. Psychosocial well-being and supportive care needs of cancer patients living in urban and rural/regional areas: a systematic review. Support Care Cancer 2012 Jan;20(1):1-22. [doi: 10.1007/s00520-011-1270-1] [Medline: 21956760]
23. Bender JL, Wiljer D, To MJ, Bedard PL, Chung P, Jewett MA, et al. Testicular cancer survivors' supportive care needs and use of online support: a cross-sectional survey. Support Care Cancer 2012 Nov;20(11):2737-2746. [doi: 10.1007/s00520-012-1395-x] [Medline: 22382589]
24. Aaronson NK, Mattioli V, Minton O, Weis J, Johansen C, Dalton SO, et al. Beyond treatment - psychosocial and behavioural issues in cancer survivorship research and practice. EJC Suppl 2014 Jun;12(1):54-64 [FREE Full text] [doi: 10.1016/j.ejcsup.2014.03.005] [Medline: 26217166]
25. Mehnert A. Employment and work-related issues in cancer survivors. Crit Rev Oncol Hematol 2011 Feb;77(2):109-130. [doi: 10.1016/j.critrevonc.2010.01.004] [Medline: 20117019]
26. Lubberding S, van Uden-Kraan CF, Te Velde EA, Cuijpers P, Leemans CR, Verdonck-de Leeuw IM. Improving access to supportive cancer care through an eHealth application: a qualitative needs assessment among cancer survivors. J Clin Nurs 2015 May;24(9-10):1367-1379. [doi: 10.1111/jocn.12753] [Medline: 25677218]
27. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital health strategies to fight COVID-19 worldwide: challenges, recommendations, and a call for papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
28. Salako O, Okunade K, Allsop M, Habeebu M, Toye M, Oluyede G, et al. Upheaval in cancer care during the COVID-19 outbreak. Ecancermedicalscience 2020 Apr 1;14:ed97 [FREE Full text] [doi: 10.3332/ecancer.2020.ed97] [Medline: 32269597]
29. Spicer J, Chamberlain C, Papa S. Provision of cancer care during the COVID-19 pandemic. Nat Rev Clin Oncol 2020 Jun;17(6):329-331 [FREE Full text] [doi: 10.1038/s41571-020-0370-6] [Medline: 32296166]
30. Young AM, Ashbury FD, Schapira L, Scotté F, Ripamonti CI, Olver IN. Uncertainty upon uncertainty: supportive Care for Cancer and COVID-19. Support Care Cancer 2020 Sep;28(9):4001-4004 [FREE Full text] [doi: 10.1007/s00520-020-05604-9] [Medline: 32613372]

31. Dickinson R, Hall S, Sinclair JE, Bond C, Murchie P. Using technology to deliver cancer follow-up: a systematic review. BMC Cancer 2014 May 03;14:311 [FREE Full text] [doi: 10.1186/1471-2407-14-311] [Medline: 24885758]
32. Agboola SO, Ju W, Elfiky A, Kvedar JC, Jethwani K. The effect of technology-based interventions on pain, depression, and quality of life in patients with cancer: a systematic review of randomized controlled trials. J Med Internet Res 2015 Mar 13;17(3):e65 [FREE Full text] [doi: 10.2196/jmir.4009] [Medline: 25793945]
33. Srivastava S, Pant M, Abraham A, Agrawal N. The technological growth in eHealth services. Comput Math Methods Med 2015;2015:894171 [FREE Full text] [doi: 10.1155/2015/894171] [Medline: 26146515]
34. Larson JL, Rosen AB, Wilson FA. The effect of telehealth interventions on quality of life of cancer patients: a systematic review and meta-analysis. Telemed J E Health 2018 Jun;24(6):397-405. [doi: 10.1089/tmj.2017.0112] [Medline: 29112484]
35. Aapro M, Bossi P, Dasari A, Fallowfield L, Gascón P, Geller M, et al. Digital health for optimal supportive care in oncology: benefits, limits, and future perspectives. Support Care Cancer 2020 Oct;28(10):4589-4612 [FREE Full text] [doi: 10.1007/s00520-020-05539-1] [Medline: 32533435]
36. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
37. EndNote homepage. EndNote. URL: https://endnote.com/ [accessed 2021-10-05]
38. Better systematic review management. Covidence. URL: https://www.covidence.org/ [accessed 2021-10-05]
39. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. Br Med J 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
40. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
41. Bruggeman-Everts FZ, Wolvers MD, van de Schoot R, Vollenbroek-Hutten MM, Van der Lee ML. Effectiveness of two web-based interventions for chronic cancer-related fatigue compared to an active control condition: results of the "Fitter na kanker" randomized controlled trial. J Med Internet Res 2017 Oct 19;19(10):e336 [FREE Full text] [doi: 10.2196/jmir.7180] [Medline: 29051138]
42. Zernicke KA, Campbell TS, Speca M, McCabe-Ruff K, Flowers S, Carlson LE. A randomized wait-list controlled trial of feasibility and efficacy of an online mindfulness-based cancer recovery program: the eTherapy for cancer applying mindfulness trial. Psychosom Med 2014 May;76(4):257-267. [doi: 10.1097/PSY.0000000000000053] [Medline: 24804884]
43. Galiano-Castillo N, Arroyo-Morales M, Lozano-Lozano M, Fernández-Lao C, Martín-Martín L, Del-Moral-Ávila R, et al. Effect of an internet-based telehealth system on functional capacity and cognition in breast cancer survivors: a secondary analysis of a randomized controlled trial. Support Care Cancer 2017 Nov;25(11):3551-3559. [doi: 10.1007/s00520-017-3782-9] [Medline: 28639097]
44. Kearney N, McCann L, Norrie J, Taylor L, Gray P, McGee-Lennon M, et al. Evaluation of a mobile phone-based, advanced symptom management system (ASyMS) in the management of chemotherapy-related toxicity. Support Care Cancer 2009 Apr;17(4):437-444. [doi: 10.1007/s00520-008-0515-0] [Medline: 18953579]
45. Ruland CM, Andersen T, Jeneson A, Moore S, Grimsbø GH, Børøsund E, et al. Effects of an internet support system to assist cancer patients in reducing symptom distress: a randomized controlled trial. Cancer Nurs 2013;36(1):6-17. [doi: 10.1097/NCC.0b013e31824d90d4] [Medline: 22495503]
46. Børøsund E, Cvancarova M, Moore SM, Ekstedt M, Ruland CM. Comparing effects in regular practice of e-communication and web-based self-management support among breast cancer patients: preliminary results from a randomized controlled trial. J Med Internet Res 2014 Dec 18;16(12):e295 [FREE Full text] [doi: 10.2196/jmir.3348] [Medline: 25525672]
47. Anderson KO, Palos GR, Mendoza TR, Cleeland CS, Liao K, Fisch MJ, et al. Automated pain intervention for underserved minority women with breast cancer. Cancer 2015 Jun 01;121(11):1882-1890 [FREE Full text] [doi: 10.1002/cncr.29204] [Medline: 25711974]
48. Wheelock AE, Bock MA, Martin EL, Hwang J, Ernest ML, Rugo HS, et al. SIS.NET: a randomized controlled trial evaluating a web-based system for symptom management after treatment of breast cancer. Cancer 2015 Mar 15;121(6):893-899 [FREE Full text] [doi: 10.1002/cncr.29088] [Medline: 25469673]
49. Badger T, Segrin C, Pasvogel A, Lopez AM. The effect of psychosocial interventions delivered by telephone and videophone on quality of life in early-stage breast cancer survivors and their supportive partners. J Telemed Telecare 2013 Jul;19(5):260-265. [doi: 10.1177/1357633X13492289] [Medline: 24163235]
50. Freeman LW, White R, Ratcliff CG, Sutton S, Stewart M, Palmer JL, et al. A randomized trial comparing live and telemedicine deliveries of an imagery-based behavioral intervention for breast cancer survivors: reducing symptoms and barriers to care. Psychooncology 2015 Aug;24(8):910-918 [FREE Full text] [doi: 10.1002/pon.3656] [Medline: 25146413]
51. Dong X, Yi X, Gao D, Gao Z, Huang S, Chao M, et al. The effects of the combined exercise intervention based on internet and social media software (CEIBISMS) on quality of life, muscle strength and cardiorespiratory capacity in Chinese postoperative breast cancer patients:a randomized controlled trial. Health Qual Life Outcomes 2019 Jun 26;17(1):109 [FREE Full text] [doi: 10.1186/s12955-019-1183-0] [Medline: 31242926]

52. Lynch BM, Nguyen NH, Moore MM, Reeves MM, Rosenberg DE, Boyle T, et al. A randomized controlled trial of a wearable technology-based intervention for increasing moderate to vigorous physical activity and reducing sedentary behavior in breast cancer survivors: the ACTIVATE Trial. Cancer 2019 Aug 15;125(16):2846-2855 [FREE Full text] [doi: 10.1002/cncr.32143] [Medline: 31012970]
53. Vallance JK, Nguyen NH, Moore MM, Reeves MM, Rosenberg DE, Boyle T, et al. Effects of the ACTIVity And TEchnology (ACTIVATE) intervention on health-related quality of life and fatigue outcomes in breast cancer survivors. Psychooncology 2020 Jan;29(1):204-211. [doi: 10.1002/pon.5298] [Medline: 31763746]
54. Cleeland CS, Wang XS, Shi Q, Mendoza TR, Wright SL, Berry MD, et al. Automated symptom alerts reduce postoperative symptom severity after cancer surgery: a randomized controlled clinical trial. J Clin Oncol 2011 Mar 10;29(8):994-1000 [FREE Full text] [doi: 10.1200/JCO.2010.29.8315] [Medline: 21282546]
55. Gustafson DH, DuBenske LL, Namkoong K, Hawkins R, Chih M, Atwood AK, et al. An eHealth system supporting palliative care for patients with non-small cell lung cancer: a randomized trial. Cancer 2013 May 01;119(9):1744-1751 [FREE Full text] [doi: 10.1002/cncr.27939] [Medline: 23355273]
56. Kroenke K, Theobald D, Wu J, Norton K, Morrison G, Carpenter J, et al. Effect of telecare management on pain and depression in patients with cancer: a randomized trial. J Am Med Assoc 2010 Jul 14;304(2):163-171 [FREE Full text] [doi: 10.1001/jama.2010.944] [Medline: 20628129]
57. Steel JL, Geller DA, Kim KH, Butterfield LH, Spring M, Grady J, et al. Web-based collaborative care intervention to manage cancer-related symptoms in the palliative care setting. Cancer 2016 Apr 15;122(8):1270-1282 [FREE Full text] [doi: 10.1002/cncr.29906] [Medline: 26970434]
58. Cheville AL, Moynihan T, Herrin J, Loprinzi C, Kroenke K. Effect of collaborative telerehabilitation on functional impairment and pain among patients with advanced-stage cancer: a randomized clinical trial. JAMA Oncol 2019 May 01;5(5):644-652 [FREE Full text] [doi: 10.1001/jamaoncol.2019.0011] [Medline: 30946436]
59. Mooney KH, Beck SL, Friedman RH, Farzanfar R, Wong B. Automated monitoring of symptoms during ambulatory chemotherapy and oncology providers' use of the information: a randomized controlled clinical trial. Support Care Cancer 2014 Sep;22(9):2343-2350 [FREE Full text] [doi: 10.1007/s00520-014-2216-1] [Medline: 24687538]
60. Sikorskii A, Given CW, Given B, Jeon S, Decker V, Decker D, et al. Symptom management for cancer patients: a trial comparing two multimodal interventions. J Pain Symptom Manage 2007 Sep;34(3):253-264 [FREE Full text] [doi: 10.1016/j.jpainsymman.2006.11.018] [Medline: 17618080]
61. Ferrell BR, Dow KH, Grant M. Measurement of the quality of life in cancer survivors. Qual Life Res 1995 Dec;4(6):523-531. [doi: 10.1007/BF00634747] [Medline: 8556012]
62. Ferrell BR, Dow KH. Quality of life among long-term cancer survivors. Oncology (Williston Park) 1997 Apr;11(4):565-8, 571; discussion 572, 575. [Medline: 9130276]
63. Laakso E, Tandy J. Use of technology as an adjunct to improve health outcomes for survivors of cancer. Physical Ther Rev 2011;16(1):39-45. [doi: 10.1179/1743288x10y.0000000017]
64. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
65. Alberts NM, Hadjistavropoulos HD, Dear BF, Titov N. Internet-delivered cognitive-behaviour therapy for recent cancer survivors: a feasibility trial. Psychooncology 2017 Jan;26(1):137-139. [doi: 10.1002/pon.4032] [Medline: 26555347]
66. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
67. Leykin Y, Thekdi SM, Shumay DM, Muñoz RF, Riba M, Dunn LB. Internet interventions for improving psychological well-being in psycho-oncology: review and recommendations. Psychooncology 2012 Sep;21(9):1016-1025 [FREE Full text] [doi: 10.1002/pon.1993] [Medline: 21608075]
68. Pingree S, Hawkins R, Baker T, duBenske L, Roberts LJ, Gustafson DH. The value of theory for enhancing and understanding e-health interventions. Am J Prev Med 2010 Jan;38(1):103-109 [FREE Full text] [doi: 10.1016/j.amepre.2009.09.035] [Medline: 20117565]
69. Phillips JL, Currow DC. Cancer as a chronic disease. Collegian 2010;17(2):47-50. [doi: 10.1016/j.colegn.2010.04.007] [Medline: 20738055]
70. Dale L, Dobson R, Whittaker R, Maddison R. The effectiveness of mobile-health behaviour change interventions for cardiovascular disease self-management: a systematic review. Eur J Prev Cardiol 2016 May;23(8):801-817. [doi: 10.1177/2047487315613462] [Medline: 26490093]
71. Greenwood DA, Gee PM, Fatkin KJ, Peeples M. A systematic review of reviews evaluating technology-enabled diabetes self-management education and support. J Diabetes Sci Technol 2017 Sep;11(5):1015-1027 [FREE Full text] [doi: 10.1177/1932296817713506] [Medline: 28560898]
72. Lee PA, Greenfield G, Pappas Y. The impact of telehealth remote patient monitoring on glycemic control in type 2 diabetes: a systematic review and meta-analysis of systematic reviews of randomised controlled trials. BMC Health Serv Res 2018 Jun 26;18(1):495 [FREE Full text] [doi: 10.1186/s12913-018-3274-8] [Medline: 29940936]

73. Lee J, Choi M, Lee SA, Jiang N. Effective behavioral intervention strategies using mobile health applications for chronic disease management: a systematic review. BMC Med Inform Decis Mak 2018 Feb 20;18(1):12 [FREE Full text] [doi: 10.1186/s12911-018-0591-0] [Medline: 29458358]
74. Brigden A, Anderson E, Linney C, Morris R, Parslow R, Serafimova T, et al. Digital behavior change interventions for younger children with chronic health conditions: systematic review. J Med Internet Res 2020 Jul 31;22(7):e16924 [FREE Full text] [doi: 10.2196/16924] [Medline: 32735227]
75. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for patient engagement: a systematic review. Front Psychol 2016 Jan 8;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
76. Hanna TP, Evans GA, Booth CM. Cancer, COVID-19 and the precautionary principle: prioritizing treatment during a global pandemic. Nat Rev Clin Oncol 2020 May;17(5):268-270 [FREE Full text] [doi: 10.1038/s41571-020-0362-6] [Medline: 32242095]
77. Chan A, Ashbury F, Fitch MI, Koczwara B, Chan RJ, MASCC Survivorship Study Group. Cancer survivorship care during COVID-19-perspectives and recommendations from the MASCC survivorship study group. Support Care Cancer 2020 Aug;28(8):3485-3488 [FREE Full text] [doi: 10.1007/s00520-020-05544-4] [Medline: 32451702]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
RoB: risk of bias

Edited by R Kukafka, G Eysenbach; submitted 02.10.20; peer-reviewed by N Wreglesworth; comments to author 30.10.20; revised version received 01.11.20; accepted 20.09.21; published 29.10.21.

Please cite as:
Marthick M, McGregor D, Alison J, Cheema B, Dhillon H, Shaw T
Supportive Care Interventions for People With Cancer Assisted by Digital Technology: Systematic Review
J Med Internet Res 2021;23(10):e24722
URL: https://www.jmir.org/2021/10/e24722
doi:10.2196/24722
PMID:

Original Paper

Detecting Parkinson Disease Using a Web-Based Speech Task: Observational Study

Wasifur Rahman[1*], MSc; Sangwu Lee[1*], BSc; Md Saiful Islam[1], BSc; Victor Nikhil Antony[1], BSc; Harshil Ratnu[1], BSc; Mohammad Rafayet Ali[1], PhD; Abdullah Al Mamun[1], BSc; Ellen Wagner[2], MSc; Stella Jensen-Roberts[2], BSc; Emma Waddell[2], BSc; Taylor Myers[2], BA; Meghan Pawlik[2], BA; Julia Soto[2], BSc; Madeleine Coffey[2], BA; Aayush Sarkar[2], BSc; Ruth Schneider[2,3], MD; Christopher Tarolli[2,3], MD; Karlo Lizarraga[2,3], MD; Jamie Adams[2,3], MD; Max A Little[4,5], PhD; E Ray Dorsey[2], MBA, MD; Ehsan Hoque[1], PhD

[1]Department of Computer Science, University of Rochester, Rochester, NY, United States

[2]Center for Health and Technology, University of Rochester Medical Center, Rochester, NY, United States

[3]Department of Neurology, University of Rochester Medical Center, Rochester, NY, United States

[4]School of Computer Science, University of Birmingham, Birmingham, United Kingdom

[5]MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA, United States

[*]these authors contributed equally

Corresponding Author:
Wasifur Rahman, MSc
Department of Computer Science
University of Rochester
250 Hutchinson Rd
Rochester, NY, 14620
United States
Phone: 1 5857487677
Email: echowdh2@ur.rochester.edu

Abstract

Background: Access to neurological care for Parkinson disease (PD) is a rare privilege for millions of people worldwide, especially in resource-limited countries. In 2013, there were just 1200 neurologists in India for a population of 1.3 billion people; in Africa, the average population per neurologist exceeds 3.3 million people. In contrast, 60,000 people receive a diagnosis of PD every year in the United States alone, and similar patterns of rising PD cases—fueled mostly by environmental pollution and an aging population—can be seen worldwide. The current projection of more than 12 million patients with PD worldwide by 2040 is only part of the picture given that more than 20% of patients with PD remain undiagnosed. Timely diagnosis and frequent assessment are key to ensure timely and appropriate medical intervention, thus improving the quality of life of patients with PD.

Objective: In this paper, we propose a web-based framework that can help anyone anywhere around the world record a short speech task and analyze the recorded data to screen for PD.

Methods: We collected data from 726 unique participants (PD: 262/726, 36.1% were women; non-PD: 464/726, 63.9% were women; average age 61 years) from all over the United States and beyond. A small portion of the data (approximately 54/726, 7.4%) was collected in a laboratory setting to compare the performance of the models trained with noisy home environment data against high-quality laboratory-environment data. The participants were instructed to utter a popular pangram containing all the letters in the English alphabet, “the quick brown fox jumps over the lazy dog.” We extracted both standard acoustic features (mel-frequency cepstral coefficients and jitter and shimmer variants) and deep learning–based embedding features from the speech data. Using these features, we trained several machine learning algorithms. We also applied model interpretation techniques such as Shapley additive explanations to ascertain the importance of each feature in determining the model’s output.

Results: We achieved an area under the curve of 0.753 for determining the presence of self-reported PD by modeling the standard acoustic features through the XGBoost—a gradient-boosted decision tree model. Further analysis revealed that the widely used mel-frequency cepstral coefficient features and a subset of previously validated dysphonia features designed for detecting PD from a verbal phonation task (pronouncing “ahh”) influence the model’s decision the most.

Conclusions: Our model performed equally well on data collected in a controlled laboratory environment and *in the wild* across different gender and age groups. Using this tool, we can collect data from almost anyone anywhere with an audio-enabled device and help the participants screen for PD remotely, contributing to equity and access in neurological care.

(J Med Internet Res 2021;23(10):e26305) doi:10.2196/26305

KEYWORDS

Parkinson's disease; speech analysis; improving access and equity in health care; mobile phone

Introduction

Parkinson disease (PD) is the fastest-growing neurological disease worldwide. Unfortunately, an estimated 20% of patients with PD remain undiagnosed. This can be largely attributed to the shortage of neurologists worldwide [1,2] and limited access to health care. Early diagnosis and continuous monitoring, which allows medication dosage adjustment, are key to managing the symptoms of this incurable disease. The current standard of diagnosis requires in-person clinic visits where an expert assesses the disease symptoms while observing the patients as they perform tasks from the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) [3]. The MDS-UPDRS includes 24 motor-related tasks to assess speech, facial expression, limb movements, walking, memory, and cognitive abilities. Although many studies have shown success by analyzing hand movements [4], limb movement patterns [5], and facial expressions [6], speech is especially important because approximately 90% of patients with PD exhibit vocal impairment [7,8], which is often one of the earliest indicators of PD [9].

For speech analysis, researchers have looked into phonations (pronouncing "ahhh") from audio recordings [10] to quantify rhythm, stress, and intonation [11]. Little et al [12] introduced pitch period entropy (PPE) as a measure of dysphonia to distinguish healthy people from people with PD with 91% accuracy. Later, Tsanas [13] expanded on this by calculating 132 dysphonia measures to classify PD versus control with almost 99% accuracy. In addition, Peker et al [14] used a novel feature selection technique with a complex-valued artificial neural network. Rueda and Krishnan [15] identified a set of mel-frequency cepstral coefficients (MFCCs) and intrinsic mode functions to represent the characteristics of PD. In the domain of analyzing real-life audio data, Wroge et al [16] analyzed verbal phonation data collected from smartphones; Vaiciukynas et al [17] used a convolutional neural network to detect PD from speech; Vásquez-Correa et al [18] collected speech samples with a handheld device in uncontrolled noise conditions; and Dubey et al [19] designed a smartwatch-based voice and speech monitoring system for people with PD receiving speech therapies from speech-language pathologists. Although the current state-of-the-art method has shown promising results, it has limitations such as small sample size [12,19,20], oversampling from the same participants [14], noise-controlled data collection [12,15], and age discrepancy between PD and control [21].

In this paper, we present our analysis of 726 audio recordings of speech from 36.1% (262/726) individuals with PD and 63.9% (464/726) without PD. The speech recordings were collected using a web-based tool called *Parkinson's Analysis with Remote Kinetic-tasks* (PARK) [22]. The PARK tool instructed the participants to utter a popular pangram containing all the letters in the English alphabet, "The quick brown fox jumps over the lazy dog," and recorded it. This allowed us to rapidly collect a data set that is more likely to contain real-world variability associated with geographical boundaries, socioeconomic status, age groups, and a wide variety of heterogeneous recording devices. The findings in this study build on this unique real-world data set; thus, we believe it could potentially be generalized for real-world deployments.

Collecting audio data from individuals often requires in-person visits to the clinic, limiting the number of data points and the diversity within the data. Recent advancements have allowed the collection of tremor data from wearable sensors [23] and sleep data from radio frequency signals [24]. The existing work with speech and audio analysis uses sophisticated equipment for collecting data that are often noise-free [12,25] and do not contain real-world variability. As a significant portion of the population has access to a mobile device with recording capability (eg, 81% of Americans own a smartphone) [26], we opted to use a framework that allows participants to record data from their homes. From the recorded audio files, we extracted acoustic features including MFCCs, which represent the short-term power spectrum of a sound, jitter or shimmer variants (representing pathological voice quality), pitch-related features, spectral power, and dysphonia-related features designed to capture PD-induced vocal impairment [12].

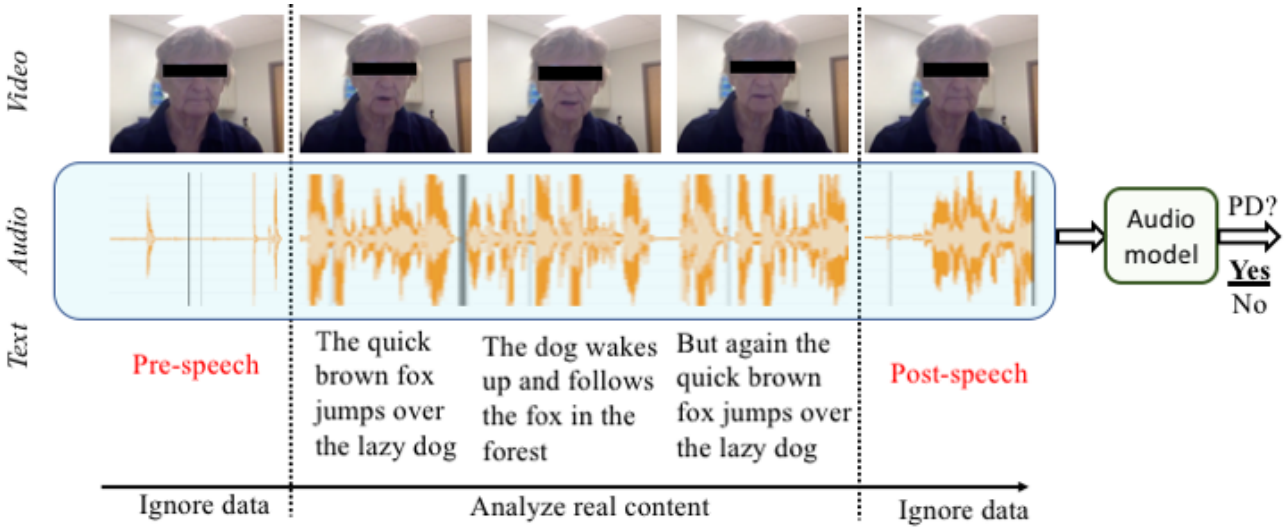
In addition, we extracted features from a deep-learning–based encoder—problem agnostic speech encoder (PASE) [27], which represents the information contained in a raw audio instance through a list of encoded vectors. These features are modeled with four different machine learning models—support vector machine (SVM), random forest, LightGBM, and XGBoost—to classify individuals with and without PD.

Figure 1 provides an outline of the data analysis system. Our contributions can be summarized as follows:

- We report findings from one of the largest data sets with real-world variability, containing 726 unique participants mostly from their homes.
- We analyzed the audio features of speech to predict PD versus non-PD with an area under the curve (AUC) score of 0.753.
- We provide evidence that our model prioritizes MFCC features and a subset of dysphonia features [12,28], which is consistent with previous literature.

- Our model performs consistently well when tested on gender- and age-stratified data collected in a controlled laboratory environment and in the wild.

Figure 1. An outline of our approach for solving the speech task of uttering "The quick brown fox...".



Methods

Recruitment and Data Collection

We collected data from 726 *unique* participants uttering the sentences, "The quick brown fox jumps over the lazy dog. The dog wakes up and follows the fox into the forest, but again the quick brown fox jumps over the lazy dog," using PARK [29] website. Figure 2 provides a brief overview of the data collection, storage, transfer, and analysis mechanisms. The users are recorded via a webcam and a microphone connected to the PC or laptop, and the recordings are uploaded to a server. Figure 3 shows images of some of the study participants, and Figure 4 shows the age distribution of the participants. The number of participants without PD is 1.8 times the number of participants with PD. Most of the participants with PD are concentrated in the age range of 40 to 80 years. However, participants without PD outnumber those with PD in the age group 20 to 40 years. Similarly, participants with PD significantly outnumber those without PD in the age group of 80 to 90 years. Table 1 provides the demographic information of the study participants.

Figure 2. An overview of our data collection, storage, and analysis pipeline.

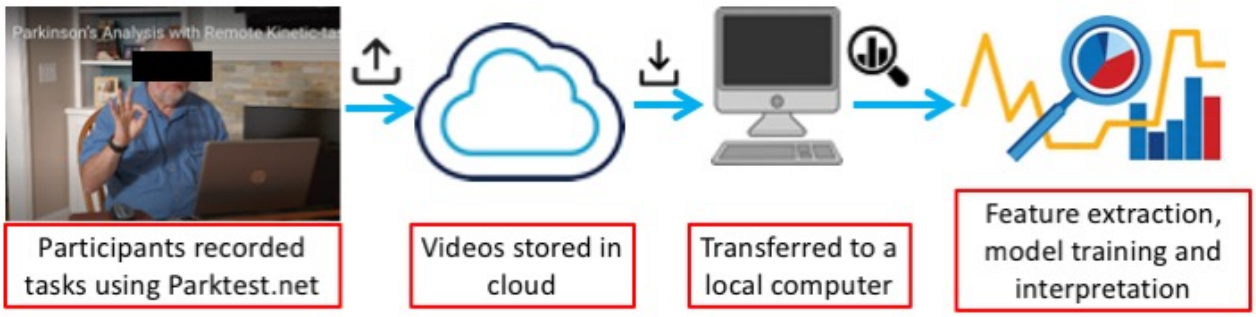


Figure 3. Some screenshots of our subjects while providing the data. All the subjects except B provided data without any supervision. B, D, E, and F have been diagnosed with Parkinson disease. Electronic informed consent was taken from the participants to use their photos for publication.

Figure 4. A bar plot showing the age distribution of participants with PD and those without PD in our data set. PD: Parkinson disease.

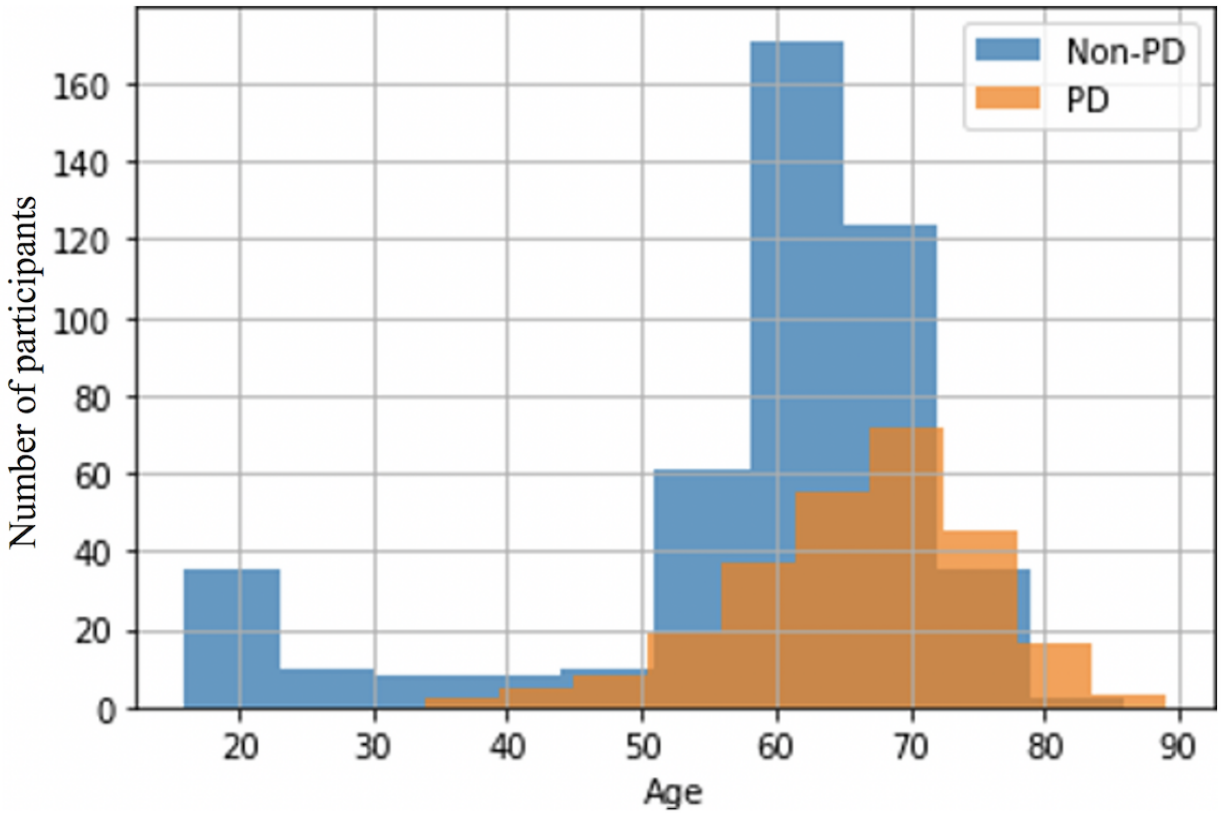


Table 1. Demographic composition of our data set (N=726).

Characteristics	Participants with PD[a]	Participants without PD
Total, n (%)	262 (36.1)	464 (63.9)
Gender, n (%)		
Female	101 (38.5)	300 (64.6)
Male	161 (61.4)	164 (22.5)
Age (years), mean (SD)	65.92 (9.2)	57.98 (14.2)
Country, n (%)		
United States	199 (75.9)	419 (90.3)
Other	63 (24)	45 (9.7)
Years since diagnosed, mean (SD)	7.88 (5.41)	N/A[b]

[a]PD: Parkinson disease.

[b]N/A: not applicable.

Among our 726 unique participants, 262 (36.1%) had received the diagnosis of PD, whereas the others were participants without PD. We obtained contact information of patients with PD from local clinics and various PD support groups. Participants without PD were recruited from Amazon Mechanical Turk. During data collection, informed consent of all participants was obtained in accordance with the protocol agreed upon between the researchers and the institutional review board of the University of Rochester. Among the 726 unique participants, only 54 (7.4%) provided data in the laboratory under the guidance of a study coordinator using the PARK tool; the rest of the 672 (92.6%) participants used the PARK system from their homes to provide data. Having participants perform the tasks at home and in the laboratory allowed us to compare the results across both conditions. No participant appeared in either set, and *all* of our participants used the identical PARK protocol.

The gender distribution in the data set was skewed, especially for female participants. Among all participants, 55.2% (401/726) were women, and 44.8% (325/726) were men. However, among participants with PD, only 38.5% (101/262) were women, and for participants without PD, 64.6% (300/464) were women. Most of the participants with PD were in the age range of 40 to 80 years, but most of the younger (20-40 years) and older (80-90 years) participants were from the group without PD and the group with PD, respectively.

As our data were collected through a web-based framework, we do not have the MDS-UPDRS scores for our participants because collecting them requires additional input from doctors. Among all the participants in our PD group, only 3 participants replied in affirmative that they had taken medication 2 hours before taking the test; others replied in the negative. Therefore, we assumed that medication effects were negligible for the purposes of this study.

Data Preprocessing

During data collection, participants often took some additional time to start to utter the task sentences and stop the recording once the sentences were uttered. Hence, we had a substantial amount of noisy and irrelevant data at the beginning and end of most data instances (Figure 1). To remove irrelevant data, we used the Penn Phonetics Lab Forced Aligner Toolkit (P2FA) [30]. Given an audio file and transcript, it attempts to predict the time boundaries in which each of the words in the transcript was pronounced. P2FA applies a combination of hidden Markov models [31] to predict the most likely sequence of phonemes for a given audio and Gaussian mixture models to combine those phonemes into words and obtain their time boundary using a predefined dictionary [32]. From the output of this system, we can obtain the starting time of the first word recognized by the P2FA and the ending time of the last word recognized by the P2FA. We used the audio segments between them for further analysis.

Acoustic Features Extraction

We extracted features by combining outputs of multiple sources: Praat features [33] obtained through the Parselmouth Python interface [34] and the previously used dysphonia features relevant for PD analysis [12,35,36]. After calculating all the features, we constructed a correlation matrix of the feature values to calculate the degree of correlation between them. Then, we iterated over each pair of features in an unordered fashion, and if the correlation coefficient between them was above 0.9, we dropped one of those features from further analysis [37]. Table 2 contains a short overview of the features used in our analysis; the feature names in italicized text are those used for building the models. We provide a more comprehensive description of the features in Multimedia Appendix 1 [12,14,36,38]. Some of our definitions were adapted from the official Praat documentation [33].

Table 2. Names of all the features, code source used for collecting them, and a short description[a].

Feature	Code source	Short description
Pitch		
MedianPitch [b]	Little et al [36]	Median principal frequency
MeanPitch	Boersma and Weenink [33]	Mean principal frequency
StdDevPitch	Little et al [36]	SD in principal frequency
Jitter		
MeanJitter	Little et al [36]	Perturbation in principal frequency (mean variation)
MedianJitter	Little et al [36]	Perturbation in principal frequency (median variation)
LocalJitter	Boersma and Weenink [33]	Average absolute difference between consecutive periods, divided by the average period
LocalAbsoluteJitter	Boersma and Weenink [33]	Average absolute difference between consecutive periods measured in seconds
RapJitter	Boersma and Weenink [33]	Average absolute difference between a period and the average of it and its two neighbors
Ppq5Jitter	Boersma and Weenink [33]	5-point period perturbation quotient
DdpJitter	Boersma and Weenink [33]	Difference of differences of periods of principal frequency
Shimmer		
MeanShimmer	Little et al [36]	Amplitude perturbation (using mean)
MedianShimmer	Little et al [36]	Amplitude perturbation (using median)
LocalShimmer	Boersma and Weenink [33]	Average absolute difference between amplitudes of consecutive period
LocaldbShimmer	Boersma and Weenink [33]	Average absolute base-10 logarithm of the difference between amplitudes of consecutive period
Apq3Shimmer	Boersma and Weenink [33]	3-point amplitude perturbation quotient
Apq5Shimmer	Boersma and Weenink [33]	5-point amplitude perturbation quotient
Apq11Shimmer	Boersma and Weenink [33]	11-point amplitude perturbation quotient
DdaShimmer	Boersma and Weenink [33]	Shimmer calculated by difference in differences of amplitude
MFCC[c]		
MeanMFCC (0-12)	Little et al [36]	13 features of mean MFCC
VariationMFCC (0-12)	Little et al [36]	13 features of mean variation of MFCC
RelBandPower (0-3)	Tsanas et al [39]	4 features capturing relative band power in 4 spectrum ranges
HNR [d]	Boersma and Weenink [33]	Signal-to-noise ratio
RPDE [e]	Little et al [36]	Pitch estimation uncertainty
DFA [f]	Little et al [36]	Measure of stochastic self-similarity in turbulent noise
PPE [g]	Little et al [36]	Measure of inability of maintaining constant pitch

[a]We collected the features using the code or methodology described in the corresponding code-source entry.

[b]The correlated features were removed, and features in italicized text were used to build the models. Feature names are preceded by the loosely defined umbrella category they belong to.

[c]MFCC: mel-frequency cepstral coefficient.

[d]HNR: harmonic-to-noise ratio.

[e]RPDE: recurrence period density entropy.

[f]DFA: detrended fluctuation analysis.

[g]PPE: pitch period entropy.

Embedding Features Extraction

We extracted deep learning–based PASE embeddings [27] for our audio files. PASE represents the information contained in a raw audio instance through a list of encoded vectors. To ensure that the encoded vectors contain the same information as the input audio file, it decodes various properties of the audio file, including the audio waveform, log power spectrum, MFCCs,

four prosody features (interpolated logarithm of the fundamental frequency, voiced or unvoiced probability, zero-crossing rate, and energy), and local InfoMax, from the encoded vectors. The encoded vectors must retain relevant information about the input audio file to decode all these properties successfully.

As these properties represent the inherent characteristics of the input audio file rather than any task-specific features, they have been used to solve a host of downstream tasks, such as speech classification, speaker recognition, and emotion recognition. Therefore, we also used them for PD detection.

Experiments

For each feature set, we applied a standard set of machine learning algorithms, such as SVM [40], XGBoost [41], LightGBM [42], and random forest [43], to classify PD versus non-PD. SVM separates the data into several classes while maintaining the maximum possible margin among the classes. A random forest is built as an ensemble of decision trees; each decision tree builds a tree using a subset of features and learns if-else type decision rules to make a prediction. We also used XGBoost and LightGBM, algorithms based on gradient boosting where they build successively better models by refining the models at hand.

We used a leave-one-out cross-validation training strategy; using this strategy, one sample of data from the data set is left out, and the other n−1 samples are used to create a model and predict the remaining sample. We used metrics such as binary accuracy and AUC to evaluate our model's performance. AUC is the area under the receiver operating characteristics (ROC) curve. The ROC curve is constructed by calculating the AUC produced by taking the ratio of the true-positive rate and the false-positive rate while varying the decision threshold. AUC has the highest value of 1, which denotes that the two classes can be separated perfectly, whereas an AUC value of 0.5 indicates that the model cannot distinguish between the two classes. Because our data set was imbalanced, AUC is a much better metric for understanding the true performance of our model. To reduce the effects of data imbalance, we used data augmentation techniques such as the synthetic minority oversampling technique [44] and SVM synthetic minority oversampling technique [45].

Model Interpretation Technique

To interpret the models, we used the Shapley additive explanations (SHAP) technique based on the Shapley value. The Shapley value is a game-theoretic concept of distributing the payouts fairly among players [46]. In the machine learning context, each individual feature of an instance can be thought of as a player, and the payout is the difference between an instance's prediction and the average prediction. We choose SHAP for two reasons: (1) it is well suited for explaining the output of any machine learning model, and (2) it is the only feature attribution method that fulfills the mathematical definition of fairness.

To use SHAP to explain gradient boosting and tree-based models such as XGBoost, Lundberg et al [47] introduced methods to provide a polynomial time model to compute optimized explanations. Their method can generate local interpretations—how the features affect one particular prediction of a data instance—and then combine those local interpretations to make global interpretations about features present in the entire data set. By setting up a class of features to condition on, they traverse the tree in the following manner: if we are traversing on a node that was split based on a feature we are conditioning on, we simply follow the decision path; otherwise, the results from the left and right subtrees originating from the current node are computed recursively, and their results are added through a weighted summation strategy, thus computing the SHAP value for the feature in question.

Results

Overview

The data were preprocessed, and both standard acoustic features (eg, pitch, jitter, shimmer, and MFCC) and deep learning–based audio embedding features, representing an audio clip as a feature vector, were extracted; henceforth, we call these *standard features* and *embedding features*. In the rest of the *Results* section, we discuss the results from the models built on the entire data set, interpretations of the best model on the entire data set, and results and interpretation from specialized models on gender-stratified and age-trimmed data sets.

Detecting PD From the Entire Data Set

Table 3 contains the AUC and accuracy scores of the four machine learning models trained on the *standard features* and *embedding features* separately. Applying XGBoost on the standard features showed the best performance of 0.75 AUC and 0.74 accuracy. We also noticed that models trained on *standard features* work better than those trained on *embedding features.*

Table 3. Performance on the entire data set. The performance of various machine learning algorithms using the standard and embedding features on a data set combining data from both home and laboratory environments[a].

Algorithm	Standard features		Embedding features	
	AUC[b]	Accuracy	AUC	Accuracy
SVM[c]	0.751	0.735	0.738	0.692
Random forest	0.745	0.720	0.726	0.708
LightGBM	0.753	0.720	0.737	0.693
XGBoost	*0.750* [d]	*0.741*	0.722	0.689

[a]Models using standard features perform better than the models using embedding features in terms of both binary accuracy and area under the curve. Although the performance of the models is almost similar in terms of area under the curve metric, XGBoost outperforms others by considering both the area under the curve and accuracy metrics simultaneously.

[b]AUC: area under the curve.

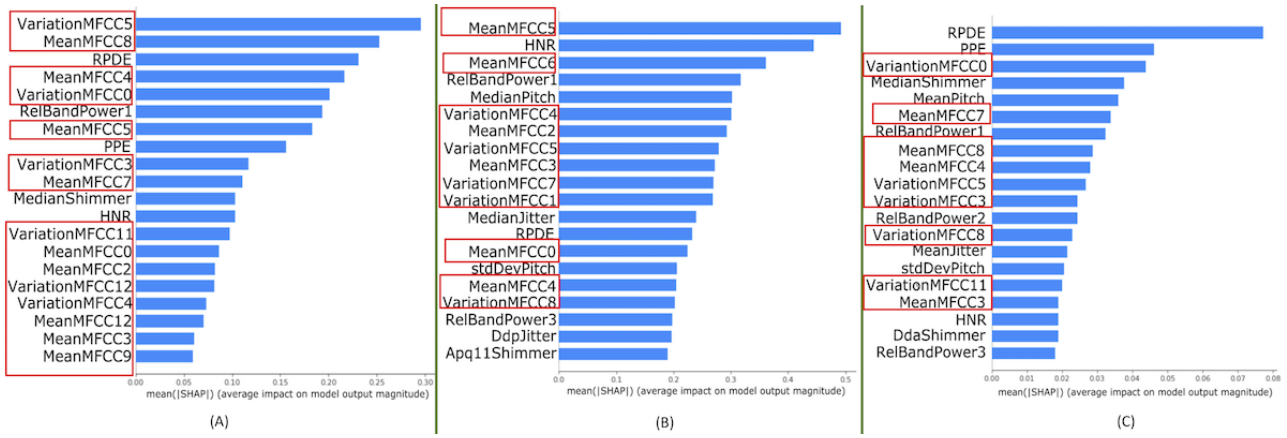
[c]SVM: support vector machine.

[d]Variable outperforms all others by taking both area under the curve and accuracy into account.

Model Interpretation

The goal of SHAP is to explain the model's prediction of any instance as a sum of contributions from its feature values; if a data instance can be thought of as $X_i=[f_1, f_2,...f_N]$, SHAP will assign a number to each of these f_j features, denoting the impact of that feature—both the magnitude and direction—on the model's prediction. Then, all these local explanations are aggregated to create a global interpretation for the entire data set. A global interpretation is presented in the first part of Figure 5; the top 10 most impactful features, ranked by having the most impact to the least, are presented. To calculate each feature's impact, all of its SHAP values across all the data instances are gathered, and then the mean of their absolute values is calculated.

Figure 5. Shapley additive explanations analysis of our best performing models on 3 data sets: (A) main model (ie, entire data set), (B) female model (ie, female only), and (C) age-trimmed model (all subjects are older than 50 years).

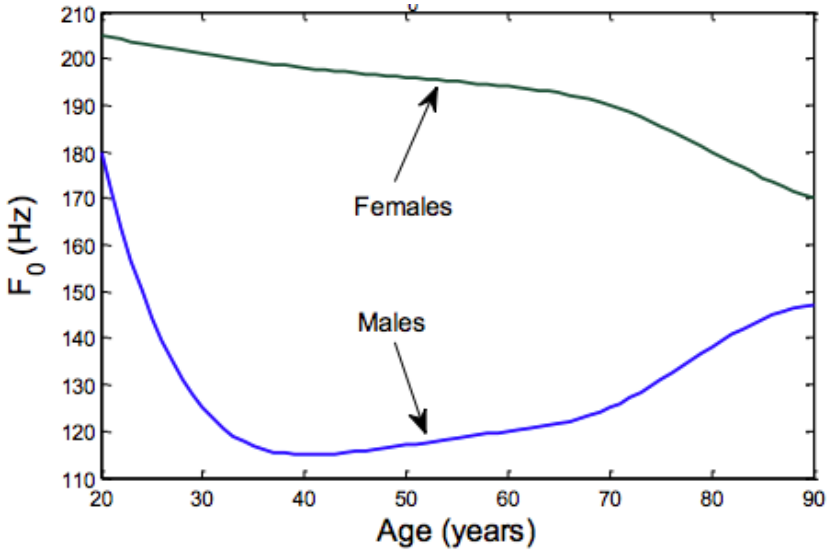


Features that have an impact on the model's performance are typically the spectral features: the mean values or the variation of MFCC in each spectrum range. Apart from that, some other complex features such as recurrence period density entropy (RPDE; a measure of uncertainty in F0 estimation), PPE (a measure of inability to maintain a constant F0), and harmonic-to-noise ratio (HNR) also affected the interpretation presented in the first part of Figure 5.

Gender- and Age-Stratified Analysis

The characteristics of a person's voice are greatly influenced by age and gender. In Figure 6, we see that men and women display a changing characteristic in their voices as they get older. Female voices have a higher F0 value, but it decreases with age. Males typically have a higher F0 value in their youth, which decreases with age and then increases roughly after the age of 45 years. Therefore, it can produce confounding effects when analyzing PD from audio, where the machine learning model uses audio features to detect PD. To minimize the effect of confounding factors, researchers trained separate models on data from male and female participants [28] or analyzed an age-trimmed data set by considering data from participants older than 50 years [4,29].

Figure 6. Changes in fundamental frequency F0 of voice as a function of gender and age (collected from Tsanas et al [13]).



Building Specialized Models for Each Gender- and Age-Trimmed Analysis

The performance metrics of the machine learning models trained on the male, female, and age-trimmed data sets are shown in Table 4. By comparing the performance with metrics presented in Table 3, we can see that the models that used male- or age-trimmed data sets performed at par or better than the models that used the whole data set to train. However, there was a performance drop in the models using the female data set. Table 1 shows that females are overrepresented in the non-PD group and underrepresented in the PD group, leading to data imbalance and possibly lowering performance for the female-only model.

Table 4. Gender- and age-stratified models. Three separate data sets were constructed: a male data set with male participants, a female data set with female participants, and age-trimmed data set by excluding the participants younger than 50 years. For each of these data sets, a separate model was constructed, and its performance is reported below (N=726).

Algorithm	Male (n=415)		Female (n=477)		Age-trimmed (male, n=366 and female, n=426)	
	AUC[a]	Accuracy	AUC	Accuracy	AUC	Accuracy
SVM[b]	0.795	0.717	0.659	0.763	0.755	0.723
Random Forest	0.758	0.702	0.699	0.788	0.739	0.713
LightGBM	0.725	0.665	0.717	0.768	0.749	0.712
XGBoost	0.762	0.717	0.682	0.771	0.742	0.704

[a]AUC: area under the curve.

[b]SVM: support vector machine.

We also analyzed the features that drive the performance of these specialized models through SHAP analysis. The second part of Figure 5 shows the most salient features ranked by their SHAP value and the distribution of the impact of each feature on the model's decision-making. The most important features are still dominated by the MFCC-related features or complex features such as the HNR, relative band power in different frequency ranges (RelBandPower1, RelBandPower3), RPDE (uncertainty in F0 estimation), perturbation in F0 (DdpJitter), or perturbation in amplitude (Apq11Shimmer). However, one noticeable fact is that three-pitch and jitter-related features—MedianPitch (median principal frequency, StdDevPitch (SD in principal frequency) and MedianJitter (median variation in F0)—also affected the model's prediction, which was not noticed in the SHAP analysis run on the all-data-model.

Similarly, we interpreted the salient features of the age-trimmed data set in the third part of Figure 5. We noticed that the most salient features usually come from the MFCC feature groups, complex features (RPDE, PPE, and HNR), relative band power (RelBandPower1, RelBandPower2, and RelBandPower3), and pitch-related features. In addition, the pitch-related features also drove the prediction of the model.

Discussion

Limitations

In all of our experiments, we chose leave-one-out cross-validation to maintain uniform experimental settings between different models, data augmentation, and data set combinations because K-fold cross-validation can show high variance in performance based on how stratified the folds are. However, we acknowledge that our choice introduces the problem of overfitting and increases the computational complexity manifold. Therefore, we could not run extensive hyperparameter tuning to improve the performance of our

models. By carefully stratifying the folds and maintaining the same fold settings for comparable experimental settings, K-fold cross-validation may enable us to achieve higher performance with lower computational complexity.

Moreover, our chosen metrics, accuracy and ROC AUC, can be overly optimistic because of the unbalanced nature of our data set. Metrics such as ROC–Precision-Recall balanced accuracy, and more information about sensitivity and specificity may shed more light on how we are performing in the minority class. In the future, we plan to apply other cross-validation techniques and better metrics to conduct our experiments and report our performance.

Detecting PD From Regular Conversation

Some of the most common voice disorders induced by PD are dysphonia (distortion or abnormality of voice), dysarthria (problems with speech articulation), and hypophonia (reduced voice volume). Two speech-related diagnostic tasks are commonly used to detect PD by exploiting the changing vocal pattern caused by these disorders: (1) sustained phonation (the participant is supposed to utter a single vowel for a long time with constant pitch) and (2) running speech (the participant speaks a standard sentence). Little et al [12] developed features for detecting dysphonia in patients with PD. Tsanas et al [25] focused on the telemonitoring of self-administered sustained vowel phonation task to predict the Unified Parkinson's Disease Rating Scale rating (on Rating Scales for Parkinson's Disease 2003), a commonly used indicator for quantifying PD symptoms. These studies trained their models with data captured by sophisticated devices (eg, wearable devices, high-resolution video recorder, and the Intel at-home–testing-device telemonitoring system) that are often not accessible to all and are difficult to scale. The performance of these models can be significantly reduced when classifying data collected in home acoustics. In addition, correctly completing the sustained phonation task requires following a specific set of guidelines, such as completing the task in one breath, which can be difficult for older individuals.

In contrast, we analyzed the running speech task from the data collected by using a web-based data collection platform that can be accessed by anyone anywhere in the world and requires only an internet-connected device with an integrated camera and microphone. In addition, the running speech task does not require conforming to specific instructions and is more similar to regular conversation; therefore, the model can be potentially augmented to predict PD from a regular conversation—a potential game-changer in PD assessment. In the future, user-consented plug-ins could be developed for apps such as Alexa, Google Home, or Zoom, where audio is transmitted between persons. Anyone who consents to download the plug-in and uses it while on the phone, over Zoom, or giving virtual or in-person presentations could benefit from receiving an informal referral to see a neurologist, when appropriate. The plug-in would not store the participants' data unless they opt to build a personalized profile to ensure privacy and ethical usage of our framework.

Validating Model Interpretation

The features that SHAP found to have an impact on modeling decisions are well supported by previous research. For example, MFCC features have already proven to be useful in a wide range of audio tasks such as speaker recognition [48], music information retrieval [38], voice activity detection [49], and most importantly, in voice quality assessment [50]. Similarly, the high impact of the HNR and the measures of uncertainty in F0 estimation (RPDE) and inability to maintain a constant F0 (PPE) on the model's output are in congruence with the findings from Little et al [12]. However, explaining the first part of Figure 5 in light of PD-induced vocal impairment is a difficult task. MFCC features are calculated by converting the audio signal into the frequency domain, and they denote how energy in the signal is distributed within various frequency ranges. Therefore, providing a physical interpretation of the SHAP values corresponding to the MFCC features is not straightforward. Similarly, Little et al [12] designed the RPDE and PPE features to model a sustained phonation task (uttering "ahh") with the assumption that healthy participants will be able to maintain a smooth and regular voice pattern. In contrast, uttering multiple sentences introduces considerable variation in the data, adding a wide set of heterogeneous patterns. Therefore, the underlying assumptions behind constructing these features do not hold for our task of uttering multiple sentences.

We conduct an empirical validation of the SHAP output shown in the first part of Figure 5. We incrementally add one feature at a time to build a dynamic feature set, train successive models on that feature set, and report the accuracy and AUC performances. We see that the performance of our model saturates after adding 7 to 8 features. Therefore, we can say that the SHAP analysis teases out the most important features driving the model's performance successfully.

Why Not Stratified Analysis Only?

We built models inclusive of all genders for several reasons. First, there are potential shared characteristics among vocal patterns of all genders that can be relevant for detecting PD. Second, dividing the data set into two portions reduces the available training data for each model, which may, in turn, reduce the generalization capability of each model. In addition, our model analyzes data from patients of all ages. Although most people diagnosed with PD are older than 60 years, approximately 10% to 20% of the population diagnosed with PD are younger than 50 years, and approximately half of them are younger than 40 years [51]. As anecdotal evidence, Michael J Fox was diagnosed with PD at the age of 29 years [52], and Muhammad Ali had PD by 42 years [53]. In our data set, there were also a minority of patients with PD who were younger than 50 years (Figure 4). On the basis of these observations, we believe that our system should provide access to all people, irrespective of age. PD does not discriminate by age while affecting a person, and an automated system should not discriminate based on age and provide equitable services to people of all ages. However, these factors can function as confounders in PD analysis. Therefore, we provided additional analysis to ensure that our model does not use the idiosyncrasies of group-specific information to make predictions.

Performance Excluding Laboratory Environment Data

When the data were collected in the laboratory, participants had access to a clinician providing support using a consistent recording setup and dedicated bandwidth. In contrast, the data collected in the home setting involved no assistance and included the real-world variability of heterogeneous recording setups and inconsistent internet speed. In theory, the data collected at the *laboratory* and *home* were very different from each other.

To ensure that our model works equally well without the *clean lab data*, we designed two experiments. In experiment 1, we removed clean lab data, which was approximately 7.4% (54/726) of the entire data set, retrained our model on the remaining 672 participants using the leave-one-out validation procedure, and calculated the performance metrics. In experiment 2, we randomly removed 7% (roughly 54 data points) of *home* data from the entire data set (while keeping the *lab* data intact), building a model with the remaining 93% data with the leave-one-out cross-validation method. Then, we conducted experiment 2 10 times and calculated the *average* performance from these 10 runs. We found that the AUC metric across these three experiments was fairly consistent, with a very small 0.015 decrease in AUC when removing lab data, demonstrating that our framework performs equally well across the *lab* and *home* data.

Label Inconsistency and Predicting Tremor Score

Using the PARK [29] protocol, we collected one of the largest data sets of participants conducting a series of motor, facial expression, and speech tasks following the MDS-UPDRS PD assessment protocol [3]. Although we analyzed only the speech task in this study, the data set can be potentially used to automate the assessment of a large set of MDS-UPDRS tasks and facilitate early-stage PD detection, thus improving the quality of life for millions of people worldwide. However, deploying the data collection protocol on the web and facilitating access to anyone anywhere around the world comes at a cost. To date, all of our participants with PD have been clinically verified to be diagnosed with PD. Therefore, the labels of the PD data points are reliable. However, the participants without PD did not undergo clinical verification. Our data collection protocol asked them appropriate questions to check whether they had been diagnosed with PD and collected data when they answered in the negative. However, we cannot discount the possibility that a small subset of our population without PD is in the very early stage of PD and is oblivious about it. At present, there are estimated to be around 1 million patients with PD in the United States, out of a population of 330 million [51], yielding a PD prevalence rate of 0.3%. However, as our non-PD data set was largely tilted toward people older than 50 years, the rate in our data set could be higher than 0.3%. Even if we consider a liberal 1% prevalence rate, the number of individuals with undiagnosed PD in our control population is likely low (at most 4.6 persons). Therefore, we believe that the non-PD data labels are generally reliable. In the future, we plan to model tremor score in (0-4 range for each task; 0 for no tremor, and 4 for severe tremor) instead of a binary label following the MDS-UPDRS protocol to address this problem more thoroughly.

Building a More Representative Data Set

The PARK protocol is web enabled, allowing anyone with access to the internet to contribute data. We plan to augment our data set by adding more non–native English speakers, females, and participants with PD. As our PD data are collected through contacts from local PD clinics and non-PD data through Amazon Mechanical Turk, most of our participants were from the United States or other English-speaking countries. To make our model more robust on data from non–native English speakers, we are in process of collecting both PD and non-PD data from non–native English-speaking countries. In addition, our current protocol can collect data from people with computer and internet access only, and therefore, it can potentially exclude underserved people. In the future, we plan to build desktop and mobile apps that can collect data offline to build a more inclusive framework.

Our best model for female data performed worse than its male counterparts, as shown in Table 4. We attribute this degraded performance PD or non-PD imbalance for female participants in our data set: the PD-to-non-PD ratio for females was 101 out of 300 (Table 1). Previous epidemiological studies have shown that both the incidence and prevalence of PD are 1.5 to 2 times higher in men than in women [54,55]. Therefore, any randomly sampled data set for PD will have a higher prevalence of males, contributing to models that are more biased toward males. Our immediate plan is to prioritize collecting balanced data from all genders, ages, and races across geographical boundaries, leading to a balanced data set.

Our data set also has the ubiquitous problem of data imbalance in diagnostic tests: the amount of data from participants without PD is 1.8 times more than their PD counterparts. Therefore, there is a risk that the model will be biased toward predicting the majority non-PD class as default and yield a high false-negative score. To address this, we plan to recruit more participants with PD in the future to make our data set more balanced. Another potential approach is to run the analysis on a subset of the data set produced by conducting proper age, gender, and matching of patients with and without PD.

As shown in Table 1, our participants with PD were diagnosed 7.88 (SD 5.41) years ago. Owing to limited data availability, we did not conduct an analysis on early-stage PD prediction. In the future, we plan to include more participants in the early stages of PD and build models that can detect them. We will especially focus on recruiting participants with PD in the age range of 20 to 40 years to analyze the very early-stage onset of PD among young adults. Furthermore, we plan to collect and analyze multiple speech instances from each person to reduce person-specific variability and obtain a more holistic view of their PD state.

Increasing Model Performance

Although we consider AUC to be a better metric for our data set, our model performs 10% better than always choosing non-PD as the prediction in terms of binary accuracy. To be practically deployable in clinical settings, the performance needs to be improved further. We will focus on four promising avenues: making the data set balanced, designing better features

capable of modeling the nuanced pattern in our data, making our model resilient to noise present in our data, and deconfounding the PD prediction from age and gender variables.

To remove noise, we plan to augment the techniques proposed in Poorjam et al [56] to automatically enhance our data quality by detecting the segments of data that conform to our experimental design. Moreover, as discussed above, gender and age can appear as confounding variables in the PD prediction task. In this paper, we have shown that our unified model and stratified sex-specific models have similar performance. However, we plan to build better models to systematically deconfound the effects of both age and gender variables while benefiting from them simultaneously. We can achieve this by incorporating the causal bootstrapping technique—a resampling method that considers the causal relationship between variables and negates the effect of spurious, indirect interactions, as outlined by Little and Badawy [57].

Acknowledgments

This research was funded by the National Institute of Neurological Disorders and Stroke of the National Institutes of Health under award number P50NS108676.

Authors' Contributions

WR, SL, MSI, and VNA worked in data analysis, feature extraction, model training, model interpretation, and manuscript preparation. HR, AAM, EW, and SJR helped build, maintain, and coordinate the data collection procedure. MRA, MAL, and RD helped improve the manuscript; suggested important experiments; and provided access to critical resources such as code and data. All the other authors helped in data collection. EH was the principal investigator of the project; he facilitated the entire project and helped shape the narrative of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of the features.
[DOCX File , 20 KB - jmir_v23i10e26305_app1.docx]

References

1. Khadilkar S. Neurology in India. Ann Indian Acad Neurol 2013 Oct;16(4):465-466 [FREE Full text] [doi: 10.4103/0972-2327.120415] [Medline: 24339561]
2. Howlett WP. Neurology in Africa. Neurology 2014 Aug 11;83(7):654-655. [doi: 10.1212/wnl.0000000000000682]
3. Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, Movement Disorder Society UPDRS Revision Task Force. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. Mov Disord 2008 Nov 15;23(15):2129-2170. [doi: 10.1002/mds.22340] [Medline: 19025984]
4. Ali M, Hernandez J, Dorsey E, Hoque E, McDuff D. Spatio-temporal attention and magnification for classification of Parkinson's disease from videos collected via the internet. In: Proceedings of the 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020). 2020 Presented at: 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020); November 16-20, 2020; Buenos Aires, Argentina. [doi: 10.1109/fg47880.2020.00008]
5. Lonini L, Dai A, Shawen N, Simuni T, Poon C, Shimanovich L, et al. Wearable sensors for Parkinson's disease: which data are worth collecting for training symptom detection models. NPJ Digit Med 2018 Nov 23;1(1):64-68 [FREE Full text] [doi: 10.1038/s41746-018-0071-z] [Medline: 31304341]
6. Bandini A, Giovannelli F, Orlandi S, Barbagallo S, Cincotta M, Vanni P, et al. Automatic identification of dysprosody in idiopathic Parkinson's disease. Biomed Signal Process Control 2015 Mar;17:47-54. [doi: 10.1016/j.bspc.2014.07.006]
7. Ho AK, Iansek R, Marigliani C, Bradshaw JL, Gates S. Speech impairment in a large sample of patients with Parkinson's disease. Behav Neurol 1999;11(3):131-137. [doi: 10.1155/1999/327643]
8. Logemann JA, Fisher HB, Boshes B, Blonsky ER. Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. J Speech Hear Disord 1978 Feb;43(1):47-57. [doi: 10.1044/jshd.4301.47] [Medline: 633872]
9. Duffy J. Motor Speech Disorders E-Book: Substrates, Differential Diagnosis, and Management. 4th Edition. Amsterdam: Elsevier; 2019:201.
10. Arora S, Venkataraman V, Zhan A, Donohue S, Biglan K, Dorsey E, et al. Detecting and monitoring the symptoms of Parkinson's disease using smartphones: A pilot study. Parkinsonism Relat Disord 2015 Jun;21(6):650-653. [doi: 10.1016/j.parkreldis.2015.02.026] [Medline: 25819808]

11. Playfer J, Hindle JV. Parkinson's Disease in the Older Patient. London: CRC Press; 2008:1-432.
12. Little M, McSharry P, Hunter E, Spielman J, Ramig L. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. Nat Prec 2008 Sep 12:1. [doi: 10.1038/npre.2008.2298.1]
13. Tsanas A. Accurate telemonitoring of Parkinson's disease symptom severity using nonlinear speech signal processing and statistical machine learning. Doctoral Dissertation, Oxford University, UK. 2012. URL: https://ora.ox.ac.uk/objects/uuid:2a43b92a-9cd5-4646-8f0f-81dbe2ba9d74 [accessed 2021-09-08]
14. Peker M, Sen B, Delen D. Computer-aided diagnosis of parkinson's disease using complex-valued neural networks and mrmr feature selection algorithm. J Healthc Eng 2015 Aug;6(3):281-302 [FREE Full text] [doi: 10.1260/2040-2295.6.3.281] [Medline: 26753436]
15. Rueda A, Krishnan S. Feature analysis of dysphonia speech for monitoring Parkinson's disease. In: Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2017 Presented at: 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 11-15, 2017; Jeju, Korea (South) p. 2308-2311. [doi: 10.1109/embc.2017.8037317]
16. Wroge T, Özkanca Y, Demiroglu C, Si D, Atkins D, Ghomi R. Parkinson's disease diagnosis using machine learning and voice. In: Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium (SPMB). 2018 Presented at: IEEE Signal Processing in Medicine and Biology Symposium (SPMB); Dec. 1, 2018; Philadelphia, PA, USA. [doi: 10.1109/spmb.2018.8615607]
17. Vaiciukynas E, Gelzinis A, Verikas A, Bacauskiene M. Parkinson's disease detection from speech using convolutional neural networks. In: Proceedings of the International Conference on Smart Objects and Technologies for Social Good. 2017 Presented at: International Conference on Smart Objects and Technologies for Social Good; November 29-30, 2017; Pisa, Italy p. 206-215. [doi: 10.1007/978-3-319-76111-4_21]
18. Vásquez-Correa JC, Orozco-Arroyave J, Bocklet T, Nöth E. Towards an automatic evaluation of the dysarthria level of patients with Parkinson's disease. J Commun Disord 2018 Nov;76:21-36. [doi: 10.1016/j.jcomdis.2018.08.002] [Medline: 30149241]
19. Dubey H, Goldberg J, Abtahi M, Mahler L, Mankodiya K. EchoWear: smartwatch technology for voice and speech treatments of patients with Parkinson's disease. In: Proceedings of the Conference on Wireless Health. 2015 Presented at: WH '15: Wireless Health 2015 Conference; October 14 - 16, 2015; Bethesda Maryland p. 1-8. [doi: 10.1145/2811780.2811957]
20. Vásquez-Correa J, Arias-Vergara T, Orozco-Arroyave J, Vargas-Bonilla J, Arias-Londoño J, Nöth E. Automatic detection of Parkinson's disease from continuous speech recorded in non-controlled noise conditions. In: Proceedings of the Sixth Annual Conferecence on International Speech Communication Association. 2015 Presented at: Sixth Annual Conferecence on International Speech Communication Association; 2015; Dresden, Germany. [doi: 10.21437/interspeech.2016-1122]
21. Chen H, Wang G, Ma C, Cai Z, Liu W, Wang S. An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson s disease. Neurocomputing 2016 Apr;184:131-144. [doi: 10.1016/j.neucom.2015.07.138]
22. Park Test. URL: https://parktest.net/ [accessed 2020-09-02]
23. Kubota K, Chen J, Little M. Machine learning for large-scale wearable sensor data in Parkinson's disease: Concepts, promises, pitfalls, and futures. Mov Disord 2016 Sep;31(9):1314-1326. [doi: 10.1002/mds.26693] [Medline: 27501026]
24. Yue S, Yang Y, Wang H, Rahul H, Katabi D. BodyCompass:Monitoring Sleep Posture with Wireless Signals. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2020 Jun 15 Presented at: ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies; June 15, 2020; New York, USA p. 1-25. [doi: 10.1145/3397311]
25. Tsanas A, Little M, McSharry P, Ramig L. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. IEEE Trans Biomed Eng 2010 Apr;57(4):884-893. [doi: 10.1109/tbme.2009.2036000]
26. Mobile fact sheet. Pew Research Center, Internet & Technology. 2019. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-09-08]
27. Pascual S, Ravanelli M, Serrà J, Bonafonte A, Bengio Y. Learning problem-agnostic speech representations from multiple self-supervised tasks. In: Proceedings of the Annual Conference on Internatioal Speech Communication Association - INTERSPEECH International. 2019 Presented at: Annual Conference on Internatioal Speech Communication Association - INTERSPEECH International; Apr 6, 2019; Graz, Austria URL: http://arxiv.org/abs/1904.03416 [doi: 10.21437/interspeech.2019-2605]
28. Tsanas A, Little M, McSharry P, Ramig L. Using the cellular mobile telephone network to remotely monitor Parkinson's disease symptom severity. Semantic Scholar. URL: https://www.semanticscholar.org/paper/Using-the-cellular-mobile-telephone-network-to-%E2%80%9F-s-Tsanas-Little/17180d351b33f13f7fe681ddf8411cf6aad74b64 [accessed 2021-09-08]
29. Langevin R, Ali MR, Sen T, Snyder C, Myers T, Dorsey ER, et al. The PARK Framework for automated analysis of Parkinson's disease characteristics. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Jun 21;3(2):1-22. [doi: 10.1145/3328925]
30. Penn Phonetics Lab Forced Aligner Toolkit (P2FA) for Python3. GitHub. URL: https://github.com/jaekookang/p2fa_py3 [accessed 2021-09-02]
31. HTK Speech Recognition Toolkit. URL: http://htk.eng.cam.ac.uk/ [accessed 2021-09-02]

32. Yuan J, Liberman M. Speaker identification on the SCOTUS corpus. J Acoust Soc Am 2008 May;123(5):3878. [doi: 10.1121/1.2935783]
33. Moltu C, Stefansen J, Svisdahl M, Veseth M. Negotiating the coresearcher mandate - service users' experiences of doing collaborative research on mental health. Disabil Rehabil 2012;34(19):1608-1616. [doi: 10.3109/09638288.2012.656792] [Medline: 22489612]
34. Jadoul Y, Thompson B, de Boer B. Introducing Parselmouth: A Python interface to Praat. J Phonetics 2018 Nov;71:1-15 [FREE Full text] [doi: 10.1016/j.wocn.2018.07.001]
35. Robust parsimonious selection of dysphonia measures for telemonitoring of Parkinson's disease symptom severity. Semantic Scholar. URL: https://www.semanticscholar.org/paper/Robust-parsimonious-selection-of-dysphonia-measures-Tsanas-Little/43b91147f0b5ab900b29ea7943d58fe761f0e26b [accessed 2019-10-14]
36. Little MA, McSharry PE, Roberts SJ, Costello DA, Moroz IM. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. BioMed Eng OnLine 2007;6(1):23. [doi: 10.1186/1475-925x-6-23]
37. Goldberger A. A Course in Econometrics. Cambridge, MA: Harvard University Press; 1991:1-426.
38. Müller M. Information Retrieval for Music and Motion. Switzerland: Springer; 2007.
39. Tsanas A, Little MA, McSharry PE, Spielman J, Ramig LO. Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. IEEE Trans Biomed Eng 2012 May;59(5):1264-1271. [doi: 10.1109/tbme.2012.2183367]
40. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995 Sep;20(3):273-297. [doi: 10.1007/bf00994018]
41. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13 - 17, 2016; San Francisco California p. 785-794. [doi: 10.1145/2939672.2939785]
42. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. Lightgbm: A highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst 2017:3154 [FREE Full text]
43. Ho T. Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition. 1995 Presented at: 3rd International Conference on Document Analysis and Recognition; Aug. 14-16, 1995; Montreal, QC, Canada. [doi: 10.1109/icdar.1995.598994]
44. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. J Artif Intell Res 2002 Jun 01;16:321-357. [doi: 10.1613/jair.953]
45. Nguyen HM, Cooper EW, Kamei K. Borderline over-sampling for imbalanced data classification. Int J Knowl Eng Soft Data Paradig Inderscience Publishers 2011;3(1):4-21. [doi: 10.1504/ijkesdp.2011.039875]
46. Shapley L. A value for n-person games. Contrib Theory Games 1953;2(28):307-317. [doi: 10.1017/cbo9780511528446.003]
47. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2020 Jan 17;2(1):56-67 [FREE Full text] [doi: 10.1038/s42256-019-0138-9] [Medline: 32607472]
48. Bhattarai K, Prasad P, Alsadoon A, Pham L, Elchouemi A. Experiments on the MFCC application in speaker recognition using Matlab. In: Seventh International Conference on Information Science and Technology (ICIST). 2017 Presented at: Seventh International Conference on Information Science and Technology (ICIST); April 16-19, 2017; Da Nang, Vietnam p. 32-37. [doi: 10.1109/icist.2017.7926796]
49. Kinnunen T, Chernenko E, Tuononen M, Fränti P, Li H. Voice activity detection using MFCC features and support vector machine. In: Proceedings of the International Conference Speech and Computer (SPECOM07). 2007 Presented at: International Conference Speech and Computer (SPECOM07); 2007; Moscow, Russia p. 556-561 URL: http://cs.uef.fi/sipu/pub/svm-vad-SPECOM07.pdf
50. Tsanas A, Little MA, McSharry PE, Ramig LO. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. J R Soc Interface 2011 Jun 06;8(59):842-855 [FREE Full text] [doi: 10.1098/rsif.2010.0456] [Medline: 21084338]
51. Kinnunen T, Chernenko E, Tuononen M, Fränti P, Li H. Voice activity detection using MFCC features and support vector machine. In: Proceedings of the International Conference Speech and Computer (SPECOM07). 2007 Presented at: International Conference Speech and Computer (SPECOM07); 2007; Moscow, Russia p. 556-561 URL: http://cs.uef.fi/sipu/pub/svm-vad-SPECOM07.pdf
52. Michael's Story. Michael J Fox. URL: https://www.michaeljfox.org/michaels-story [accessed 2021-09-02]
53. McCallum K. Muhammad Ali's advocacy for Parkinson's disease endures with boxing legacy. Parkinson's News Today. 2016. URL: https://parkinsonsnewstoday.com/2016/06/10/muhammad-alis-advocacy-parkinsons-disease-endures-boxing-legacy/ [accessed 2021-09-02]
54. Van Den Eeden SK, Tanner C, Bernstein A, Fross R, Leimpeter A, Bloch D, et al. Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity. Am J Epidemiol 2003 Jun 01;157(11):1015-1022. [doi: 10.1093/aje/kwg068] [Medline: 12777365]
55. Haaxma CA, Bloem BR, Borm GF, Oyen WJ, Leenders KL, Eshuis S, et al. Gender differences in Parkinson's disease. J Neurol Neurosurg Psychiatry 2007 Aug 01;78(8):819-824 [FREE Full text] [doi: 10.1136/jnnp.2006.103788] [Medline: 17098842]

56. Poorjam AH, Kavalekalam MS, Shi L, Raykov JP, Jensen JR, Little MA, et al. Automatic quality control and enhancement for voice-based remote Parkinson's disease detection. Speech Commun 2021 Mar;127:1-16. [doi: 10.1016/j.specom.2020.12.007]
57. Little M, Badawy R. Causal bootstrapping. arXiv. 2019. URL: https://arxiv.org/abs/1910.09648 [accessed 2021-09-08]

Abbreviations

AUC: area under the curve
HNR: harmonic-to-noise ratio
MDS-UPDRS: Movement Disorder Society-Unified Parkinson's Disease Rating Scale
MFCC: mel-frequency cepstral coefficient
P2FA: Penn Phonetics Lab Forced Aligner Toolkit
PARK: Parkinson's Analysis with Remote Kinetic-tasks
PASE: problem agnostic speech encoder
PD: Parkinson disease
PPE: pitch period entropy
ROC: receiver operating characteristics
RPDE: recurrence period density entropy
SHAP: Shapley additive explanations
SVM: support vector machine

Edited by R Kukafka, G Eysenbach; submitted 19.02.21; peer-reviewed by M Goni, C Fincham, D Zhai; comments to author 03.04.21; revised version received 13.04.21; accepted 07.08.21; published 19.10.21.

Please cite as:
Rahman W, Lee S, Islam MS, Antony VN, Ratnu H, Ali MR, Mamun AA, Wagner E, Jensen-Roberts S, Waddell E, Myers T, Pawlik M, Soto J, Coffey M, Sarkar A, Schneider R, Tarolli C, Lizarraga K, Adams J, Little MA, Dorsey ER, Hoque E
Detecting Parkinson Disease Using a Web-Based Speech Task: Observational Study
J Med Internet Res 2021;23(10):e26305
URL: https://www.jmir.org/2021/10/e26305
doi:10.2196/26305
PMID:34665148

Original Paper

Learning About the Current State of Digital Mental Health Interventions for Canadian Youth to Inform Future Decision-Making: Mixed Methods Study

Jessica Kemp[1,2], BSc; Jill Chorney[3,4], PhD, RPsych; Iman Kassam[1], BSc, MHI; Julie MacDonald[5], PhD, RPsych; Tara MacDonald[6], MSc; Lori Wozney[7], PhD; Gillian Strudwick[1,2], PhD, RN, FAMIA

[1]Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada

[2]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada

[3]Mental Health and Addictions, IWK Health, Halifax, NS, Canada

[4]Department of Psychiatry, Dalhousie University, Halifax, NS, Canada

[5]Mental Health and Addictions, Nova Scotia Health, Sydney, NS, Canada

[6]Mental Health and Addictions, Nova Scotia Health, Port Hawkesbury, NS, Canada

[7]Mental Health and Addictions, Policy and Planning, Nova Scotia Health, Dartmouth, NS, Canada

Corresponding Author:
Gillian Strudwick, PhD, RN, FAMIA
Campbell Family Mental Health Research Institute
Centre for Addiction and Mental Health
1001 Queen St W
Toronto, ON, M6J 1H4
Canada
Phone: 1 416 535 8501
Email: gillian.strudwick@camh.ca

Abstract

Background: The COVID-19 pandemic has increased the demand for youth mental health services in Canada as disruptions to clinical care continue to persist due to the risk of transmission and exposure to the virus. Digital mental health interventions, including web-based resources and mobile apps, have provided opportunities to support youth mental health remotely across Canada. There is a need to better understand how these digital interventions are being selected, recommended, and used in various regions across Canada.

Objective: A national jurisdictional scan was completed to (1) determine what web-based programs, apps, and websites are promoted and licensed in Canada for youth mental health; (2) identify criteria and decision-making processes that Canadian jurisdictions use to select web-based programs, apps, and websites for youth mental health; and (3) identify upcoming trends, innovations, and digital mental health possibilities that are emerging in the youth sector.

Methods: The aims of the jurisdictional scan were addressed through a review of related academic and grey literature; stakeholder interviews, including individuals involved in various areas of the youth mental health sector; and a social media review of pertinent Twitter content.

Results: A total of 66 web-based resources and apps were identified for use by youth in Canada. 16 stakeholder interviews were completed and included discussions with researchers, clinicians, youth organizations, and others involved in digital interventions for youth mental health. These discussions identified a limited use of frameworks used to guide decision-making processes when selecting digital interventions. Many clinicians agreed on a similar set of eligibility requirements for youth mental health apps and digital resources, such as the evidence base and cultural relevance of the intervention. Stakeholders also identified upcoming trends and innovations in the youth digital mental health space, including artificial intelligence, digital phenotyping, and personalized therapy. Over 4 weeks, 2184 tweets were reviewed to identify and compare global and national trends and innovations involving digital mental health and youth. Key trends included the promotion of regional chat services as well as the effects of the COVID-19 pandemic on youth mental health and access to care.

Conclusions: As organizations begin to plan for the delivery of mental health care following the pandemic, there are concerns about the sustainability of these digital mental health interventions as well as a need for services to be more informed by the experiences and preferences of youth.

(J Med Internet Res 2021;23(10):e30491) doi:10.2196/30491

KEYWORDS

youth mental health; digital mental health; COVID-19; digital mental health interventions; e-mental health

Introduction

Canadian jurisdictions have significantly altered their delivery of mental health care for youth over the past year, given the rising need for mental health services related to the COVID-19 pandemic [1]. A Canadian study indicated that 67% to 70% of children and adolescents aged 6-18 years have experienced a deterioration in one or more mental health domains because of changes such as stress and social isolation during the pandemic [1]. Organizations have broadened their use of virtual care and digital mental health interventions, including web-based programs, apps, and websites [2]. In consideration of the rapid shift to the use of digital mental health services and interventions, this review can support Canadian decision-makers and leaders in understanding these changes and how youth mental health is supported through digital interventions. To accomplish this, a jurisdictional scan was completed to address current needs and future directions in the youth mental health space.

To identify digital mental health resources, criteria, and decision-making processes used to select resources, and upcoming trends and innovations, the following aims were established:

1. Determine what web-based programs, apps, and websites are promoted and/or licensed by Canadian jurisdictions (eg, health authorities, national youth mental health organizations, provinces, and territories) for youth mental health.
2. Identify criteria and decision-making processes that Canadian jurisdictions use to select web-based programs, apps, and websites for youth mental health.
3. Identify upcoming trends, innovations, and digital mental health possibilities that are emerging in the youth sector.

Methods

A review of current academic and grey literature was completed, including searches of databases (eg, PubMed, University of Toronto Libraries), reports, working papers, and presentations from Canadian digital health organizations. Titles and abstracts were reviewed to determine if information about youth mental health and digital interventions was included. Sources that met this inclusion criterion were then subjected to rapid content analysis and were organized based on identified themes. App stores, including the Google Play Store and Apple App Store, were also reviewed to identify digital mental health apps currently available to youth in Canada.

A series of key stakeholder interviews were conducted with representatives across Canada. Single interviews were conducted by two members of the research team via videoconferencing or telephone and followed a semistructured design based on a set of six questions. Participants were asked to discuss the criteria their organization uses to select resources for youth or decision-making processes/appraisals that are used to evaluate digital resources such as websites and mobile apps. Trending and emerging innovations were also discussed, and all participants were able to provide additional information relevant to the topic. One research member took notes during each interview, and this information was later used for thematic analysis to identify key themes.

Data mining techniques and content analysis were used to retrieve and analyze information collected from the Twitter application programming interface over 4 weeks using a Twitter Archiving Google Sheet (TAGS). As Twitter is a platform commonly used by researchers, clinicians, youth organizations, health care leaders, and others involved in youth mental health care, Twitter data provided a snapshot of in-the-moment insights into trending topics and innovations related to digital mental health interventions for youth during the COVID-19 pandemic. These trends were identified to understand the similarities or disparities between the experiences of national and international stakeholders.

Further information about the methodology used for this research is available upon request from the corresponding author.

Results

Review of Web-Based Resources, Apps, and Related Literature

A total of 66 digital interventions were identified as promoted or licensed by Canadian jurisdictions and organized based on their delivery methods, including web-based resources or mobile apps. Information was recorded for each resource based on the American Psychiatric Associations App Evaluation Model [3]. A complete list of the digital interventions, including web-based resources, and associated criteria and tags are available in Multimedia Appendix 1, and the mobile apps are provided in Multimedia Appendix 2.

The review of related literature identified current trends in the youth digital mental health space. A common theme that was highlighted throughout the review of the literature was the limited availability of evidence-based apps. Numerous reviews revealed that formal evaluation methods were often not used to support the use of mobile apps for youth mental health, raising concerns about the efficacy of digital interventions used in clinical settings [4-8]. Researchers recommended that more

rigorous testing be done to provide high-quality evidence for the use of digital interventions for youth mental health [8].

Research studies also reviewed the use of evaluation methods for digital mental health interventions for youth and provided various metrics used to quantify effectiveness and improvements to mental health [9-11]. A systematic review included metrics such as mental health outcome data, health service outcome data, technology use, and feedback from children and youth [9]. Other evaluation methods included clinical outcomes of randomized control trials to determine the effectiveness of digital mental health interventions [10].

Co-design was identified as a powerful tool in the development of digital mental health interventions for children and youth. Researchers indicated that there is a need for co-design with youth, including "creative collaboration, varying stages of involvement, collaboration activities, user-engagement techniques, prototype development, and consideration for diversity within youth populations" [12].

Most of the articles focused on the development and use of mobile apps targeting child and youth experiences with anxiety and depression [13-16]. There was a paucity of research on other types of digital mental health interventions as well as any digital interventions for eating disorders and many complex mental illnesses, with only one article identified on youth experiences with psychosis [17].

Gamification was also identified as a prominent topic in the results of the review. Some considerations were discussed that can address the needs of specific youth groups who may benefit from the use of personalized game-based interventions [18-21]. Some considerations include the integration of video games with clinical care processes and programming, the use of gender-diverse characters in game-based digital mental health interventions, and a focus on the prevention of mental illness as opposed to treatment-based games [18].

Stakeholder Interviews

A total of 16 stakeholders, including clinicians, researchers, youth organizations, online therapists, business analysts, and others involved in the youth mental health space, took part in the interviews. Organizations included representation from numerous regions within Canada, including Ontario, British Columbia, Alberta, and Nova Scotia, as well as the US state of Massachusetts. Stakeholders discussed the digital mental health interventions they currently use, that their organization endorses, or that they are involved in for research purposes.

Additionally, the stakeholders discussed the need for improved guidance when selecting and appraising digital mental health interventions for youth. Overall, most respondents relied on individual selection criteria to inform their decision-making processes as opposed to standardized frameworks. Table 1 includes a list of specific appraisal tools referenced by participants as well as individual criteria cited by respondents as informing their decision-making processes.

Table 1. Decision-making processes and appraisal methods used for selecting digital mental health interventions for youth identified during stakeholder interviews.

Process or appraisal method	The rationale for process or appraisal
Frameworks	
Mental Health Commission of Canada Toolkit	The toolkit can reduce clinician workload, ensure resources are evidence-based, increase interoperability, and provide valuable information for those with limited experience with digital mental health.
Homewood Research Institute App Evaluation Framework	Used to evaluate the effectiveness of mental health apps and identifies the highest quality apps. Focuses on measuring outcomes, design and transparency, design methodology, and operational considerations of each application.
Selection criteria	
Evidence base	Resources that are evidence-based and have supporting literature are often more accepted for use in clinical settings.
Accessibility	Providers and researchers must consider the cost and technology requirements associated with recommending a digital intervention.
Visual appeal	Mobile apps and web-based resources should host bright and catchy graphics, including music and characters, to grab the attention of youth.
Feedback from youth	Recommendations should be developed based on feedback from youth about what they are looking for and the apps they are already using.
Cultural relevance	When serving youth, including BIPOC,[a] further considerations must be made to ensure that the digital resource/intervention considers specific needs of various cultural and racial minority groups.

[a]BIPOC: Black, Indigenous, and people of color.

The discussions with stakeholders also identified new trends and innovations these individuals and organizations have seen in the youth digital mental health space. Table 2 highlights the most frequently cited innovations considered in response to the COVID-19 pandemic and other trends currently being studied in Canada.

Table 2. Upcoming trends and innovations in digital youth mental health identified through stakeholder interviews.

Trend or innovation	Description
Artificial intelligence/ machine learning	Including mobile apps that use artificial intelligence–powered chatbots.
Digital phenotyping	Tracking the activities of users to determine how digital activity might indicate a change in behavior or mental state.
Personalized therapy	Interventions that consider culture, language, geography, and more.
Interoperability	Using data from mobile apps alongside clinical data from health information systems.
Online peer support groups	Online groups moderated by health professionals to create safer digital environments for youth.
Trauma-informed interventions	Emerging mobile apps that target specific experiences with trauma.
Increasing accessibility	Potential to reach a greater number of children and youth through digital means alongside traditional in-person care.
Gamification	Development of video and digital games used to provide support and education about mental health and wellness.
Social media	The use of platforms such as Instagram and Twitter to help organizations reach youth and promote the availability of digital health interventions and resources. Instagram was also used in specific cases by youth organizations to provide mental health support through the application's direct messaging feature.
TikTok	Includes TikTok accounts that provide short and accessible videos about mental health and wellness that can easily reach youth who are already using the app.

Social Media Review

A total of 2184 tweets containing content or hashtags relating to digital interventions for youth mental health were included. Of these tweets, themes included promotion of regional chat services, such as text and telephone lines available 24 hours per day, 7 days per week for youth mental health support. Effects of isolation due to COVID-19 on youth mental health were highlighted, primarily concerning families who have experienced suicide and self-harm in youth [22]. A greater need for access to mental health care and reduction of hospital beds available for youth mental health due to staff shortages in New Brunswick was also a prominent topic [23]. Web-based resource hubs, webinars, and virtual panel discussions provided by youth mental health organizations were frequently promoted. Advocacy for government funding due to an increasing national need for youth mental health care was also discussed. Lastly, resources specifically designed to meet the needs of lesbian, gay, bisexual, transgender, queer, and two-spirit (LGBTQ2S+) youth were also promoted on the platform through youth-friendly infographics.

Discussion

Future Directions

Although digital mental health interventions can provide great benefit during the COVID-19 pandemic, there is concern among some clinicians and researchers about the sustainability of these interventions once in-person care resumes across Canada. As digital mental health interventions are developed for youth, there needs to be increased awareness of how these interventions are being selected and the opportunities to better integrate these interventions with clinical care. One consideration includes improved transparency and quality considerations in how digital mental health interventions are selected, as shown in Table 1. Individuals involved in the development of digital resources for youth mental health must consider these selection criteria, such as cost, technology, privacy requirements, evidence base, visual appeal, and cultural relevance. Engagement with youth in planning and improving the integration of digital health with organizations across the mental health system continuum is vital to ensuring the longevity of these resources following the pandemic. Additionally, further research should be done to review content related to digital mental health on other social media platforms that are frequently used by youth, including Instagram and TikTok.

Conclusions

Current literature, key stakeholder interviews, and mining of a social networking platform all point to the need for a greater understanding of digital intervention use for youth mental health during and beyond the COVID-19 pandemic. The use of digital interventions such as web-based resources and apps can be driven using suitable frameworks and selection criteria. Upcoming trends and innovations indicate an opportunity to improve and expand the use of digital interventions for youth in Canada.

Acknowledgments

The authors would like to acknowledge the funding support provided by Nova Scotia Health and their involvement alongside IWK Health in sharing their perspectives and thoughts throughout the completion of this jurisdictional scan.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Web-based resources identified for youth mental health in Canada, including criteria and tags.
[XLSX File (Microsoft Excel File), 25 KB - jmir_v23i10e30491_app1.xlsx]

Multimedia Appendix 2
Mobile apps identified for youth mental health in Canada, including criteria and tags.
[XLSX File (Microsoft Excel File), 20 KB - jmir_v23i10e30491_app2.xlsx]

References

1. Cost KT, Crosbie J, Anagnostou E, Birken CS, Charach A, Monga S, et al. Mostly worse, occasionally better: impact of COVID-19 pandemic on the mental health of Canadian children and adolescents. Eur Child Adolesc Psychiatry 2021 Feb 26:1-14 [FREE Full text] [doi: 10.1007/s00787-021-01744-3] [Medline: 33638005]
2. Davenport TA, Cheng VWS, Iorfino F, Hamilton B, Castaldi E, Burton A, et al. Flip the Clinic: a digital health approach to youth mental health service delivery during the COVID-19 pandemic and beyond. JMIR Ment Health 2020 Dec 15;7(12):e24578 [FREE Full text] [doi: 10.2196/24578] [Medline: 33206051]
3. Torous JB, Chan SR, Gipson SYT, Kim JW, Nguyen T, Luo J, et al. A hierarchical framework for evaluation and informed decision making regarding smartphone apps for clinical care. Psychiatr Serv 2018 May 01;69(5):498-500. [doi: 10.1176/appi.ps.201700423] [Medline: 29446337]
4. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
5. Punukollu M, Marques M. Use of mobile apps and technologies in child and adolescent mental health: a systematic review. Evid Based Ment Health 2019 Oct 23;22(4):161-166. [doi: 10.1136/ebmental-2019-300093] [Medline: 31358537]
6. Lo B, Shi J, Hollenberg E, Abi-Jaoudé A, Johnson A, Wiljer D. Surveying the role of analytics in evaluating digital mental health interventions for transition-aged youth: scoping review. JMIR Ment Health 2020 Jun 25;7(6):e15942 [FREE Full text] [doi: 10.2196/15942] [Medline: 32348261]
7. Babbage C, Jackson GM, Nixon E. Desired features of a digital technology tool for self-management of well-being in a nonclinical sample of young people: qualitative study. JMIR Ment Health 2018 Dec 18;5(4):e10067 [FREE Full text] [doi: 10.2196/10067] [Medline: 30563820]
8. Clarke AM, Kuosmanen T, Barry MM. A systematic review of online youth mental health promotion and prevention interventions. J Youth Adolesc 2015 Jan 01;44(1):90-113. [doi: 10.1007/s10964-014-0165-0] [Medline: 25115460]
9. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual Research Review: digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Mar 27;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
10. Bidargaddi N, Musiat P, Winsall M, Vogl G, Blake V, Quinn S, et al. Efficacy of a web-based guided recommendation service for a curated list of readily available mental health and well-being mobile apps for young people: randomized controlled trial. J Med Internet Res 2017 May 12;19(5):e141 [FREE Full text] [doi: 10.2196/jmir.6775] [Medline: 28500020]
11. Alvarez-Jimenez M, Rice S, D'Alfonso S, Leicester S, Bendall S, Pryor I, et al. A novel multimodal digital service (moderated online social therapy+) for help-seeking young people experiencing mental ill-health: pilot evaluation within a national youth e-mental health service. J Med Internet Res 2020 Aug 13;22(8):e17155 [FREE Full text] [doi: 10.2196/17155] [Medline: 32788151]
12. Bevan Jones R, Stallard P, Agha SS, Rice S, Werner-Seidler A, Stasiak K, et al. Practitioner review: co-design of digital mental health technologies with children and young people. J Child Psychol Psychiatry 2020 Jun 22;61(8):928-940. [doi: 10.1111/jcpp.13258] [Medline: 32572961]
13. Thabrew H, Stasiak K, Hetrick SE, Wong S, Huss JH, Merry SN. E-Health interventions for anxiety and depression in children and adolescents with long-term physical conditions. Cochrane Database Syst Rev 2018 Aug 15;8:CD012489 [FREE Full text] [doi: 10.1002/14651858.CD012489.pub2] [Medline: 30110718]
14. Lim MH, Rodebaugh TL, Eres R, Long KM, Penn DL, Gleeson JFM. A pilot digital intervention targeting loneliness in youth mental health. Front Psychiatry 2019 Aug 23;10:604 [FREE Full text] [doi: 10.3389/fpsyt.2019.00604] [Medline: 31507469]
15. Litvin S, Maier M. How mHealth programmes can treat depression: a randomized controlled trial. Zenodo. 2019 Jun 01. URL: https://doi.org/10.5281/zenodo.3236691 [accessed 2021-09-21]
16. Rebedew D. Five mobile apps to help patients with anxiety and depression. American Academy of Family Physicians. 2018 Mar 01. URL: https://www.aafp.org/fpm/2018/0300/fpm20180300oa1.pdf [accessed 2021-09-21]

17. Peck CE, Lim MH, Purkiss M, Foley F, Hopkins L, Thomas N. Development of a lived experience-based digital resource for a digitally-assisted peer support program for young people experiencing psychosis. Front Psychiatry 2020 Jul 02;11:635 [FREE Full text] [doi: 10.3389/fpsyt.2020.00635] [Medline: 32714223]
18. Strauss P, Morgan H, Wright Toussaint D, Lin A, Winter S, Perry Y. Trans and gender diverse young people's attitudes towards game-based digital mental health interventions: a qualitative investigation. Internet Interv 2019 Sep 05;18:100280 [FREE Full text] [doi: 10.1016/j.invent.2019.100280] [Medline: 31890628]
19. Fox LG, Barnes D. BoosterBuddy: using gamification as a compensatory strategy for motivational deficits. Psychiatr Serv 2016 Jan 01;67(1):141-142. [doi: 10.1176/appi.ps.661101] [Medline: 26725495]
20. Ferrari M, Suzanne A. Youth mental health, family practice, and knowledge translation video games about psychosis: family physicians' perspectives. J Can Acad Child Adolesc Psychiatry 2017 Oct 01;26(3):184-189 [FREE Full text] [Medline: 29056980]
21. Litvin S, Saunders R, Maier MA, Lüttke S. Gamification as an approach to improve resilience and reduce attrition in mobile mental health interventions: a randomized controlled trial. PLoS One 2020 Sep 02;15(9):e0237220 [FREE Full text] [doi: 10.1371/journal.pone.0237220] [Medline: 32877425]
22. Cribb R. They had to write an obituary for their 16-year-old son. This is what Saul's parents want you to know about the pandemic's effect on youth mental health. Toronto Star. 2021 Mar 15. URL: https://www.thestar.com/news/investigations/2021/03/15/they-had-to-write-an-obituary-for-their-16-year-old-son-this-is-what-sauls-parents-want-you-to-know-about-the-pandemics-effect-on-youth-mental-health.html [accessed 2021-03-29]
23. Steeves S. Staff shortages force bed closures at youth mental health unit at Moncton Hospital. Global News. 2021 Mar 12. URL: https://globalnews.ca/news/7693606/nb-youth-mental-health-unit-closure/ [accessed 2021-03-29]

Abbreviations

LGBTQ2S+: lesbian, gay, bisexual, transgender, queer, and two-spirit
TAGS: Twitter Archiving Google Sheet

Edited by R Kukafka; submitted 17.05.21; peer-reviewed by E Stafford, A Alturkistani, R Poss-Doering; comments to author 30.07.21; revised version received 17.08.21; accepted 12.09.21; published 19.10.21.

Please cite as:
Kemp J, Chorney J, Kassam I, MacDonald J, MacDonald T, Wozney L, Strudwick G
Learning About the Current State of Digital Mental Health Interventions for Canadian Youth to Inform Future Decision-Making: Mixed Methods Study
J Med Internet Res 2021;23(10):e30491
URL: https://www.jmir.org/2021/10/e30491
doi:10.2196/30491
PMID:34665141

Original Paper

An Enhanced SMS Text Message–Based Support and Reminder Program for Young Adults With Type 2 Diabetes (TEXT2U): Randomized Controlled Trial

Timothy Middleton[1,2,3], BSc, MBBS, MPhil; Maria Constantino[1,2,3], PhD; Margaret McGill[1,2,3], MSc; Mario D'Souza[4], PhD; Stephen M Twigg[1,2,3], PhD; Ted Wu[1,2,3], PhD; Aravinda Thiagalingam[3,5,6], PhD; Clara Chow[3,5,6,7], PhD; Jencia Wong[1,2,3], PhD

[1]Diabetes Centre, Royal Prince Alfred Hospital, Camperdown, Australia

[2]Royal Prince Alfred Clinic, Charles Perkins Centre, The University of Sydney, Camperdown, Australia

[3]Faculty of Medicine and Health, University of Sydney, Camperdown, Australia

[4]Sydney Local Health District Clinical Research Centre, Camperdown, Australia

[5]Department of Cardiology, Westmead Hospital, Westmead, Australia

[6]Westmead Applied Research Centre, The University of Sydney, Westmead, Australia

[7]Cardiovascular Division, The George Institute for Global Health, Sydney, Australia

Corresponding Author:
Timothy Middleton, BSc, MBBS, MPhil
Diabetes Centre
Royal Prince Alfred Hospital
Missenden Road
Camperdown, 2050
Australia
Phone: 61 295155888
Fax: 61 295155820
Email: timothy.middleton@health.nsw.gov.au

Abstract

Background: Clinic attendance, metabolic control, engagement in self-management, and psychological health are suboptimal in young-onset (age of onset <40 years) type 2 diabetes.

Objective: We examined the effectiveness of an enhanced SMS text message–based support and reminder program in improving clinic attendance, metabolic control, engagement in self-management, and psychological health in young-onset type 2 diabetes.

Methods: A 12-month, parallel-arm, randomized controlled trial comparing an enhanced, semipersonalized SMS text message–based intervention (incorporating 1-8 supportive and/or informative text messages per month) against standard care was conducted in a specialized clinic for young adult type 2 diabetes. The primary outcome was maintenance of 100% attendance at scheduled quarterly clinical appointments. Secondary outcomes included (1) metabolic indices, (2) pathology and self-monitored blood glucose (SMBG) data availability, and (3) psychosocial well-being.

Results: A total of 40 participants were randomized, and 32 completed their 12-month study visit. The average participant age was 32.7 (SD 5.1) years, 50% (20/40) were male, and baseline glycated hemoglobin A_{1c} (HbA_{1c}) was 7.3% (SD 1.9%) (56 mmol/mol, SD 20). A higher proportion of the intervention group achieved 100% attendance (12/21, 57%, vs 5/19, 26%, for the control group); Kaplan-Meier analysis demonstrated significantly greater cumulative attendance in the intervention group (*P*=.04). There were no between-group differences in HbA_{1c}, BMI, lipids, or availability of pathology and SMBG data. Odds of recording an improvement in the Diabetes Empowerment Scale–Short Form score were higher in the intervention group at 6 months (odds ratio [OR] 4.3, 95% CI 1.1-17), with attenuation of this effect at study end (OR 3.1, 95% CI 0.9-11). Program acceptability was high; >90% of participants would recommend the program to new patients.

Conclusions: An enhanced SMS text message–based support and reminder program doubled scheduled clinic attendance rates for patients with young-onset type 2 diabetes. The program was highly acceptable and provided early support for patient empowerment but had no significant effect on measures of metabolic control or self-management.

Trial Registration: Australian and New Zealand Clinical Trials Registry (ACTRN12618000479202); https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=373579

(J Med Internet Res 2021;23(10):e27263) doi:10.2196/27263

KEYWORDS

young-onset type 2 diabetes; SMS; clinic attendance; engagement; diabetes; digital health; mobile health; adolescents

Introduction

Over the past 20 years, young-onset type 2 diabetes (age of onset <40 years) has emerged as a significant clinical problem. The worldwide prevalence of young-onset type 2 diabetes more than doubled during the decade spanning 2003 to 2013, and today, well in excess of 60 million young people are living with type 2 diabetes [1,2]. This is particularly concerning given the aggressive nature of young-onset type 2 diabetes, with high rates of complications, established premature mortality, and the poorest comparative outcomes [3-6].

Suboptimal clinic attendance and poor engagement in self-management represent significant barriers to improving long-term outcomes [7,8]. There is specific evidence that clinic attendance associates with favorable outcomes in young-onset type 2 diabetes; during the landmark Treatment Options for Type 2 Diabetes in Adolescents and Youth (TODAY) study of adolescent type 2 diabetes, it was established that participants who attended ≥75% of scheduled lifestyle education sessions achieved significantly greater weight reductions than those who attended less often [9]. An audit of clinic attendance at a metropolitan, hospital-based diabetes center clearly demonstrated that young adults with type 2 diabetes were the subgroup that had the highest rate of nonattendance; median loss to follow-up time was approximately 4 months [10]. As each clinical encounter provides an excellent opportunity to review a patient's understanding of diabetes and its management, convey important messages relating to self-management, adjust pharmacotherapy, perform complications screening, and foster a therapeutic rapport, novel ideas are needed to motivate young adults with type 2 diabetes and encourage more reliable clinic attendance.

In recent years, innovative text messaging interventions have yielded promising results in high-risk groups. The landmark TEXTME study demonstrated that a structured, lifestyle-based SMS text messaging program could improve cardiovascular risk factor management after myocardial infarction [11]. In adolescent medicine, SMS text message interventions have been used to address important issues, including teenage pregnancy and binge drinking of alcohol [12,13]. In the field of diabetes, the SMS4BG study established that an SMS text messaging program could facilitate modest improvement in glycemic control [14]. However, the SMS4BG study did not specifically examine the impact of the program on clinic attendance, nor did it specifically target individuals with young-onset type 2 diabetes. A study from Hong Kong did identify a benefit of an SMS text message reminder for the attendance rate within a type 2 diabetes population at a primary care clinic [15]. However, participants in the Hong Kong study had a mean age of 63 years; as a result, the benefit of SMS text message reminders in the young-onset type 2 diabetes setting remains to be established. Overall, young adults with type 2 diabetes have been underrepresented in most clinical studies to date [16]. Consequently, high-quality evidence to guide management of young-onset type 2 diabetes remains scarce.

It is widely understood that young people are frequent users of information and communications technology. In 2012 and 2017, young adult respondents to our diabetes and technology survey reported the highest rates of mobile phone ownership and a high degree of comfort with the use of technology in diabetes management [17]. Familiarity with SMS text messaging technology as well as its low cost and ready scalability provide a great impetus for an exploration of the utility of SMS text message interventions in the young adult demographic. We hypothesized that an enhanced SMS text message–based support and reminder program would improve clinic attendance, metabolic health, and engagement in diabetes self-management and psychosocial well-being within our young-onset type 2 diabetes cohort. We set out to test this hypothesis in the TEXT2U randomized controlled trial (RCT).

Methods

Study Design

TEXT2U was a parallel group, investigator-blinded, 12-month RCT involving participants with young-onset type 2 diabetes (age of onset 18-40 years). The primary aim was to assess the effect of an enhanced SMS text message–based support and reminder program on clinic attendance. Secondary aims were to assess (1) metabolic outcomes, (2) diabetes self-management practices as evidenced by availability of self-monitoring of blood glucose (SMBG) data and pathology results, and (3) the psychosocial impact of the intervention as measured by change in validated diabetes-specific questionnaire scores (Problem Areas in Diabetes 5-item Short Form (PAID-5) [18], Diabetes Empowerment Scale–Short Form (DES-SF) [19], and Type 2 Diabetes Stigma Assessment Scale (DSAS-2) [20]). The study was approved by the Sydney Local Health District Ethics Review Committee and registered with the Australian and New Zealand Clinical Trials Registry (ACTRN12618000479202).

Study Setting and Participant Recruitment

The Royal Prince Alfred Hospital (RPAH) Diabetes Centre is a secondary and tertiary referral center that services an ethnically diverse population in metropolitan Sydney, Australia. Individuals were eligible to enter the study if they had an established type 2 diabetes diagnosis, had access to a mobile phone, and were aged 18-40 years at time of enrolment. Individuals with insufficient proficiency in English to read the text messages in the study were excluded. All participants

provided written informed consent prior to study entry. Recruitment was performed between March and August 2018.

Randomization

Block randomization was employed with a standard block size of 4 using a customized, computer-based program. Participants were allocated 1:1 to either the standard care control group or the standard care plus enhanced SMS text message–based support and reminder intervention group. Medical staff delivering patient care were blinded to study group allocation.

Interventions

Control: Clinic Standard of Care

The control group received our young-onset type 2 diabetes clinic standard of care, which is aligned with and realizes current recommendations for youth and young adults with type 2 diabetes [21]. It includes quarterly follow-up appointments with a diabetes educator and endocrinologist. Glycemic control was assessed via glycated hemoglobin A_{1c} (HbA_{1c}) and capillary-based SMBG results, weight and blood pressure (BP) were measured, and treatment adjustments were made at the discretion of the treating endocrinologist. Dietary review with an accredited dietitian was provided as required. All participants were given a personal blood glucose meter with a starter pack of glucose monitoring strips and a personal lancing device for SMBG. All individuals in the young adult type 2 diabetes clinic are enrolled in the Australian National Diabetes Services Scheme, which provides ongoing access to subsidized glucose testing strips and lancets.

After each clinic visit, participants were provided with details of their next follow-up appointment. A generic SMS text message reminder was sent prior to scheduled appointments via the hospital's outpatient management system. In the event of a missed clinic appointment, clinic staff made attempts to contact the relevant individual (via telephone and email) to reschedule a makeup appointment.

Intervention: Enhanced SMS Text Message–Based Support and Reminder Program

The enhanced SMS text message–based support and reminder program involved receipt of clinic standard of care as outlined above, plus a structured program of semipersonalized text messages. Message content was individualized on the basis of key baseline characteristics, including gender and smoking status. Message delivery was managed by an automated system developed and programmed by coauthor AT (TextQStream, Python 3.6, using the Pycap 1.02 library). This system was developed previously [22], and customization was performed in-house for the TEXT2U trial. Computer software was run through the University of Sydney Research Electronic Data Capture (REDCap) system [23]. Messages to the study participants were sent through a gateway interface over Australian telephone networks at no cost to individual participants. All SMS text messages sent during the study were logged by the system; the study log included the date and time that each message was delivered. All participants received an introductory message in the week following their baseline visit and then two messages per week for the first 2 months of the study. Text message frequency decreased to one message per week during the third month and then one text message per month thereafter. All study messages were sent at random times during standard business hours: Monday to Friday, 9 AM to 5 PM. Instructions on how to opt out of the messaging program were provided.

The SMS text messaging program was developed by a focus group of 10 diabetes specialists associated with the Royal Prince Alfred Hospital Diabetes Centre; adjustments were made following consumer review by young adults with diabetes. In particular, 15 young adults with diabetes (aged 18-40 years), who did not form part of the young adult type 2 diabetes clinic, were asked to complete a brief survey and each provide feedback on different subsets of messages that were to be featured in the study. Program messages were designed to contain a mix of supportive and informative content. Messages were entirely text-based; emojis and multimedia messaging were not incorporated into the program. A personalized appointment reminder was sent to participants in the enhanced SMS group in the week before each follow-up appointment. Examples of program text messages are included in Multimedia Appendix 1 (Table S1).

An option was provided for enhanced SMS group participants to engage with the study team via the study-specific SMS portal. Questions relating to diabetes and its management could be sent directly to the study team at a participant's convenience. The SMS portal was actively monitored by a research assistant, and all text messages received from participants were reviewed and replied to within 1 business day.

Trial Procedures

All participants underwent baseline assessment and were asked to attend quarterly follow-up visits for 12 months. Weight, BP, SMBG records, and routine pathology test results were assessed at each study visit. At baseline, all participants completed the Health Literacy Questionnaire [24]. Validated measures of psychosocial status, including the PAID-5 [18], DES-SF [19], and DSAS-2 [20], were self-completed by participants at baseline, 6 months, and 12 months. Participants assigned to the enhanced SMS intervention group were invited to complete an evaluation of the program via a standardized, study-specific questionnaire at the end of the study.

Statistical Analysis

From prior analyses [10], the median time to clinic dropout associated with our standard care approach in the young-onset type 2 diabetes clinic was 4 months. Previous studies comparing attendance rates in SMS text message reminder vs control interventions have demonstrated odds ratios (ORs) of 1.7-4.3 in favor of text messaging [25]. Given this information, we estimated that an enhanced SMS text message–based support and reminder system would increase the median time to clinic dropout in our young-onset type 2 diabetes cohort from 4 to 12 months. We determined that a 12-month study with quarterly clinic follow-up for 40 individuals (randomized 1:1 to the intervention and control groups) would have 80% power (α=.05) to enable us to reject the null hypothesis of no difference in loss to follow-up.

The primary outcome was evaluated using the principles of time to event analysis. For secondary outcomes, data collected at each follow-up visit (regardless of whether the visit was attended as scheduled or as a make-up appointment) were considered in the analyses. For those participants who failed to attend a follow-up appointment and did not attend a makeup appointment, missing data were not imputed. Participants were analyzed by original group assignment. Data from descriptive analyses are reported as mean (SD) or n (%) values. Comparisons between groups were made using analysis of variance for continuous variables and Pearson chi-squared tests for categorical variables.

The primary outcome of attendance at all scheduled follow-up appointments was analyzed using a Kaplan-Meier approach. Difference in scheduled follow-up attendance between the intervention and control groups was compared using the log-rank test. In addition to our primary attendance outcome, overall attendance (ie, attendance at scheduled appointments and rescheduled makeup appointments) was assessed.

With respect to secondary outcomes, metabolic indices at 12 months in the control and intervention groups were compared using the independent samples *t* test. Mean changes (from baseline) for metabolic indices in both the control and intervention groups were evaluated using the paired samples *t* test. Binary logistic regression analyses were performed to explore differences between the intervention and control groups in the log odds of observing the other secondary outcomes of interest. For the PAID-5, DES-SF, and DSAS-2 analyses, the outcome of interest was positive change in questionnaire score (ie, a score change that would indicate less diabetes distress and stigma or greater self-efficacy). The dependent variable for each model was the outcome of interest, and the independent variables included group assignment, follow-up time, and the interaction between group assignment and follow-up time. A general estimating equations framework was employed to allow within-participant correlations (in the context of repeated measures) to be taken into account. Data collected from participants who attended make-up appointments were considered in secondary outcome analyses. These data were ascribed to the nearest scheduled follow-up appointment.

Statistical analyses were performed using SPSS, version 24.0 (IBM Corporation). All statistical tests were two-tailed and were conducted at the .05 significance level.

Results

Between March and August 2018, 41 patients of the RPAH Diabetes Centre young-onset type 2 diabetes clinic were approached regarding study participation; 40 were enrolled and randomized, and 32 participants completed a 12-month follow-up appointment (Multimedia Appendix 2, Figure S1). The mean age of the study cohort was 32.7 years, 50% (20/40) of participants were male, mean HbA_{1c} was 7.3% (56 mmol/mol) and mean duration of diabetes at the time of study entry was 6.4 years. Baseline characteristics were similar in the enhanced SMS and standard of care groups (Table 1). Notably, there were no between-group differences in glycemic control, diabetic pharmacotherapy use, or health literacy measures.

Table 1. Baseline characteristics of the study cohort (N=40).

Characteristic	Values		*P* value
	Intervention group (n=21)	Control group (n=19)	
Male gender, n (%)	10 (48)	10 (53)	.99
Age (years), mean (SD)	33.0 (5.8)	32.4 (4.4)	.71
Ethnic background, n (%)			.38
East/Southeast Asian	8 (38)	3 (16)	
Subcontinental	5 (24)	7 (37)	
European	4 (19)	6 (32)	
Other	4 (19)	3 (16)	
Nonsmoker, n (%)	11 (52)	14 (74)	.17
Education, n (%)			.14
TAFE[a]/university	19 (91)	13 (68)	
High school	2 (10)	5 (26)	
Not disclosed	0 (0)	1 (5)	
Duration of diabetes (years), mean (SD)	7.6 (6.2)	5.0 (5.9)	.18
Diabetes treatment, n (%)			.32
Diet alone	2 (10)	5 (26)	
Oral hypoglycemic medication	14 (67)	8 (42)	
Insulin oral hypoglycemic agent	5 (24)	6 (32)	
HbA_{1c}[b] (%), mean (SD)	7.2 (1.6)	7.3 (2.1)	
HbA_{1c} range (%), n (%)			.94
<6.5	9 (43)	9 (47)	
6.5-8.5	8 (38)	8 (42)	
>8.5	4 (19)	2 (10)	
BMI (kg/m^2), mean (SD)	31.8 (8.6)	31.6 (5.1)	.92
Systolic blood pressure (mm Hg), mean (SD)	118 (11)	118 (9)	.94
Antihypertensive medication, n (%)	7 (33)	4 (21)	.39
Total cholesterol (mmol/L), mean (SD)	4.6 (1.2)	4.7 (0.9)	.87
Triglycerides (mmol/L), mean (SD)	2.1 (1.6)	2.7 (2.1)	.32
Lipid-lowering medication, n (%)	6 (29)	7 (37)	.58
eGFR[c] (mL/min/1.73 m^2), mean (SD)	111 (19)	112 (13)	.88
Abnormal UACR[d], n (%)	8 (38)	6 (32)	.67
HLQ[e] Scale 2 score[f], mean (SD)	3.1 (0.3)	3.3 (0.5)	.14
HLQ Scale 6 score[g], mean (SD)	4.0 (0.7)	4.1 (0.6)	.40

[a]TAFE: Technical and Further Education.

[b]HbA_{1c}: glycated hemoglobin A_{1c}.

[c]eGFR: estimated glomerular filtration rate.

[d]UACR: urine albumin to creatinine ratio.

[e]HLQ: Health Literacy Questionnaire.

[f]HLQ Scale 2 score reflects an individual's confidence in their knowledge to manage their health. Higher scores equate to higher confidence levels.

[g]HLQ Scale 6 score reflects an individual's confidence in their ability to actively engage with health care providers. Higher scores equate to higher confidence levels.

Primary Endpoint: 100% Attendance at Scheduled Follow-up Appointments

The proportion of participants who maintained 100% attendance at scheduled clinic appointments for successive follow-up appointments is presented in Figure 1. At 12 months, more participants in the enhanced SMS intervention group achieved 100% attendance at scheduled clinic appointments (12/21, 57%, vs 5/19, 26%). A statistically significant difference between the intervention and control groups was observed with respect to 100% attendance (Figure 1, Table 2; log-rank *P*=.04).

Figure 1. Kaplan-Meier curve illustrating the proportion of participants who maintained 100% attendance as scheduled throughout the study.

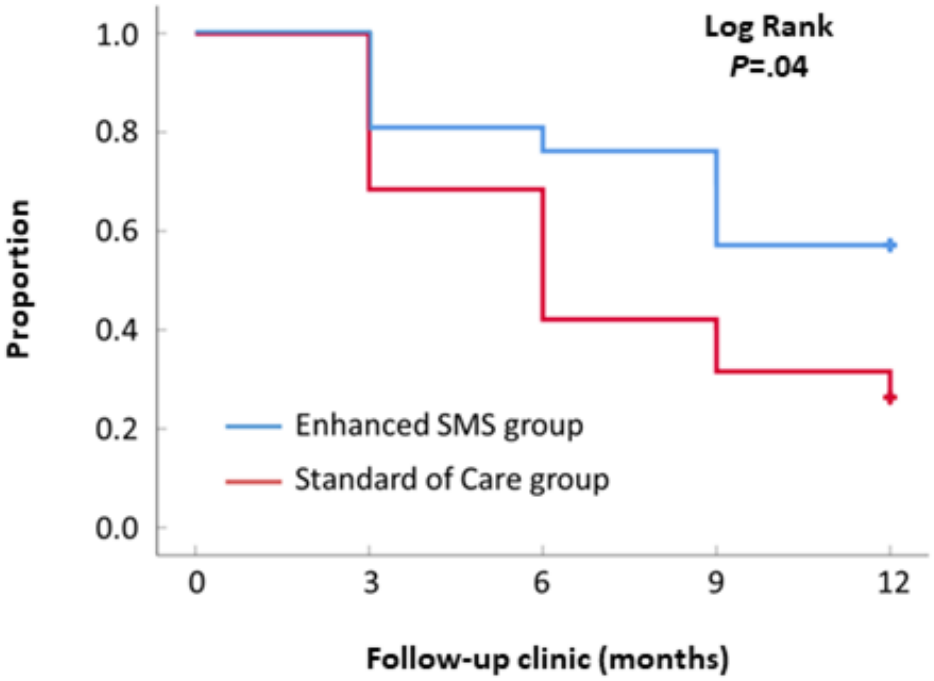


Table 2. Participants in the enhanced SMS and control groups who maintained 100% attendance in the TEXT2U study (N=40).

Group	Value, n (%)				
	Baseline	3 months	6 months	9 months	12 months
Intervention (n=21)	21 (100)	17 (81)	16 (76)	12 (57)	12 (57)
Control (n=19)	19 (100)	13 (68)	8 (42)	6 (32)	5 (26)

Secondary Endpoints

Overall Clinic Attendance

The proportion of clinic appointments that were attended as scheduled was greater in the intervention group than in the control group (64/84, 76%, vs 44/76, 58%; Multimedia Appendix 3, Table S2). It was possible to boost clinic attendance to 80% in both the intervention and control groups, but only after proactive contact by administrative staff and rescheduling of missed appointments (Multimedia Appendix 3, Table S2). The requirement for this extra administrative burden was lower in the intervention group; 18% fewer rescheduling interventions were necessary in this group.

Clinical Metabolic Outcomes

Key end-of-study metabolic indices for the intervention and control groups are presented in Table 3. For those with 12-month follow-up data (20/21 intervention and 15/19 control, respectively), there was no difference in mean HbA_{1c} between groups (mean 7.1%, SD 1.1%, for the intervention group vs mean 6.6%, SD 1.7%, for the control group; *P*=.37). For both groups, clinic attendance was associated with glycemic stability; there was no statistically significant change in mean 12-month HbA_{1c} (relative to baseline) for either the intervention or control group (Multimedia Appendix 4, Table S3). If the analysis was restricted to those with 100% attendance, the change in HbA_{1c} from baseline to study end was –0.58% (95% CI –1.70 to 0.59) (intervention) and –0.18% (95% CI –0.46 to 0.10) (control). Although the physicians working in the young adult type 2 diabetes clinic ensure that all patients have access to valid prescriptions for their diabetic pharmacotherapy at each clinic visit, absolute adherence to pharmacotherapy was not actively assessed during this study. Analogously to HbA_{1c}, there were no statistically significant between-group differences with respect to BMI, total cholesterol, or triglycerides.

Table 3. Mean (SD) values of selected metabolic indices after 12 months for the intervention and control groups.

Metabolic index	Values, mean (SD)			*P* value
	Intervention group	Control group	Mean difference (95% CI)	
HbA_{1c}[a] (%)	7.1 (1.1)	6.6 (1.7)	0.4 (–0.5 to +1.4)	.37
BMI[b] (kg/m^2)	30.4 (8.4)	31.8 (5.8)	–1.5 (–6.7 to +3.8)	.57
Total cholesterol[c] (mmol/L)	4.6 (1.4)	4.7 (0.9)	–0.1 (–1.0 to +0.7)	.75
Triglycerides[c] (mmol/L)	2.1 (1.7)	2.9 (4.1)	–0.8 (–3.3 to +1.6)	.50

[a]HbA_{1c}: glycated hemoglobin A_{1c}. 12-month HbA_{1c} data were available for 20/21 intervention group participants and 15/19 control group participants.

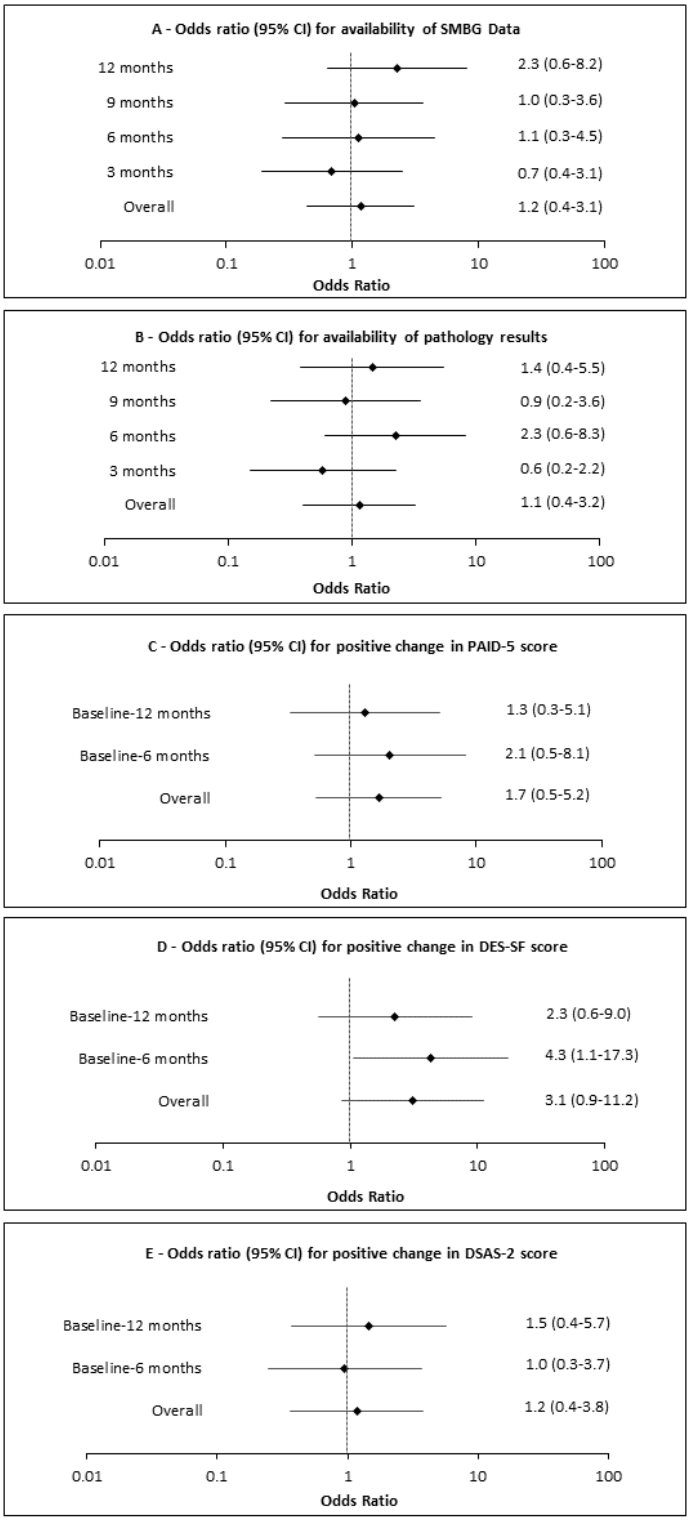
[b]12-month BMI data were available for 18/21 intervention group participants and 15/19 control group participants.

[c]12-month total cholesterol and triglyceride data were available for 20/21 intervention group participants and 14/19 control group participants.

Diabetes Self-management Practices

Overall, SMBG data were available at 40% of the intervention group's appointments, versus 37% of the control group's appointments (Multimedia Appendix 5, Figure S2). The corresponding figures for the pathology results were 69% and 66%, respectively (Multimedia Appendix 6, Figure S3). Binary logistic regression modeling revealed no between-group differences with respect to availability of SMBG or pathology results at follow-up. The ORs for availability of SMBG data and pathology results in the intervention group were 1.2 (95% CI 0.4-3.1) and 1.1 (95% CI 0.4-3.2), respectively (Figure 2, Panels A and B).

Figure 2. Forest plots illustrating the odds ratios and 95% confidence intervals for positive changes in selected secondary outcomes (A, SMBG; B, pathology results; C, PAID-5 score; D, DES-SF score; E, DSAS-2 score) for the TEXT2U intervention. DES-SF: Diabetes Empowerment Scale-Short Form; DSAS-2: Type 2 Diabetes Stigma Assessment Scale; PAID-5: Problem Areas in Diabetes 5-item Short Form; SMBG: self-monitored blood glucose.



Psychosocial Impact (Baseline Versus 6 Months and 12 Months)

An overview of the PAID-5, DES-SF, and DSAS-2 scores for the study cohort is presented in Multimedia Appendix 7 (Table S4). Analysis of the changes in DES-SF score at 6 months revealed a significant between-group difference favoring the intervention (Multimedia Appendix 7, Table S4); 68% (13/19) of members of the intervention group recorded a favorable change in DES-SF score at 6 months, versus 31% (5/16) in the control group (P=.03). At 12 months, this difference was no longer statistically significant. In the analyses of the changes in PAID-5 and DSAS-2 scores, there were no significant

between-group differences for either of the scores at 6 or 12 months (Multimedia Appendix 7, Table S4).

Binary logistic regression modeling revealed no evidence of a difference in the odds of recording a positive change in PAID-5 or DSAS-2 scores for the intervention and control groups (Figure 2, panels C and E). Although not statistically significant at 12 months, the OR for positive change in DES-SF score at 6 months (4.3, 95% CI 1.1-17.3) was significantly higher for those who participated in the enhanced SMS text messaging program (Figure 2, panel D).

Evaluation of the Enhanced SMS Text Message–Based Support and Reminder Program

Of the 21 participants randomized to the enhanced SMS text messaging intervention group, 20 received all study text messages; 1 participant asked to stop receiving text messages midstudy. Of the 21 participants randomized to the enhanced SMS group, 11 chose to use the study-specific SMS portal (Multimedia Appendix 8, Table S5). The median number of messages sent by participants was 3 (range 1-29). Analysis of message content revealed that 28% (21/74) of the messages sent by participants expressed gratitude (for a message received or a service provided by the clinic), 25% (19/74) of messages directed a diabetes-specific question to the study team or sought clarification regarding some aspect of diabetes management, and 24% (18/74) of messages were concerned with rescheduling a clinic appointment. With respect to diabetes-related questions, participants inquired about a variety of different issues, including management of hypoglycemia, dietary recommendations for people with type 2 diabetes, guidelines for dental follow-up, diabetic foot care, and the effects of stress and alcohol on blood glucose.

Following study completion, 16/21 (76%) of the SMS group participants provided feedback on their experience (Multimedia Appendix 9, Table S6). Of the 21 SMS group participants, 5 did not attend their scheduled 12-month follow-up clinic visit (or a makeup appointment for their missed clinic visit); therefore, these study participants did not provide feedback on the program. Those who did not provide feedback on the program were slightly younger (age 29.2 years, SD 6.1 years, vs 34.2 years, SD 5.4 years) and heavier (40.4 kg/m^2, SD 11.3 kg/m^2, vs 29.1 kg/m^2, SD 5.7 kg/m^2) with slightly lower baseline HbA_{1c} (6.8%, SD 0.9%, vs 7.4%, SD 1.8%).

All respondents to the end-of-study survey found that the study messages were easy to understand, supportive in nature, and delivered at appropriate times. In addition, 94% (15/16) of respondents reported that the messages contained practical information for people with diabetes and motivated them to think about their diabetes management. Overall, 94% (15/16) of respondents indicated that they would recommend the program to all new patients of the clinic.

Discussion

Principal Results

The TEXT2U study established that an enhanced SMS text message–based support and reminder program was effective at improving scheduled clinic attendance rates. Relative to the control group, the enhanced SMS text messaging intervention more than doubled the number of participants who attended all follow-up appointments. For those who attended all scheduled clinic appointments, the overall finding was stability of HbA_{1c}, weight, and lipid profiles during the 12-month study period. With respect to the psychological impact of the intervention, we observed an improvement in psychosocial self-efficacy; a statistically significant favorable change in DES-SF score was observed at 6 months. However, we did not observe any between-group differences in measures of diabetes-specific distress nor in measures of perceived and experienced stigma related to living with type 2 diabetes. Overall, the enhanced SMS text message–based support and reminder program was well received and facilitated between-visit engagement with the health care team; >90% of respondents to the end-of-study survey said they would recommend the program to new patients. Although the completion rate of the end-of-study survey was 76% (5 participants did not attend a 12-month study visit), it remains reassuring that >70% of all SMS program recipients found utility in the SMS text messaging program employed in this study. In this setting, one anticipates that the majority of future consumers of TEXT2U will derive benefit from participation in the program.

Research in Context

Our findings align with existing evidence that indicates text messaging interventions are an effective method to aid in the modification of health behaviors [11,13]. In a diabetes-specific context, the Sweet Talk RCT [26] identified that a text messaging system to support young people with type 1 diabetes was successful in improving measures of diabetes self-efficacy but not glycemic control. It is well recognized that the needs of youth with type 1 diabetes differ from those with type 2 diabetes [27]. Consequently, it is informative that this is the first study to show clear benefit of an enhanced SMS text message–based intervention over and above simple SMS reminder text messaging on clinic attendance in young-onset type 2 diabetes.

The issues of suboptimal clinic attendance and loss to follow-up are important problems in clinical practice. Maintaining clinic attendance is particularly important during young adulthood given the increased risk of diabetes complications [6] and the competing demands of family, work, study, and socializing, which interfere with maintenance of optimal care. Recurrent nonattendance prevents timely identification of changes in an individual's health status and eliminates the opportunity for diabetes complications screening, early intervention, and ongoing education; moreover, it hinders the development of a therapeutic relationship [28]. Various strategies, ranging from written, verbal, and electronic reminders to expensive case management approaches, have been used by health service providers in an attempt to improve clinic attendance and patient outcomes [29,30]. This study adds to the literature by demonstrating that a low-cost, enhanced SMS text message–based intervention is effective at improving clinic attendance over and above a standard SMS appointment reminder system for those with young-onset type 2 diabetes, a cohort that is arguably at greatest risk for poor diabetes outcomes.

Nevertheless, maintenance of optimal clinic attendance remained a significant challenge, with just over one half of the enhanced SMS group and one-quarter of the control group recording a 100% attendance record. The factors that influence clinic nonattendance in youth with type 2 diabetes are poorly understood, likely to be complex, and an important area for further study.

In addition to adverse effects for the individual, nonattendance reduces clinic efficiency and places a significant burden on administrative staff, who are required to contact nonattenders and reschedule make-up appointments [31]. Notably, the enhanced SMS text messaging intervention resulted in improved clinic attendance rate without additional administrative burden. Overall clinic attendance rates were boosted to 80% in both groups; however, with additional administrative intervention to contact patients after missed appointments and then reschedule makeup appointments, this requirement was much higher for the control group. In a resource-constrained setting, it is highly likely that youth who miss clinic appointments will be lost to follow-up and therefore at increased risk of poor long-term outcomes. Moving forward, in-built automation of this program will allow the intervention to be scaled with minimal demand on clinic resources. Although a formal cost-benefit analysis was not undertaken, the relatively low cost of SMS text messaging ensures that this program is not prohibitively expensive.

It is well accepted that improving engagement of patients with their own health care can improve outcomes, and the concept of an informed and empowered patient is now central to many models of care [32-34]. The DES-SF measures diabetes self-efficacy across three main domains: (1) managing psychosocial aspects of diabetes care, (2) assessing dissatisfaction and readiness to change, and (3) setting and achieving goals [19]. For those who attended follow-up appointments in our study, we observed a more favorable pattern of response with respect to serial DES-SF scores in the intervention group. Conceivably, delivery of supportive text messages and the opportunity to clarify queries by SMS text messaging between clinic visits in our study facilitated an increased sense of empowerment. Furthermore, incorporation of knowledge-based messages allowed important clinical information (including dietary and lifestyle recommendations) to be reinforced outside the clinic setting. This is particularly relevant given that postconsultation retention of medical information is poor [35].

The lack of a statistically significant improvement in DES-SF score at 12 months may indicate that the support provided by the enhanced SMS text messaging intervention has limited durability or that there may be a threshold to any improvement seen. Furthermore, we did not demonstrate improvement in measures of distress or stigma; plausibly, these more negative psychological elements respond best to in-person encounters and are not amenable to a text-based intervention. Although diabetes empowerment was supported by the enhanced SMS text message–based intervention, we did not see improvement in objective measures of self-care as evidenced by the availability of the SMBG/pathology results.

Although the intervention increased scheduled clinic attendance, we did not observe a difference in mean HbA_{1c}, weight, or lipids between the intervention and control groups at study completion. For the most part, this is unsurprising given that the overall attendance (ie, scheduled attendance plus attendance at makeup appointments) was equal in both study arms. Irrespective of whether a participant attended a scheduled or makeup appointment, those who attended were afforded the same opportunity to have their management adjusted by the treating clinician. Nonetheless, stability of glycemia, close to target HbA_{1c}, over 12 months of follow-up is not an insignificant achievement. The TODAY study clearly demonstrated that many adolescents with type 2 diabetes struggle to maintain glycemic control despite comprehensive care in a well-supervised setting [36].

Study Strengths

The age of the study cohort and their familiarity with text messaging allowed for immediate engagement with the SMS text messaging program; individuals were not hampered by a learning curve in deriving benefit from study participation. In designing this study, we were mindful of the need to avoid overwhelming participants with too many messages; prevention of participant withdrawal due to message fatigue was an important consideration. In previous studies of SMS interventions, investigators regularly sent participants multiple text messages every week. The length of most previous studies has been comparatively short (usually <6 months), and participant attrition due to message fatigue in the short term is less likely. Given our 12-month study duration, a conscious decision was made to taper message frequency over the course of the study. Certainly, there is evidence to suggest that text messaging interventions with a tapering frequency have higher efficacy [37]. Personalization of the text messages ensured that the messages received by each participant were relevant to their personal circumstances, and the intervention had high acceptability. Feedback revealed that 88% of participants felt they had received an appropriate number of messages and 100% felt they received messages at appropriate times. Of the 21 enhanced SMS text messaging program participants, 20 opted to receive all study text messages; study message fatigue was not a significant issue.

Limitations

There are several limitations of this study. First, it should be noted that our study was powered to detect a difference in clinic attendance but not to detect small differences in secondary outcomes. Secondly, our study was conducted at a secondary and tertiary referral center with a modest number of participants; testing in other settings will be required to confirm generalizability. In addition, our study population comprised a well-educated cohort (80% had attended, or were attending, Technical and Further Education [TAFE] or university). Although the message program employed plain language, it would be worthwhile to consider further testing in groups with lower literacy levels to establish better generalizability of the results. Furthermore, we specifically focused on the younger demographic; therefore, applicability in older age groups remains uncertain. Although all study participants had sufficient

English proficiency, given our multicultural population, provision of the program in a participant's native language may well have resulted in greater engagement. By design, it was not possible to blind both participants and treating clinicians in this study. However, blinding of treating clinicians and the use of an objectively measured primary endpoint were important aspects of the study design that help to mitigate observer bias.

Conclusions

In young-onset type 2 diabetes, a low-cost, automated, enhanced SMS text message–based support and reminder system improved scheduled clinic attendance and diabetes self-efficacy over 12 months of follow-up. These achievements were made with high acceptability and low administrative burden. This text-based intervention has the potential to have substantial impact on the care of youth with type 2 diabetes at an especially vulnerable time of life.

Acknowledgments

This study was supported by an Australian Federal Government Medical Research Future Fund Rapid Applied Research Translation Grant. TM was supported by an Australian Postgraduate Research Scholarship. Elements of this study were presented in abstract form at the Australasian Diabetes Congress, Sydney, Australia, August 21-23, 2019.

Authors' Contributions

TM was involved in the study design; contributed to data acquisition, analysis, and interpretation; and drafted the manuscript. MC and MM assisted in the acquisition of data and provided critical review of the manuscript. MDS contributed to the statistical analysis and critical review of the manuscript. SMT, TW, and JW were involved in the study design, contributed to data analysis and interpretation, and provided critical review of the manuscript. AT and CC were involved in the study design, provided technical support, contributed to analysis and interpretation of data, and provided critical review of the manuscript. All authors approved the manuscript for submission. TM is the guarantor of this work and, as such, had access to all study data and takes responsibility for data integrity and accuracy of the analyses.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Examples of text messages included in the TEXT2U study.
[DOC File , 28 KB - jmir_v23i10e27263_app1.doc]

Multimedia Appendix 2
Enrollment and participant flow in the TEXT2U randomized controlled trial.
[DOC File , 75 KB - jmir_v23i10e27263_app2.doc]

Multimedia Appendix 3
Overall attendance patterns for the intervention and control groups of the TEXT2U study.
[DOC File , 32 KB - jmir_v23i10e27263_app3.doc]

Multimedia Appendix 4
Mean change in selected metabolic indices (95% CI) after 12 months for the intervention and control groups of the TEXT2U study.
[DOC File , 37 KB - jmir_v23i10e27263_app4.doc]

Multimedia Appendix 5
Proportion of participants with SMBG data available at clinic study visits.
[DOC File , 43 KB - jmir_v23i10e27263_app5.doc]

Multimedia Appendix 6
Proportion of participants with pathology data available at clinic study visits.
[DOC File , 44 KB - jmir_v23i10e27263_app6.doc]

Multimedia Appendix 7
Changes in validated measures of psychosocial status throughout the study.

[DOC File , 37 KB - jmir_v23i10e27263_app7.doc]

Multimedia Appendix 8
Overview of participant utilization of the optional SMS portal.
[DOC File , 35 KB - jmir_v23i10e27263_app8.doc]

Multimedia Appendix 9
Utility and acceptability of the enhanced SMS program to the intervention participants.
[DOC File , 36 KB - jmir_v23i10e27263_app9.doc]

Multimedia Appendix 10
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1136 KB - jmir_v23i10e27263_app10.pdf]

References

1. Diabetes Atlas, 2nd Edition. International Diabetes Federation. URL: https://diabetesatlas.org/en/resources/ [accessed 2021-09-24]
2. Diabetes Atlas, 6th Edition. International Diabetes Federation. URL: https://diabetesatlas.org/en/resources/ [accessed 2021-09-24]
3. Dabelea D, Stafford JM, Mayer-Davis EJ, D'Agostino R, Dolan L, Imperatore G, SEARCH for Diabetes in Youth Research Group. Association of type 1 diabetes vs type 2 diabetes diagnosed during childhood and adolescence with complications during teenage years and young adulthood. JAMA 2017 Feb 28;317(8):825-835 [FREE Full text] [doi: 10.1001/jama.2017.0686] [Medline: 28245334]
4. Constantino MI, Molyneaux L, Limacher-Gisler F, Al-Saeed A, Luo C, Wu T, et al. Long-term complications and mortality in young-onset diabetes: type 2 diabetes is more hazardous and lethal than type 1 diabetes. Diabetes Care 2013 Dec;36(12):3863-3869 [FREE Full text] [doi: 10.2337/dc12-2455] [Medline: 23846814]
5. Al-Saeed AH, Constantino MI, Molyneaux L, D'Souza M, Limacher-Gisler F, Luo C, et al. An inverse relationship between age of type 2 diabetes onset and complication risk and mortality: the impact of youth-onset type 2 diabetes. Diabetes Care 2016 May;39(5):823-829. [doi: 10.2337/dc15-0991] [Medline: 27006511]
6. Middleton TL, Constantino MI, Molyneaux L, D'Souza M, Twigg SM, Wu T, et al. Young-onset type 2 diabetes and younger current age: increased susceptibility to retinopathy in contrast to other complications. Diabet Med 2020 Jun;37(6):991-999 [FREE Full text] [doi: 10.1111/dme.14238] [Medline: 31968129]
7. Schectman JM, Schorling JB, Voss JD. Appointment adherence and disparities in outcomes among patients with diabetes. J Gen Intern Med 2008 Oct;23(10):1685-1687 [FREE Full text] [doi: 10.1007/s11606-008-0747-1] [Medline: 18661189]
8. Fisher E, Lazar L, Shalitin S, Yackobovitch-Gavan M, de Vries L, Oron T, et al. Association between glycemic control and clinic attendance in emerging adults with type 1 diabetes: a tertiary center experience. J Diabetes Res 2018;2018:9572817 [FREE Full text] [doi: 10.1155/2018/9572817] [Medline: 30116747]
9. Berkowitz RI, Marcus MD, Anderson BJ, Delahanty L, Grover N, Kriska A, TODAY Study Group. Adherence to a lifestyle program for youth with type 2 diabetes and its association with treatment outcome in the TODAY clinical trial. Pediatr Diabetes 2018 Mar;19(2):191-198 [FREE Full text] [doi: 10.1111/pedi.12555] [Medline: 28664624]
10. Griffiths C, Brooks B, Landy L, Shi Y, Gauld A, Constantino M, et al. Young onset type 2 diabetes: out of touch and out of time. Poster. 2016 Presented at: Australian Diabetes Society and the Australian Diabetes Educators Association Annual Scientific Meeting; August 24-26, 2016; Gold Coast, Australia p. 24-26 URL: http://ads-adea-2016.m.asnevents.com.au/schedule/session/9237/abstract/36455
11. Chow CK, Redfern J, Hillis GS, Thakkar J, Santo K, Hackett ML, et al. Effect of lifestyle-focused text messaging on risk factor modification in patients with coronary heart disease: a randomized clinical trial. JAMA 2015;314(12):1255-1263. [doi: 10.1001/jama.2015.10945] [Medline: 26393848]
12. Devine S, Leeds C, Shlay JC, Leytem A, Beum R, Bull S. Methods to assess youth engagement in a text messaging supplement to an effective teen pregnancy program. J Biomed Inform 2015 Aug;56:379-386 [FREE Full text] [doi: 10.1016/j.jbi.2015.07.003] [Medline: 26173038]
13. Suffoletto B, Kristan J, Chung T, Jeong K, Fabio A, Monti P, et al. An interactive text message intervention to reduce binge drinking in young adults: a randomized controlled trial with 9-month outcomes. PLoS One 2015;10(11):e0142877 [FREE Full text] [doi: 10.1371/journal.pone.0142877] [Medline: 26580802]
14. Dobson R, Whittaker R, Jiang Y, Maddison R, Shepherd M, McNamara C, et al. Effectiveness of text message based, diabetes self management support programme (SMS4BG): two arm, parallel randomised controlled trial. BMJ 2018 May 17;361:k1959 [FREE Full text] [doi: 10.1136/bmj.k1959] [Medline: 29773539]

15. Dao MC, Fu SN, Wong CKH, Luk W. Use of short message service (SMS) reminders to improve the attendance rate of diabetes-related multidisciplinary services in a primary care clinic a randomized controlled trial. Research Square Preprint posted online on November 2, 2019 [FREE Full text] [doi: 10.21203/rs.2.16495/v1]
16. Ke C, Shah BR, Luk AO, Di Ruggiero E, Chan JCN. Cardiovascular outcomes trials in type 2 diabetes: time to include young adults. Diabetes Obes Metab 2020 Jan;22(1):3-5. [doi: 10.1111/dom.13874] [Medline: 31486276]
17. Middleton TL, Constantino MI, Molyneaux L, AlMogbel T, McGill M, Yue DK, et al. Secular trends in information communications technology: access, use, and attitudes of young and older patients with diabetes. Diabetes Spectr 2020 Feb;33(1):66-73. [doi: 10.2337/ds18-0082] [Medline: 32116456]
18. McGuire BE, Morrison TG, Hermanns N, Skovlund S, Eldrup E, Gagliardino J, et al. Short-form measures of diabetes-related emotional distress: the Problem Areas in Diabetes Scale (PAID)-5 and PAID-1. Diabetologia 2010 Jan;53(1):66-69. [doi: 10.1007/s00125-009-1559-5] [Medline: 19841892]
19. Anderson RM, Fitzgerald JT, Gruppen LD, Funnell MM, Oh MS. The Diabetes Empowerment Scale-Short Form (DES-SF). Diabetes Care 2003 May;26(5):1641-1642. [Medline: 12716841]
20. Browne JL, Ventura AD, Mosely K, Speight J. Measuring the stigma surrounding type 2 diabetes: development and validation of the type 2 Diabetes Stigma Assessment Scale (DSAS-2). Diabetes Care 2016 Dec;39(12):2141-2148. [doi: 10.2337/dc16-0117] [Medline: 27515964]
21. Arslanian S, Bacha F, Grey M, Marcus MD, White NH, Zeitler P. Evaluation and management of youth-onset type 2 diabetes: a position statement by the American Diabetes Association. Diabetes Care 2018 Dec;41(12):2648-2668 [FREE Full text] [doi: 10.2337/dci18-0052] [Medline: 30425094]
22. Stevenson J, Campbell KL, Brown M, Craig J, Howard K, Howell M, et al. Targeted, structured text messaging to improve dietary and lifestyle behaviours for people on maintenance haemodialysis (KIDNEYTEXT): study protocol for a randomised controlled trial. BMJ Open 2019 May 05;9(5):e023545 [FREE Full text] [doi: 10.1136/bmjopen-2018-023545] [Medline: 31061013]
23. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
24. Osborne RH, Batterham RW, Elsworth GR, Hawkins M, Buchbinder R. The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). BMC Public Health 2013 Jul 16;13:658 [FREE Full text] [doi: 10.1186/1471-2458-13-658] [Medline: 23855504]
25. Lin H, Wu X. Intervention strategies for improving patient adherence to follow-up in the era of mobile information technology: a systematic review and meta-analysis. PLoS One 2014;9(8):e104266 [FREE Full text] [doi: 10.1371/journal.pone.0104266] [Medline: 25100267]
26. Franklin VL, Waller A, Pagliari C, Greene SA. A randomized controlled trial of Sweet Talk, a text-messaging system to support young people with diabetes. Diabet Med 2006 Dec;23(12):1332-1338. [doi: 10.1111/j.1464-5491.2006.01989.x] [Medline: 17116184]
27. Zeitler P, Arslanian S, Fu J, Pinhas-Hamiel O, Reinehr T, Tandon N, et al. ISPAD Clinical Practice Consensus Guidelines 2018: type 2 diabetes mellitus in youth. Pediatr Diabetes 2018 Oct;19 Suppl 27:28-46. [doi: 10.1111/pedi.12719] [Medline: 29999228]
28. Kelley JM, Kraft-Todd G, Schapira L, Kossowsky J, Riess H. The influence of the patient-clinician relationship on healthcare outcomes: a systematic review and meta-analysis of randomized controlled trials. PLoS One 2014;9(4):e94207 [FREE Full text] [doi: 10.1371/journal.pone.0094207] [Medline: 24718585]
29. Nuti L, Turkcan A, Lawley MA, Zhang L, Sands L, McComb S. The impact of interventions on appointment and clinical outcomes for individuals with diabetes: a systematic review. BMC Health Serv Res 2015 Sep 02;15:355 [FREE Full text] [doi: 10.1186/s12913-015-0938-5] [Medline: 26330299]
30. Shojania KG, Ranji SR, McDonald KM, Grimshaw JM, Sundaram V, Rushakoff RJ, et al. Effects of quality improvement strategies for type 2 diabetes on glycemic control: a meta-regression analysis. JAMA 2006 Jul 26;296(4):427-440. [doi: 10.1001/jama.296.4.427] [Medline: 16868301]
31. Griffin SJ. Lost to follow-up: the problem of defaulters from diabetes clinics. Diabet Med 1998 Nov;15 Suppl 3:S14-S24. [doi: 10.1002/(sici)1096-9136(1998110)15:3+<s14::aid-dia725>3.3.co;2-9] [Medline: 9829764]
32. Greenfield S, Kaplan S, Ware JE. Expanding patient involvement in care. Effects on patient outcomes. Ann Intern Med 1985 Apr;102(4):520-528. [Medline: 3977198]
33. Charles C, Gafni A, Whelan T. Decision-making in the physician-patient encounter: revisiting the shared treatment decision-making model. Soc Sci Med 1999 Sep;49(5):651-661. [Medline: 10452420]
34. Hibbard JH. Engaging health care consumers to improve the quality of care. Med Care 2003 Jan;41(1 Suppl):I61-I70. [doi: 10.1097/00005650-200301001-00007] [Medline: 12544817]
35. Kessels RPC. Patients' memory for medical information. J R Soc Med 2003 May;96(5):219-222 [FREE Full text] [Medline: 12724430]

36. TODAY Study Group, Zeitler P, Hirst K, Pyle L, Linder B, Copeland K, et al. A clinical trial to maintain glycemic control in youth with type 2 diabetes. N Engl J Med 2012 Jun 14;366(24):2247-2256 [FREE Full text] [doi: 10.1056/NEJMoa1109333] [Medline: 22540912]
37. Head KJ, Noar SM, Iannarino NT, Grant HN. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]

Abbreviations

BP: blood pressure
DES-SF: Diabetes Empowerment Scale–Short Form
DSAS-2: Type 2 Diabetes Stigma Assessment Scale
HbA_{1c}: hemoglobin A_{1c}
OR: odds ratio
PAID-5: Problem Areas in Diabetes 5-item Short Form
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
RPAH: Royal Prince Alfred Hospital
SMBG: self-monitoring of blood glucose
TAFE: Technical and Further Education
TODAY: Treatment Options for Type 2 Diabetes in Adolescents and Youth

Edited by R Kukafka; submitted 19.01.21; peer-reviewed by N Allen, E Villalba-Mora, K Rawlings; comments to author 10.05.21; revised version received 24.06.21; accepted 21.07.21; published 21.10.21.

Please cite as:
Middleton T, Constantino M, McGill M, D'Souza M, Twigg SM, Wu T, Thiagalingam A, Chow C, Wong J
An Enhanced SMS Text Message–Based Support and Reminder Program for Young Adults With Type 2 Diabetes (TEXT2U): Randomized Controlled Trial
J Med Internet Res 2021;23(10):e27263
URL: https://www.jmir.org/2021/10/e27263
doi:10.2196/27263
PMID:34524102

Original Paper

The Development of a Web-Based Tobacco Tracker Tool to Crowdsource Campus Environmental Reports for Smoke and Tobacco–Free College Policies: Mixed Methods Study

Sabrina F Loureiro[1], BS; Kim Pulvers[2], MPH, PhD; Melissa M Gosdin[1], PhD; Keavagh Clift[3], MPH, CHES, CHWC; Myra Rice[2], MA; Elisa K Tong[4], MA, MD

[1]Center for Healthcare Policy and Research, University of California Davis, Sacramento, CA, United States

[2]Department of Psychology, California State University San Marcos, San Marcos, CA, United States

[3]Department of Occupational Health Services, University of California Davis, Sacramento, CA, United States

[4]Department of Internal Medicine, University of California Davis, Sacramento, CA, United States

Corresponding Author:
Elisa K Tong, MA, MD
Department of Internal Medicine
University of California Davis
4150 V Street, Suite 2400
Sacramento, CA, 95817
United States
Phone: 1 (916) 734 7005
Email: ektong@ucdavis.edu

Abstract

Background: College campuses in the United States have begun implementing smoke and tobacco–free policies to discourage the use of tobacco. Smoke and tobacco–free policies, however, are contingent upon effective policy enforcement.

Objective: This study aimed to develop an empirically derived web-based tracking tool (Tracker) for crowdsourcing campus environmental reports of tobacco use and waste to support smoke and tobacco–free college policies.

Methods: An exploratory sequential mixed methods approach was utilized to inform the development and evaluation of Tracker. In October 2018, three focus groups across 2 California universities were conducted and themes were analyzed, guiding Tracker development. After 1 year of implementation, users were asked in April 2020 to complete a survey about their experience.

Results: In the focus groups, two major themes emerged: barriers and facilitators to tool utilization. Further Tracker development was guided by focus group input to address these barriers (eg, information, policing, and logistical concerns) and facilitators (eg, environmental motivators and positive reinforcement). Amongst 1163 Tracker reports, those who completed the user survey (n=316) reported that the top motivations for using the tool had been having a cleaner environment (212/316, 79%) and health concerns (185/316, 69%).

Conclusions: Environmental concerns, a motivator that emerged in focus groups, shaped Tracker's development and was cited by the majority of users surveyed as a top motivator for utilization.

(J Med Internet Res 2021;23(10):e26280) doi:10.2196/26280

KEYWORDS

tobacco cessation; college smoke and tobacco–free policies; crowdsourcing; environmental reporting; public health; smoke and tobacco research

Introduction

Smoke and tobacco–free (STF) policies among college and university campuses have become a high priority nationally to denormalize the use of tobacco [1,2]. The average smoker begins before the age of 26 years, making the transition into university life a significant period during which young people can either begin or avoid tobacco use [3]. Benefits of STF policies include decreased tobacco use and exposure to second-hand smoke, reduced tobacco litter, and greater well-being [4-6]. However, the benefits are contingent upon effective policy implementation and enforcement. Only 15% of US colleges detail methods for

reporting an STF policy violation and over a third do not formally reprimand individuals who smoke on campus [7].

Low STF policy compliance is widespread at universities and poses a critical barrier to reducing tobacco use and related disease risk for the estimated 19.9 million college students in the United States [8-10]. The majority (62%) of colleges rely on social enforcement of the policy, with typical methods including education and outreach [7]. Given the responsibility placed with campus community members for enforcement, innovative approaches to engage them with the STF policy are needed [11]. One innovative approach, crowdsourcing, is a method of involving a large group of people on a collective mission and is being used to address a number of environmental and public health problems [12,13]. A crowdsourcing tool for tobacco activity surveillance can help address a major barrier to social enforcement: discomfort confronting tobacco users [7,14,15]. Further, it provides a sustainable solution to an infrastructure issue faced by most campuses lacking resources to collect data on campus tobacco use and related litter.

The goal of this study is to develop and implement a web-based tracking tool (Tobacco Tracker hereafter "Tracker") to engage users with STF policies. This study aims to (1) describe the empirically driven process of developing and implementing Tracker using feedback from focus groups and (2) evaluate survey feedback from Tracker users about their motivations and experience using the tool.

Methods

Methods Overview

An exploratory sequential mixed methods approach was utilized to develop Tracker with guidance from campus focus groups and to evaluate Tracker with a user survey.

Study Setting

The 2 universities in the study are located in northern and southern California. University 1, in northern California, has been STF since 2014. Its campus has over 39,000 students, 21,000 staff, and 2000 faculty, and its campus encompass 5300 acres including a separate health campus. University 2, located in southern California, has been STF since 2017, has over 14,000 students, over 300 administrative staff, and 600 academic staff, and its campus encompasses 304 acres. Approval from human subjects was obtained through the institutional review board at each university.

Exploratory Focus Groups for Tracker Development

Focus groups aimed to explore ideas for prototype refinement were conducted in October 2018. A focus group comprising students was conducted at each university (n=7 at University 1 and n=10 at University 2). An additional focus group comprising employees was conducted at University 1 given its larger campus and staff size (n=6). Participants at University 1 were recruited from student or staff health and wellness listservs and a general staff electronic newsletter. Participants were selected in the order they responded to the recruitment advertisements and were offered a US $40 gift card. Participants at University 2 opted into the study from a web-based research portal and received credit toward their course requirement for participation. Eligibility included being at least 18 years old and a current university student or employee. Written informed consent was obtained from each participant prior to study participation. Each focus group lasted approximately 90 minutes.

To explore ideas for the Tracker prototype's content and design, guiding questions for the focus groups explored the (1) appearance and functionality of the tool prototype, (2) communication channels and strategies for promoting the tool, and (3) images and messages that may raise awareness or increase motivation to use the tool. Educational information was presented from the California Tobacco Control Program's public advertisements about how cigarette butts are the top litter problem and are not biodegradable with their plastic filters. All focus groups were conducted by the same moderator, a psychologist and tobacco control researcher (KP), audio recorded, and transcribed verbatim.

Two independent researchers, one of whom is a sociologist and qualitative researcher (MMG), conducted a thematic analysis [16]. Emergent codes were developed from the data inductively in addition to deductive priori codes. After the initial analysis, a codebook was constructed to create categories and develop themes. Two themes and 5 subthemes were developed collaboratively between the 2 independent researchers. No software programs were used to aid the focus group qualitative data analysis.

Tracker Development

The Tracker prototype's content, design, and promotion was refined on the basis of focus group findings, and themes linked to refined Tracker components are described below. Prototype components such as real-time web-based reporting and using visual maps to geographically locate environmental incidents were based on an existing crowdsourcing tool to measure energy overconsumption at University 1, TherMOOstat [17]. Tracker development in regard to content and design began in November 2018 and was promoted at both universities beginning in February 2019.

Survey of Tracker Users

Individuals who used Tracker at either university between February 2019 and February 2020 and provided contact information for entry into an opportunity drawing were invited by email to participate in a survey in April 2020. A 2-week period was provided for a response and 1 reminder email was sent. An opportunity drawing for 4 US $25 gift cards at each university was offered. Survey questions developed by the study team assessed user experience: (1) how participants accessed the tool, (2) motivation for using the tool, (3) barriers to tool use and for future study directions, and (4) ideal actions following a report. Survey questions allowed for multiple responses and data were coded as a selection, no selection, or as missing. Data were reported as n (%) values in total and by university.

Results

Focus Group Discussions

Focus group analysis yielded 2 distinct themes: (1) barriers and (2) facilitators to tool utilization (Table 1). Barriers to tool utilization contained 3 subthemes: logistical barriers limit tool use, information barriers, and framing as a policing tool. Facilitators for tool use contained 2 subthemes: environmental motivators and positive reinforcement to further engage with the tool. Minimal differences in responses were noted among the student participants, and differences between staff and student responses were minor (eg, less familiar with technology and less preference for a mascot or gamifying the tool). The results of thematic analysis are presented as an aggregate, given the goal of creating a tool for students and staff at both universities.

Table 1. Themes, subthemes, and representative quotes.

Themes, subthemes, and context	Representative quotes
Theme: Barriers to tool utilization	
Subtheme: Logistical barriers limit tool use	
Access to Tracker with QR[a] codes	• "What is it?" (University 1) • "I'm in IT and we mock them" (University 1) • "That's kind of scary. I don't know, that's crazy. I've never done that." (University 1)
Access to Tracker with campus app	• "I think the campus app would just be the best solution, because a lot of [inaudible] have it installed already, but I think a lot of people have it installed. Rather than going and getting a whole new account or having to go find it somewhere else, it's just an update and it's easy" (University 2) • "I really prefer the app a lot more… It seems more accessible in a sense. Like the online survey, you sometimes can't access it or it doesn't work. Obviously, that happens with apps too. It's just more reliable." (University 2)
Identifying location with automated GIS[b] location services	• "Yeah, it's really useful, because you can find yourself. At least to me, it seems kind of like a combination of the dropdown and the map, because it's just our campus and it has all the names of the buildings" (University 2) • "Sometimes you're like not at a building, you're in between. But I do think sometimes people feel weird about sharing their locations." (University 1)
Identifying location with a dropdown menu and a campus map	• "It would be best to have both options available in the tool" (University 1) • "I can see what you mean about the challenge of having it as a dropdown menu just from this example where you're like 'where's my place', it would be easier to type" (University 1)
Internet connection	• "I feel like this might be going too out of the scope of this. The internet connection on campus is awful, and if it's really slow, then I would just give up and not do it. That many reloading of pages would be really agonizing." (University 1)
Subtheme: Informational barriers	
Tobacco products	• "What's vaping?" (University 1) • "I don't know what Juul or blunts means" (University 1)
Tobacco Waste	• "Are vaping things becoming garbage?" (University 1) • "What is litter from vaping?" (University 2)
Image preference	• "Just because it's more identifiable because I'm on the go and I'm reporting this, I need to just see an actual picture." (University 1) • "It would be more helpful to have the different images of what you're actually saying" (University 1)
Subtheme: Framing as a policing tool	
	• "It also avoids the policing idea" (University 1) • "It just likes a very, it's an activity that's very much in a specific location, especially since it's forbidden to do it on campus, though. Again, I don't feel it's threatening, I don't feel like it's about policing people but rather about seeing signs of this activity. I can relate it to fire, but it could also be a consequence of smoking" (University 1) • "I just think maybe a title that plays along with being environmentally conscious, instead of 'Oh, I'm reporting somebody doing something bad and illegal,' it's more like, 'I'm taking care of the environment.' So it's less of like a whistle-blower thing" (University 1) • "Yeah. Because it almost seems like a tattling" (University 1)
Theme: Facilitators to tool utilization	

Themes, subthemes, and context	Representative quotes
Subtheme: Environmental motivators	
Design	• "I think you should have a blue theme… That kind of ties in the blue sky" (University 1) • "Well you don't necessarily think of the trees, so I think that, and making the art with it, I think that was very effective. Because I never thought about trees before" (University 1) • "You could have a bird with like a cigarette butt in its beak, like a cartoony bird. And maybe there could be like a friend bird with a fish in its beak, or something, next to it. I don't know. But make sure they're really cute" (University 1) • "I said the squirrel with the [University 1] shirt and binoculars, because I think it directly ties into the community on campus, and squirrels are really [University 1]. And I think it's good to have a cute little icon like the Thermoostat has because I think people want to use things that are cute, so I think it's just good to have a main character" (University 1) • "I said [University 1 mascot] Clean just because I think there's the school pride, and then the clean is, I don't know. Even with the air, I feel like that all goes back to like… I like that it's hinting at the environment, but it's still hinting at the smoking aspect versus the littering" (University 1)
Environmental concerns	• "Yeah, simple, but I also learned something. I did not know it took up to a decade to decompose. That would make me want to report that" (University 1)
Subtheme: Positive reinforcement	
Incentives	• "I think if you provide incentives, people would definitely be up for that" (University 2) • "Because I think it goes beyond this, and it's more like why we recycle right, because it makes us feel good. Right I kind of see this tool as the same thing" (University 1)
Gamification	• "Gamification is the way you're going to win this" (University 1) • "I think like also with stuff like this that's so widespread, it can be hard to see any impact and nobody wants to do something, but it's kind of hard to like want to be super engaged and do something all the time if you can't really tell what you're doing. I don't know. Like making kind of like mini games or something where it's like you've reported five times, and every time you report, you're cleaning up lungs or something" (University 2)

[a]QR: quick response.

[b]GIS: geographic information system.

Barriers to Tracker utilization

Logistical Barriers

Participants identified logistical barriers to tool utilization, such as technical and access barriers. Participants cited the utilization of a dropdown menu and geographic information system (GIS) location–enabling services as potential technical barriers to tool utilization. Participants had mixed feelings about the utility of an automated GIS location–enabling service compared to the utilization of a dropdown menu of buildings from a campus map. Some individuals were reluctant to share their location with Tracker, while others felt it was more efficient and convenient. Participants concluded they would like both choices.

Participants also discussed various access barriers such as quick response (QR) codes, mobile apps, and internet connectivity. Some participant concerns about QR codes, a barcode that can be photographed with a phone to redirect a user to a webpage, included how to use them or negative attitudes about their utility (Table 1). Furthermore, participants discussed the use of mobile apps instead of accessing the tool through a website. However, some participants believed students would be less inclined to download another app that would take up storage space. Therefore, one recommendation was to integrate Tracker with an existing campus tool. Lastly, 1 participant expressed concerns that slow internet connectivity for devices or phones themselves could discourage the use of Tracker.

Informational Barriers

To identify tobacco use and waste, many participants expressed concern that they were unaware of what vaping products or vaping waste looked like. Participants recommended presenting visual examples of such tobacco products and waste (Table 1).

Furthermore, participants preferred that these images look realistic rather than being a cartoon drawing.

Framing as a Policing Tool

Participants preferred the name and branding strategies to portray a neutral crowdsourcing tool. Specifically, participants were concerned about whether Tracker's name would be viewed as public surveillance by the campus community's smokers and nonsmokers. One participant suggested the use of an environmentally themed name to avoid portraying Tracker as one for policing fellow students and staff (Table 1). The participants preferred to de-emphasize the idea of policing smokers on campus.

Facilitators for Tracker Utilization

Environmental Motivators

When participants were asked about Tracker's design, environmentally themed color schemes such as blue or green and a name that suggested "clean" were preferred. Branding Tracker with an animal mascot, such as a dog or squirrel, was also recommended. Moreover, when presented with the California Tobacco Control Program's public advertisements, participants found information about tobacco waste's impact on the environment as a highly motivational reason to use Tracker.

Positive Reinforcement

To encourage continued use of Tracker, participants were asked about possible incentives. Participants agreed that incentives such as gift cards would encourage Tracker utilization. Some participants remarked that simply submitting a report would be rewarding enough for them to stay engaged. One user compared submitting a report to recycling and explained that the 2 actions made the individual feel motivated enough to continue.

To increase engagement with Tracker, student participants recommended that it be gamified to provide positive affirmation for Tracker reports on campus. Participants recommended that Tracker show a mascot dancing or interacting with the user for increased gratification and utilization. Other suggestions included using a reward system for reports to allow dressing the mascots.

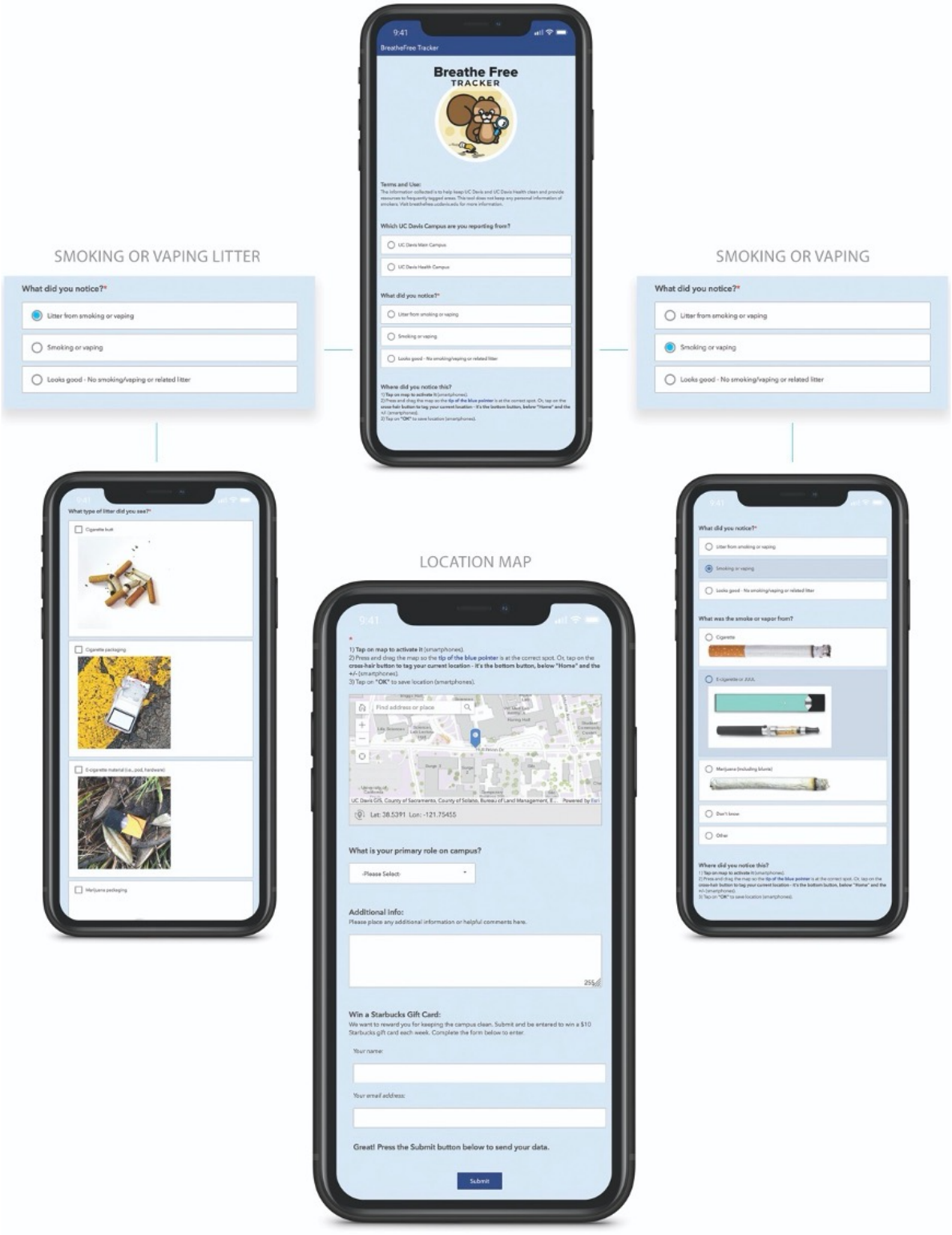
Empirically Driven Tracker Development

Based on focus group discussions, Tracker was designed for users to document tobacco use (smoking and vaping) or waste. Additionally, "no smoking or vaping" and "looks good" options were added for Tracker to have the capacity to reflect environmental improvements (Figure 1). Realistic images of tobacco products and waste were provided, in response to subtheme *Information Barriers*. Tracker instructions were framed to document observations of tobacco products and waste, not people, and the word "report" was purposefully avoided; instead, users were asked "what do you notice" and "where did you notice this" in response to subtheme *Framing as Policing Tool* as a barrier. Tracker was branded with environmentally themed names affiliated with the STF programs on each campus in response to subtheme *Environmental Motivator*. A squirrel mascot was adopted at University 1 and a hawk mascot at University 2, keeping with the environmental design theme.

The study team created a Tracker website for each university with features responsive to subtheme *Logistical Barriers*. The website-based platform Esri Survey 123 supported accessibility and subsequent integration with existing campus technological infrastructure, such as the campus app and websites. While both universities have a personal campus website account, mostly used by students, only University 1 allowed integration with Tracker. Multiple options for location tracking were provided to users: automatic GIS-enabled mobile location services, manual designation on the campus map, or manual selection of a building location from a drop-down menu. Additionally, graphic designers helped design images or promotional messages. Promotion consisted of flyers or advertisements and items (eg, stickers and bags) distributed at campus events.

Incentives for positive reinforcement, in response to subtheme *Positive Reinforcement* were also incorporated in Tracker. A positive affirmation message after the user pushed the submit button states, "Great! Your data was sent successfully. Thanks." To provide a reinforcing feedback loop, the message further stated, "Feedback is used to identify areas of concern and inform future policy efforts" where users may click on a link to a live map of submitted reports. The "Looks good. No smoking/vaping or related litter" option to report problem-free areas created an additional opportunity to build positive reinforcement in messages emphasizing a clean campus. Additionally, a weekly gift card raffle incentive was offered. The study team decided that gamification would need further development for future efforts.

Figure 1. Images of the final Tracker tool and key components.



Survey of Tracker Users

In total, 1163 discreet responses were documented in total from both universities' Tracker, and 717 individuals who provided an email ID were invited to participate in the user survey (n=605 users at University 1 and n=112 users at University 2). Tracker user survey responses by university are shown below (Table 2). The response rate for University 1 was 40% (242/605 invitations) and that for University 2 was 70% (78/112 invitations). Four duplicate surveys were omitted (University 2), yielding a total of 316 survey responses. Over half of the survey participants at both universities reported accessing Tracker through a website. Approximately a quarter of respondents at University 2 also reported using a QR code or campus app; however, few participants (10 of 242 participants, <5%) at University 1 reported using these methods.

Table 2. Tracker user survey results (N=316).

	Total, n (%)[a]	University 1 (n=242), n (%)	University 2 (n=74), n (%)
How did you access the Tracker?			
Personal campus website portal	N/A[b]	102[c] (45.9)	N/A
STF[d] or Tracker-specific website	148 (54.4)	114 (51.4)	34 (68.0)
Campus app	21 (7.7)	7 (3.2)	14 (28.0)
QR[e] code	18 (6.6)	5 (2.3)	13 (26.0)
Other	11 (4.0)	10 (4.5)	1 (2.0)
What motivates you to use the Tracker?			
A clean environment is important to me	212 (78.8)	172 (78.2)	40 (81.6)
Exposure to smoke or vapor is harmful to health	185 (68.8)	152 (69.1)	33 (67.3)
Smoking/vaping on campus encourages others to smoke/vape	110 (40.9)	87 (39.5)	23 (46.9)
I support the campus smoke and tobacco free policy	181 (67.3)	148 (67.3)	33 (67.3)
I believe campus policies should be followed	137 (50.9)	108 (49.1)	29 (59.2)
Smoking/vaping does not fit the campus image	87 (32.3)	72 (32.7)	15 (30.6)
Using the Tracker gives me an opportunity to win a gift card	98 (36.4)	82 (37.3)	16 (32.7)
Which of the following might discourage you from using the Tracker?			
I intend to use the Tracker, but forget	131 (50.6)	111 (51.9)	20 (44.4)
It takes too long to access the tool	108 (41.7)	87 (40.7)	21 (46.7)
There are too many questions	46 (17.8)	36 (16.8)	10 (22.2)
The instructions are not clear	23 (8.9)	19 (8.9)	4 (8.9)
The smoking/vaping or litter does not bother me very much	2 (0.8)	2 (0.9)	0
I am concerned people would know I reported	31 (12.0)	25 (11.7)	6 (13.3)
I do not want people to get in trouble	20 (7.7)	18 (8.4)	2 (4.4)
I feel uncomfortable using the Tracker	8 (3.1)	7 (3.3)	1 (2.2)
Other	34 (13.1)	31 (14.5)	3 (6.7)
Ideally, what would you like to happen after you report smoking/vaping or related litter?			
The litter will be cleaned up	205 (77.9)	171 (78.8)	34 (73.9)
The violator will be assigned community service	117 (44.5)	98 (45.2)	19 (41.3)
The violator will be fined	109 (41.4)	94 (43.3)	15 (32.6)
The violator will be assigned a diversion training	70 (26.6)	58 (26.7)	12 (26.1)
A new sign will be installed	69 (26.2)	61 (28.1)	8 (17.4)
The violator will be sanctioned	49 (18.6)	42 (19.4)	7 (15.2)
Other	24 (9.1)	19 (8.8)	5 (10.9)

[a]Total number of responses differed between questions and universities. In response to “How did you access the Tracker”, a total of 272 responses were recorded (222 at University 1; 50 at University 2); “What motivates you to use the Tracker”, a total of 269 responses (220 at University 1; 49 at University 2); “Which of the following might discourage you from using the Tracker” a total of 259 responses (214 at University 1; 45 at University 2); and “Ideally, what would you like to happen after you report smoking/ vaping or related litter?” a total of 263 responses (217 at University 1; 46 at University 2).

[b]N/A: not applicable.

[c]Percent sums over 100% are due to select-all type responses.

[d]STF: smoke and tobacco–free.

[e]QR: quick response.

The top motivations to use Tracker included environmental concerns, health concerns, and policy support. The majority (212/269, 78.8%) cited the importance of a clean environment, 185 (69%) participants were concerned about health hazards of exposure to smoke or vapor, and 181 (67%) voiced personal support for the campus STF policy.Approximately half (137/269, 50.9%) of the participants also expressed personal support that campus policies should be followed, and less than half (110/269, 40.9%) reported concerns that tobacco use encouraged others to smoke. Approximately one-third of participants (98/269, 36.4%) reported being motivated by campus images or the opportunity to win a gift card.

In regard to utilization barriers to Tracker, logistical issues were the leading barriers at both universities: of 259 participants, 131 (50.6%) intended but forgot to use Tracker, 108 (41.7%) were discouraged because it took too long to access Tracker, and 46 (17.8%) thought that Tracker had too many questions.Information barriers were infrequently cited, and few reported that the instructions were unclear or that they were unbothered by the tobacco use or waste. Other infrequently cited responses included concerns about a policing tool, concerns that others would know about their reporting, not wanting others to get in trouble, and being uncomfortable using Tracker.

When asked about users' ideal outcome following a Tracker report, the majority (205/263, 77.9%) of participants across both universities voiced environmental concerns with having the litter cleaned up. There was also interest in the consequences for policy violators including community service (117/263, 44.5%) and fines (109/263, 41.4%), followed by diversion trainings (70/263, 26.6%). Additional signage installation was selected by one-fourth (69/263) of respondents.

Discussion

Principal Findings

To our knowledge, this is the first reported study to describe the development of an empirically derived web-based Tracker for crowdsourcing campus environmental reports of tobacco use and waste to support STF college policy compliance. Campus focus group discussions identified barriers and facilitators for Tracker utilization, which informed Tracker development. Over 1 year, Tracker at both universities had been used by over 1000 people. In response to the survey, Tracker users cited environmental concerns as the leading motivator for utilization, complementary to focus group findings, followed by health concerns and STF policy support. Tracker users also cited having litter cleaned up as the ideal outcome, further highlighting their environmental concerns. Most potential barriers identified in the focus groups were not cited by Tracker users, although refinements should still be considered.

Health concerns from environmental and second-hand smoke/vapor, cited by Tracker users as top motivators for using Tracker, is consistent with broader population concerns, which may be due in part to existing public health educational campaigns. Environmental motives are consistent with a national study of 8000 students, which revealed that the majority (66%) reported that a college's commitment to environmental issues contributed to their decision to apply or attend a school [18]. Instead of focusing on individual smokers' behaviors and beliefs, which can create defensiveness, focusing on extrinsic issues, such as environmental concerns and second-hand smoke, has been recommended to promote a more receptive audience [19]. Public health education campaigns about the hazards of tobacco waste and second-hand smoke/vapor exposure are already found in the national Truth Campaign and state campaigns through the California Department of Public Health's Tobacco Free California [20,21] Future efforts incorporating environmentally focused public health campaign messages on campus could activate and reinforce crowdsourcing participation with Tracker.

Tracker was designed to minimize accessibility issues. The use of a web-based platform appears to have been a good choice, as accessing Tracker though campus websites was more common than through the Campus App or from QR codes. However, users cited the leading barrier as the time it took to access Tracker. Further study is needed to understand and resolve this issue. For example, it could be a convenience issue, such that users may be too focused on their commute to stop and access Tracker. This could also be related to Tracker users' second most frequently reported barrier: "intended to use the Tracker but forgot." Additional user research could inform modifications to make accessing Tracker faster and more efficient.

Positive reinforcers were built into the Tracker design to drive activation, crowdsourcing participation, and STF policy engagement. In response to focus group suggestions to provide a positive affirmation message, the Tracker message after a report submission stated its intended goal: to identify areas of concern and inform policy efforts. However, the user survey finding that the top desired action following the submission of a report was to "clean up litter." The positive affirmation messaging could be refined for greater specificity. For example, the message could provide more tangible feedback, such as "Your report has contributed to the clean-up of x number of tobacco litter items on campus." Interestingly, the opportunity to win a gift card was one of the least common reasons for using Tracker, suggesting that this incentive was not a strong reinforcement and could be scaled back or removed.

Our findings show mixed support for formal enforcement of Tracker. Focus group participants expressed concerns about policing, whereas Tracker users reported some preference for consequences for policy violators with community service or fines. Although punitive approaches such as fines may be perceived as a solution, they could decrease perceptions of university support among those who smoke [22]. A combined approach, including social and formal enforcement, may be warranted [9]. For example, the information gathered through crowdsourcing can be used to complement existing campus outreach programs, such as peer-led Ambassador programs, to target outreach or clean-up efforts in areas of concentrated tobacco activity [23]. For colleges without the resources for an ambassador or comparable enforcement program, the Tobacco Tracker provides a surveillance mechanism to monitor tobacco activity on campus and demonstrate to campus stakeholders the need for intervention [24].

There is growing interest in using crowdsourcing tools to effect change for the environment, including tracking tobacco use and litter. Ohio State's "Cleaner U" app focuses on all types of litter on their college campus [25] instead of just smoke and tobacco litter, similar to Tracker. Meanwhile, the World Health Organization's "Tobacco Spotter" app focuses on reporting compliance or noncompliance with a broad array of tobacco control policies in different countries, spanning tobacco retail, advertisement, and nonsmoking public policies [26]. Lastly, "Litterati," a smartphone app aimed at encouraging users to pick up litter in their respective communities, maps litter reports, including smoke and tobacco litter. Litterati's user community cleaned up approximately 50,000 pieces of litter within 1 week [27]. While these tools have effectively used crowdsourcing to report environmental issues, Tobacco Tracker uses crowdsourcing to record environmental reports of STF policies, namely tobacco use and waste, on college campuses.

Limitations

Our results must be interpreted with caution, owing to certain limitations. While both universities studied herein represent public systems with 100% STF policies, only 1 university per system was included, and the results may not be generalizable to those of other universities or outside of California. Furthermore, focus group sizes were small and participant-level characteristics were not assessed in the focus groups or user surveys. Therefore, it is not known how representative the respondents are of the university populations. Furthermore, Tracker user experience is limited to those who provided an email and responded to the survey and may not represent others, including those who decided not to use Tracker.

Conclusions

Environmental concerns successfully shaped Tracker's development and were commonly cited as the top motivators of use. Educating the campus community about the environmental hazards of tobacco waste may increase support for STF policies, especially among young adults [28]. Tracker is a promising tool that addresses environmental and health concerns to help support STF college policy compliance. Using Tracker, the campus community can be engaged to uphold STF policies by crowdsourcing data to help monitor campus areas in need of improvement. Future efforts may include refining Tracker for increased user accessibility or engagement and to target outreach and clean-up efforts. Additional research is needed to measure Tracker's effect on improving compliance with STF campus policies.

Acknowledgments

This study was funded by the California Tobacco Related Disease Research Program (#27IP-0041). No financial disclosures were reported by the authors of this paper. We would like to thank Raeann Davis, Allen Risley, Hiba Naz, Mirella Orozco, Ashley Reyes, Megan Hammerschmidt, Raul Raygoza, Jessica Contreras, Kiernan Salmon, and the California Youth Advocacy Network for their support, expertise, and invaluable contributions.

Authors' Contributions

SFL conducted the qualitative analysis and drafted the manuscript. KP designed and conducted the study and analyses, drafted the manuscript, and secured funding support. MMG conducted the qualitative analyses and drafted the manuscript. KC conducted the study and edited the final manuscript. MR conducted the quantitative analyses and the study, and edited the final manuscript. EKT designed and conducted the study and analyses, drafted the manuscript, and secured funding support.

Conflicts of Interest

None declared.

References

1. American CS, Center FTC. Tobacco-Free Generation Campus Initiative. Research. URL: https://www.tobaccofreecampus.org/ [accessed 2020-04-02]
2. American College Health Association. Position statement on tobacco on college and university campuses. J Am Coll Health 2012 Mar 15;60(3):266-267. [doi: 10.1080/07448481.2012.660440] [Medline: 22420706]
3. United States. Public Health Service. Office of the Surgeon General. Preventing Tobacco Use Among Youth and Young Adults: A Report of the Surgeon General. Washington, DC: U.S. Department of Health and Human Services, Public Health Service, Office of the Surgeon General; 2012.
4. Fichtenberg C, Glantz S. Effect of smoke-free workplaces on smoking behaviour: systematic review. BMJ 2002 Jul 27;325(7357):188 [FREE Full text] [doi: 10.1136/bmj.325.7357.188] [Medline: 12142305]
5. Sawdey M, Lindsay RP, Novotny TE. Smoke-free college campuses: no ifs, ands or toxic butts. Tob Control 2011 Feb 23;20 Suppl 1:i21-i24 [FREE Full text] [doi: 10.1136/tc.2010.040139] [Medline: 21504920]
6. Seo D, Macy JT, Torabi MR, Middlestadt SE. The effect of a smoke-free campus policy on college students' smoking behaviors and attitudes. Prev Med 2011 Aug 09;53(4-5):347-352. [doi: 10.1016/j.ypmed.2011.07.015] [Medline: 21851836]
7. Seitz CM, Kabir Z, Greiner BA, Davoren MP, McIntyre SC. Enforcement of Smoke/Tobacco-free Policies at Public Universities in the US. tobacco reg sci 2018 Jul 01;4(4):22-31. [doi: 10.18001/trs.4.4.3]

8. Jancey J, Bowser N, Burns S, Crawford G, Portsmouth L, Smith J. No smoking here: examining reasons for noncompliance with a smoke-free policy in a large university. Nicotine Tob Res 2014 Jul;16(7):976-983. [doi: 10.1093/ntr/ntu012] [Medline: 24532382]
9. Fallin-Bennett A, Roditis M, Glantz SA. The carrot and the stick? Strategies to improve compliance with college campus tobacco policies. J Am Coll Health 2017;65(2):122-130 [FREE Full text] [doi: 10.1080/07448481.2016.1262380] [Medline: 27869568]
10. Russette HC, Harris KJ, Schuldberg D, Green L. Policy compliance of smokers on a tobacco-free university campus. J Am Coll Health 2014;62(2):110-116. [doi: 10.1080/07448481.2013.854247] [Medline: 24456513]
11. Bennett BL, Deiner M, Pokhrel P. College anti-smoking policies and student smoking behavior: a review of the literature. Tob Induc Dis 2017;15:11 [FREE Full text] [doi: 10.1186/s12971-017-0117-z] [Medline: 28163669]
12. Brabham DC, Ribisl KM, Kirchner TR, Bernhardt JM. Crowdsourcing applications for public health. Am J Prev Med 2014 Mar;46(2):179-187. [doi: 10.1016/j.amepre.2013.10.016] [Medline: 24439353]
13. Schootman M, Nelson EJ, Werner K, Shacham E, Elliott M, Ratnapradipa K, et al. Emerging technologies to measure neighborhood conditions in public health: implications for interventions and next steps. Int J Health Geogr 2016 Jun 23;15(1):20 [FREE Full text] [doi: 10.1186/s12942-016-0050-z] [Medline: 27339260]
14. Cheung E, Romero T, Crespi C, Perez C, Huang JE, Pechmann C, et al. Undergraduate support for university smoke-free and vape-free campus policies and student engagement: a quasi-experimental intervention. J Am Coll Health 2020 Jul 01:1-9. [doi: 10.1080/07448481.2020.1782920] [Medline: 32610034]
15. Vardavas CI, Dimitrakaki C, Schoretsaniti S, Patelarou E, Filippidis FT, Connolly GN, et al. The role of the non-smoker in enforcing smoke-free laws. J Public Health Policy 2011 Mar;32(1):46-59. [doi: 10.1057/jphp.2010.45] [Medline: 21150943]
16. Gavin H. Understanding Research Methods and Statistics in Psychology. New York, NY: Sage Publications; 2008.
17. Sanguinetti A, Pritoni M, Salmon K, Morejohn J. TherMOOstat: Occupant Feedback to Improve Comfort and Efficiency on a University Campus. URL: https://www.aceee.org/files/proceedings/2016/data/papers/8_442.pdf [accessed 2021-08-31]
18. The Princeton Review 2017 College Hopes & Worries Survey Report. URL: https://www.princetonreview.com/cms-content/cohowo-survey-report-2017.pdf [accessed 2021-08-31]
19. Creating a tobacco-free campus: A policy guide. URL: http://tobaccofreecampus.Org/sites/default/files/resources/sea-king_collegestoolkit_112912.Pdf [accessed 2021-08-31]
20. Growing Wave of Quitters. Truth. URL: https://www.thetruth.com/home [accessed 2020-07-02]
21. California Tobacco Control Branch: Media Campaign Information. California Department of Public Health. URL: https://www.cdph.ca.gov/Programs/CCDPHP/DCDIC/CTCB/Pages/MediaCampaignInformation.aspx [accessed 2020-07-02]
22. Lee J, Purcell C, Chaney B. An Experiment Assessing Punitive versus Wellness Framing of a Tobacco-Free Campus Policy on Students' Perceived Level of University Support. Int J Environ Res Public Health 2017 Aug 20;14(8):938 [FREE Full text] [doi: 10.3390/ijerph14080938] [Medline: 28825650]
23. Ickes MJ, Rayens MK, Wiggins AT, Hahn EJ. A tobacco-free campus ambassador program and policy compliance. J Am Coll Health 2015;63(2):126-133 [FREE Full text] [doi: 10.1080/07448481.2014.990972] [Medline: 25612060]
24. Fallin A, Murrey M, Johnson AO, Riker CA, Rayens MK, Hahn EJ. Measuring compliance with tobacco-free campus policy. J Am Coll Health 2012;60(7):496-504. [doi: 10.1080/07448481.2012.670676] [Medline: 23002797]
25. Giannetto I. App to report litter, smokers at Ohio State. The Lantern. URL: https://www.thelantern.com/2014/02/app-seeks-create-cleaner-ohio-state-campus/ [accessed 2020-07-28]
26. Tobacco Spotter. URL: https://tobaccospotter.org/ [accessed 2020-07-02]
27. Spaen B. This App Encourages People To Take Pictures Of Their Trash To Make Recycling Easier. Green Matters. URL: https://www.greenmatters.com/community/2018/05/22/1PcHff/litterati-app-update [accessed 2020-07-28]
28. Epperson AE, Novotny TE, Halpern-Felsher B. Perceptions About the Impact of Cigarette Filters on the Environment and Smoking-Related Behaviors. J Adolesc Health 2021 Apr;68(4):823-826. [doi: 10.1016/j.jadohealth.2020.10.022] [Medline: 33288455]

Abbreviations

GIS: geographic information system
QR: quick response
STF: smoke and tobacco–free

Edited by R Kukafka; submitted 04.12.20; peer-reviewed by T Novotny, MA Styn; comments to author 01.02.21; revised version received 12.02.21; accepted 19.07.21; published 29.10.21.

Please cite as:
Loureiro SF, Pulvers K, Gosdin MM, Clift K, Rice M, Tong EK
The Development of a Web-Based Tobacco Tracker Tool to Crowdsource Campus Environmental Reports for Smoke and Tobacco–Free College Policies: Mixed Methods Study
J Med Internet Res 2021;23(10):e26280
URL: https://www.jmir.org/2021/10/e26280
doi:10.2196/26280
PMID:

Original Paper

Factors Associated With Behavioral and Psychological Symptoms of Dementia: Prospective Observational Study Using Actigraphy

Eunhee Cho[1], PhD; Sujin Kim[2], PhD; Sinwoo Hwang[3], MSN; Eunji Kwon[3], MSN; Seok-Jae Heo[4], MSc; Jun Hong Lee[5], MD, PhD; Byoung Seok Ye[6*], MD, PhD; Bada Kang[1*], PhD

[1]Mo-Im Kim Nursing Research Institute, College of Nursing, Yonsei University, Seoul, Republic of Korea

[2]College of Nursing, Yonsei University, Seoul, Republic of Korea

[3]Korea Armed Forces Nursing Academy, Daejeon, Republic of Korea

[4]Department of Biostatistics and Computing, Yonsei University Graduate School, Yonsei University, Seoul, Republic of Korea

[5]National Health Insurance Service, Ilsan Hospital, Goyang, Republic of Korea

[6]College of Medicine, Yonsei University, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Bada Kang, PhD
Mo-Im Kim Nursing Research Institute
College of Nursing
Yonsei University
50-1, Yonsei-Ro, Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 222283274
Email: bdkang@yuhs.ac

Abstract

Background: Although disclosing the predictors of different behavioral and psychological symptoms of dementia (BPSD) is the first step in developing person-centered interventions, current understanding is limited, as it considers BPSD as a homogenous construct. This fails to account for their heterogeneity and hinders development of interventions that address the underlying causes of the target BPSD subsyndromes. Moreover, understanding the influence of proximal factors—circadian rhythm–related factors (ie, sleep and activity levels) and physical and psychosocial unmet needs states—on BPSD subsyndromes is limited, due to the challenges of obtaining objective and/or continuous time-varying measures.

Objective: The aim of this study was to explore factors associated with BPSD subsyndromes among community-dwelling older adults with dementia, considering sets of background and proximal factors (ie, actigraphy-measured sleep and physical activity levels and diary-based caregiver-perceived symptom triggers), guided by the need-driven dementia-compromised behavior model.

Methods: A prospective observational study design was employed. Study participants included 145 older adults with dementia living at home. The mean age at baseline was 81.2 (SD 6.01) years and the sample consisted of 86 (59.3%) women. BPSD were measured with a BPSD diary kept by caregivers and were categorized into seven subsyndromes. Independent variables consisted of background characteristics and proximal factors (ie, sleep and physical activity levels measured using actigraphy and caregiver-reported contributing factors assessed using a BPSD diary). Generalized linear mixed models (GLMMs) were used to examine the factors that predicted the occurrence of BPSD subsyndromes. We compared the models based on the Akaike information criterion, the Bayesian information criterion, and likelihood ratio testing.

Results: Compared to the GLMMs with only background factors, the addition of actigraphy and diary-based data improved model fit for every BPSD subsyndrome. The number of hours of nighttime sleep was a predictor of the next day's sleep and nighttime behaviors (odds ratio [OR] 0.9, 95% CI 0.8-1.0; *P*=.005), and the amount of energy expenditure was a predictor for euphoria or elation (OR 0.02, 95% CI 0.0-0.5; *P*=.02). All subsyndromes, except for euphoria or elation, were significantly associated with hunger or thirst and urination or bowel movements, and all BPSD subsyndromes showed an association with environmental change. Age, marital status, premorbid personality, and taking sedatives were predictors of specific BPSD subsyndromes.

Conclusions: BPSD are clinically heterogeneous, and their occurrence can be predicted by different contributing factors. Our results for various BPSD suggest a critical window for timely intervention and care planning. Findings from this study will help devise symptom-targeted and individualized interventions to prevent and manage BPSD and facilitate personalized dementia care.

(J Med Internet Res 2021;23(10):e29001) doi:10.2196/29001

KEYWORDS

behavioral and psychological symptoms; dementia; older adults; actigraphy; sleep; activity; risk factors

Introduction

Behavioral and psychological symptoms of dementia (BPSD) constitute a core and prevalent feature of Alzheimer disease and related dementia [1], with most patients experiencing one or more types of symptoms over the course of the disease [2]. BPSD refer to frequently occurring symptoms of disturbed perception through content, mood, or behavior [3], which manifest into a wide range of forms, such as agitation, aggression, depression, apathy, wandering, and socially inappropriate behaviors [1]. Increasingly recognized as the most challenging aspect of dementia, BPSD present adverse outcomes, including decreased functioning and accelerated disease progression [4], and, if poorly managed, an increased risk of nursing home placement [5] and hospitalization [4,6,7]. BPSD are also associated with an increased burden on, and decreased quality of life of, caregivers [8-10].

Although BPSD are associated with neurological mechanisms of neurocognitive disease to some extent, previous studies have revealed that the actual occurrence of symptoms can be attributed to diverse personal factors (eg, medical conditions, premorbid personality, and physical and psychological unmet needs), social factors (eg, communication with caregivers, caregivers' stress and depression, and lack of social activities), and environmental factors (eg, overstimulation and lack of established routines), rather than to neurocognitive impairment alone [11,12]. While a few studies have provided evidence that supports frequent co-occurrence of individual BPSD, others have contended that BPSD are distinct and heterogeneous and have different determinants and consequences [13-15]. Determining key factors that predict the different types of BPSD is the first step, since it would guide the determination of which strategies should be chosen to target the underlying causes and ultimately prevent or manage the symptoms [12,16,17].

We based this study on the need-driven dementia-compromised behavior (NDB) model [18]. In the NDB model, BPSD arise from the interaction of two types of factors: (1) relatively stable background factors, including sociodemographic characteristics; neurological, cognitive, and functional status; underlying health; and personality traits, and (2) proximal factors, which are fluctuating and changing states of physical and psychological unmet needs and immediate environmental conditions. To further explain stress-related proximal factors, we incorporated the progressively lowered stress threshold (PLST) model [19], which posits that individuals with dementia are increasingly unable to manage stress, as the threshold for stress tolerance lowers as the disease progresses. If heightened perceived stressors accumulate, then exceed the stress threshold, the person with dementia starts to exhibit BPSD [19-21]. The PLST model recognizes circadian rhythms as a factor that influences the stress threshold level and accordingly postulates that impaired sleep and inadequate physical activity level are among the stressors that consequently trigger BPSD [22,23].

Previous studies have recommended person-centered nonpharmacological interventions as the first-line treatment modality for managing BPSD given the limited efficacy and undesired adverse effects of antipsychotics [24-26]. However, the effect size of the existing nonpharmacological interventions for BPSD has been small [27]. Thus, caregivers and providers continue to struggle to implement the most effective interventions targeting the underlying cause of certain types of BPSD due to limited evidence for target symptoms [28]. Moreover, current knowledge of BPSD is limited, since most existing studies considered BPSD imprecisely as a unitary construct by using measures that aggregated symptoms, which failed to account for the heterogeneity of different types of BPSD [12,14,29]. Although a range of factors have been associated with BPSD occurrence in the literature, most studies focused on only one aspect by testing the effect of a single intervention, such as music therapy, environmental modification, or structured recreational activities [30,31]. A recent scoping review highlighted a gap in the research, namely that diverse factors, including personal, social, and environmental, have been studied only for depression [12]. Surprisingly, few predictive models in existing studies accounted conjointly for the diverse factors [32,33].

A few studies found that sleep problems are associated with depression [34,35], apathy [35,36], agitation [37,38], and aggression [39]; however, efforts have been limited by the use of proxy-rated instruments for disturbed sleep [34,35,39]. Since caregiver-reported sleep measures require systemic and continuous observation of sleep behaviors, such a resource-intensive and time-consuming assessment may be infeasible in real-time care settings [40]. Further, as Blytt et al found significant discrepancies between proxy-rated sleep and actigraphy measures, proxy raters may underreport and fail to recognize sleep disturbances as compared with those measured by actigraphy [41].

Although physical activity has increasingly gained attention as a nonpharmacological approach to managing BPSD [42], few studies have examined the influence of physical activity on BPSD using observational rating instruments to measure activity level [43,44]. A study reported that physical activity objectively measured using actigraphy was significantly correlated with agitation and aggression but did not account for other potential factors [45]. In summary, surprisingly little is known about the

influence of circadian rhythms, including sleep and physical activity levels, on BPSD among community-dwelling older adults with dementia. This is due to the limited number of methodologically rigorous studies using objective measures, such as actigraphy, and accounting for personal, social, and environmental factors within a single study.

Understanding factors that predict BPSD subsyndromes is important not only for establishing early care planning for symptom prevention and management but for selecting the most effective person-centered interventions. Thus, this study explored factors associated with BPSD subsyndromes among community-dwelling older adults with dementia, considering background and proximal factors (ie, actigraphy-measured sleep and physical activity levels and diary-based caregiver-perceived symptom triggers), guided by the NDB model.

Methods

Design

This exploratory study employed a prospective observational design with two waves of data collection. Within each wave, background factors were collected at baseline, and repeated measures were collected for proximal factors over approximately 14 days. Reporting of this study adheres to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement [46].

Participants and Setting

Older adults with dementia living at home were recruited from outpatient neurology clinics at two tertiary hospitals and daycare centers in Seoul and the Gyeonggi region in Korea. Inclusion criteria for older adults with dementia were as follows: aged at least 65 years, a diagnosis of dementia by a physician, a Mini–Mental State Examination (MMSE) score of less than 24, and exhibition of BPSD at least once a week, as screened using the Korean version of the Neuropsychiatric Inventory at baseline [47]. The primary caregivers who provided the majority of care for the recruited older adults with dementia, lived in the same home, and were able to read and write in Korean were also included.

Procedure

The Institutional Review Board of the affiliated institutions approved this study. Participants were recruited via on-site visits between June 2018 and June 2019. We conducted the second-wave data collection for participants who agreed to continue in the study between July 2019 and June 2020. The assessment and data collection were conducted by research staff that included registered nurses with a master's degree and registered nurse research assistants with a bachelor's degree. The research staff with a master's degree taught the data collection protocol and trained the research assistants in on-site data collection. The research staff then contacted potential participants and explained the study purpose and procedures. All data collection was conducted at the participants' homes. After obtaining written informed consent at baseline from caregivers who were screened for eligibility, family caregivers completed a structured questionnaire consisting of sociodemographic information and standardized scales for physical and neuropsychological assessments, with assistance from the research staff. Following the baseline assessments, we placed an actigraphy device on the participants' wrists, and primary caregivers logged the BPSD manifestations into the BPSD diary on a daily basis for 14 consecutive days.

Measures

Outcome Measure: BPSD

Although the Neuropsychiatric Inventory [48] is a widely used tool to measure BPSD, its 2-week retrospective rating likely results in recall bias, since its rating reliability is dependent on caregiver training [49]. Therefore, we developed a BPSD diary, adapted from the Neuropsychiatric Inventory, to detect 12 behavioral and psychological symptoms commonly observed in patients with dementia on a daily basis: delusions, hallucinations, agitation or aggression, depression or dysphoria, anxiety, elation or euphoria, apathy or indifference, disinhibition, irritability or lability, aberrant motor behavior, sleep and nighttime behavior, and appetite or eating disorders. The BPSD diary is a structured, easy-to-use checklist that allows a caregiver to record the presence and severity (ie, mild, moderate, and severe) of individual symptoms.

Although BPSD encompass heterogeneous symptoms, clustering a number of individual symptoms that are highly correlated and contingently co-occur allows for a more meaningful interpretation of the study findings [17,50]. Moreover, use of subsyndromes rather than 12 individual symptoms can increase power, as the number of participants who endorse the symptom cluster will increase [51]. Since euphoria or elation, aberrant motor behavior, sleep and nighttime behavior, and appetite or eating disorders did not load on any clusters in previous studies [50-54], we included them as individual symptoms. The remaining eight symptoms were clustered as follows, based on results from factor analyses reported in the literature: (1) psychotic symptoms (hallucination and delusion), (2) affective symptoms (depression, anxiety, and apathy), and (3) hyperactivity (agitation or aggression, disinhibition, and irritability) [51,53-55].

Proximal Factors

Sleep and Physical Activity

Sleep and physical activity were objectively measured using actigraphy (the wGT3X-BT activity monitor; ActiGraph, LLC). Wrist actigraphy has been shown to be a reliable method by which to objectively measure sleep-wake cycles and is suitable for older adults with dementia [56,57]. Participants were asked to wear actigraphy devices on their nondominant wrist for 14 consecutive days starting on the day of the first visit. ActiLife software (version 6.13.3; ActiGraph, LLC) was used to evaluate actigraphy data and provide standard indices of sleep duration and fragmentation using vector magnitude counts in 60-second epoch data. We applied the Cole-Kripke algorithm to score a 1-minute epoch as asleep or awake [58]. We defined nighttime sleep as the time between 8 PM and 8 AM. We used total sleep time (TST) at night and wake time after sleep onset (WASO) at night as sleep parameters. The previous night's sleep parameters were used as potential predictors of BPSD occurring the following day. We also used actigraphy-derived energy

expenditure (kcal burned × 100/hour) as a parameter of activity level. Energy expenditure measured for 24 hours was used as a potential predictor of BPSD occurring on the same day, which reflected the physical conditions during the day when BPSD were exhibited.

Physiological Unmet Needs States and Interpersonal and Environmental Triggers

We used a checklist, embedded within the BPSD diary, to assess contributing factors that family caregivers considered to be immediate triggers for symptoms that encompassed physiological unmet needs states and interpersonal and environmental triggers. On the checklist, the triggers were specifically listed as follows: hunger or thirst, urination or bowel movement, pain or discomfort, sleep disturbance, noise, light, temperature, interpersonal trigger related to a person or persons who were present when the symptom occurred, environmental change, and other. Caregivers were asked to check all contributing factors that were present on the same day when the BPSD were present on this daily-basis symptom checklist. When caregivers checked the "other" option, they were asked to provide a brief description of the triggers. Diary-based caregiver-reported contributing factors that were observed immediately prior to BPSD occurrence were used as potential predictors of BPSD occurring on the same day.

Background Factors

Sociodemographic and Health Information

Participants provided demographic data, including age, gender, marital status, and education level. Dementia diagnosis and neurological and psychiatric medications were obtained via medical chart review or interviews with family caregivers and staff at the recruitment sites.

Cognitive and Functional Status

The Korean version of the MMSE (K-MMSE) was administered to evaluate cognitive functioning at baseline. The highest total score is 30 points, with a higher score indicating better cognitive function [59]. The Cronbach α score for Korean older adults with dementia was .91 [60]. The Korean version of the Activities of Daily Living scale (K-ADL) was administered to measure baseline functional status. This measure consists of seven items that are rated on a 3-point Likert scale, with higher scores indicating more severe dependency. The K-ADL has been validated for Korean older adults with dementia, exhibiting good psychometric properties (eg, Cronbach α score of .94) [61].

Personality Type

A family caregiver informant was asked to rate his or her family member's premorbid personality using the Korean version of the Big Five Inventory (BFI-K) [62]. The BFI-K consists of 15 items that are rated on a 5-point Likert scale, which assess five domains of personality features, namely openness, conscientiousness, neuroticism, extroversion, and agreeableness. The instrument is known to be reliable, with Cronbach α scores ranging from .67 to .82 in a Korean sample [62].

Statistical Analysis

For descriptive statistics, categorical variables were expressed as the number of dementia patients by percentage, while continuous variables were presented as means and SDs. We used two-sample independent *t* tests and Fisher exact tests to compare the difference between data from waves 1 and 2 (Tables S1-S3 in Multimedia Appendix 1). We used generalized linear mixed models (GLMMs) to explore factors predictive of BPSD subsyndromes, with estimation of odds ratios (ORs) and 95% CIs. The GLMM, a commonly used random-effects model, was well suited to this analysis because the approach permits random effects and is suitable for nonnormal or discrete outcomes that are repeatedly measured for each subject [63,64]. All models included a random effect of participant with a random intercept to account for heterogeneity among individuals; all other factors were modeled as fixed effects [65]. Once we decided on the variables to be included in the final model based on clinical and theoretical relevance, we then calculated variance inflation factors to assess multicollinearity among variables. We found that the data were suitable for regression analyses, given that all variance inflation factors were <5, which indicated that no multicollinearity could be detected (Multimedia Appendix 2). The GLMMs to predict each BPSD subsyndrome (ie, psychotic symptoms, affective symptoms, hyperactivity, euphoria or elation, aberrant motor behavior, sleep and nighttime behaviors, and appetite or eating disorders) were tested and compared. Predictors were included in the models in a block-wise manner: (1) background factors (age, gender, marital status, education, K-ADL score, K-MMSE score, premorbid personality type, taking sedative, and dementia type) and (2) proximal factors (TST at night, WASO at night, and caregiver-reported symptom triggers). The order in which the blocks entered the model was determined based on clinical and theoretical considerations: background factors entered first (Model 1), followed by proximal factors (Model 2) for each BPSD subsyndrome. Given that all available data from each participant in the data sets from waves 1 and 2 were used to fit the GLMM, a time variable that indicated the wave (wave 1 vs wave 2) was also present in the models. Fit of the models was assessed and compared using the Akaike information criterion (AIC), the Bayes information criterion (BIC), and the likelihood ratio test. Results were statistically significant if $P<.05$. All analyses were performed using R (version 3.6.3; The R Foundation).

Results

Background Characteristics

During the first wave of data collection, we initially recruited 166 eligible participants. Out of these participants, 18 (10.8%) were lost to follow-up due to refusal to wear the actigraphy devices (n=10), hospitalization or emergency room visits (n=6), death (n=1), and absence of family caregivers at home during the study period (n=1). A total of 3 (1.8%) participants were excluded from the analysis due to no available actigraphy data. During the second wave of data collection, out of 64 eligible participants, 5 (8%) were excluded from the analysis due to no available actigraphy data (n=3) and no available BPSD diary data (n=2). Consequently, this study included a total of 145 older adults with dementia who participated in the first wave.

Of the 145 participants, 59 (40.7%) older adults with dementia continued to participate in the second wave.

At baseline for the first wave, the participants' mean age was 81.23 (SD 6.01) years, with a female to male ratio of approximately 3:2 (86/145, 59.3% female). Education level for most participants was elementary school or below. Participants had moderate cognitive impairment, as indicated by the K-MMSE mean score of 17.28 (SD 5.5); the mean score of the K-ADL was 10.57 (SD 3.6). Background characteristics of the 145 participants in the first wave at baseline are illustrated in Table 1; these characteristics are presented separately for the first and second waves in Table S1 in Multimedia Appendix 1.

Table 1. Background characteristics of older adults with dementia in the first wave at baseline.

Variable	Value (N=145)
Age (years), mean (SD)	81.23 (6.01)
Gender, n (%)	
Male	59 (40.7)
Female	86 (59.3)
Marital status, n (%)	
Married	86 (59.3)
Bereaved or divorced	59 (40.7)
Education level, n (%)	
Elementary school or below	73 (50.3)
Middle school	14 (9.7)
High school	34 (23.4)
College or above	24 (16.6)
Total K-ADL[a] score, mean (SD)	10.57 (3.63)
Total K-MMSE[b] score, mean (SD)	17.28 (5.51)
Big Five Inventory score, mean (SD)	
Openness	8.60 (2.96)
Conscientiousness	11.66 (2.73)
Neuroticism	7.74 (2.87)
Extroversion	8.49 (1.87)
Agreeableness	10.92 (2.96)
Sedative (yes), n (%)	51 (35.2)
Dementia type, n (%)	
Alzheimer disease	71 (49.0)
Lewy body dementia	60 (41.4)
Vascular dementia	23 (15.9)
Other dementia	31 (21.4)

[a]K-ADL: Korean version of the Activities of Daily Living scale.

[b]K-MMSE: Korean version of the Mini–Mental State Examination.

Descriptive Statistics for BPSD Occurrence and Proximal Factors

Summary statistics of BPSD and proximal factors for the 2354 days that encompassed the first and second waves are presented in Table 2; these factors are presented separately for the first and second waves in Table S2 in Multimedia Appendix 1. Caregivers completed the BPSD diaries and participants wore actigraphy devices for a mean of 11.5 (SD 3.5) days. Of the 2354 days during which the symptoms were measured, the most frequently occurring BPSD subsyndrome was affective symptoms (n=548, 23.3%), followed by hyperactivity symptoms (n=350, 14.9%) and sleep and nighttime behaviors (n=275, 11.7%). Euphoria or elation (n=108, 4.6%) and aberrant motor behaviors (n=103, 4.4%) were relatively infrequent. The prevalence of BPSD across participants is also presented in Table S3 in Multimedia Appendix 1. The prevalence of affective symptoms was the highest (82/145, 56.6%), followed by hyperactivity (70/145, 48.3%).

The mean TST at night was 6.4 (SD 2.5) hours, and the mean WASO at night was 0.4 (SD 0.4) hours. The mean energy

expenditure (100 kcal/hour) was 0.21 (SD 0.14). Among the diverse set of caregiver-reported contributing factors, the most frequently reported factor was sleep disturbance (276/2354, 11.7%). Urination or bowel movement (220/2354, 9.3%), pain or discomfort (190/2354, 8.1%), and interpersonal triggers (171/2354, 7.3%) were also relatively frequent.

Table 2. Summary statistics of BPSD subsyndromes and proximal factors.

BPSD[a] subsyndromes and proximal factors	Value (N=2354 days)
Days of recording BPSD and actigraphy per person, mean (SD)	11.54 (3.50)
BPSD subsyndromes, n (%)	
Psychotic symptoms	234 (9.9)
Affective symptoms	548 (23.3)
Hyperactivity symptoms	350 (14.9)
Euphoria or elation	108 (4.6)
Aberrant motor behavior	103 (4.4)
Sleep and nighttime behavior	275 (11.7)
Appetite or eating disorders	193 (8.2)
Proximal factors	
Sleep- and energy-related factors, mean (SD)	
Total sleep time (hours at night)	6.44 (2.52)
Wake time after sleep onset (hours at night)	0.43 (0.36)
Energy expenditure (100 kcal/hour)	0.21 (0.14)
Caregiver-reported contributing factors, n (%)	
Hunger or thirst	149 (6.3)
Urination or bowel movement	220 (9.3)
Pain or discomfort	190 (8.1)
Sleep disturbance	276 (11.7)
Noise	76 (3.2)
Light	69 (2.9)
Temperature	103 (4.4)
Interpersonal trigger	171 (7.3)
Environmental change	87 (3.7)
Other	238 (10.1)

[a]BPSD: behavioral and psychological symptoms of dementia.

GLMMs Predicting BPSD Subsyndromes From Background and Proximal Factors

In the first step, the GLMMs for each BPSD subsyndrome were assessed with only background factors included in the models (Model 1; see results in Tables S1 and S2 in Multimedia Appendix 3). Based on Model 1, proximal factors—actigraphy-measured sleep and activity levels and diary-based caregiver-reported symptom triggers—were entered into Model 2. Table 3 presents *P* values of the likelihood ratio tests as well as the respective AIC and BIC scores for Models 1 and 2 for each BPSD subsyndrome. Likelihood ratio testing comparing Model 1, which consisted of only background factors, with Model 2 (ie, the full model), which included both background and proximal factors, showed that the full model had superior fit for every subsyndrome ($P<.001$). Goodness-of-fit statistics (ie, lower AIC and BIC values) indicated that models for every subsyndrome improved with the addition of proximal factors.

Table 3. Comparisons of generalized linear mixed models.

Outcomes (BPSD[a] subsyndromes)	Background factors (Model 1)		Background factors + proximal factors (Model 2)		*P* value[b]
	AIC[c]	BIC[d]	AIC	BIC	
Psychotic symptoms	872.94	993.98	756.84	952.81	<.001
Affective symptoms	1683.85	1804.89	1379.99	1575.96	<.001
Hyperactivity symptoms	1278.41	1399.46	1064.87	1260.84	<.001
Euphoria or elation	689.12	810.17	587.68	783.65	<.001
Aberrant motor behaviors	525.33	646.37	442.19	638.16	<.001
Sleep and nighttime behaviors	1215.43	1336.47	957.22	1153.20	<.001
Appetite or eating disorders	882.84	1003.88	792.87	988.84	<.001

[a]BPSD: behavioral and psychological symptoms of dementia.

[b]*P* values were calculated by the likelihood ratio test.

[c]AIC: Akaike information criterion.

[d]BIC: Bayesian information criterion.

Factors Predictive of BPSD Subsyndromes

The results of the GLMMs of background and proximal factors as predictors of BPSD subsyndromes are depicted in forest plots of ORs in Figures 1 and 2 and displayed in Tables S1 and S2 in Multimedia Appendix 4. The results are also summarized below for each subsyndrome.

Figure 1. Forest plots of odds ratios (with 95% CIs shown as whiskers) for the influence of background and proximal factors on psychotic symptoms, affective symptoms, hyperactivity symptoms, and euphoria or elation. ADL: Activities of Daily Living scale; MMSE: Mini–Mental State Examination; BFI: Big Five Inventory; ref.: reference; WASO: wake time after sleep onset.

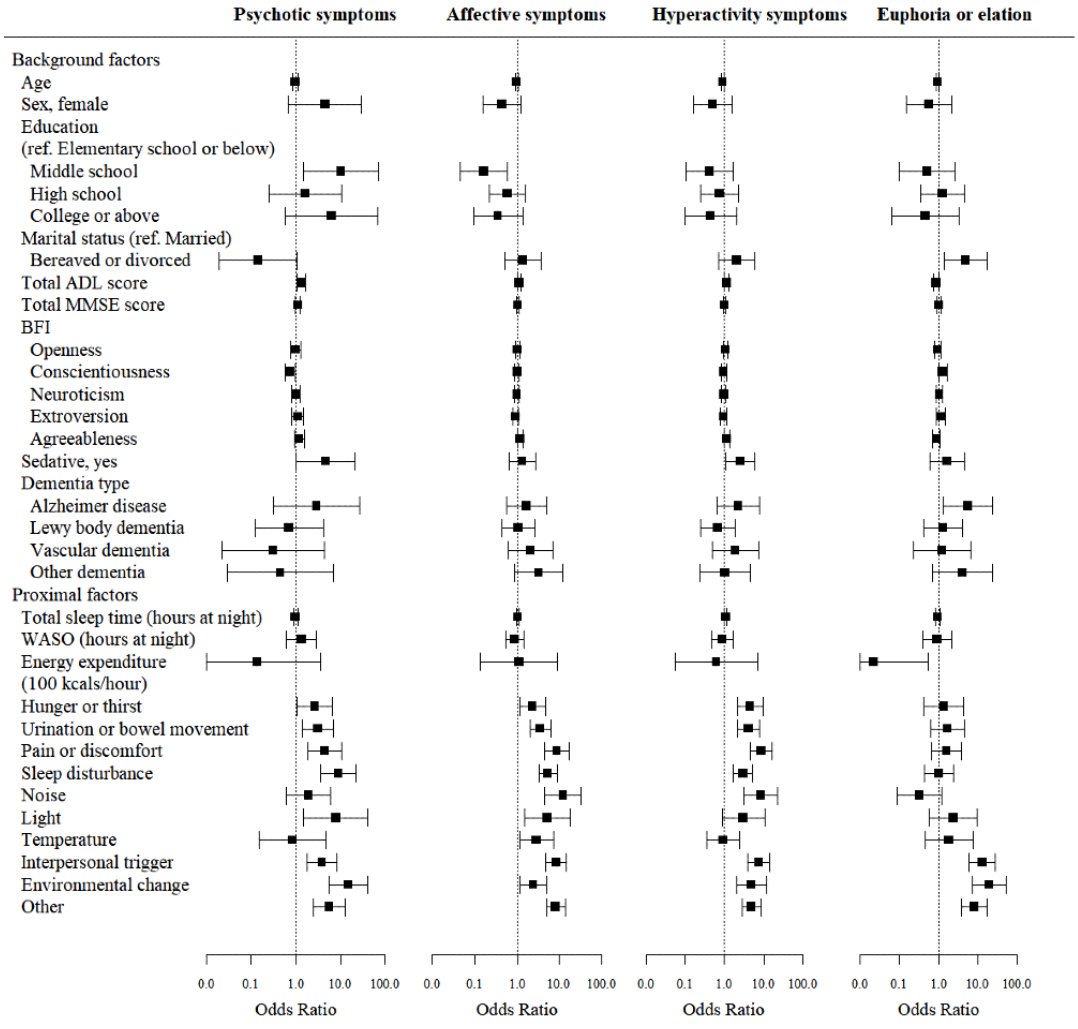
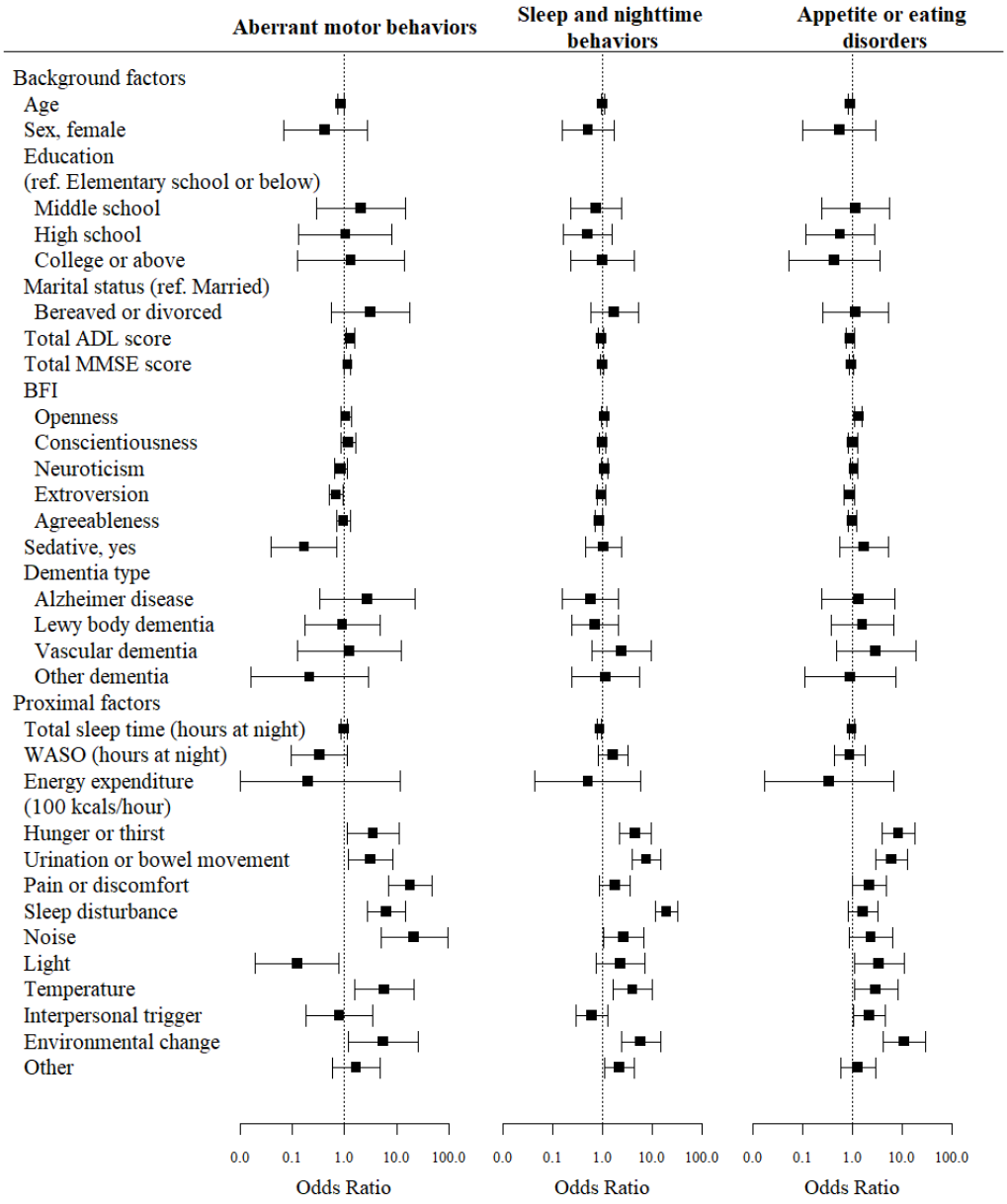


Figure 2. Forest plots of odds ratios (with 95% CIs shown as whiskers) for the influence of background and proximal factors on aberrant motor behavior, sleep and nighttime behaviors, and appetite or eating disorders. ADL: Activities of Daily Living scale; MMSE: Mini–Mental State Examination; BFI: Big Five Inventory; ref.: reference; WASO: wake time after sleep onset.



Psychotic Symptoms

Regarding proximal factors, patients were more likely to exhibit psychotic symptoms on the same day if they were exposed to environmental change (OR 14.7, 95% CI 5.3-40.8; *P*<.001) or inadequate light (OR 7.8, 95% CI 1.5-40.7; *P*=.02). Patients were more likely to exhibit psychotic symptoms (OR 8.8, 95% CI 3.6-21.7; *P*<.001) if they had sleep disturbance observed and reported by caregivers as well as pain or discomfort (OR 4.4, 95% CI 1.9-10.3; *P*<.001), urination or bowel movement–related problems (OR 3.1, 95% CI 1.4-6.9; *P*=.005), or hunger or thirst (OR 2.6, 95% CI 1.1-6.4; *P*=.03).

Background factors significantly associated with psychotic symptoms were educational attainment at the middle school level (OR 10.2, 95% CI 1.5-70.1; *P*=.02), greater impairment in activities of daily living (ADLs) (OR 1.3, 95% CI 1.1-1.7; *P*=.01), taking sedatives (OR 4.6, 95% CI 1.0-21.1; *P*=.049), and high conscientiousness traits (OR 0.8, 95% CI 0.6-1.0; *P*=.02).

Affective Symptoms

Of environmental condition–related proximal factors, increased likelihood of affective symptoms was significantly associated with noise (OR 12.1, 95% CI 4.5-32.4; *P*<.001), inadequate light (OR 5.1, 95% CI 1.5-17.4; *P*=.01), inadequate temperature (OR 2.9, 95% CI 1.2-7.1; *P*=.02), and being exposed to environmental change (OR 2.4, 95% CI 1.2-4.9; *P*=.02). Of proximal factors associated with physical unmet needs states, affective symptoms were more likely to be exhibited by patients with pain or discomfort (OR 8.6, 95% CI 4.6-16.4; *P*<.001), sleep disturbance (OR 5.3, 95% CI 3.3-8.7; *P*<.001), urination or bowel movement problems (OR 3.5, 95% CI 2.0-6.1; *P*<.001), or hunger or thirst (OR 2.3, 95% CI 1.2-4.6; *P*=.02). Interpersonal triggers related to a person or persons who were

present with the patients with dementia were also significantly associated with an increased likelihood of affective symptoms (OR 8.4, 95% CI 4.8-14.6; *P*<.001).

Background factors significantly associated with affective symptoms included agreeable personality traits (OR 1.2, 95% CI 1.0-1.3; *P*=.03), greater impairment in ADLs (OR 1.1, 95% CI 1.0-1.3; *P*=.03), and educational attainment at the middle school level (OR 0.2, 95% CI 0.0-0.6; *P*=.005).

Hyperactivity Symptoms

With respect to proximal factors, patients were more likely to exhibit hyperactivity symptoms if they experienced pain or discomfort (OR 8.8, 95% CI 4.6-16.9; *P*<.001), hunger or thirst (OR 4.5, 95% CI 2.1-9.5; *P*<.001), urination or bowel movement problems (OR 4.2, 95% CI 2.2-8.0; *P*<.001), or sleep disturbance (OR 3.0, 95% CI 1.7-5.2; *P*<.001). Increased likelihood of hyperactivity symptoms was also predicted by being exposed to noise (OR 8.5, 95% CI 3.2-22.6; *P*<.001) and environmental change (OR 4.9, 95% CI 2.1-11.8; *P*<.001). Interpersonal triggers also increased the predicted odds of hyperactivity symptoms (OR 7.6, 95% CI 4.0-14.4; *P*<.001).

Background factors that increased the predicted odds of hyperactivity symptoms were greater impairment in ADLs (OR 1.1, 95% CI 1.0-1.3; *P*=.04) and taking sedatives (OR 2.5, 95% CI 1.1-5.9; *P*=.03).

Euphoria or Elation

Regarding proximal factors, patients with greater levels of energy expenditure (100 kcal/hour) measured by actigraphy were less likely to exhibit euphoria or elation on the same day (OR 0.02, 95% CI 0.0-0.5; *P*=.02). In contrast, patients who were exposed to environmental change (OR 19.7, 95% CI 7.5-51.7; *P*<.001) or interpersonal triggers (OR 13.0, 95% CI 6.0-28.4; *P*<.001) were more likely to show symptoms of euphoria or elation.

Background factors significantly associated with euphoria or elation included diagnosis of Alzheimer disease (OR 5.7, 95% CI 1.4-23.6; *P*=.02), bereavement or divorced status (OR 4.9, 95% CI 1.4-17.1; *P*=.01), and high conscientiousness traits (OR 1.3, 95% CI 1.0-1.6; *P*=.046).

Aberrant Motor Behaviors

Of proximal factors, environmental condition–related factors that increased the likelihood of aberrant motor behaviors were being exposed to noise (OR 21.6, 95% CI 5.0-93.5; *P*<.001), inadequate temperature (OR 5.8, 95% CI 1.6-21.3; *P*=.008), and environmental change (OR 5.6, 95% CI 1.2-26.0; *P*=.03). In contrast, light was significantly associated with decreased likelihood of such behaviors (OR 0.1, 95% CI 0.0-0.8; *P*=.03). Patients were more likely to exhibit BPSD behaviors if they had pain or discomfort (OR 18.0, 95% CI 6.9-47.1; *P*<.001), sleep disturbance (OR 6.4, 95% CI 2.7-14.9; *P*<.001), hunger or thirst (OR 3.6, 95% CI 1.1-11.2; *P*=.03), or urination or bowel movement problems (OR 3.2, 95% CI 1.2-8.3; *P*=.02).

Background factors significantly associated with aberrant motor behaviors were greater impairment in ADLs (OR 1.3, 95% CI 1.1-1.6; *P*=.01), higher K-MMSE score (OR 1.2, 95% CI 1.0-1.3; *P*=.048), older age (OR 0.9, 95% CI 0.8-1.0; *P*=.02), high extroversion traits (OR 0.7, 95% CI 0.5-0.9; *P*=.01), and taking sedatives (OR 0.2, 95% CI 0.0-0.7; *P*=.02).

Sleep and Nighttime Behaviors

Among proximal factors, greater numbers of total nighttime sleep hours measured by actigraphy were significantly associated with decreased likelihood of sleep and nighttime behaviors occurring the next day (OR 0.9, 95% CI 0.8-1.0; *P*=.005). In contrast, patients were more likely to exhibit sleep and nighttime behaviors given sleep disturbance observed and reported by caregivers (OR 19.4, 95% CI 11.6-32.7; *P*<.001), urination or bowel movement problems (OR 7.7, 95% CI 4.0-14.7; *P*<.001), or hunger or thirst (OR 4.6, 95% CI 2.2-9.6; *P*<.001). Increased odds of sleep and nighttime behaviors were also associated with environmental change (OR 5.9, 95% CI 2.4-14.6; *P*<.001), inadequate temperature (OR 4.0, 95% CI 1.6-10.2; *P*=.003), and noise (OR 2.7, 95% CI 1.0-6.8; *P*=.04). None of the background factors were significantly associated with the likelihood of sleep and nighttime behaviors.

Appetite or Eating Disorders

Of proximal factors, increased likelihood of appetite or eating disorders was associated with exposure to environmental change (OR 11.1, 95% CI 4.2-29.3; *P*<.001), inadequate lighting (OR 3.4, 95% CI 1.1-10.8; *P*=.03), and inadequate temperature (OR 2.9, 95% CI 1.1-8.0; *P*=.04). Patients were also more likely to exhibit appetite or eating disorders on the same day if they experienced hunger or thirst (OR 8.5, 95% CI 4.0-18.1; *P*<.001) or urination or bowel movement problems (OR 6.0, 95% CI 2.9-12.6; *P*<.001). Interpersonal triggers also predicted an increased likelihood of appetite or eating disorders (OR 2.2, 95% CI 1.0-4.6; *P*=.04). The openness trait was the only background factor that significantly predicted appetite or eating disorders (OR 1.3, 95% CI 1.1-1.6; *P*=.003).

Discussion

Principal Findings

Although BPSD are recognized to arise from background and proximal factors, it is unclear exactly how the different BPSD subsyndromes are influenced by specific proximal factors, such as sleep and physical activity levels and physical and psychosocial unmet needs states. This study explored factors associated with BPSD subsyndromes among community-dwelling older adults with dementia, including sets of background and proximal factors (ie, actigraphy-measured sleep and physical activity levels and diary-based caregiver-perceived symptom triggers) guided by the NDB model. This research expanded upon the limited previous research by using actigraphy and diary-based assessments, thereby enabling the collection of objective and/or continuous time-varying proximal data.

The Effects of Proximal Factors on the Occurrence of BPSD Subsyndromes

Our results demonstrated that BPSD subsyndromes were predictable based on proximal factors, including actigraphy-measured sleep and activity levels and diary-based

caregiver-reported symptom triggers. The models to predict BPSD subsyndromes demonstrated better predictive ability when both background and proximal factors served as predictors in the models compared to those with only background factors. While BPSD have been largely considered unpredictable [17] and few previous studies have considered the impact of preventative approaches to BPSD [11], our results support a paradigm shift to an individualized approach to BPSD subsyndromes through prediction and early prevention. In particular, caregiver-reported physical, psychosocial, and environmental triggers largely influenced most subsyndromes, except for euphoria or elation. This suggests that, despite the unmodifiable nature of background factors, most BPSD can be prevented through early identification of at-risk individuals and timely assessment of targeted physical and psychosocial unmet needs.

However, it remains unclear as to which BPSD subsyndromes are under circadian control (ie, influenced by sleep and activity levels). In this study, greater nighttime sleep hours were associated with a lower likelihood of sleep and nighttime behaviors, manifested as nighttime awakening or excessive daytime napping, occurring the next day. This finding is consistent with previous studies, in that sleep problems are both a risk factor and a symptom outcome [66,67]. In previous studies, the association between sleep disturbance and BPSD has been inconsistent and variable among different BPSD. For example, a previous study of a Korean sample found that more severe sleep impairment, measured by the Korean version of the Pittsburgh Sleep Quality Index (PSQI) [68], was significantly associated with apathy or indifference, but not with other types of symptoms [69]. In another study conducted in China, the total scores of the PSQI and its subscales were significantly correlated with depression, apathy, and sleep and nighttime behaviors [35]. Moreover, one study of a hospitalized sample noted that the average number of sleep minutes significantly predicted agitation and irritability [37]; another study of nursing home residents with severe dementia found that a greater number of nighttime sleep hours was a predictor of daytime aggressive behaviors [70]. One possible explanation for the nonsignificant relationship between sleep parameters and aggression in our study could lie in the differences in sample characteristics. Given that the mean K-MMSE score in our sample indicated moderate cognitive impairment, nighttime sleep perhaps affects aggression when cognitive impairment is severe. Regarding the activity level, euphoria or elation was associated with energy expenditure measured by actigraphy. A previous study showed that an increased physical activity level was associated with an increased likelihood of agitation and aggression in hospitalized patients with severe dementia [45]. This may be because increased physical activities could reflect heightened agitation and aggressive symptoms per se rather than physical exercise. To summarize, it still remains unclear as to which BPSD subsyndromes are under circadian control. The inconsistent results among prior studies call for further research to build on extant knowledge.

Implications for Different Approaches to Heterogeneous BPSD Subsyndromes

While the NDB model explains the effect of the interplay between background and proximal factors on the broad construct of BPSD [71], the findings from this study extend the NDB model by indicating which specific background and proximal factors exert their effects on which specific BPSD subsyndromes. For example, while a greater number of nighttime sleep hours was associated with attenuated likelihood of sleep and nighttime behaviors occurring the next day, nighttime sleep hours did not influence other BPSD subsyndromes. While physical unmet needs states (eg, pain or discomfort, sleep disturbance, and hunger or thirst) and physical environmental conditions (eg, noise and lighting) were significant factors associated with affective symptoms, they were not found to influence symptoms of euphoria or elation. Instead, actigraphy-measured energy expenditure was a proximal factor that predicted euphoria or elation, whereas it did not influence any other symptoms. Thus, the results underscore that etiological-based differentiated interventions are needed to focus on the factors underlying the target BPSD subsyndromes [17].

Accordingly, our study provides empirical evidence regarding the heterogeneity of BPSD [13-15] by revealing the distinct underlying predictors for the seven BPSD subsyndromes. Results showed that there were multiple predictors of all BPSD; however, the mechanisms underlying how a set of neurobiological, psychosocial, and environmental factors interplay and subsequently result in specific symptoms remain elusive [72]. Since the etiopathogenesis of BPSD is complex, multifactorial, and variable depending on the BPSD subsyndromes or individual symptoms [12,72], it is challenging for health care providers and caregivers to decide which interventions should be used to manage certain types of symptoms. Several interventional algorithms have been proposed to simplify complex information on the relationships among multifactorial factors and make it usable, such as the Describe, Investigate, Create, and Evaluate (DICE) approach [11,73] and the BPSD–Describe and Measure, Analyze, Treat, and Evaluate (DATE) algorithm [72]. Although the algorithms have been helpful for caregivers to simplify the complex nature of heterogeneous BPSD trajectories, further research is needed to develop predictive models that will guide health care providers and caregivers to move toward personalized BPSD care. Our study represents a starting point for developing more data-driven predictive algorithms tailored to BPSD subsyndromes to advance more precise BPSD care.

Advancing BPSD Research and Practice by Leveraging Digital Health Technologies

The results also facilitate the use of technological advances to achieve personalized BPSD care. Assessment of BPSD has largely relied on information reported by informal or formal caregivers, which has resulted in information bias and caregiver burden [74]. While the traditional nomothetic approach to assessment of BPSD has limitations in terms of capturing the unique trajectories of BPSD manifestations and their contexts, human behavior and psychology research has considered idiographic approaches to capture prospective and

time-dependent variation within individuals [75]. Our study demonstrates the potential of using real-world data to measure fluctuating contextual information by leveraging digital health technologies to advance BPSD research and practice using an idiographic and personalized approach. Although concerns may exist regarding applying actigraphy to patients with dementia, our results demonstrated the feasibility of assessing sleep and activity levels using actigraphy in individuals with dementia, as evidenced by relatively low attrition rates and low amounts of missing data.

Along with the objectively measured actigraphy data, the results of the GLMMs revealed the large magnitude of the associations between diary-based caregiver-perceived symptom triggers and BPSD subsyndromes, which indicates the importance of collateral information from caregivers for predicting BPSD subsyndromes. Information and communications technologies (ICTs) can be a solution to complement paper-based symptom diaries by allowing the collection of high-frequency contextual information on BPSD that is episodic and evolves over time [76]. Future studies could adopt mobile app–based symptom diaries, which would facilitate the tracking and monitoring of BPSD by using push notifications to ensure measurements are not missed and to assist caregivers in easily checking symptoms anywhere and anytime with their mobile phones. Ecological momentary assessment (EMA) [77] is a promising tool that could be incorporated into mobile app–based symptom diaries. EMA data collected using electronic diaries could capture the momentary aspects of BPSD occurrence, rather than relying on recall over time and how BPSD manifestations vary over time and situations through repeated assessments [77]. Moreover, integrating prediction models into digital devices, such as smartphones, or connecting them to electronic health records could further improve the clinical utility of prediction models.

Limitations

Several limitations should be noted. First, we used a proxy measure of BPSD based on caregivers' reporting, rather than direct observation of symptoms, which may have introduced observer and recall bias. Second, because BPSD and caregiver-reported contributing factors were checked on a daily rather than episodic basis, caregivers could check multiple types of symptoms and related triggers if different types of symptoms were observed within a day. This may have made it difficult to identify the caregiver-reported factors that contributed to specific types of BPSD that occurred during the day. Future mobile app–based symptom diaries that collect episodic data would make it possible to disentangle aggregated proximal factors and symptoms that occur within a day. Finally, although it was not within the scope of this study, the bidirectionality of the relationships between proximal factors and BPSD is unclear in the current literature. A more thorough understanding of the reciprocal effects of sleep and physical activity levels with BPSD is needed to understand the mechanisms that underly the associations.

Conclusions

In conclusion, this prospective study demonstrated that BPSD are clinically heterogeneous, and their occurrence can be predicted by different contributing factors. Our results for various BPSD suggest a critical window for timely intervention and care planning. These findings represent the starting point for the establishment of a prediction model tailored to specific symptoms. However, further studies with larger samples are needed to confirm these findings. Nonetheless, these findings will assist in devising symptom-targeted and individualized interventions to prevent and manage different types of BPSD and to ultimately facilitate personalized dementia care.

Acknowledgments

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT) (grant NRF-2018R1A2B6003506) and supported by the Basic Science Research Program through the NRF funded by the Ministry of Education (grant NRF-2020R1A6A1A03041989). The funders had no role in the study design, data collection and interpretation, or preparation of the manuscript.

Authors' Contributions

EC contributed to the study concept and design. SK, SH, EK, JHL, and BSY were responsible for acquisition of the data. EC, BK, and S-JH analyzed and interpreted the data and drafted the manuscript. All authors participated in the interpretation of the results and revision of the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Background characteristics of older adults with dementia, summary statistics of behavioral and psychological symptoms of dementia (BPSD) and proximal factors, and summary statistics for the prevalence of BPSD subsyndromes.
[DOCX File , 23 KB - jmir_v23i10e29001_app1.docx]

Multimedia Appendix 2
Summary statistics for the prevalence of behavioral and psychological symptoms of dementia (BPSD) subsyndromes.

[DOCX File , 17 KB - jmir_v23i10e29001_app2.docx]

Multimedia Appendix 3
Results of generalized linear mixed models (GLMMs) for psychotic, affective, and hyperactivity symptoms and euphoria or elation (Model 1) and results of GLMMs for aberrant motor behaviors, sleep and nighttime behaviors, and appetite or eating disorders (Model 1).
[DOCX File , 21 KB - jmir_v23i10e29001_app3.docx]

Multimedia Appendix 4
Results of generalized linear mixed models (GLMMs) for psychotic, affective, and hyperactivity symptoms and euphoria or elation (Model 2, full model) and results of GLMMs for aberrant motor behaviors, sleep and nighttime behaviors, and appetite or eating disorders (Model 2, full model).
[DOCX File , 26 KB - jmir_v23i10e29001_app4.docx]

References

1. Lyketsos CG, Carrillo MC, Ryan JM, Khachaturian AS, Trzepacz P, Amatniek J, et al. Neuropsychiatric symptoms in Alzheimer's disease. Alzheimers Dement 2011 Sep;7(5):532-539 [FREE Full text] [doi: 10.1016/j.jalz.2011.05.2410] [Medline: 21889116]
2. Zhao Q, Tan L, Wang H, Jiang T, Tan M, Tan L, et al. The prevalence of neuropsychiatric symptoms in Alzheimer's disease: Systematic review and meta-analysis. J Affect Disord 2016 Jan 15;190:264-271. [doi: 10.1016/j.jad.2015.09.069] [Medline: 26540080]
3. Finkel SI, Costa e Silva J, Cohen G, Miller S, Sartorius N. Behavioral and psychological signs and symptoms of dementia: A consensus statement on current knowledge and implications for research and treatment. Int Psychogeriatr 1996;8 Suppl 3:497-500. [doi: 10.1017/s1041610297003943] [Medline: 9154615]
4. Wancata J, Windhaber J, Krautgartner M, Alexandrowicz R. The consequences of non-cognitive symptoms of dementia in medical hospital departments. Int J Psychiatry Med 2003;33(3):257-271. [doi: 10.2190/ABXK-FMWG-98YP-D1CU] [Medline: 15089007]
5. Zuidema S, Koopmans R, Verhey F. Prevalence and predictors of neuropsychiatric symptoms in cognitively impaired nursing home patients. J Geriatr Psychiatry Neurol 2007 Mar;20(1):41-49. [doi: 10.1177/0891988706292762] [Medline: 17341770]
6. Herrmann N, Lanctôt KL, Sambrook R, Lesnikova N, Hébert R, McCracken P, et al. The contribution of neuropsychiatric symptoms to the cost of dementia care. Int J Geriatr Psychiatry 2006 Oct;21(10):972-976. [doi: 10.1002/gps.1594] [Medline: 16955429]
7. O'Brien JA, Caro JJ. Alzheimer's disease and other dementia in nursing homes: Levels of management and cost. Int Psychogeriatr 2001 Sep;13(3):347-358. [doi: 10.1017/s1041610201007736] [Medline: 11768381]
8. Kunik ME, Snow AL, Davila JA, McNeese T, Steele AB, Balasubramanyam V, et al. Consequences of aggressive behavior in patients with dementia. J Neuropsychiatry Clin Neurosci 2010;22(1):40-47. [doi: 10.1176/jnp.2010.22.1.40] [Medline: 20160208]
9. Ornstein K, Gaugler JE. The problem with "problem behaviors": A systematic review of the association between individual patient behavioral and psychological symptoms and caregiver depression and burden within the dementia patient-caregiver dyad. Int Psychogeriatr 2012 Oct;24(10):1536-1552 [FREE Full text] [doi: 10.1017/S1041610212000737] [Medline: 22612881]
10. Baharudin AD, Din NC, Subramaniam P, Razali R. The associations between behavioral-psychological symptoms of dementia (BPSD) and coping strategy, burden of care and personality style among low-income caregivers of patients with dementia. BMC Public Health 2019 Jun 13;19(Suppl 4):447 [FREE Full text] [doi: 10.1186/s12889-019-6868-0] [Medline: 31196141]
11. Kales HC, Gitlin LN, Lyketsos CG. Assessment and management of behavioral and psychological symptoms of dementia. BMJ 2015 Mar 02;350:h369 [FREE Full text] [doi: 10.1136/bmj.h369] [Medline: 25731881]
12. Kolanowski A, Boltz M, Galik E, Gitlin LN, Kales HC, Resnick B, et al. Determinants of behavioral and psychological symptoms of dementia: A scoping review of the evidence. Nurs Outlook 2017;65(5):515-529 [FREE Full text] [doi: 10.1016/j.outlook.2017.06.006] [Medline: 28826872]
13. Volicer L. Review of programs for persons facing death with dementia. Healthcare (Basel) 2019 Apr 15;7(2):1-12 [FREE Full text] [doi: 10.3390/healthcare7020062] [Medline: 30991668]
14. Volicer L, Galik E. Agitation and aggression are 2 different syndromes in persons with dementia. J Am Med Dir Assoc 2018 Dec;19(12):1035-1038. [doi: 10.1016/j.jamda.2018.07.014] [Medline: 30197272]
15. Kang B, Karel MJ, Corazzini KN, McConnell ES. A mixed methods study on the manifestations of behavioural symptoms of dementia among veterans with and without posttraumatic stress disorder. J Adv Nurs 2021 May 10:3176-3188. [doi: 10.1111/jan.14864] [Medline: 33969916]

16. Gitlin LN, Kales HC, Lyketsos CG. Nonpharmacologic management of behavioral symptoms in dementia. JAMA 2012 Nov 21;308(19):2020-2029 [FREE Full text] [doi: 10.1001/jama.2012.36918] [Medline: 23168825]
17. Cerejeira J, Lagarto L, Mukaetova-Ladinska EB. Behavioral and psychological symptoms of dementia. Front Neurol 2012;3:73 [FREE Full text] [doi: 10.3389/fneur.2012.00073] [Medline: 22586419]
18. Algase DL, Beck C, Kolanowski A, Whall A, Berent S, Richards K, et al. Need-driven dementia-compromised behavior: An alternative view of disruptive behavior. Am J Alzheimers Dis Other Demen 2016 Sep 04;11(6):10-19. [doi: 10.1177/153331759601100603]
19. Hall GR, Buckwalter KC. Progressively lowered stress threshold: A conceptual model for care of adults with Alzheimer's disease. Arch Psychiatr Nurs 1987 Dec;1(6):399-406. [Medline: 3426250]
20. Smith M, Hall GR, Gerdner L, Buckwalter KC. Application of the progressively lowered stress threshold model across the continuum of care. Nurs Clin North Am 2006 Mar;41(1):57-81, vi. [doi: 10.1016/j.cnur.2005.09.006] [Medline: 16492454]
21. Smith M, Gerdner LA, Hall GR, Buckwalter KC. History, development, and future of the progressively lowered stress threshold: A conceptual model for dementia care. J Am Geriatr Soc 2004 Oct;52(10):1755-1760. [doi: 10.1111/j.1532-5415.2004.52473.x] [Medline: 15450057]
22. Richards KC, Beck CK. Progressively lowered stress threshold model: Understanding behavioral symptoms of dementia. J Am Geriatr Soc 2004 Oct;52(10):1774-1775. [doi: 10.1111/j.1532-5415.2004.52477.x] [Medline: 15450062]
23. Volicer L, Hurley AC. Management of behavioral symptoms in progressive degenerative dementias. J Gerontol A Biol Sci Med Sci 2003 Sep;58(9):M837-M845. [doi: 10.1093/gerona/58.9.m837] [Medline: 14528041]
24. Lyketsos CG, Colenda CC, Beck C, Blank K, Doraiswamy MP, Kalunian DA, Task Force of American Association for Geriatric Psychiatry. Position statement of the American Association for Geriatric Psychiatry regarding principles of care for patients with dementia resulting from Alzheimer disease. Am J Geriatr Psychiatry 2006 Jul;14(7):561-572. [doi: 10.1097/01.JGP.0000221334.65330.55] [Medline: 16816009]
25. APA Work Group on Alzheimer's Disease and Other Dementias, Rabins PV, Blacker D, Rovner BW, Rummans T, Schneider LS, Steering Committee on Practice Guidelines, et al. American Psychiatric Association practice guideline for the treatment of patients with Alzheimer's disease and other dementias. Second edition. Am J Psychiatry 2007 Dec;164(12 Suppl):5-56. [Medline: 18340692]
26. Reus VI, Fochtmann LJ, Eyler AE, Hilty DM, Horvitz-Lennon M, Jibson MD, et al. The American Psychiatric Association practice guideline on the use of antipsychotics to treat agitation or psychosis in patients with dementia. Am J Psychiatry 2016 May 01;173(5):543-546. [doi: 10.1176/appi.ajp.2015.173501] [Medline: 27133416]
27. Dyer SM, Harrison SL, Laver K, Whitehead C, Crotty M. An overview of systematic reviews of pharmacological and non-pharmacological interventions for the treatment of behavioral and psychological symptoms of dementia. Int Psychogeriatr 2018 Mar;30(3):295-309. [doi: 10.1017/S1041610217002344] [Medline: 29143695]
28. Kales HC, Gitlin LN, Lyketsos CG. When less is more, but still not enough: Why focusing on limiting antipsychotics in people with dementia is the wrong policy imperative. J Am Med Dir Assoc 2019 Sep;20(9):1074-1079 [FREE Full text] [doi: 10.1016/j.jamda.2019.05.022] [Medline: 31399358]
29. Gitlin L, Winter L, Dennis M, Hodgson N, Hauck W. Targeting and managing behavioral symptoms in individuals with dementia: A randomized trial of a nonpharmacological intervention. J Am Geriatr Soc 2010 Aug;58(8):1465-1474 [FREE Full text] [doi: 10.1111/j.1532-5415.2010.02971.x] [Medline: 20662955]
30. Majić T, Pluta JP, Mell T, Treusch Y, Gutzmann H, Rapp MA. Correlates of agitation and depression in nursing home residents with dementia. Int Psychogeriatr 2012 Nov;24(11):1779-1789. [doi: 10.1017/S104161021200066X] [Medline: 22591584]
31. Scales K, Zimmerman S, Miller SJ. Evidence-based nonpharmacological practices to address behavioral and psychological symptoms of dementia. Gerontologist 2018 Jan 18;58(suppl_1):S88-S102 [FREE Full text] [doi: 10.1093/geront/gnx167] [Medline: 29361069]
32. Beck C, Richards K, Lambert C, Doan R, Landes RD, Whall A, et al. Factors associated with problematic vocalizations in nursing home residents with dementia. Gerontologist 2011 Jun;51(3):389-405 [FREE Full text] [doi: 10.1093/geront/gnq129] [Medline: 21292752]
33. Morgan RO, Sail KR, Snow AL, Davila JA, Fouladi NN, Kunik ME. Modeling causes of aggressive behavior in patients with dementia. Gerontologist 2013 Oct;53(5):738-747. [doi: 10.1093/geront/gns129] [Medline: 23103521]
34. Arbus C, Gardette V, Cantet CE, Andrieu S, Nourhashémi F, Schmitt L, REAL.FR Group. Incidence and predictive factors of depressive symptoms in Alzheimer's disease: The REAL.FR study. J Nutr Health Aging 2011 Aug;15(8):609-617. [doi: 10.1007/s12603-011-0061-1] [Medline: 21968854]
35. Zhou G, Liu S, Yu X, Zhao X, Ma L, Shan P. High prevalence of sleep disorders and behavioral and psychological symptoms of dementia in late-onset Alzheimer disease: A study in Eastern China. Medicine (Baltimore) 2019 Dec;98(50):e18405 [FREE Full text] [doi: 10.1097/MD.0000000000018405] [Medline: 31852160]
36. Mulin E, Zeitzer JM, Friedman L, Le Duff F, Yesavage J, Robert PH, et al. Relationship between apathy and sleep disturbance in mild and moderate Alzheimer's disease: An actigraphic study. J Alzheimers Dis 2011;25(1):85-91. [doi: 10.3233/JAD-2011-101701] [Medline: 21335662]

37. Tanev KS, Winokur A, Pitman RK. Sleep patterns and neuropsychiatric symptoms in hospitalized patients with dementia. J Neuropsychiatry Clin Neurosci 2017;29(3):248-253 [FREE Full text] [doi: 10.1176/appi.neuropsych.16090166] [Medline: 28294708]
38. Rose K, Beck C, Tsai P, Liem P, Davila D, Kleban M. Sleep disturbances and nocturnal agitation behaviors in older adults with dementia. Sleep 2011;34(6):779-786. [doi: 10.5665/sleep.1048]
39. Moran M, Lynch CA, Walsh C, Coen R, Coakley D, Lawlor BA. Sleep disturbance in mild to moderate Alzheimer's disease. Sleep Med 2005 Jul;6(4):347-352. [doi: 10.1016/j.sleep.2004.12.005] [Medline: 15978517]
40. Hjetland GJ, Nordhus IH, Pallesen S, Cummings J, Tractenberg RE, Thun E, et al. An actigraphy-based validation study of the sleep disorder inventory in the nursing home. Front Psychiatry 2020;11:173 [FREE Full text] [doi: 10.3389/fpsyt.2020.00173] [Medline: 32231600]
41. Blytt KM, Bjorvatn B, Husebo B, Flo E. Clinically significant discrepancies between sleep problems assessed by standard clinical tools and actigraphy. BMC Geriatr 2017 Oct 27;17(1):253 [FREE Full text] [doi: 10.1186/s12877-017-0653-7] [Medline: 29078755]
42. Rolland Y, Pillard F, Klapouszczak A, Reynish E, Thomas D, Andrieu S, et al. Exercise program for nursing home residents with Alzheimer's disease: A 1-year randomized, controlled trial. J Am Geriatr Soc 2007 Feb;55(2):158-165. [doi: 10.1111/j.1532-5415.2007.01035.x] [Medline: 17302650]
43. Christofoletti G, Oliani MM, Bucken-Gobbi LT, Gobbi S, Beinotti F, Stella F. Physical activity attenuates neuropsychiatric disturbances and caregiver burden in patients with dementia. Clinics (Sao Paulo) 2011;66(4):613-618 [FREE Full text] [doi: 10.1590/s1807-59322011000400015] [Medline: 21655755]
44. Woodhead EL, Zarit SH, Braungart ER, Rovine MR, Femia EE. Behavioral and psychological symptoms of dementia: The effects of physical activity at adult day service centers. Am J Alzheimers Dis Other Demen 2005;20(3):171-179 [FREE Full text] [doi: 10.1177/153331750502000314] [Medline: 16003933]
45. Ishimaru D, Tanaka H, Nagata Y, Takabatake S, Nishikawa T. Physical activity in severe dementia is associated with agitation rather than cognitive function. Am J Alzheimers Dis Other Demen 2020;35:1-7 [FREE Full text] [doi: 10.1177/1533317519871397] [Medline: 31533445]
46. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. Lancet 2007 Oct 20;370(9596):1453-1457. [doi: 10.1016/S0140-6736(07)61602-X] [Medline: 18064739]
47. Choi SH, Na DL, Kwon HM, Yoon SJ, Jeong JH, Ha CK. The Korean version of the neuropsychiatric inventory: A scoring tool for neuropsychiatric disturbance in dementia patients. J Korean Med Sci 2000 Dec;15(6):609-615 [FREE Full text] [doi: 10.3346/jkms.2000.15.6.609] [Medline: 11194184]
48. Cummings JL. The Neuropsychiatric Inventory: Assessing psychopathology in dementia patients. Neurology 1997 May;48(5 Suppl 6):S10-S16. [Medline: 9153155]
49. Morganti F, Soli A, Savoldelli P, Belotti G. The Neuropsychiatric Inventory-Diary rating scale (NPI-Diary): A method for improving stability in assessing neuropsychiatric symptoms in dementia. Dement Geriatr Cogn Dis Extra 2018;8(3):306-320 [FREE Full text] [doi: 10.1159/000490380] [Medline: 30323831]
50. Liew TM. Symptom clusters of neuropsychiatric symptoms in mild cognitive impairment and their comparative risks of dementia: A cohort study of 8530 older persons. J Am Med Dir Assoc 2019 Aug;20(8):1054.e1-1054.e9 [FREE Full text] [doi: 10.1016/j.jamda.2019.02.012] [Medline: 30926409]
51. Liew TM. Neuropsychiatric symptoms in cognitively normal older persons, and the association with Alzheimer's and non-Alzheimer's dementia. Alzheimers Res Ther 2020 Mar 31;12(1):35 [FREE Full text] [doi: 10.1186/s13195-020-00604-7] [Medline: 32234066]
52. Canevelli M, Adali N, Voisin T, Soto ME, Bruno G, Cesari M, et al. Behavioral and psychological subsyndromes in Alzheimer's disease using the Neuropsychiatric Inventory. Int J Geriatr Psychiatry 2013 Aug;28(8):795-803. [doi: 10.1002/gps.3904] [Medline: 23147419]
53. van der Linde RM, Dening T, Matthews FE, Brayne C. Grouping of behavioural and psychological symptoms of dementia. Int J Geriatr Psychiatry 2014 Jun;29(6):562-568 [FREE Full text] [doi: 10.1002/gps.4037] [Medline: 24677112]
54. van der Linde RM, Dening T, Stephan BCM, Prina AM, Evans E, Brayne C. Longitudinal course of behavioural and psychological symptoms of dementia: Systematic review. Br J Psychiatry 2016 Nov;209(5):366-377 [FREE Full text] [doi: 10.1192/bjp.bp.114.148403] [Medline: 27491532]
55. Aalten P, de Vugt ME, Lousberg R, Korten E, Jaspers N, Senden B, et al. Behavioral problems in dementia: A factor analysis of the Neuropsychiatric Inventory. Dement Geriatr Cogn Disord 2003;15(2):99-105. [doi: 10.1159/000067972] [Medline: 12566599]
56. Camargos EF, Louzada FM, Nóbrega OT. Wrist actigraphy for measuring sleep in intervention studies with Alzheimer's disease patients: Application, usefulness, and challenges. Sleep Med Rev 2013 Dec;17(6):475-488. [doi: 10.1016/j.smrv.2013.01.006] [Medline: 23669093]
57. Figueiro MG, Hunter CM, Higgins P, Hornick T, Jones GE, Plitnick B, et al. Tailored lighting intervention for persons with dementia and caregivers living at home. Sleep Health 2015 Dec 01;1(4):322-330 [FREE Full text] [doi: 10.1016/j.sleh.2015.09.003] [Medline: 27066526]

58. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. Sleep 1992 Oct;15(5):461-469. [Medline: 1455130]
59. Kang Y, Na D, Hahn S. A validity study on the Korean Mini-Mental State Examination (K-MMSE) in dementia patients [Article in Korean]. J Korean Neurol Assoc 1997;15(2):300-308 [FREE Full text]
60. Song JA, Park JW, Kim H. Impact of behavioral and psychological symptoms of dementia on caregiver burden in nursing homes [Article in Korean]. J Korean Gerontol Nurs 2013;15(1):62-74 [FREE Full text]
61. Won CW, Rho YG, Kim SY, Cho BR, Lee YS. The validity and reliability of Korean Activities of Daily Living (K-ADL) scale. J Korean Geriatr Soc 2002;6(2):98-106 [FREE Full text]
62. Kim J, Kim B, Ha M. Validation of a Korean version of the Big Five Inventory. J Hum Underst Couns 2011;32(1):47-65 [FREE Full text]
63. Sciandra M, Muggeo VMR, Lovison G. Subject-specific odds ratios in binomial GLMMs with continuous response. Stat Methods Appt 2007 Jul 6;17(3):309-320. [doi: 10.1007/s10260-007-0060-x]
64. Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. J Am Stat Assoc 2012 Dec 20;88(421):9-25. [doi: 10.1080/01621459.1993.10594284]
65. Snijders TAB, Bosker RJ. Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. 2nd edition. London, UK: SAGE Publications Ltd; 2011.
66. Bubu OM, Brannick M, Mortimer J, Umasabor-Bubu O, Sebastião YV, Wen Y, et al. Sleep, cognitive impairment, and Alzheimer's disease: A systematic review and meta-analysis. Sleep 2017 Jan 01;40(1):1-18. [doi: 10.1093/sleep/zsw032] [Medline: 28364458]
67. Ooms S, Ju Y. Treatment of sleep disorders in dementia. Curr Treat Options Neurol 2016 Sep;18(9):40 [FREE Full text] [doi: 10.1007/s11940-016-0424-3] [Medline: 27476067]
68. Sohn SI, Kim DH, Lee MY, Cho YW. The reliability and validity of the Korean version of the Pittsburgh Sleep Quality Index. Sleep Breath 2012 Sep;16(3):803-812. [doi: 10.1007/s11325-011-0579-9] [Medline: 21901299]
69. Shin H, Han HJ, Shin D, Park H, Lee Y, Park KH. Sleep problems associated with behavioral and psychological symptoms as well as cognitive functions in Alzheimer's disease. J Clin Neurol 2014 Jul;10(3):203-209 [FREE Full text] [doi: 10.3988/jcn.2014.10.3.203] [Medline: 25045372]
70. Whall AL, Colling KB, Kolanowski A, Kim H, Son Hong G, DeCicco B, et al. Factors associated with aggressive behavior among nursing home residents with dementia. Gerontologist 2008 Dec;48(6):721-731 [FREE Full text] [doi: 10.1093/geront/48.6.721] [Medline: 19139246]
71. Kolanowski AM. An overview of the need-driven dementia-compromised behavior model. J Gerontol Nurs 1999 Sep;25(9):7-9. [doi: 10.3928/0098-9134-19990901-05] [Medline: 10776138]
72. Tible OP, Riese F, Savaskan E, von Gunten A. Best practice in the management of behavioural and psychological symptoms of dementia. Ther Adv Neurol Disord 2017 Aug;10(8):297-309 [FREE Full text] [doi: 10.1177/1756285617712979] [Medline: 28781611]
73. Kales HC, Gitlin LN, Lyketsos CG, Detroit Expert Panel on Assessment and Management of Neuropsychiatric Symptoms of Dementia. Management of neuropsychiatric symptoms of dementia in clinical settings: Recommendations from a multidisciplinary expert panel. J Am Geriatr Soc 2014 Apr;62(4):762-769 [FREE Full text] [doi: 10.1111/jgs.12730] [Medline: 24635665]
74. Lyketsos CG. Neuropsychiatric symptoms in dementia: Overview and measurement challenges. J Prev Alzheimers Dis 2015 Sep;2(3):155-156 [FREE Full text] [doi: 10.14283/jpad.2015.60] [Medline: 26779454]
75. Qassem T, Tadros G, Moore P, Xhafa F. Emerging technologies for monitoring behavioural and psychological symptoms of dementia. In: Proceedings of the 9th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing. 2014 Presented at: 9th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing; November 8-10, 2014; Guangdong, China p. 8-10. [doi: 10.1109/3pgcic.2014.82]
76. Piau A, Rumeau P, Nourhashemi F, Martin MS. Information and communication technologies, a promising way to support pharmacotherapy for the behavioral and psychological symptoms of dementia. Front Pharmacol 2019;10:1122 [FREE Full text] [doi: 10.3389/fphar.2019.01122] [Medline: 31632271]
77. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [Medline: 18509902]

Abbreviations

ADL: activity of daily living
AIC: Akaike information criterion
BFI-K: Korean version of the Big Five Inventory
BIC: Bayes information criterion
BPSD: behavioral and psychological symptoms of dementia
DATE: Describe and Measure, Analyze, Treat, and Evaluate
DICE: Describe, Investigate, Create, and Evaluate

EMA: ecological momentary assessment
GLMM: generalized linear mixed model
ICT: information and communications technology
K-ADL: Korean version of the Activities of Daily Living scale
K-MMSE: Korean version of the Mini–Mental State Examination
MMSE: Mini–Mental State Examination
NDB: need-driven dementia-compromised behavior
NRF: National Research Foundation of Korea
OR: odds ratio
PLST: progressively lowered stress threshold
PSQI: Pittsburgh Sleep Quality Index
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
TST: total sleep time
WASO: wake time after sleep onset

Edited by G Eysenbach, R Kukafka; submitted 22.03.21; peer-reviewed by J Antoun, W Eikelboom; comments to author 14.04.21; revised version received 07.06.21; accepted 27.09.21; published 29.10.21.

Please cite as:
Cho E, Kim S, Hwang S, Kwon E, Heo SJ, Lee JH, Ye BS, Kang B
Factors Associated With Behavioral and Psychological Symptoms of Dementia: Prospective Observational Study Using Actigraphy
J Med Internet Res 2021;23(10):e29001
URL: https://www.jmir.org/2021/10/e29001
doi:10.2196/29001
PMID:

Original Paper

Validation of Fitbit Charge 2 Sleep and Heart Rate Estimates Against Polysomnographic Measures in Shift Workers: Naturalistic Study

Benjamin Stucky[1,2*], PhD; Ian Clark[1*], PhD; Yasmine Azza[3,4,5], PhD; Walter Karlen[2,6], PhD; Peter Achermann[2,7], PhD; Birgit Kleim[2,3,4], PhD; Hans-Peter Landolt[1,2], PhD

[1]Institute of Pharmacology and Toxicology, University of Zurich, Zurich, Switzerland

[2]Sleep & Health Zurich, University Center of Competence, University of Zurich, Switzerland

[3]Department of Experimental Psychopathology and Psychotherapy, University of Zurich, Zurich, Switzerland

[4]Department of Psychiatry, Psychotherapy and Psychosomatics, University Hospital for Psychiatry, University of Zurich, Zurich, Switzerland

[5]Department of Psychiatry and Psychotherapy, Translational Psychiatry Unit, University of Lubeck, Lubeck, Germany

[6]Mobile Health Systems Lab, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland

[7]The Key Institute for Brain-Mind Research, Department of Psychiatry, Psychotherapy and Psychosomatics, University Hospital for Psychiatry, University of Zurich, Zurich, Switzerland

[*]these authors contributed equally

Corresponding Author:
Hans-Peter Landolt, PhD
Institute of Pharmacology and Toxicology
University of Zurich
Winterthurerstrasse 190
Zurich, 8057
Switzerland
Phone: 41 44 635 59 53
Email: landolt@pharma.uzh.ch

Abstract

Background: Multisensor fitness trackers offer the ability to longitudinally estimate sleep quality in a home environment with the potential to outperform traditional actigraphy. To benefit from these new tools for objectively assessing sleep for clinical and research purposes, multisensor wearable devices require careful validation against the gold standard of sleep polysomnography (PSG). Naturalistic studies favor validation.

Objective: This study aims to validate the Fitbit Charge 2 against portable home PSG in a shift-work population composed of 59 first responder police officers and paramedics undergoing shift work.

Methods: A reliable comparison between the two measurements was ensured through the data-driven alignment of a PSG and Fitbit time series that was recorded at night. Epoch-by-epoch analyses and Bland-Altman plots were used to assess sensitivity, specificity, accuracy, the Matthews correlation coefficient, bias, and limits of agreement.

Results: Sleep onset and offset, total sleep time, and the durations of rapid eye movement (REM) sleep and non–rapid-eye movement sleep stages N1+N2 and N3 displayed unbiased estimates with nonnegligible limits of agreement. In contrast, the proprietary Fitbit algorithm overestimated REM sleep latency by 29.4 minutes and wakefulness after sleep onset (WASO) by 37.1 minutes. Epoch-by-epoch analyses indicated better specificity than sensitivity, with higher accuracies for WASO (0.82) and REM sleep (0.86) than those for N1+N2 (0.55) and N3 (0.78) sleep. Fitbit heart rate (HR) displayed a small underestimation of 0.9 beats per minute (bpm) and a limited capability to capture sudden HR changes because of the lower time resolution compared to that of PSG. The underestimation was smaller in N2, N3, and REM sleep (0.6-0.7 bpm) than in N1 sleep (1.2 bpm) and wakefulness (1.9 bpm), indicating a state-specific bias. Finally, Fitbit suggested a distribution of all sleep episode durations that was different from that derived from PSG and showed nonbiological discontinuities, indicating the potential limitations of the staging algorithm.

Conclusions: We conclude that by following careful data processing processes, the Fitbit Charge 2 can provide reasonably accurate mean values of sleep and HR estimates in shift workers under naturalistic conditions. Nevertheless, the generally wide

limits of agreement hamper the precision of quantifying individual sleep episodes. The value of this consumer-grade multisensor wearable in terms of tackling clinical and research questions could be enhanced with open-source algorithms, raw data access, and the ability to blind participants to their own sleep data.

(J Med Internet Res 2021;23(10):e26476) doi:10.2196/26476

KEYWORDS

wearables; actigraphy; polysomnography; validation; multisensory; mobile phone

Introduction

Highly sensitive and precise instruments are necessary for the accurate measurement of sleep in healthy and clinical populations. Polysomnography (PSG), the prevailing gold standard in clinical and research settings [1], reliably reflects the physiological processes underlying sleep with high temporal resolution [2]. The PSG recordings are conducted to capture sleep macrostructure (eg, stages and cycles) and microstructure (eg, K-complexes, spindles, and arousals), and to quantify different variables such as power density spectra of the electroencephalogram and heart rate (HR) variability, to estimate an individual's sleep quality and health. Despite the many strengths of PSG, attendant disadvantages include high cost, the need for personnel trained in technical aspects and interpretation of data, and the highly technical recording system itself, which usually necessitates a dedicated sleep laboratory, although ambulatory systems also exist [3]. Inexpensive, practical, and portable alternatives that are equally accurate and reliable as PSG in measuring sleep would be welcome for clinicians and researchers.

Currently, the only validated and United States Food and Drug Administration–approved alternative to PSG in ambulatory settings is actigraphy [4]. Actigraphy measures movement using a multiaxis accelerometer in a device resembling a wristwatch, sometimes accompanied by an embedded light sensor. Actigraphy captures rest-activity behaviors such as sleep habits, bedrest, rise times, and light exposure [5,6]. The basic assumption of actigraphy is that motion implies wakefulness, whereas no motion implies sleep. Fully disclosed algorithms [7,8] are used to compute sleep variables with some precision, but performance compared with PSG varies because of the inherent limitation in discriminating sleep from waking that is not accompanied by movement [3,4]. Actigraphy is a dedicated scientific instrument in clinical and research contexts and depends on specialists for setting up and interpreting data [9].

Recently, there has been greater acceptance, but also controversy, among the scientific community about using commercially available wearable devices such as fitness trackers in research [10]. Fitness trackers are multisensor, consumer-grade devices that represent a cost-efficient, practical, and convenient means of objectively collecting rest-activity data longitudinally under ambulatory conditions [4]. Fitbit is a market leader [11], and efforts have been made to validate its devices, such as the Fitbit Charge 2, against PSG [4,12-15] and the portable single-channel electroencephalogram sleep scope device [16]. Such devices not only rely on movement but also measure HR via photoplethysmography. Changes in the activity of the autonomic nervous system regulating HR are coupled to changes in electroencephalogram patterns [17,18], and various HR measures are correlated with electroencephalogram-defined sleep states [19]. These relationships potentially permit a multisensor fitness tracker to estimate an array of sleep variables above and beyond that of conventional actigraphy [20-22].

A recent laboratory-based validation study suggested that the proprietary algorithm of Fitbit Charge 2 (Fitbit Inc) to estimate different sleep variables performed reasonably well [3]. More specifically, the device displayed a 9-minute overestimation of total sleep time (TST), whereas sleep onset (S_{on}) latency was underestimated by 4 minutes. Furthermore, Fitbit's *light* stage was overestimated by 34 minutes, and Fitbit's *deep* sleep stage, assumed to be equivalent to the N3 sleep stage, was underestimated by 24 minutes compared with the PSG-derived sleep stages N1+N2 and N3, respectively. No bias was observed in wakefulness after S_{on} (WASO) or the duration of rapid eye movement (REM) sleep stage. Findings in patients with periodic limb movements during sleep revealed comparable results [3]. In contrast, a study in patients with obstructive sleep apnea contradicted the unbiasedness of WASO for 2 Fitbit devices, Fitbit Charge 2 and Fitbit Alta HR. Both devices underestimated WASO, possibly indicating variable performance in different clinical populations [23]. Other work performed at participants' homes compared Fitbit Charge 2 with a portable single-channel electroencephalogram sleep monitor [16]. This study showed 86.9% agreement; however, there was an underestimation of TST by 12.3 minutes, of *light* sleep by 42.4 minutes, and of REM sleep by 11.6 minutes. Conversely, WASO was overestimated by 24.5 minutes and deep sleep by 39.8 minutes. These estimates also showed a large SD.

Regarding HR, a study found a moderate underestimation of 5.9 beats per minute (bpm) with Fitbit Charge 2 compared with the electrocardiogram, whereas precision for individual measurements was poor as reflected by wide limits of agreement (LoA) [24]. Another study found that this device tended to slightly overestimate HR in ranges <50 bpm (bias=0.51 bpm) and underestimate HR in ranges >80 bpm (bias=0.63 bpm) compared with the electrocardiogram [13]. The Fitbit Charge HR model displayed a general underestimation (bias=0.88 bpm) in a similar range [25].

Therefore, the findings of previous sleep and HR validation studies of Fitbit Charge 2 are rather inconsistent and warrant further research. It was previously concluded that apart from the sample population studied, inaccurate temporal synchronization between Fitbit wearables and PSG is an important challenge in some validation studies [26]. In addition, consumer-grade wearables need to be validated under naturalistic conditions and in diverse populations, as such factors

may affect their performance. We attempted to validate Fitbit Charge 2 against gold-standard PSG in a healthy study sample, but one that regularly performed shift work and exhibited an elevated risk of occupational stressors, which likely interfered with and attenuated the quality of sleep. With these objectives in mind, we seek to validate the usefulness of Fitbit Charge 2 to evaluate sleep quality in first responder shift workers under naturalistic conditions, with a special focus on rigorous data preprocessing and time alignment of the data recordings.

Methods

Study Sample

The participants of this study were recruited from July 2017 to November 2019 by various informational media, emails, and presentations at shift change as part of a larger study investigating sleep and resilience to psychological stress and trauma. They completed 1 month of monitoring of wrist-derived rest-activity behavior with a Fitbit Charge 2 that was worn continuously by all individuals on their nondominant wrist.

The Ethics Commission of the Canton of Zurich approved (2016-01357) all study protocols and experimental procedures, and written informed consent was obtained before participation. Participants invited to participate fulfilled all inclusion criteria: aged between 18 and 65 years, BMI ≤26 (or if exceeding a BMI of 26, which is typical of very athletic participants, an absence of sleep problems, such as sleep breathing disorders, was reported), current employment in 1 of 2 participating emergency rescue stations and a police station in the greater Zurich area of Switzerland, possession of a smartphone, and German language fluency. Exclusion criteria included the presence of a neurological disorder diagnosis or head injury with the potential to affect electroencephalogram variables, reported intake of >5 alcoholic beverages per week, or if a urine drug screen (Drug Screen Multi 12-AE; Nal von Minden GmbH) revealed drug abuse. All participants were shift workers, although specific shift schedules varied among individuals by occupation, such that emergency medical rescue workers and emergency doctors worked cycles of two 12-hour days followed by two 12-hour nights, terminating in 4 free days. Police officers worked four contiguous shifts with varying individual activities and bedrest times. Data on individual shifts were not collected or analyzed. Individuals received monetary compensation for participating in the study. Participants additionally received a report on their sleep derived from their own sleep data derived from Fitbit Charge 2 and PSG. This report was explained to them by a study staff member.

Validated German translations of questionnaires administered at meetings at the start and upon completion of 1 month of monitoring were used to assess lifestyle and psychological and sleep variables. The Pittsburgh Sleep Quality Index (PSQI) [27], Posttraumatic Stress Disorder Checklist for Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition [28], and the Perceived Stress Scale 10 (PSS-10) [29] were used to assess subjective sleep quality, posttraumatic stress symptoms, and stress in the past month. Cutoff scores of ≥5 on the PSQI, >31 on the Posttraumatic Stress Disorder Checklist for Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, and substantial deviations from the normative values—12.1 (SD 5.9) for men and 13.7 (SD 6.6) for women—indicate poor sleep quality [27], a probable posttraumatic stress disorder diagnosis [28], and elevated perceived stress [29], respectively. The Horne-Östberg Morningness-Eveningness Questionnaire-A Reduced Scale (rMEQ) was used to assess the participants' preferred rest-activity behavior or *chronotype*, with higher scores indicating increased morning activity preference. Scores on the rMEQ have a range of 4-25. A previous study found that most individuals (60%; scores: 12-17) show neither a pronounced evening (20% of individuals; scores: 4-11) nor morning (20% of individuals; scores: 18-25) activity preference [30].

Polysomnographic Recordings

A total of 62 individuals (43 emergency medical rescue workers, 16 police officers, and 3 emergency doctors), of whom 56% (35/62) were women, completed 2 nights of ambulatory PSG recordings in their homes. The PSG recordings were always made of nocturnal sleep following a day work shift and consisted of an adaptation night and then a baseline night the following evening. Individuals were free to determine their bedtime and sleep duration. The adaptation night served as a combined adaptation and screening night, whereas the baseline night provided the data analyzed in this report, with the exception of 8 individuals, whose data originated from the adaptation night because the PSG data of the baseline nights were of poor quality. The PSG data from one individual were excluded from the analyses because the data were of poor quality on both nights. Therefore, the total PSG sample consisted of 61 individuals. On 2 nights, the Fitbit Charge 2 data sets for 2 individuals were not obtained, reducing the sample to 59 individuals who had both PSG and Fitbit Charge 2 data for comparison. All PSG data were acquired using dedicated ambulatory polysomnographic amplifiers (SOMNOscreen Plus, SOMNOmedics GmbH). All electrodes and sensors for PSG recordings were applied by trained members of the research team. The overall PSG montage consisted of scalp electrode sites Fz, Cz, Pz, Oz, C3, C4, A1, and A2 applied according to the International 10-20 System [31] and electrooculogram, submental electromyogram, and electrocardiogram and grounding electrode according to the American Academy of Sleep Medicine standards [32]. The Cz electrode served as the reference during recording, and the opposite mastoid was used for the rereferenced display. The sampling rate for all the sites was 256 Hz. For recording, high-pass (0.2 Hz) and low-pass filters (128.0 Hz) were used. High-pass (0.3 Hz) and low-pass (35.0 Hz) filters in addition to a powerline filter were applied for visual sleep scoring. Sleep stages were scored visually by an experienced individual in 20-second epochs according to the American Academy of Sleep Medicine (2007) criteria.

The electrocardiogram trace in the PSG recordings was examined visually for one epoch at a time for all wake epochs before S_{on} and all epochs of sleep and wake stages after S_{on} (performed by an experienced individual). Artifacts and ectopic beats present in the electrocardiogram trace that had the potential to interfere with the quantification of interbeat intervals (IBIs), defined as the time interval between the normal R peaks of the QRS complex, were manually marked and removed before data processing and analysis.

Fitbit Charge 2 Recordings

All participants wore the Fitbit Charge 2 continuously during the PSG recorded nights. The device records wrist activity using accelerometry and pulses via photoplethysmography. It produces two types of sleep data depending on whether certain criteria are fulfilled during data collection. These criteria are sufficient battery charge, a sleep episode >3 hours in duration, and sufficient skin contact with the photoplethysmography sensor. If these criteria are not fulfilled, then *classic* sleep data are generated, comprising *asleep*, *awake*, and *restless* variables at a 1-minute data granularity. If these criteria are fulfilled, then *stages* data are produced, comprising *wake*, *light*, *REM*, and *deep* sleep at a 30-second data granularity. If *stages* data are obtained for a given sleep episode, then users receive two data sets, that is, (1) sleep data, which is composed of stages, and (2) wake data, which is composed exclusively of wake episodes <30 seconds. Both data sets are present in a single JSON file for a given data collection date. However, there were also wake episodes contained within the sleep data set. This data structure is especially relevant for researchers who wish to extract entire hypnogram data and information not provided by Fitbit, such as REM sleep latency (REML). The variable WASO was created in this study by merging these two data sets contained within the *stages* data type output. The Fitbit sleep staging algorithm occasionally scores the first stage after S_{on} and the last stage before sleep offset (S_{off}) as *wake*. This runs counter to the intuitive definition of S_{on} and S_{off} as the first occurrence of sleep and the last occurrence of sleep, respectively.

We manually omitted such bordering wake epochs and adjusted the S_{on}, S_{off}, TST (ie, $S_{off} - S_{on}$), and WASO values accordingly. S_{on}, S_{off}, and REML are variables that are not provided directly by Fitbit; hence, we calculated them from the sleep staging information provided by Fitbit. All other variables were standard Fitbit variables. Adjustments only affected the Bland-Altman analyses. The results of the analyses without adjustment for the standard Fitbit variables can be found in Figure S1 and Table S1 in Multimedia Appendix 1. A sleep sensitivity setting is needed to be set for Fitbit's sleep recordings, with options *sensitive* and *normal*. When set to *normal*, only major body movements, such as rolling over, will register as *wake*, whereas when set to *sensitive*, more subtle movements will additionally be registered as *wake*. We set the setting to *sensitive* throughout the data collection.

Statistical Analyses

All analyses and data processing steps were performed in the programming language *R* (version 4.0.0; R Foundation for Statistical Computing) [33]. Fitbit intraday HR measures were used. For electrocardiogram R peak detection, the Pan-Tompkins algorithm [34] was used as implemented in the *rsleep* package (version 1.0.3) [35]. However, the algorithm could not distinguish sharp T waves from R peaks on various occasions. Thus, a modification of the algorithm had to be made. The signal can sometimes be inverted in the sign, and for this reason, we changed the signal to have positive R peaks (which was revealed by the mean of the detected peak values by the Pan-Tompkins algorithm). Sometimes, the peak can be slightly misaligned with the actual R peak maximum. Therefore, after running the Pan-Tompkins algorithm, the detected peak was aligned with the actual maximum ±200 ms around the detected peak. Furthermore, in cases where two peaks were observed within less than 360 ms, we checked if the subsequent peak was a mistakenly detected T wave or an actual R peak. This was done by examining the signal in a small window of ±28 ms around the detected and maximally aligned peak and taking its second derivative. T waves generally display slower changes in the tangents of the electrocardiogram signal as compared to faster tangent changes found in R peaks. The 60% quantile of the absolute value of the second derivative ($QAVSD_{60}$) was then compared with a cutoff point specific to an individual participant derived from the density function of the $QAVSD_{60}$ values from all the detected peaks. The cutoff point was defined as the first local minimum of the density within the hard limits of 35/256 $\mu V/s^2$ and 120/256 $\mu V/s^2$. If no local minimum was present, 35/256 $\mu V/s^2$ was used instead. The density of $QAVSD_{60}$ revealed a multimodal distribution of nearly no overlap between the T wave characteristic $QAVSD_{60}$ values compared with those originating from R peaks. Erroneously detected T waves were omitted, thereby rescuing the affected segments of the electrocardiogram data sets for subsequent analyses. This small T wave check and alignment of the peak to the local maximum significantly improved the algorithm performance on visual inspection. From PSG IBIs, a transformation into bpm was made with 60 seconds divided by the IBI duration in seconds.

The internal clock times of the Fitbit and PSG systems were misaligned. This is a common problem in studies involving multiple measurement instruments, as they often do not share the same clock and thus require temporal alignment [36,37]. Hence, we estimated a time shift for each individual to ensure good time alignment. For this, linear interpolation was used to estimate values between two data points in either the PSG beat-per-beat data or the lower-resolution Fitbit data. We resampled both the Fitbit and PSG interpolated time series of a given night at 0.2-second intervals. The cross-correlation function was used to extract the lag with the maximal correlation between the time series.

Bland-Altman plots were constructed with the *blandr* package (version 0.5.1) [38] for all the sleep variables, two tailed *t* tests, and LoA defined as b (SD 1.96), where b denotes the bias and SD is the standard deviation of the bias. A variable is termed as *unbiased* if bias b is not significantly different from 0 from the corresponding *t* test. The differences in the Bland-Altman analyses were set to denote PSG *minus* Fitbit. Thus, a positive difference corresponds to an underestimation of Fitbit compared with PSG, and a negative difference corresponds to an overestimation. Concerning the repeated measurements of the 10%-trimmed HR average (HR_{10}) and 10%-trimmed HR variance average ($HRvar_{10}$) as measured at 1-minute intervals, a linear mixed effects regression with the *nlme* package (version 3.1-147) was estimated [39]. The dependent variable was set to be the PSG-Fitbit value, and just a single intercept without a slope was considered the independent variable. For the random effect, a random intercept per subject was included. Owing to the consecutive 1-minute HR_{10} and $HRvar_{10}$ measurements with potential time correlations, an autocorrelation structure of order

one was added. The *t* tests and LoAs were estimated using a mixed model.

Epoch-by-epoch (EBE) analyses were performed through the following statistical measures:

Sensitivity = TP/P **(1)**

Specificity = TN/N **(2)**

Accuracy = (TP + TN)/(P + N) **(3)**

Matthews correlation coefficient (MCC) = (TP * TN - FP * FN)/ **(4)**

Positive predictive value (PPV) = TP/(TP + FP) **(5)**

Negative predictive value (NPV) = TN/(TN + FN) **(6)**

In these equations, TP represents true positives (number of Fitbit epochs that share a given PSG stage), TN represents true negatives (the number of Fitbit epochs that are not in a given stage and where the according PSG epoch is also not labeled as that stage), FP represents false positives (number of Fitbit epochs that do not share a given PSG stage), and FN represents false negatives (number of Fitbit epochs that did detect a given stage, whereas PSG did not detect it). Sensitivity measures the proportion of epochs of a given PSG-derived sleep state that was correctly identified by Fitbit (eg, for REM sleep, it is the percentage of Fitbit *REM* sleep stages among all PSG REM sleep stages). Specificity, however, describes the percentage of Fitbit correctly identifying the nonoccurrence of a given sleep state. Accuracy is a combined measure of the true discoveries and true negatives of Fitbit divided by all positives and negatives in the PSG sample. MCC is more informative than the measure accuracy, because it considers all true positive, true negative, false positive, and false negative. This can be interpreted as a correlation coefficient, that is, the more positive, the better Fitbit predicts the PSG epochs, such that 0 would be random guessing, and negative values indicate disagreement. PPV, often called precision, describes the proportion of Fitbit correctly identifying a given stage among the number of times Fitbit assigned that stage, and NPV describes the equivalent for correctly identifying an epoch that is not a given stage. In our sample, the epoch length was defined as 20 seconds, but Fitbit's algorithm has an epoch length of 30 seconds. Thus, a direct EBE analysis was not possible. Therefore, we looked at all PSG-derived epochs and compared them with the dominating Fitbit stage (>50%) in the same interval. In cases where one PSG epoch contained two different Fitbit stages of equal length, we chose the first stage.

Results

Demographic Characteristics of the Study Sample

The demographics of the 59 individuals studied as well as their mean PSG- and Fitbit-derived sleep and HR measures are summarized in Tables 1 and 2. The mean values on the PSQI and PSS-10 indicated slightly impaired subjective sleep quality and a slightly elevated perceived stress level [27]; however, no diagnostic criteria for possible posttraumatic stress disorder have been met [28].

Table 1. Demographics of study sample (N=59).

	Value
Female, n (%)	33 (56)
Police, n (%)	15 (25)
Age (years), mean (SD)	33.5 (8.1)
BMI, mean (SD)	23.9 (2.9)
PSQI[a], mean (SD)	5.8 (2.7)
PCL-5[b], mean (SD)	6.2 (7.9)
PSS-10[c], mean (SD)	12.2 (4.9)
rMEQ[d], mean (SD)	14.4 (3.5)

[a]PSQI: Pittsburgh Sleep Quality Index.

[b]PCL-5: Posttraumatic Stress Disorder Checklist for Diagnostic and Statistical Manual of Mental Disorders Fifth Edition.

[c]PSS-10: Perceived Stress Scale 10.

[d]rMEQ: Horne-Östberg Morningness-Eveningness Questionnaire-A Reduced Scale.

Table 2. Sleep and heart rate variables (N=59).

	Value, mean (SD)	
	Polysomnography	Fitbit
$N1_{so}$[a] (clock time)	23.4 (0.9)	23.4 (2.4)
TST[b] (hours)	8.0 (1.7)	7.8 (2.6)
REM_d[c] (hours)	1.7 (0.8)	1.7 (0.7)
$light_d$[d] (hours)	4.2 (1.1)	4.4 (1.3)
$deep_d$[e] (hours)	1.5 (0.6)	1.3 (0.5)
WASO[f] (hours)	0.4 (0.5)	1.0 (1.1)
REML[g] (minutes)	76.3 (30.6)	103.9 (59.7)
REM[h] in the first cycle (%)	11.6 (8.1)	15 (8.7)
HR_{10}[i] REM (bpm[j])	60.9 (9.1)	59.9 (8.2)
HR_{10} N1[k] (bpm)	61.8 (9.2)	59.2 (7.5)
HR_{10} N2[l] (bpm)	56.6 (7.7)	55.7 (7.0)
HR_{10} N3[m] (bpm)	58.8 (8.8)	57.2 (7.2)
$HRvar_{10}$[n] REM (bpm)	28.1 (90.8)	6.4 (16.1)
$HRvar_{10}$ N1 (bpm)	48.7 (110.1)	6.8 (16.7)
$HRvar_{10}$ N2 (bpm)	22.0 (76.7)	4.7 (24.3)
$HRvar_{10}$ N3 (bpm)	25.4 (111)	2.9 (12.9)

[a]$N1_{so}$: sleep onset with non–rapid eye movement (NREM) sleep stages 1 criteria.

[b]TST: total sleep time.

[c]REM_d: rapid eye movement sleep duration.

[d]$light_d$: light sleep or NREM sleep stages 1+NREM sleep stages 2 duration, respectively.

[e]$deep_d$: deep sleep or NREM sleep stages 3 duration, respectively.

[f]WASO: wakefulness after sleep onset.

[g]REML: rapid eye movement sleep latency.

[h]REM: rapid eye movement.

[i]HR_{10}: 10%-trimmed heart rate average.

[j]bpm: beats per minute.

[k]N1: NREM sleep stages 1.

[l]N2: NREM sleep stages 2.

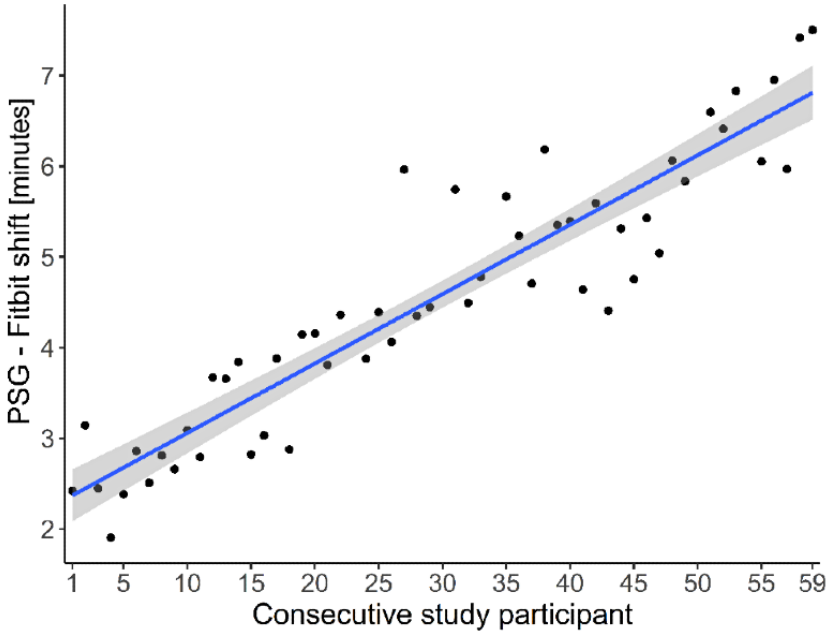
[m]N3: NREM sleep stages 3.

[n]$HRvar_{10}$: 10%-trimmed heart rate variability.

Time Alignment

Accurate temporal synchronization between the PSG system and the wearable Fitbit device often poses a methodological challenge in validation studies [4]. This was also the case in this study. When scrutinizing our data, we noticed that the time discrepancies between the PSG system's and the Fitbit app's clocks increased as the study progressed. In other words, the later the participant entered the study, the higher the time difference between PSG and Fitbit recordings. This relationship can be seen in Figure 1 as a linear association between the individual participant identifier number and the estimated time shift between the two measurement instruments.

Figure 1. The consecutive study participant numbers (higher numbers indicate chronologically later entry into the study) from the entire study sample are shown on the x-axis; the data-driven timeshift between polysomnography and Fitbit is shown on the y-axis. There was a significant linear relationship between the identifier and the shift (P<.001; adjusted R^2=0.85). Thus, the times drifted apart as the study went on, with a minimum time misalignment of 1.9 minutes and a maximum of 7.5 minutes. PSG: polysomnography.



To align the time series, we computed the cross-correlation function for each participant and corrected the time shift by the emergent maximum. Our time alignment efforts produced good correspondence in our data between the two instruments, as evident in the simultaneous occurrences of HR bursts in the two time series (Figure 2). Nevertheless, the variability and amplitude of the Fitbit curve were reduced compared with PSG because only between 4 and 12 measurements per minute were made available by Fitbit. The analysis of the entire Fitbit sample revealed that an average of 7.48 HR counts per minute was available (Figure 3). In contrast, PSG HR data were sampled at a frequency of 1/256 Hz.

Figure 2. Data on the validation night of the first participant in the study with identifying number 004 (left column) and the last participant in the study with number 104 (right column) are shown. Row A displays the cross-correlation function, which displays a large visible maximum at the orange vertical line representing the best alignment between the two devices (PSG and Fitbit). The dashed vertical reference line shows a lag of 0 minutes. Rows B-D share the same x-axis, which denotes hours after PSG-derived sleep onset with criteria. For each hour in the recording, a vertical dashed gray line was added. Row B shows the HR in bpm derived from PSG (red) and Fitbit (black) that were seen before any time alignment was applied, whereas row C presents the HR data after the data-driven shift from panel A was applied. The time-aligned time series visually shows good agreement after correcting for the time difference. Fitbit shows reduced variability in the signal but fairly good average correspondence. In panel D, the top row shows PSG-derived hypnograms for both participants, whereas in the bottom row, the Fitbit-derived hypnograms are displayed. All hypnograms have been time-corrected according to panel A. The overall sleep structure is captured reasonably well by Fitbit, but Fitbit detects more wake and REM episodes compared with PSG, and the distinction of light (N1+N2) and deep (N3) sleep often seems to be particularly challenging for Fitbit. bpm: beats per minute; HR: heart rate; PSG: polysomnography; REM: rapid eye movement; W: wake.

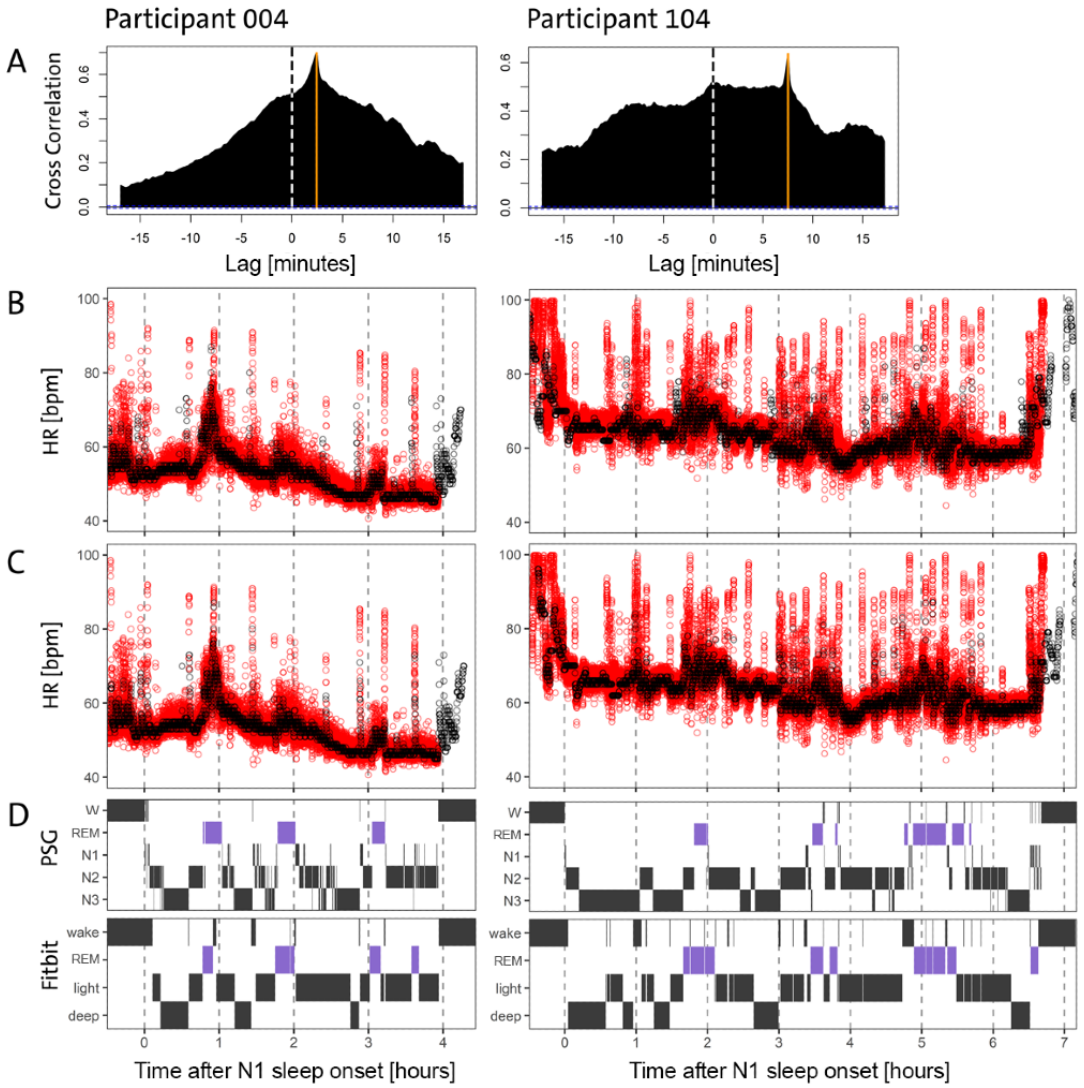
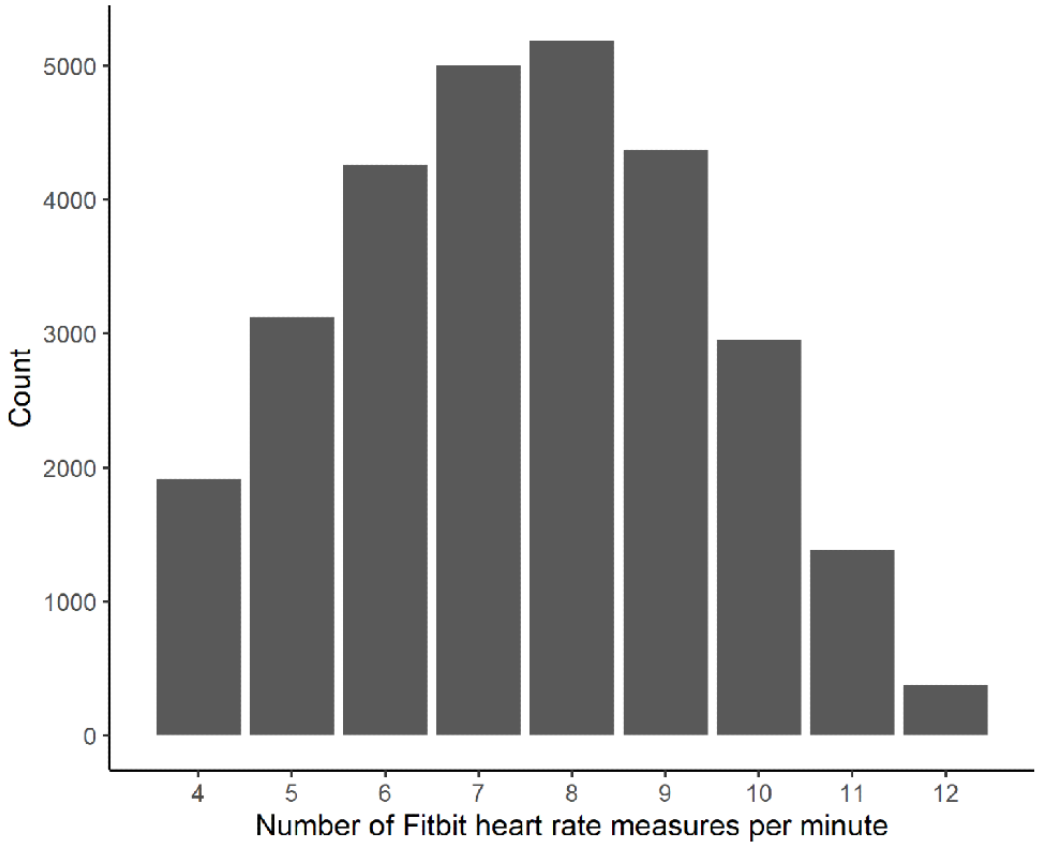


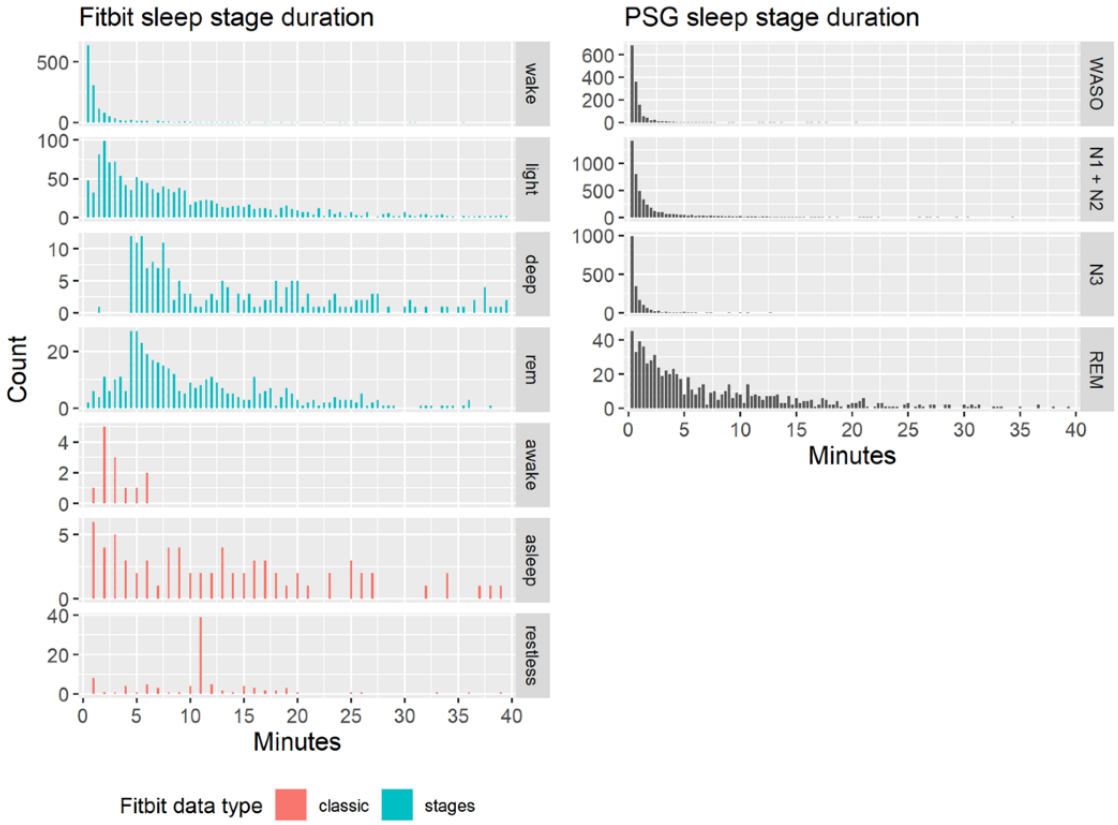
Figure 3. The available data of all nights (n=59) were extracted and counted for the number of heart rate measures contained. A total of roughly 28,320 minutes (corresponding to 59 study participants who, on average, spent 8×60 minutes asleep) were expected. In fact, 28,601 individual minutes of data were recorded; this figure displays the distribution of all heart rate measures, yielding an average of 7.48 measures per minute. Count data for >12 measures per minute and <4 measures per minute are not displayed because their occurrences were so small that they are not visible on the plot.



Distribution of Sleep Stage Durations

Next, we compared the distribution of sleep stage durations between the Fitbit and PSG data (Figure 4). Duration was defined as the duration of consecutive epochs with the same sleep stage until interrupted by any other stage, independent of its duration. We observed that Fitbit uses 30-second intervals to classify the *stages* data, whereas the *classic* data are presented with less time-resolved, 1-minute resolutions. With respect to wake episodes, the Fitbit data resembled the PSG distribution, with mostly short uninterrupted wake episodes and much rarer longer episodes. The *awake* category in the *classic* datatype had higher tails, possibly owing to having a resolution of 1 minute instead of 30 seconds, thus potentially missing certain stage changes that occur faster. On the basis of the inspection of the data distributions, we assumed that Fitbit's *light* sleep stage in the *stages* datatype might capture PSG-defined N1+N2 sleep stages, whereas *deep* sleep might capture PSG-defined N3 sleep. However, these assumptions need to be treated with caution because no information is provided by Fitbit. In general, *light* and *deep* sleep showed longer tails than the PSG-defined non–rapid eye movement (NREM) sleep stages, possibly owing to different temporal resolutions or slower changes in HR and HR variability compared with the more sudden changes in brain states. Furthermore, the *deep* sleep distribution showed a pronounced discontinuity at around 4 minutes and 30 seconds, which could also be observed in the Fitbit *REM* sleep stage duration. The distributions of *light*, *deep*, and *REM* sleep showed discrepancies to the PSG-derived durations, indicating that the algorithm does not fully reflect PSG-derived data and may miss brief stage changes and stage interruptions. Furthermore, the *restless* stage in the *classic* datatype is unknown. This stage displayed a peak at approximately 11 minutes, with an unknown origin.

Figure 4. The distribution of sleep stage durations for Fitbit (left panel) and PSG (right panel). Both were computed on the sample of the nights used for validation. Here, the plot has been cut off at 40 minutes for visual purposes; the tails continue to decrease as one would expect. The Fitbit sleep staging data types "classic" (red) and "stages" (blue) show large deviations compared with PSG sleep stages (black). Of note, deep and REM sleep show nonbiological discontinuity at around 4.5 minutes, and all Fitbit stages have larger tails. The stage "restless" has a peak at 11 minutes with unknown meaning. PSG: polysomnography. REM: rapid eye movement; WASO: wakefulness after sleep onset.



Bland-Altman Analyses of Sleep Variables

We split our validation into two analyses, one with the PSG-determined first occurrence of N1 sleep as the criterion for S_{on} (N1 S_{on} [$N1_{on}$]) and the other with the first occurrence of N2 sleep as the criterion for S_{on} (N2 S_{on} [$N2_{on}$]). This was done because it is unknown how Fitbit estimates S_{on}. In Figure 5, we plotted the variables computed with $N1_{on}$, and Table 3 provides the associated statistics. The $N2_{on}$ analyses revealed systematically higher biases. These data are presented in Figure S2 and Table S2 in Multimedia Appendix 1. S_{on}, defined as $N1_{on}$, was unbiased (–1.6 minutes; P=.73). S_{off}, TST, REM sleep duration (REM_d), the duration of Fitbit's *light* sleep duration ($light_d$) in minutes (as recorded by the Fitbit; interpreted as N1+N2), and the *deep* sleep duration ($deep_d$) in minutes (as recorded by the Fitbit) did not display significant bias. Nevertheless, $deep_d$ showed a trend toward a bias of 11.2 minutes with $N1_{on}$ (P=.08), likely pointing to a slight underestimation with Fitbit of N3 sleep. REML and WASO both exhibited a significant overestimation with Fitbit—REML was overestimated by 29.4 minutes and WASO by 37.1 minutes (P_{all}<.001). Although the marginal densities of the differences for S_{on}, S_{off}, and TST were quite narrow, indicating a good estimator in general, some occasional sleep episodes disagreed strongly between the Fitbit and PSG instruments, as reflected in the large LoA (Table 3). The marginal distributions of REM_d, $light_d$, and $deep_d$ showed higher variance, even if outliers were neglected. This observation may indicate that the estimation of stages of sleep is challenging for Fitbit's algorithm and a source of variability, although being unbiased. The data on the standard Fitbit variables without the bordering wake epoch adjustment revealed very similar results, however, with slightly larger biases for TST and WASO, as shown in Figure S1 and Table S1 in Multimedia Appendix 1.

Figure 5. Bland-Altman plots for various sleep variables are shown with sleep onset defined as the first occurrence of N1. The dashed lines denote lower limits of agreement, bias, and upper limits of agreement. The dotted lines are the respective 95% CI of limits of agreement. On the top and right of each panel, the marginal densities are plotted. The x-axis displays the PSG variables, and the y-axis denotes the differences between the two devices (PSG-Fitbit). N1-derived sleep onset is unbiased. Sleep offset, total sleep time, light sleep or N1+N2 sleep duration, deep sleep or N3 sleep duration, and REMd do not have significant bias. WASO and REML display a significant deviation of the difference between the devices from 0. $deep_d$: deep sleep duration; $light_d$: light sleep duration; PSG: polysomnography; REM_d: rapid eye movement sleep duration; REML: rapid eye movement sleep latency; S_{off}: sleep offset; S_{on}: sleep onset; TST: total sleep time; WASO: wake after sleep onset.

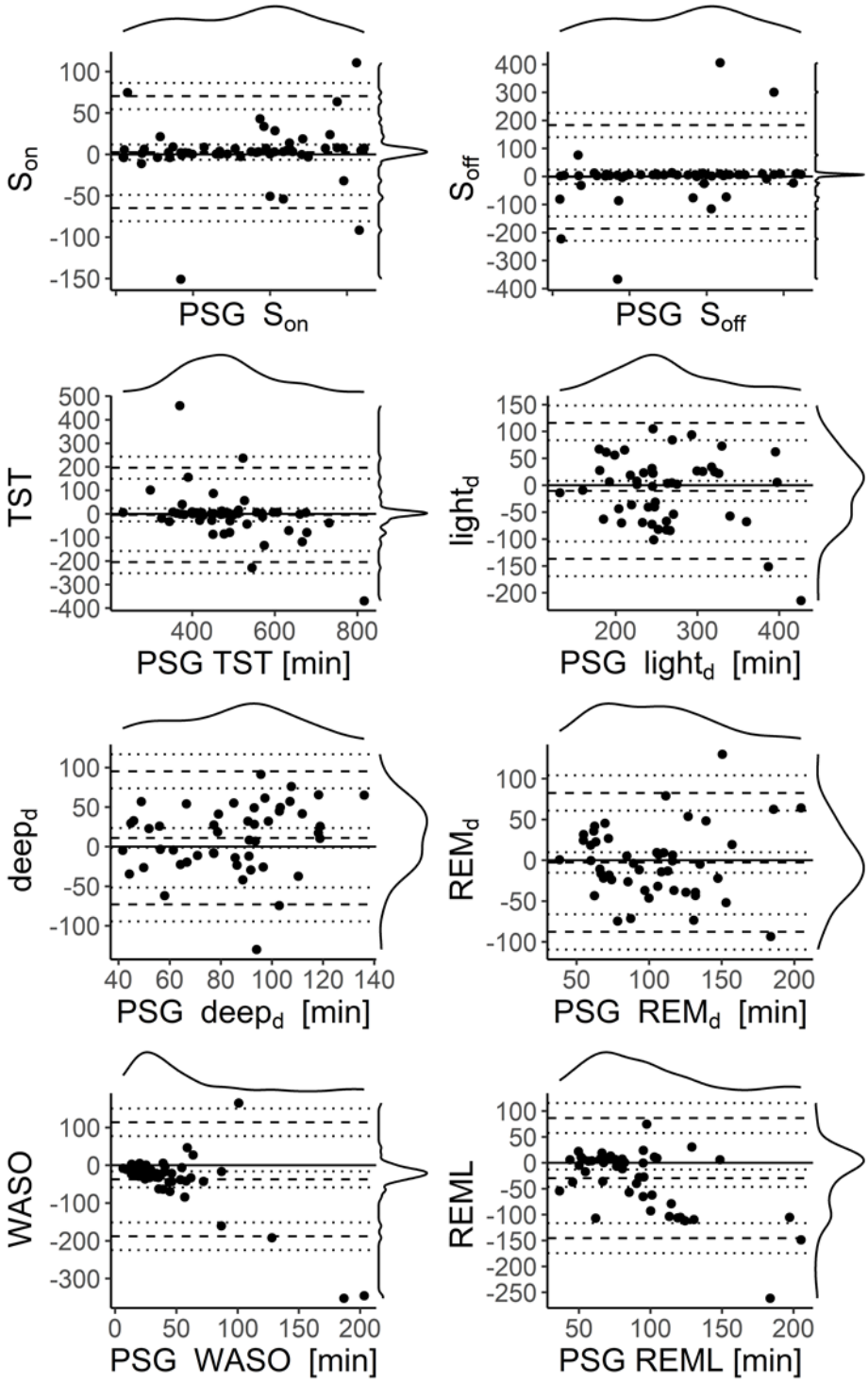


Table 3. Bland-Altman statistics[a].

Variable	PSG[b]-Fitbit	Lower LoA[c]	Upper LoA	*P* value
S_{on}[d] (minutes)	–1.6	–68.8	65.6	.73
S_{off}[e] (minutes)	–5.6	–189.3	178.2	.66
TST[f] (minutes)	–4.0	–204.3	196.3	.77
REM_d[g] (minutes)	–2.7	–87.8	82.4	.67
$light_d$[h] (minutes)	–10.4	–136.8	116.0	.27
$deep_d$[i] (minutes)	11.2	–72.9	95.2	.08
WASO[j] (minutes)	–37.1	188.1	113.8	.001
REML[k] (minutes)	–29.4	–145.4	86.6	.001
HR_{10}[l] (bpm[m])				
Overall	0.9	–6.9	8.6	<.001
WASO	1.9	–5.4	9.2	.03
N1[n]	1.2	–8.9	11.3	.14
N2[o]	0.6	–4.7	6.0	.001
N3[p]	0.6	–6.4	7.6	.008
REM[q]	0.7	–4.7	6.0	<.001

[a]Statistics accompanying the Bland-Altman plots (Figure 5). Sleep onset and rapid eye movement (REM) sleep latency were calculated using the non–rapid eye movement (NREM) sleep stages 1 sleep onset criteria. The average 10%-trimmed heart rate and 10%-trimmed heart rate variance values in various sleep states are presented in the columns below the sleep variables. The average difference between polysomnography and Fitbit measures bias can be found in the first column. The lower and upper limits of agreement describe 1.96 times the SD around the bias and can be found in the subsequent columns. In the last column, the *P* values for the paired *t* test are reported; we tested whether the bias was significantly different from 0.

[b]PSG: polysomnography.

[c]LoA: limit of agreement.

[d]S_{on}: sleep onset.

[e]S_{off}: sleep offset.

[f]TST: total sleep time.

[g]REM_d: REM sleep duration.

[h]$light_d$: light sleep duration.

[i]$deep_d$: deep sleep duration.

[j]WASO: wakefulness after sleep onset.

[k]REML: REM sleep latency.

[l]HR_{10}: 10%-trimmed heart rate average.

[m]bpm: beats per minute.

[n]N1: NREM stage 1 sleep.

[o]N2: NREM stage 2 sleep.

[p]N3: NREM stage 3 sleep.

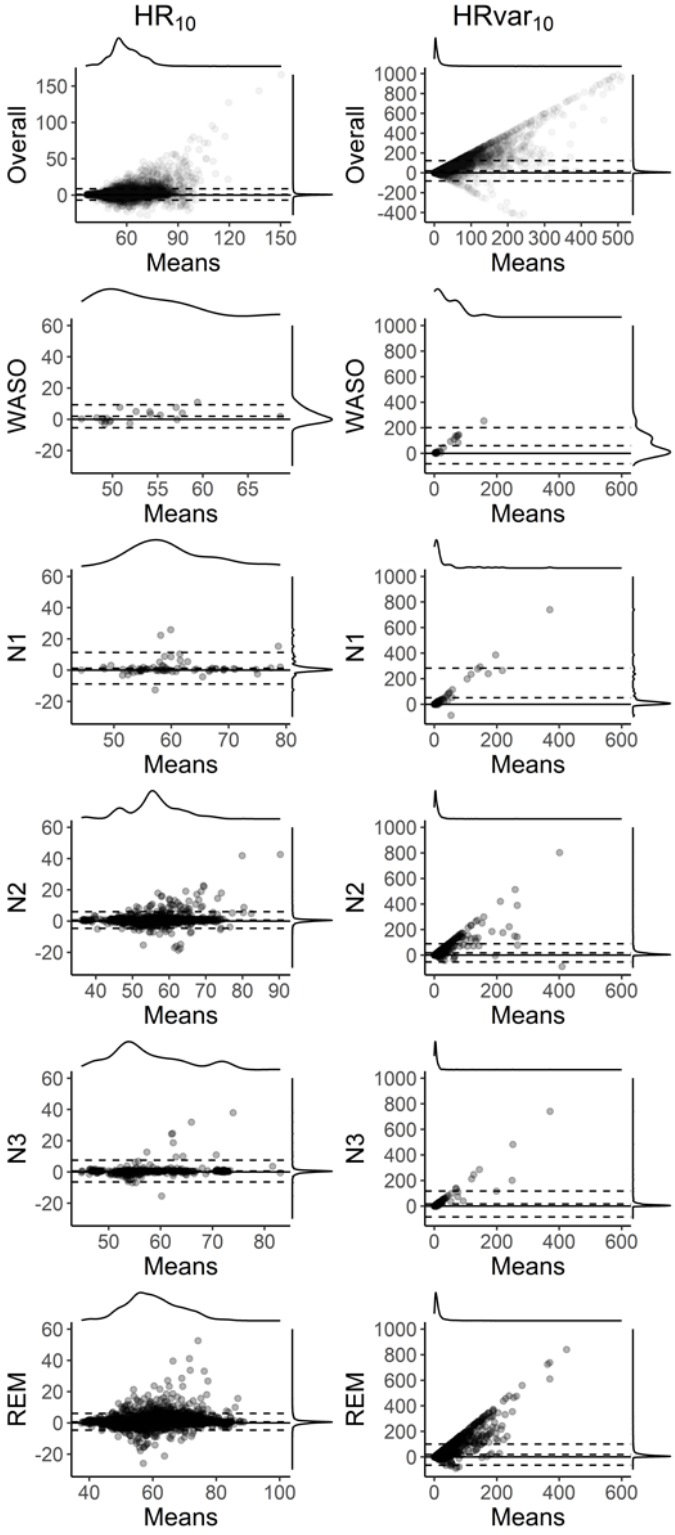
[q]REM: rapid eye movement.

Bland-Altman Analyses of HR Variables

The Bland-Altman plots for the HR variables are shown in Figure 6. When computing the interval between 30 minutes before $N1_{on}$ until S_{off} without considering the different wakefulness and sleep states, HR_{10} and $HRvar_{10}$ measures both appeared biased. More specifically, Fitbit underestimated HR_{10} overall by 0.9 bpm and displayed LoA of –6.9 and 8.6 bpm (Table 3). This underestimation was rather small, with a relatively narrow marginal distribution of the differences. When focusing on 1-minute HR_{10} values restricted to the time interval between S_{on} and S_{off} and dividing among the PSG-derived states N1, N2, N3, REM sleep, and wake, HR_{10} displayed a higher bias in the wake (1.9 bpm; *P*=.03) and N1 (1.2 bpm; *P*=.14) stages compared with the sleep stages N2 (0.6 bpm; *P*=.001), N3 (0.6 bpm; *P*=.008), and REM sleep (0.7 bpm; *P*<.001).

When analyzing overall HR variance, Fitbit strongly underestimated $HRvar_{10}$ with a bias of 20.3 bpm (P<.001), which was associated with higher LoA –82.1 and 122.7. When $HRvar_{10}$ was divided among the different sleep stages, we observed behavior similar to HR_{10}, such that $HRvar_{10}$ wake and N1 had a higher bias (60.2 and 51.1 bpm) than N2, N3, and REM sleep (17.6, 16.3, and 18.5 bpm), all with low P values and considerably large LoA.

Figure 6. Bland-Altman plots for heart rate–derived variables. The dashed lines denote lower limits of agreement, bias, and upper limits of agreement for a mixed model dealing with the repeated measures. On the top and right of each panel are the marginal densities. The x-axis displays the means of both devices (ie, [polysomnography + Fitbit]/2), and the y-axis denotes the differences between the two devices (polysomnography-Fitbit). Overall average 10%-trimmed heart rate and 10%-trimmed heart rate variance values are calculated for 1-minute intervals between 30 minutes before sleep onset with N1 criteria and 30 minutes after sleep offset. All other variables are calculated between sleep onset and sleep offset, only extracting the designated variable, in 1-minute intervals. HR_{10}: 10%-trimmed heart rate average; $HRvar_{10}$: 10%-trimmed heart rate variance average; REM: rapid eye movement; WASO: wake after sleep onset.



EBE Analysis

The EBE analysis results are displayed in Table 4. The EBE comparison between Fitbit and PSG revealed that Fitbit displayed better specificity (WASO: 0.898; light sleep as N1+N2: 0.574; deep sleep as N3: 0.92; REM sleep: 0.889) than sensitivity (WASO: 0.428; light sleep as N1+N2: 0.534; deep sleep as N3: 0.279; REM sleep: 0.548). The sensitivity for REM

sleep was worse during the initial 120 minutes of sleep (0.432) when compared with REM episodes beginning 120 minutes or more after S_{on} (0.57). In contrast, for specificity, this relationship was reversed (REM<120 minutes: 0.963; REM>120 minutes: 0.864). Accuracy was best for WASO (0.898) and REM sleep (0.880) and worse for deep sleep N3 (0.776) and light sleep N1+N2 (0.553). A similar relationship was reflected in the MCC, ranging from weak to moderate correlation (REM sleep: 0.339; WASO: 0.329; deep sleep as N3: 0.25; light sleep as N1+N2: 0.108). The MCC measure is preferable to accuracy as it only leads to higher scores if the prediction is simultaneously accurate in all confusion matrix categories (true positive, false positive, true negative, and false negative) [40]. PPV, the probability that an episode with a given Fitbit stage will also have the same PSG stage, was generally lower (WASO: 0.438; light sleep as N1+N2: 0.592; deep sleep as N3: 0.501; REM sleep 0.306) compared with NPV, the probability that an episode that does not have a certain Fitbit stage will also not have that PSG stage (WASO: 0.894; light sleep as N1+N2: 0.516; deep sleep as N3: 0.815; REM sleep: 0.956).

Table 4. Epoch-by-epoch analysis[a].

State	Sensitivity	Specificity	Accuracy	MCC[b]	PPV[c]	NPV[d]
WASO[e]	0.428	0.898	0.824	0.329	0.438	0.894
Light sleep	0.534	0.574	0.553	0.108	0.592	0.516
Deep sleep	0.279	0.920	0.776	0.250	0.501	0.815
REM[f] sleep	0.548	0.889	0.861	0.339	0.306	0.956
REM sleep <120 minute	0.432	0.963	0.934	0.383	0.403	0.967
REM sleep >120 minute	0.570	0.864	0.837	0.329	0.296	0.953

[a]Epoch-by-epoch comparison of Fitbit and polysomnography stages.Each stage—wakefulness after sleep onset, light sleep (non–rapid eye movement [REM] stage 1 [N1] sleep+NREM stage 2 sleep), deep sleep (NREM stage 3 sleep), and REM sleep—was analyzed. REM sleep was divided into analyses with REM sleep episodes occurring during the first 120 minutes after sleep onset with N1 sleep criteria (N1 sleep onset) and REM sleep episodes occurring later than 120 minutes after N1 sleep onset. Various performance measures were used, including sensitivity, specificity, accuracy, the Matthews correlation coefficient, the positive predictive value, and the negative predictive value. More information on these measures can be found in the *Methods* section. Fitbit showed mostly good specificity but poor sensitivity. The accuracy was relatively high except for the light sleep stage. The Matthews correlation coefficient displayed a moderately positive relationship, with light and deep sleep being considerably less good. The negative predictive value was usually higher than the positive predictive value.

[b]MCC: Matthews correlation coefficient.

[c]PPV: positive predictive value.

[d]NPV: negative predictive value.

[e]WASO: wakefulness after sleep onset.

[f]REM: rapid eye movement.

Discussion

Principal Findings

We evaluated the performance of the multisensor wearable Fitbit Charge 2 against PSG of the sleep macrostructure and HR in a sample of first responder shift workers under naturalistic conditions. We observed that S_{on}, S_{off}, TST, REM_d, N1+N2 sleep duration, and N3 sleep duration showed unbiased estimates but nonnegligible LoA. Fitbit overestimated REML by –29.4 minutes, possibly because the proprietary algorithm failed to detect very short first REM sleep episodes. This hypothesis is supported by the right shift in the maximum duration of stages and larger tails (Figure 4) and a cluster of REML data points occurring at approximately –100 minutes (Figure 5), indicating that Fitbit cannot capture short-lasting stage durations well. Not only REML but also other sleep variables often exhibited a wide LoA. In addition, despite performing a careful, data-driven time alignment between the Fitbit and PSG time series, which differed from 1.9 minutes to 7.5 minutes depending on the participants' entrance into the study, Fitbit overestimated WASO by as much as 37.1 minutes. We concluded that the unbiased sleep variables allow average estimations of important sleep quality characteristics in ecological conditions. However, the wide LoA in most variables and the large biases in REML and WASO limited the meaningfulness of quantifying individual sleep episodes. These findings highlight the considerable challenges still present when relying on consumer-grade technology to address clinical and research questions.

One of our most striking and novel findings is that the distribution of all sleep episode durations differs between the Fitbit Charge 2 and PSG. Fitbit's sleep staging algorithm probably treats *REM* and *deep* sleep states of less than 4.5 minutes differently than sleep stages exceeding this duration. This introduces a nonbiological discontinuity, indicating the potential limitations of the tracker's staging algorithm. Furthermore, it is not clear what PSG measurement corresponds to the Fitbit stage *restless*, which renders meaningful comparisons impossible. Our findings in the sleep episode duration distribution are consistent with recent work [41], which also revealed an underestimation of sleep stage transition dynamics.

The S_{on} measures from Fitbit were unbiased concerning the $N1_{on}$ criteria, whereas there was a higher but nonsignificant underestimation for $N2_{on}$. Thus, it is likely that Fitbit's definition

of S_{on} time roughly corresponds to PSG-derived $N1_{on}$. S_{on} criteria should be reported in future validation studies because whatever criterion one selects (eg, $N1_{on}$, $N2_{on,}$ or alternatively any stage of sleep) will impact many sleep variables, such as TST, REML, and WASO, whose operational definition and calculation depend upon the criterion of S_{on}. This may be one of the reasons for discrepancies reported in the validation literature. A peculiarity of the staging information provided by Fitbit is that the first stage after the S_{on} time and the last stage before S_{off} time is sometimes staged as *awake* or *wake*. We manually adjusted the S_{on} and S_{off} times to be delineated by the first and last occurring stages of sleep rather than including stages of wake at the border of sleep. In a large Fitbit data set collected in 89 individuals for 1 month capturing roughly 3000 sleep episodes [42], 69.8% of all sleep episodes in the first stage after S_{on} and in 50% of all cases, the last stage before S_{off} was not coded as a sleep stage. In other words, an appreciable proportion of Fitbit sleep episodes are *bookended* by a stage of wake. This is an inconspicuous but important caveat. Our adjustment of these data could be a reason why we found $N1_{on}$, S_{off}, and TST to be unbiased when comparing Fitbit data with PSG data, whereas Liang et al [16], de Zambotti et al [43], and Morena-Pino et al [23] found TST biases in different directions. More specifically, a previous study [43] found unbiasedness; another reported an overestimation of WASO [16], whereas Moreno-Pino [23] found an underestimation of WASO when validating Fitbit Charge 2 against PSG. The study by Liang et al [16] with a WASO bias of 24.5 minutes is closest to our results of 37.1 minutes.

Overall, EBE analyses revealed better specificity than sensitivity for all sleep states. This might have been expected. For example, there are much fewer *deep* sleep epochs than epochs labeled as any other sleep stage, which is why a single misclassification carries more weight for sensitivity than specificity. We found *light* sleep to have 0.55 accuracy, whereas de Zambotti et al [43] found an accuracy of 0.81. However, the same study found an accuracy of 0.49 for *deep* sleep, whereas we found a higher respective value of 0.78. Furthermore, REM sleep showed an accuracy of 0.86, similar to that of 0.74 found by de Zambotti et al [43]. A recent systematic review (Haghayegh [14] on various Fitbit devices including Alta, Alta HR, Charge 2, Charge HR, Classic, Flex, One, Surge, Ultra and Versa models) found accuracy values in the range of 0.69-0.81 for *light* sleep, 0.36-0.89 for *deep* sleep, and 0.62-0.89 for REM sleep. Thus, our results for *light* sleep are slightly lower than the range suggested previously, whereas, for *deep* sleep and REM sleep, the accuracy in our study was in the upper range reported. The MCC value, which can be interpreted as a usual correlation coefficient, ranged from 0.11 in *light* sleep to 0.34 in REM sleep. These numbers indicate low to medium strength of correlation, pointing toward room for improvement in the estimation of sleep stages by Fitbit.

The information Fitbit provides on the sleep sensitivity setting, with options *sensitive* and *normal*, may have an influence on the amount of stages that are scored as wake [44]. We set the setting to *sensitive* when data were collected, which might have led to an overestimation of WASO, as seen in Figure 2. However, Fitbit states that this setting has no impact on devices utilizing HR to track sleep [45]. Consistent with our results, REM_d was also found to be unbiased by [3]. In addition, we found $light_d$ and $deep_d$ to be unbiased. As the algorithm is not open source, we do not know with certainty whether our study was running on an updated version of the algorithm compared with other validation studies. This limitation makes it difficult to compare the validation study outcomes of consumer fitness trackers in general [4] and could contribute to the discrepancies with the previous literature. Another reason might stem from the different populations sampled or recording conditions. For example, the algorithm might be better suited to assess sleep in healthy individuals than in patients or shift workers or may perform better in a sleep laboratory than in a naturalistic environment. The discrepancies among studies underscore the necessity to define standardized procedures to test consumer sleep technology to benefit from their potential to collect large-scale sleep data in ecological conditions [21,22,26].

Regarding the HR data, Fitbit slightly underestimated overall HR_{10} by 0.9 bpm with a limited capability to capture sudden HR changes. This underestimation was smaller in N2, N3, and REM sleep stages (0.6, 0.6, and 0.7 bpm, respectively) compared with N1 sleep and wake (1.2 and 1.9 bpm), thus indicating a sleep stage–specific bias. The bias was low and probably not biologically relevant. The low *P* values of biases in differences in the HR measures between the devices arise from the repeated measure design as a vast number of 1-minute values during the whole night for each subject was calculated, thereby increasing the statistical power to detect small biases as significant. The evident HR bias of 0.9 bpm is strikingly similar to the HR bias of 0.88 bpm found in de Zambotti et al [25] in the related Fitbit Charge HR device. As mentioned in the report by Haghayegh et al [13], Fitbit Charge HR and Fitbit Charge 2 share the same hardware and software, thus making a comparison feasible, software updates notwithstanding. We found a stage-dependent bias with lower underestimation in deeper sleep stages sharing lower HR on average and a larger underestimation in wake state and a more transitory sleep stage N1, which share higher HR values on average, a finding compatible with the HR-dependent bias reported by Haghayegh et al [13]. For an HR during sleep of <50 bpm, these authors found an overestimation of 0.51 bpm, and for an HR during sleep >80 bpm, an underestimation of 0.63 bpm. These values are comparable with our findings. On the other hand, Benedetto et al [24] found an HR underestimation of 5.9 bpm during wake state. We also found a larger underestimation during wake episodes of 1.2 bpm, but not as high as 5.9 bpm. In the study by Benedetto et al [24], no time alignment between the two instruments was reported. The method of capturing HR via video recording of live values displayed on the Fitbit app was innovative but could be a source of error. Hence, the results could potentially be influenced by a timing misalignment between the instruments and data collection methods.

Fitbit HR variance was reduced owing to the inaccessibility of raw data and showed higher LoA than the LoA for HR. The differences between the assessments are not surprising, as Fitbit only provided 7.4 measurements per minute on average (Figure 3). This is probably owing to their algorithm providing some

averaged values in preferably 5 seconds, 10 seconds, and 15 seconds measurement intervals (but other interval lengths, eg, 2-second or 7-second intervals, can also be found in the data). For comparison, a PSG-derived HR value can be computed for each IBI. Thus, receiving preprocessed data from Fitbit instead of raw data naturally leads to a considerably higher variance in PSG recordings. Moreover, all HR values from Fitbit are integers, whereas the values from the PSG are real values. This difference in the nature of the values (rounded to integers) additionally leads to slightly different behaviors of the HR_{10} and $HRvar_{10}$ measures. The Fitbit photoplethysmography would be able to capture brief bursts in HR, as evidenced by a study on exercising awake individuals [24]. Data with approximately 1-second time resolutions are only made available in the device's *exercise* mode, which prevents sleep tracking. Nevertheless, Fitbit may still be able to detect variability changes for longer periods during sleep with a reasonable degree of accuracy even without providing users with high resolution or raw HR data (as seen in Table 1, where the ordering of the variance per sleep stage remains nearly intact between Fitbit and PSG).

Limitations

The missing information regarding an objective marker of *lights out* is a limitation of our study, which prevented us from estimating sleep latency. In addition, the number of measurements per minute provided by Fitbit varied, potentially owing to variable signal quality and other internal decision-making processes in Fitbit's proprietary data preprocessing algorithms. Updates to software or firmware could have occurred without notice, harboring a great potential to confound research or clinical undertaking, particularly in longitudinal scenarios. Individual sleep episodes can vary appreciably even within an individual, and caution should be exercised when interpreting results from a Fitbit device. Not being able to blind participants to their own sleep data after collection could influence their behavior in subsequent sleep episodes. This concern is particularly pressing when clinical or otherwise vulnerable populations are involved, and device output is interpreted, which may impact treatment options or health outcomes. For this reason, it is crucial that these devices be validated in more clinically diverse populations.

Conclusions

In a study conducted at home in a relatively large sample validating Fitbit Charge 2 against PSG, compared with most previous validation studies (n=15 [24]; n=25 [16]; n=35 [43]; n=35 [14]; and n=65 [23]), we found unbiased mean estimates of various sleep and HR variables, although the data generally exhibited wide LoA. In addition, we noticed problems in capturing the first REM sleep episodes. The naturalistic design of the study in a heterogeneous sample in terms of age and sex and regularly performing shift work increased the external validity and benefited our understanding of the Fitbit Charge 2's performance in a minimally controlled home environment. Nevertheless, for the reliable use of consumer-grade sleep technology for clinical and research purposes, access to raw data, the use of open-source data analysis algorithms, more control of the data flow to blind users, and compliance with all regulatory aspects are indispensable. Furthermore, future validation studies should also be conducted in populations with sleep disorders, such as narcolepsy, who often present with S_{on} REM sleep episodes that appear particularly difficult to detect. Such studies can help identify the factors that determine the accuracy of Fitbit's sleep and HR measures.

Acknowledgments

The authors thank all participants and staff members of Schutz & Rettung Zürich, Rettungsdienst Winterthur and Stadtpolizei Winterthur. The authors also thank Maria Dimitriu, Giulia Haller, Zilla Huber, Zora Kaiser, Nora Krucker, Josefine Meier, Gioia Peterhans, Daniel Prossnitz, Sinja Rosemann-Niedrig, Nora Werner, Rafael Wespi, Laura van Bommel, and Chantal Wey for assisting with data collection and data processing. This study was supported by the Clinical Research Priority Program Sleep and Health of the University of Zurich, the University of Zurich, and the Swiss National Science Foundation (grant 320030_163439 to HPL).

Authors' Contributions

HPL, BK, YA, and IC conceived the study. YA and IC contributed to implementation and data acquisition. BS, HPL, IC, PA, and WK contributed to data acquisition, analysis, and interpretation. BS, IC, and HPL wrote the manuscript. All authors contributed to manuscript revisions. All authors approved the submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental figures and tables.
[DOCX File , 661 KB - jmir_v23i10e26476_app1.docx]

References

1. Hirshkowitz M. Polysomnography and beyond. In: Kryger MH, Roth T, Dement WC, editors. Principles and Practice of Sleep Medicine. 6th edition. Amsterdam: Elsevier; 2017:1564-1566.

2. Hirshkowitz M. The history of polysomnography: tool of scientific discovery. In: Chokroverty S, Billiard M, editors. Sleep Medicine. New York, NY: Springer; 2015:91-100.
3. de Zambotti M, Goldstone A, Claudatos S, Colrain IM, Baker FC. A validation study of Fitbit Charge 2™ compared with polysomnography in adults. Chronobiol Int 2018 Apr;35(4):465-476. [doi: 10.1080/07420528.2017.1413578] [Medline: 29235907]
4. de Zambotti M, Cellini N, Goldstone A, Colrain IM, Baker FC. Wearable sleep technology in clinical and research settings. Med Sci Sports Exerc 2019 Jul;51(7):1538-1557 [FREE Full text] [doi: 10.1249/MSS.0000000000001947] [Medline: 30789439]
5. Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. Sleep 2003 May 01;26(3):342-392. [doi: 10.1093/sleep/26.3.342] [Medline: 12749557]
6. Borbély AA, Rusterholz T, Achermann P. Three decades of continuous wrist-activity recording: analysis of sleep duration. J Sleep Res 2017 Apr;26(2):188-194 [FREE Full text] [doi: 10.1111/jsr.12492] [Medline: 28093825]
7. Sadeh A, Alster J, Urbach D, Lavie P. Actigraphically based automatic bedtime sleep-wake scoring: validity and clinical applications. J Ambul Monit 1989;2(3):209-216 [FREE Full text]
8. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. Sleep 1992 Oct;15(5):461-469. [doi: 10.1093/sleep/15.5.461] [Medline: 1455130]
9. Walch O, Huang Y, Forger D, Goldstein C. Sleep stage prediction with raw acceleration and photoplethysmography heart rate data derived from a consumer wearable device. Sleep 2019 Dec 24;42(12):zsz180 [FREE Full text] [doi: 10.1093/sleep/zsz180] [Medline: 31579900]
10. Inderkum A, Tarokh L. High heritability of adolescent sleep-wake behavior on free, but not school days: a long-term twin study. Sleep 2018 Mar 01;41(3). [doi: 10.1093/sleep/zsy004] [Medline: 29329461]
11. Vailshery LS. Number of Fitbit devices sold worldwide from 2010 to 2020. Statista. 2021. URL: https://www.statista.com/statistics/472591/fitbit-devices-sold/ [accessed 2020-06-11]
12. Scott H, Lack L, Lovato N. A systematic review of the accuracy of sleep wearable devices for estimating sleep onset. Sleep Med Rev 2020 Feb;49:101227. [doi: 10.1016/j.smrv.2019.101227] [Medline: 31901524]
13. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR. Accuracy of PurePulse photoplethysmography technology of Fitbit Charge 2 for assessment of heart rate during sleep. Chronobiol Int 2019 Jul;36(7):927-933. [doi: 10.1080/07420528.2019.1596947] [Medline: 30990098]
14. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Performance comparison of different interpretative algorithms utilized to derive sleep parameters from wrist actigraphy data. Chronobiol Int 2019 Dec;36(12):1752-1760. [doi: 10.1080/07420528.2019.1679826] [Medline: 31658822]
15. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Performance assessment of new-generation Fitbit technology in deriving sleep parameters and stages. Chronobiol Int 2020 Jan;37(1):47-59. [doi: 10.1080/07420528.2019.1682006] [Medline: 31718308]
16. Liang Z, Chapa Martell MA. Validity of consumer activity wristbands and wearable EEG for measuring overall sleep parameters and sleep structure in free-living conditions. J Healthc Inform Res 2018 Apr 20;2(1-2):152-178. [doi: 10.1007/s41666-018-0013-1]
17. Trinder J, Waloszek J, Woods MJ, Jordan AS. Sleep and cardiovascular regulation. Pflugers Arch 2012 Jan;463(1):161-168. [doi: 10.1007/s00424-011-1041-3] [Medline: 22038322]
18. Cajochen C, Pischke J, Aeschbach D, Borbély AA. Heart rate dynamics during human sleep. Physiol Behav 1994 Apr;55(4):769-774. [doi: 10.1016/0031-9384(94)90058-2] [Medline: 8190808]
19. Ako M, Kawara T, Uchida S, Miyazaki S, Nishihara K, Mukai J, et al. Correlation between electroencephalography and heart rate variability during sleep. Psychiatry Clin Neurosci 2003 Feb;57(1):59-65 [FREE Full text] [doi: 10.1046/j.1440-1819.2003.01080.x] [Medline: 12519456]
20. What should I know about sleep stages? Fitbit. 2020. URL: https://help.fitbit.com/articles/en_US/Help_article/2163 [accessed 2020-07-10]
21. Karlen W, Floreano D. Adaptive sleep-wake discrimination for wearable devices. IEEE Trans Biomed Eng 2011 Apr;58(4):920-926. [doi: 10.1109/TBME.2010.2097261] [Medline: 21172750]
22. Karlen W, Mattiussi C, Floreano D. Improving actigraph sleep/wake classification with cardio-respiratory signals. In: Proceedings of the 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2008 Presented at: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Aug 20-25, 2008; Vancouver, BC, Canada. [doi: 10.1109/iembs.2008.4650401]
23. Moreno-Pino F, Porras-Segovia A, López-Esteban P, Artés A, Baca-García E. Validation of Fitbit Charge 2 and Fitbit Alta HR against polysomnography for assessing sleep in adults with obstructive sleep apnea. J Clin Sleep Med 2019 Nov 15;15(11):1645-1653 [FREE Full text] [doi: 10.5664/jcsm.8032] [Medline: 31739855]
24. Benedetto S, Caldato C, Bazzan E, Greenwood DC, Pensabene V, Actis P. Assessment of the Fitbit Charge 2 for monitoring heart rate. PLoS One 2018 Feb 28;13(2):e0192691 [FREE Full text] [doi: 10.1371/journal.pone.0192691] [Medline: 29489850]

25. de Zambotti M, Baker FC, Willoughby AR, Godino JG, Wing D, Patrick K, et al. Measures of sleep and cardiac functioning during sleep using a multi-sensory commercially-available wristband in adolescents. Physiol Behav 2016 May 01;158:143-149 [FREE Full text] [doi: 10.1016/j.physbeh.2016.03.006] [Medline: 26969518]
26. Menghini L, Cellini N, Goldstone A, Baker FC, de Zambotti M. A standardized framework for testing the performance of sleep-tracking technology: step-by-step guidelines and open-source code. Sleep 2021 Feb 12;44(2):zsaa170. [doi: 10.1093/sleep/zsaa170] [Medline: 32882005]
27. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
28. Blevins CA, Weathers FW, Davis MT, Witte TK, Domino JL. The posttraumatic stress disorder checklist for DSM-5 (PCL-5): development and initial psychometric evaluation. J Trauma Stress 2015 Dec;28(6):489-498. [doi: 10.1002/jts.22059] [Medline: 26606250]
29. Cohen S, Williamson G. Perceived stress in a probability sample of the United States. In: Spacapan S, Oskamp S, editors. The Social Psychology of Health: Claremont Symposium on Applied Social Psychology. Newbury Park, CA: Sage Publications Inc; 1988:1-256.
30. Adan A, Almirall H. Horne and Ostberg morningess-eveningness questionnaire: a reduced scale. Pers Indiv Differ 1991;12(3):241-253. [doi: 10.1016/0191-8869(91)90110-w]
31. Jasper HH. The ten-twenty electrode system of the International Federation. Electroen Clin Neuro 1958;10:371-375 [FREE Full text]
32. The AASM manual for the scoring of sleep and associated events. American Academy of Sleep Medicine. URL: https://aasm.org/clinical-resources/scoring-manual/ [accessed 2021-08-05]
33. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.r-project.org/ [accessed 2021-08-05]
34. Pan J, Tompkins WJ. A real-time QRS detection algorithm. IEEE Trans Biomed Eng 1985 Mar;BME-32(3):230-236. [doi: 10.1109/tbme.1985.325532]
35. Bouchequet P. rsleep: analysis of sleep data. R package version 1.0.3. 2020. URL: https://cran.r-project.org/web/packages/rsleep/index.html [accessed 2021-08-05]
36. Khaleghi B, Khamis A, Karray FO, Razavi SN. Multisensor data fusion: a review of the state-of-the-art. Inform Fusion 2013 Jan;14(1):28-44. [doi: 10.1016/j.inffus.2011.08.001]
37. Rhudy M. Time alignment techniques for experimental sensor data. Int J Comput Sci Eng Surv 2014 Apr 30;5(2):1-14. [doi: 10.5121/ijcses.2014.5201]
38. Datta D. blandr: a Bland-Altman method comparison package for R. GitHub. 2018. URL: https://github.com/deepankardatta/blandr [accessed 2021-08-05]
39. Pinheiro J, Bates D, DebRoy S, Sarkar D. nlme: linear and nonlinear mixed effects models. R Core Team. URL: https://cran.r-project.org/web/packages/nlme/index.html [accessed 2021-08-05]
40. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 2020 Jan 02;21(1):6 [FREE Full text] [doi: 10.1186/s12864-019-6413-7] [Medline: 31898477]
41. Liang Z, Chapa-Martell MA. Accuracy of Fitbit wristbands in measuring sleep stage transitions and the effect of user-specific factors. JMIR Mhealth Uhealth 2019 Jun 06;7(6):e13384 [FREE Full text] [doi: 10.2196/13384] [Medline: 31172956]
42. Clark I, Stucky B, Azza Y, Schwab P, Müller SJ, Weibel D, et al. Diurnal variations in multi-sensor wearable-derived sleep characteristics in morning- and evening-type shift workers under naturalistic conditions. Chronobiol Int 2021 Jul 18:1-12 (forthcoming). [doi: 10.1080/07420528.2021.1941074] [Medline: 34278901]
43. de Zambotti M, Goldstone A, Claudatos S, Colrain IM, Baker FC. A validation study of Fitbit Charge 2™ compared with polysomnography in adults. Chronobiol Int 2018 Apr 13;35(4):465-476. [doi: 10.1080/07420528.2017.1413578] [Medline: 29235907]
44. How do I track my sleep with my Fitbit device? 2020. Fitbit Inc. URL: https://help.fitbit.com/articles/en_US/Help_article/1314.htm [accessed 2020-12-08]
45. How do I set Fitbit sleep sensitivity? Fitbit Community. 2019. URL: https://community.fitbit.com/t5/Inspire-Inspire-HR/How-do-I-set-Fitbit-sleep-sensitivity/td-p/3556808 [accessed 2021-08-05]

Abbreviations

bpm: beats per minute
$deep_d$: deep sleep duration
EBE: epoch-by-epoch
HR: heart rate
HR_{10}: 10%-trimmed heart rate average
$HRvar_{10}$: 10%-trimmed heart rate variance average

IBI: interbeat interval
$light_d$: light sleep duration
LoA: limits of agreement
MCC: Matthews correlation coefficient
$N1_{on}$: N1 sleep onset
$N2_{on}$: N2 sleep onset
NPV: negative predictive value
PPV: positive predictive value
PSG: polysomnography
PSQI: Pittsburgh Sleep Quality Index
PSS-10: Perceived Stress Scale 10
QAVSD60: 60% quantile of the absolute value of the second derivative
REM: rapid eye movement
REM_d: rapid eye movement sleep duration
REML: rapid eye movement sleep latency
rMEQ: Horne-Östberg Morningness-Eveningness Questionnaire-A Reduced Scale
S_{off}: sleep offset
S_{on}: sleep onset
TST: total sleep time
WASO: wakefulness after sleep onset

Edited by R Kukafka; submitted 14.12.20; peer-reviewed by J Zhang, M Pobiruchin, J Golbus; comments to author 15.03.21; revised version received 08.05.21; accepted 14.06.21; published 05.10.21.

Please cite as:
Stucky B, Clark I, Azza Y, Karlen W, Achermann P, Kleim B, Landolt HP
Validation of Fitbit Charge 2 Sleep and Heart Rate Estimates Against Polysomnographic Measures in Shift Workers: Naturalistic Study
J Med Internet Res 2021;23(10):e26476
URL: https://www.jmir.org/2021/10/e26476
doi:10.2196/26476
PMID:34609317

Original Paper

Feasibility of a Mobile Health App for Routine Outcome Monitoring and Feedback in SMART Recovery Mutual Support Groups: Stage 1 Mixed Methods Pilot Study

Peter J Kelly[1,2], BSc, GradDipAppPsych, PhD; Alison K Beck[1], DClin, BPsych (Hons); Frank P Deane[1,2], BSc, PGDipClinPsych, MSc, PhD; Briony Larance[1,2], BSc (Hons), PhD; Amanda L Baker[3], BA (Hons), MPsych, PhD; Leanne Hides[4], BBehSc (Hons), PhD; Victoria Manning[5], BSc (Hons), MSc, PhD; Anthony Shakeshaft[6], BA, MA, PhD; Joanne Neale[7], BA (Hons), MA, CQSW, DPhil; John F Kelly[8], BSc, PhD; Christopher Oldmeadow[9], BMath (Hons), PhD; Andrew Searles[10], BEc, MMedStat, DipEd, PhD; Kerrin Palazzi[9], BBiomedSc, MPH; Kenny Lawson[10], BA (Hons), MSc, PhD; Carla Treloar[11], BSc (Hons), PhD; Rebecca M Gray[11], BA (Hons), MA, PhD; Angela Argent[12], BA (Hons), PhD; Ryan McGlaughlin[12], PGCert (Comm Mgt), MNFPSocEntM

[1]School of Psychology, Faculty of Arts, Social Sciences and Humanities, University of Wollongong, Wollongong, Australia

[2]Illawarra Health and Medical Research Institute, University of Wollongong, Wollongong, Australia

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

[4]Centre for Youth Substance Abuse Research, Lives Lived Well Group, School of Psychology, University of Queensland, Brisbane St Lucia, Australia

[5]Eastern Health Clinical School, Faculty of Medicine, Nursing and Health Sciences, Monash University, Box Hill, Australia

[6]National Drug and Alcohol Research Centre, University of New South Wales, Sydney, Australia

[7]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[8]Harvard Medical School, Harvard University, Boston, MA, United States

[9]Clinical Research Design, IT and Statistical Support Unit, Hunter Medical Research Institute, New Lambton, Australia

[10]Hunter Medical Research Institute Health Research Economics, Hunter Medical Research Institute, New Lambton, Australia

[11]Centre for Social Research in Health, Faculty of Arts and Social Sciences, University of New South Wales, Sydney, Australia

[12]SMART Recovery Australia, Sydney, Australia

Corresponding Author:
Alison K Beck, DClin, BPsych (Hons)
School of Psychology
Faculty of Arts, Social Sciences and Humanities
University of Wollongong
Northfields Ave
Wollongong, 2522
Australia
Phone: 61 2 4239 2382
Email: alisonbe@uow.edu.au

Abstract

Background: Mutual support groups are an important source of long-term help for people impacted by addictive behaviors. Routine outcome monitoring (ROM) and feedback are yet to be implemented in these settings. SMART Recovery mutual support groups focus on self-empowerment and use evidence-based techniques (eg, motivational and behavioral strategies). Trained facilitators lead all SMART Recovery groups, providing an opportunity to implement ROM.

Objective: The aim of this stage 1 pilot study is to explore the feasibility, acceptability, and preliminary outcomes of a novel, purpose-built mobile health ROM and feedback app (*SMART Track*) in mutual support groups coordinated by SMART Recovery Australia (SRAU) over 8 weeks.

Methods: *SMART Track* was developed during phase 1 of this study using participatory design methods and an iterative development process. During phase 2, 72 SRAU group participants were recruited to a nonrandomized, prospective, single-arm trial of the *SMART Track* app. Four modes of data collection were used: ROM data directly entered by participants into the app; app data analytics captured by Amplitude Analytics (number of visits, number of unique users, visit duration, time of visit, and

user retention); baseline, 2-, and 8-week follow-up assessments conducted through telephone; and qualitative telephone interviews with a convenience sample of study participants (20/72, 28%) and facilitators (n=8).

Results: Of the 72 study participants, 68 (94%) created a *SMART Track* account, 64 (88%) used *SMART Track* at least once, and 42 (58%) used the app for more than 5 weeks. During week 1, 83% (60/72) of participants entered ROM data for one or more outcomes, decreasing to 31% (22/72) by the end of 8 weeks. The two main screens designed to provide personal feedback data (*Urges* screen and *Overall Progress* screen) were the most frequently visited sections of the app. Qualitative feedback from participants and facilitators supported the acceptability of *SMART Track* and the need for improved integration into the SRAU groups. Participants reported significant reductions between the baseline and 8- week scores on the Severity of Dependence Scale (mean difference 1.93, SD 3.02; 95% CI 1.12-2.73) and the Kessler Psychological Distress Scale-10 (mean difference 3.96, SD 8.31; 95% CI 1.75-6.17), but no change on the Substance Use Recovery Evaluator (mean difference 0.11, SD 7.97; 95% CI –2.02 to 2.24) was reported.

Conclusions: Findings support the feasibility, acceptability, and utility of *SMART Track*. Given that sustained engagement with mobile health apps is notoriously difficult to achieve, our findings are promising. *SMART Track* offers a potential solution for ROM and personal feedback, particularly for people with substance use disorders who attend mutual support groups.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12619000686101; https://anzctr.org.au/Trial/Registration/TrialReview.aspx?id=377336

International Registered Report Identifier (IRRID): RR2-10.2196/15113

(J Med Internet Res 2021;23(10):e25217) doi:10.2196/25217

KEYWORDS

mHealth; SMART Recovery; mutual support group; mutual aid; routine outcome monitoring; treatment progress feedback; addiction; mobile phone

Introduction

Background

Routine outcome monitoring (ROM) is central to evidence-based health care for a range of chronic conditions [1], including addictive behaviors [2,3]. ROM is central to understanding, evaluating, and improving service delivery [4-6]. A range of clinical benefits have been identified [7-9], particularly for those people identified as *not on track* early in the course of treatment [10,11]. Emerging evidence suggests that providing clients with tailored feedback may be central to demonstrated improvements in client outcomes [12].

To date, ROM and feedback have been implemented in a range of mental health [13] and addiction [14,15] treatment settings but not in mutual support groups. Mutual support groups offer an important source of fee-free, accessible support to people experiencing a range of addictive behaviors. Mutual support is particularly important for people experiencing addictive behaviors, given the often long-term and nonlinear process of recovery [16]. Mutual support groups may be attended before, during, after, or in lieu of engagement with formal treatment services, providing the potential for continuity across the recovery process. Although accumulating evidence highlights the importance and benefits of participating in mutual support [17-21], a major limitation is the lack of systematically collected data evaluating the outcomes. Unlike other clinically endorsed [2,3] models of mutual support for addictive behaviors (eg, 12-step approaches), SMART Recovery groups use a trained facilitator. This provides a unique opportunity to work with group facilitators to embed ROM and personal feedback as a standard component of the groups.

Integrating ROM and tailored feedback into routine service provision is not without challenges [22,23]. Common barriers include the *time burden* associated with completing, scoring, interpreting, or discussing outcome assessments [22,24], as well as skepticism regarding the perceived relevance of the outcomes assessed and feedback generated [25,26]. Additional limitations include the traditionally clinician-centric nature of ROM (see studies by Carlier and van Eden [7] and Thompson et al [13] for a discussion and studies by Lambert et al [8], Goodman et al [12], and Burgess et al [27] for common instruments) and accompanying feedback [28,29]. Improved acknowledgment of the client perspective during assessment [30] and greater client involvement in the feedback process [31] are both important clinical and research priorities.

The idea of using technology to track progress within health care settings is not new, but current approaches are limited [32]. Unlike other health information technology approaches (eg, web-based platforms), mobile health (mHealth [33]) apps offer a quick, easy, interactive, and engaging platform for tracking and accessing information about health and health-related behaviors [34]. A key benefit of mHealth apps is their ability to provide timely, individualized feedback [35]. Given the ubiquity of smartphone ownership [36,37], smartphone apps can engage individuals in real time and in their natural environment and by offering moment-to-moment support as needed [38]. Indeed, a recent systematic review of digital support services highlighted that their *on-demand* nature is a key benefit [39].

Although not specifically designed for the purposes of ROM and feedback, mHealth apps with the capability to track a variety of health behaviors, conditions, or outcomes [40-45], including alcohol consumption, substance use, and other addictive behaviors [39,44-52], have been developed. However, a key limitation is the ever-increasing gap between the availability of mHealth apps and their scientific validation [40,52-55].

Moreover, the level of end-user involvement throughout the development process is often unclear. This is important because inadequate consideration of the needs and preferences of the end user has been implicated in mHealth attrition [56-58]. Accordingly, we worked alongside end users to develop a purpose-built mHealth app for ROM and feedback in SMART Recovery Australia (SRAU) mutual support groups (*SMART Track*), which was then evaluated in this study.

Objective

The aim of this stage 1 nonrandomized, single-arm pilot study is to explore the feasibility, acceptability, and preliminary outcomes of a novel mHealth ROM and feedback app (*SMART Track*) in mutual support groups coordinated by SRAU.

Methods

Overview

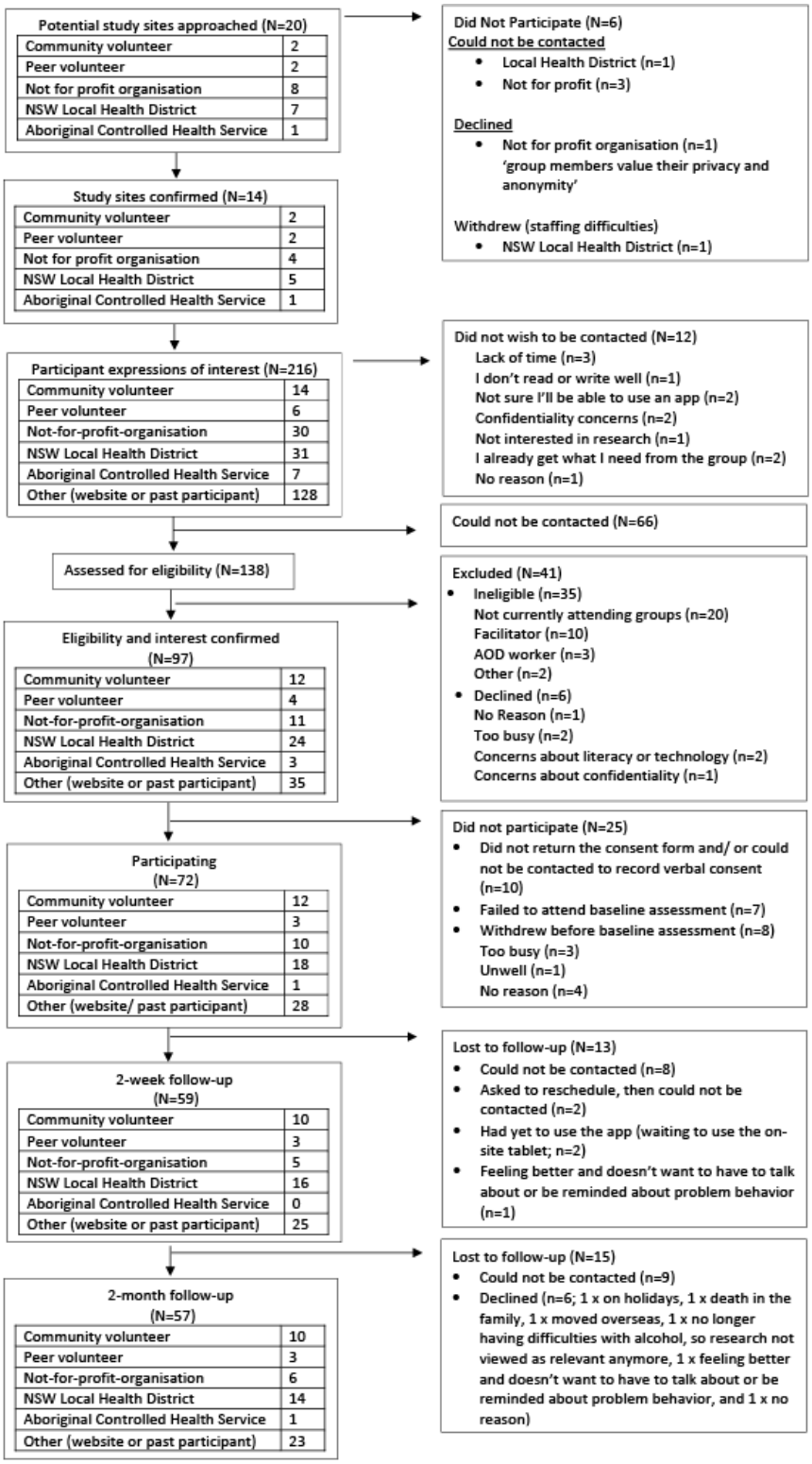
Approval was granted by the University of Wollongong and Illawarra Shoalhaven Local Health District Health and Medical Human Research Ethics Committee (2018/099; HREC/18/WGONG/34). The study was registered with the Australian New Zealand Clinical Trials Registry (ACTRN12619000686101), and a protocol was published [59]. The reporting of this study follows the CONSORT (Consolidated Standards of Reporting Trials)-EHEALTH checklist [60].

Setting

Participants were recruited from the SMART Recovery groups registered with SRAU. Detailed accounts of SMART Recovery groups have been published [61]. Briefly, SMART Recovery groups originated in the United States and are now available across 23 countries. They offer support for people experiencing a range of addictive behaviors, including substance- and non–substance-related behaviors. SMART Recovery groups focus on self-empowerment and use evidence-based techniques (eg, cognitive behavioral therapy and motivational interviewing) [62]. These groups are held in a variety of community, inpatient, outpatient, residential rehabilitation, and clinical health settings. Online support groups are also available.

We invited 20 sites in New South Wales, Australia, to participate in this study, and 14 (70%) agreed (Figure 1). To enhance generalizability, the invited sites were selected to reflect a range of geographical locations and service providers. We sought to recruit 100 study participants. A sample of this size was selected to allow estimation of the recruitment rate and 95% CI with a margin of error of no more than 7%.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram depicting the number of participants referred, lost, and retained according to the referral source. AOD: alcohol and other drugs; NSW: New South Wales.



Participants

Participants were eligible if they were aged at least 18 years, were currently participating in SRAU groups (either face-to-face or online), had (or were willing to obtain) an email address, and comprehended English at a level sufficient to complete the study requirements. Participants were eligible irrespective of self-reported computer or smartphone literacy, and they did not have to own a smartphone. The study sites were provided with an Android (Samsung Galaxy Tab A) tablet for on-site participant use. No restrictions were placed on concomitant care or the frequency or duration of SMART Recovery group participation. The only exclusion criterion was inability or unwillingness to provide informed consent.

Recruitment

A group facilitator or member of the research team provided potential participants with standardized written and verbal information at the beginning of the SMART Recovery group session. Potential participants were asked to provide their preferred contact details, and they were contacted directly by a member of the research team. To avoid any potential coercion or desirability bias arising from the working relationship between facilitators and participants, the researcher (not the facilitator) was responsible for confirming participant interest and seeking informed consent. To boost accrual, during the final month of recruitment, a web-based expression-of-interest form (displayed prominently on the SRAU website) was introduced. Potential participants could contact the research team directly through email, phone, or the web-based expression-of-interest form. All participants provided verbal or written informed consent. The participants were reimbursed (Aus $30 [US $22.02] supermarket voucher) for their time, travel, and effort associated with each interview: baseline and 8-week assessments, as well as qualitative interviews (up to a total of Aus $90 [US $66.06]).

SMART Track: ROM and Feedback mHealth App

Development

Overview

The preparatory qualitative work [63] and development process [64] have been reported separately, and further details are available in the published protocol [59]. Briefly, three frameworks [56,65,66] informed the design, development, and content of *SMART Track*. Although each framework can be used in isolation, we chose to combine these approaches to ensure that app development was informed by a more comprehensive set of guidelines that included foci related to the end user (ie, *person*; *person-based* [56]); best practice recommendations for mHealth development (Behavioral Intervention Technology Model [65]); and a collaborative, iterative development process involving the research team, app developers, and participants (Integrate, Design, Assess, and Share Framework [66]). *SMART Track* is grounded in behavioral theory (Self-Determination Theory [67] and Social Control Theory [68]) and the guiding principles of SMART Recovery (self-management, mutual aid, and choice [69]). Behavioral strategies are drawn from the Behavior Change Taxonomy (self-monitoring, feedback, action planning, prompts or cues, and nonspecific reward [70]) and process motivators (choice or control, competence, context, curiosity, personalization, and reframing [66]). The agency contracted for app development and design was GHO, Sydney [71].

Beta-Testing

The initial beta version of the app was submitted to the Apple App Store and Google Play Store for approval in March 2019. The functionality of the app was initially tested with 3 members of the research team (beginning April 5, 2019). Several bugs were identified and fixed before the emended beta version was released (June 4, 2019) for further testing to a convenience sample comprising 40 members of the SRAU Research Advisory Committee, SRAU steering committee, and SMART Recovery board, as well as SMART Recovery facilitators. Further refinements were made in line with the feedback (bug fixes and minor amendments to functionality and content). The participant version of *SMART Track* was available in the Google Play Store (version 0.0.7) and Apple App Store (version 0.7) on July 15, 2019. *SMART Track* is freely available for Android [72] and Apple [73] devices.

Revisions and Updating

The time frame of the weekly period of ROM data collection was emended in July 2019 (from closing 24 hours after the nominated meeting began to closing 30 minutes after the nominated meeting began). This was to enable the next week of data collection to begin during the meeting such that the participants could set a new 7-day plan at the end of the meeting (rather than having to wait 24 hours). Cloud functions were updated in September 2019 to fix to participant reports (4/72, 6%) that they had not received the expected prompt from *SMART Track* to complete the ROM items.

Overview

Summary

The *SMART Track* app is designed for participants attending SMART Recovery groups (either face-to-face or online). *SMART Track* comprises core ROM and feedback functionality and several additional features to enhance engagement (resources, customizable supports, personal motivations, interactive urge log, and pop-up motivations and self-management strategies, as described below). The content is distributed across five main screens (Figure S1 of Multimedia Appendix 1).

ROM Domains and Items

Consistent with clinical guidelines [2,3] and published recommendations [7,31], *SMART Track* provides multidimensional assessment and feedback. The items included in the app are detailed in Table S1 of Multimedia Appendix 2 [74-83] as a function of target domain and assessment frequency. Further details are available in the published protocol [59]. Briefly, the participants were prompted each week to answer a set of questions, and their responses were used to provide tailored progress feedback.

Progress Feedback

Feedback consists of tailored visual and written feedback across eight domains (7-day plan, behavior of concern, effect of substance use, self-care, relationships, outlook on life, resources, and mental health; see Tables S2 and S3 of Multimedia Appendix 2 for the scoring algorithms).

Resources

The *Resources* screen is able to deliver a maximum of 10 pieces of content. This was distributed across seven self-management resources (including SMART Recovery resources) and three motivational stories (extracted with permission from the *Lives of Substance* website [84]). Content upload was managed by the research team using *WordPress* according to the schedule outlined in Table S4 of Multimedia Appendix 2.

Customizable Supports and Personal Motivations

Participants have the option of tailoring app content by uploading one or more contact numbers, support services or personal motivations for change (photo, audio, video, or text) into the *Me* section of the app.

Interactive Urge Log

In addition to tracking the number, frequency, and strength of urges, when the participant reports an urge, this interactive tool prompts them to manage their urges, log triggers, and reflect on how to maintain or improve effective urge self-management. The interactive urge log contains a range of urge management strategies or motivational content (Table S5 of Multimedia Appendix 2). The content was derived from SMART Recovery manuals [85,86] and transcripts of participants' qualitative interviews [63] and presented to the participants in random order. The participants could also use the *Me* section of the app to enter their own personal strategies and motivations. Participant-entered content is always shown before prespecified content, and it is not accessible to other participants.

Pop-up Motivations and Self-management Strategies

The participants received *pop*-up messages when they opened the app for the first time each day (Table S6 of Multimedia Appendix 2). This content is derived from transcripts of qualitative interviews [63]. A combination of direct excerpts and emended content (modified for clarity) was used.

Implementation

Orientation

After completion of the baseline assessment, the researcher asked the participants to use *SMART Track* at least once a week to complete the ROM questions and enter their 7-day plan and use the other app functions *as needed*. This was reinforced in an introductory email, which also contained the Google Play Store and Apple App Store links to download the app. SMART Recovery facilitators were asked to prompt the participants at the beginning and end of each group session. No additional training or support was provided (outside of what may have been naturally provided by facilitators and peers as part of the group session). A *walk-through* is included in the app to orient participants to the app (Figure S1 of Multimedia Appendix 1).

Prompts and Reminders

The 7-day plan and ROM questions were linked to the day and time that the participant used *SMART Track* in their regular SMART Recovery group session. Tasks were set for 7 days after the meeting. The 7-day plan notifications were customizable. For each task, the participants elected whether and when to receive a reminder notification. A notification to complete the ROM questions was automatically sent 24 hours before the nominated group session. If the questions were not answered, additional reminders were sent 12 hours and again 30 minutes before the group session. The ROM notifications could not be *switched off* by the participant.

Privacy and Confidentiality

To allow participants to reset their password, *SMART Track* captures the email addresses of all end users. However, it is up to the participant to decide whether the email address they choose to use contains any element of personal information (eg, their name). Given the potential impact of privacy- and confidentiality-related concerns on participant engagement with *SMART Track*, a comprehensive privacy and confidentiality policy is available.

Data Collection Procedures

Overview

The study activities are outlined in Figure 2. The four modes of data collection included (1) participant-completed ROM data collected through *SMART Track* (Table S1 of Multimedia Appendix 2); (2) app data analytics captured using Amplitude Analytics (Amplitude, Inc; number of visits, number of unique users, visit duration, time of visit, and user retention) [87]; (3) baseline, 2-week, and 8-week follow-up assessments conducted over the telephone by AKB; and (4) qualitative interviews with the study participants and group facilitators (conducted over the telephone by RMG). The primary and secondary objectives, measures, and indicator variables are summarized in Table 1.

Figure 2. Flowchart of study activities.

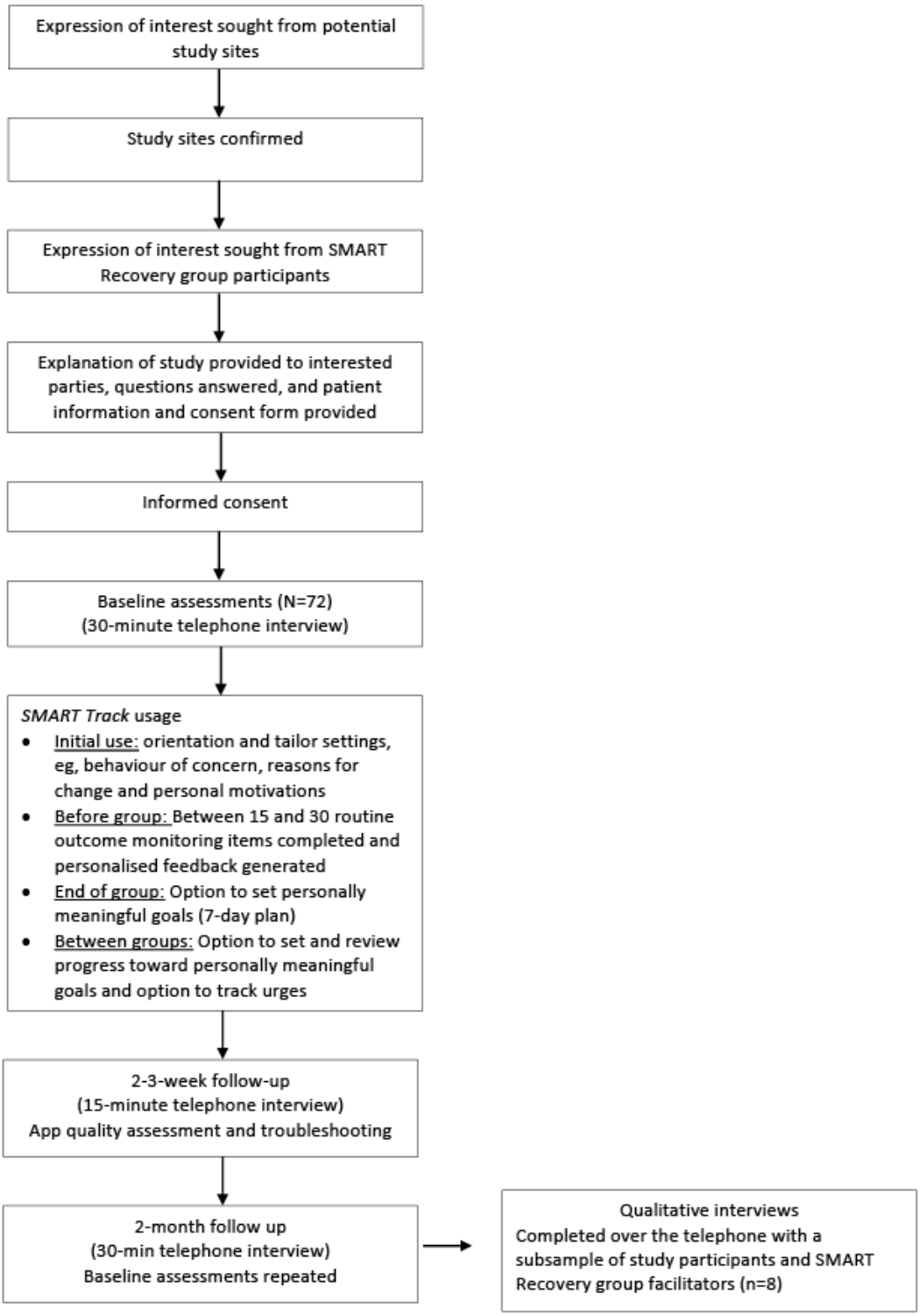


Table 1. Primary and secondary objectives, measures, and indicators.

Objectives	Variables
Primary objectives (measures and indicators)	
To explore the feasibility of using *SMART Track* as part of SMART Recovery groups	• Proportion of eligible participants who consent to the study, create an account, and use SMART Track • Proportion of missing data for each of the routine outcome monitoring items and instruments at each week of administration across the 8-week period of SMART Track use • Engagement and use patterns indexed through data analytics captured daily across the data collection period • Costs associated with developing SMART Track and maintaining the app until the completion of data collection
To explore the acceptability of using *SMART Track* as part of SMART Recovery groups	• Detailed qualitative feedback from SMART Recovery group members and facilitators to explore their experience of, and satisfaction with, SMART Track (8-week follow-up) • Quality ratings as assessed by participant ratings of the end-user version [87] of the Mobile App Rating Scale [88] at 8-week follow-up • Digital therapeutic alliance ratings as assessed by participant ratings of the Digital Working Alliance Inventory at 8-week follow-up
Secondary objective (secondary end points)	
To provide preliminary evidence for participant-reported outcomes	• Participant-reported progress across the 8-week period of app use in (1) substance dependence (Severity of Dependence Scale [89]), (2) addiction recovery (Substance Use Recovery Evaluator [74]), and (3) mental health (Kessler Psychological Distress Scale [75,90])

Key Measures and Assessment Instruments

Overview

The study measures and assessment instruments are detailed in the published protocol [59] and summarized in Table 2. Feasibility and acceptability were informed by data analytics captured using Amplitude Analytics (number of visits, number of unique users, visit duration, time of visit, and user retention) [87]; qualitative interviews; quality assessment conducted using the simplified, end-user version [88] of the Mobile App Rating Scale (MARS) [89] and the Digital Working Alliance Inventory (DWAI) [92]; and a cost analysis informed by a *cost capture template* [93-95] and an adapted version of the Client Service Receipt Inventory—*Generic* UK Mental Health [96]. Preliminary evidence for participant-reported outcomes after the use of *SMART Track* in conjunction with SMART Recovery groups was captured using the Severity of Dependence Scale (SDS) [90], Kessler Psychological Distress Scale-10 (K-10) [97], and the Substance Use Recovery Evaluator (SURE) [74].

Table 2. Schedule of data collection.

	Baseline	Daily	Weekly	2-week follow-up	8-week follow-up
SMART Recovery participants					
***SMART Track* app**					
Data analytics		✓[a]			
ROM[b] items[c]			✓		
Demographics	✓				✓
NADA[d] COMS[e]					
Severity of Dependence Scale	✓				✓
Drug and Alcohol Use	✓				✓
Kessler 10+	✓				✓
The World Health Organization Quality of Life 8	✓				✓
NSW[f] Minimum Data Set items (living arrangements and income)	✓				✓
BTOM-C[g] items on arrests	✓				✓
BTOM-C items on risky drug using practices	✓				✓
Substance Use Recovery Evaluator	✓				✓
Client Services Receipt Inventory	✓				✓
Mobile Application Rating Scale–User Version				✓	
Digital Working Alliance Inventory				✓	✓
Qualitative interview (n=20)					✓
SMART Recovery facilitators					
Demographics					✓
Mobile App Rating Scale					✓
Qualitative interview (n=8)					✓

[a]Data collected.

[b]ROM: routine outcome monitoring.

[c]See Multimedia Appendix 2 (Table S1) for a detailed description of routine outcome monitoring items as a function of assessment domain and frequency of administration.

[d]NADA: Network of Alcohol and Other Drugs Agencies.

[e]COMS: Client Outcomes Management System.

[f]NSW: New South Wales.

[g]BTOM-C: Brief Treatment Outcome Measure—Concise.

Nested Qualitative Evaluation

Qualitative interviews were conducted by RMG after the 8-week period of app use to explore the experiences and opinions of participants with diverse engagement with *SMART Track*. The participants were sampled to reflect the diversity of their characteristics (gender and primary behavior of concern), referral source, and pattern of *SMART Track* use. An independent qualitative researcher (RMG) used a topic guide (Table S7 of Multimedia Appendix 2) to ask additional open-ended questions of a selection of participants (n=20) and facilitators (n=8). The participants and facilitators were sampled to reflect diversity in gender, geographical location, and (for participants only) behavior of concern. For the app users, this included the pathway to SMART Recovery groups (opening and warm-up), perceptions and experiences of app use, initiation circumstances for the app, motivation to join the trial and use the app, and suggestions for improvements. The facilitators were asked similar questions, but the focus was on their professional capacity rather than on their personal experience with apps. The interview started by eliciting information about how they came to be a facilitator, how the app was initiated with their group, their perceptions and experiences in implementing the app with the service users, their motivation levels related to the implementation of the app, and suggestions for improvements. All interviews were audio recorded and transcribed by a professional transcriber working under a confidentiality agreement.

Analysis

Feasibility Indicators

Enrollment and Engagement

Data analytics were captured daily from the time the app was launched until the last participant completed their 8-week follow-up interview. The first participant was given the download details on July 15, 2019, and follow-up data collection was completed on December 2, 2019. Weekly summaries for the total number of unique users and the average number of visits per user were downloaded from Amplitude Analytics. Unique user codes were linked, and the number of weeks that each study participant used the app was calculated.

Use Patterns

To explore how the participants engaged with the various features of *SMART Track*, weekly analytics (total number of visits, total number of unique users, and total duration) for each of the *SMART Track* features (*Urge button*, *Urges* screen, *Resources* screen, *Me* screen, and *Overall Progress* screen) and the time of day that the app was used were downloaded from Amplitude Analytics and descriptive statistics calculated. Retention was characterized using weekly summaries from the *User Lifecycle* feature of Amplitude Analytics. This feature categorizes participants into the following mutually exclusive categories:

1. New users (used the app for the first time that week).
2. Current users (used the app at least once that week and at least once during the preceding week).
3. Resurrected users (used the app at least once during the week after being dormant during at least the previous week).
4. Dormant users (did not use the app that week but did use the app at least once during the preceding week).

Proportion of Missing ROM Data

ROM use (yes or no) for each week of the 8-week follow-up period was defined as participant entry of *SMART Track* data for at least one outcome domain (7-day plan, ROM questionnaire, and Urges). This was used to calculate the weekly proportion of participants who entered the data.

Research and Development Costs

An economic costing analysis was conducted to assess the research and development (R&D) costs related to the creation of the *SMART Track* app. This included both the costs of developing the technology and the research costs (mainly time) that were integral to the development of the app, such as workshops to assess development and testing. Furthermore, an estimate of the total time spent in meetings across the R&D process was estimated, from steering group meetings to the conduct of focus groups. The number of hours were estimated to provide additional context of the time invested in the R&D process in developing a comprehensive and user-friendly app.

Acceptability Indicators

Nested Qualitative Evaluation

The qualitative analysis component of the study was undertaken through two processes: first, as a thematic study to provide insights into the acceptability of the app and the meetings more broadly, which was described in detail in a previous paper [63], and second, as part of the nested evaluation process [98], where the qualitative data were used to support app development and contribute to the experimental nature of the study. Unlike strictly triangulated studies, nested research studies use a combination of data to enrich insights and provide points of comparison to generate new hypotheses [98]. Research questions and interview topics informed the first more deductive coding frame. Categories were summarized and presented with pertinent quotes to the broader team for discussion, which continued until consensus was reached. The analysis sought to shed light on specific questions about the feasibility and acceptability of the app from the perspective of end users. We also explored accounts of the experience of submitting ROMs. Although thematic saturation was not the aim of this part of the analysis, we noticed recurring themes in 12 interviews.

Quality Ratings and Digital Therapeutic Alliance

The MARS–User Version (uMARS) [88] and DWAI [92] domain as well as overall mean scores were calculated at the 2-week follow-up.

Preliminary Outcomes

Paired sample two-tailed *t* tests were used to compare participant-reported outcomes on the SDS, K-10, and SURE between baseline and the 8-week follow-up.

Results

Sample Characteristics

A total of 72 participants were enrolled in this study (Figure 1). The participant characteristics were comparable with prior accounts of SRAU group characteristics [99]; the average age of the participants was 44 years (SD 11), with more men (44/72, 61%) than women (28/72, 39%; Table 3). Most of the participants were born in Australia (59/72, 81%), and 6% (4/72) reported being of Aboriginal, Torres Strait Islander, or both Aboriginal and Torres Strait Islander descent. Employment was the main source of income for almost half of the participants (35/72, 48%).

At baseline, the participants reported attending an average of 6.63 (SD 5.44) SMART Recovery meetings in the preceding 12 weeks (range 0-24). Excessive alcohol consumption was the most common primary behavior of concern, endorsed by 68% (49/72) of participants over the preceding 4 weeks. Injecting drug use (ever) was reported by 25% (18/72), and 11% (8/72) of the sample reported a recent arrest (past 3 months).

Table 3. Participant characteristics (n=72).

Variables	Values
Age (years), mean (SD)	44 (11)
Gender, n (%)	
Male	44 (61)
Female	28 (39)
Born in Australia, n (%)	59 (81)[a]
Aboriginal, Torres Strait, or both Aboriginal and Torres Strait Islander descent, n (%)	4 (6)
Primary source of income[b], n (%)	
Employment (full-time, part-time, or self-employed)	35 (48)
Temporary benefit (eg, unemployment)	10 (13)
Pension (eg, aged and disability)	13 (18)
Other (eg, retirement fund, savings, and investment)	7 (9)
No income or dependent on others	5 (6)
Highest completed level of education or training, n (%)	
High school or less	19 (26)
Certificate, diploma, or trade	26 (36)
Bachelor's degree	16 (22)
Postgraduate degree	11 (15)
Usual accommodation[b], n (%)	
Own home	33 (45)
Rented home (public or private)	33 (45)
Other (eg, friends, family, and rehab)	4 (5)
Marital status[a], n (%)	
Single or unmarried	27 (38)
Married or defacto	28 (39)
Separated	7 (9)
Divorced	8 (11)
Widow or widower	1 (1)
Ever received treatment for a mental health problem	54 (75)
Self-reported diagnosis received, n (%)	
Depression	9 (12)
Anxiety	8 (11)
Depression and anxiety	22 (30)
Other (eg, posttraumatic stress disorder, bipolar disorder, borderline personality disorder, and schizophrenia)	15 (20)
Addictive behavior causing the greatest concern, n (%)	
Alcohol	49 (68)
Amphetamines	7 (9)
Cannabis	6 (8)
Another drug (eg, cocaine, ecstasy, γ hydroxybutyrate, benzodiazepines)	5 (6)
Another behavior (eg, gambling and food)	5 (6)
Injecting drug use, n (%)	
Within the last 3 months	3 (4)

Variables	Values
More than three but less than 12 months ago	3 (4)
12 months ago or more	10 (13)
Never injected	54 (75)
Arrested in the last 3 months?[b], n (%)	8 (11)
Overdose (any drug) in the last 3 months[b], n (%)	1 (1)
Service use (preceding 3 months), n (%)	
Detoxification or withdrawal management	11 (15)
Residential rehabilitation	4 (5)
Alcohol or other clinic	13 (18)
Psychiatrist	13 (18)
General practitioner	47 (65)
Psychologist	30 (41)
Other allied health care provider (nurse, social worker, or counsellor)	17 (23)
SMART Recovery	69 (95)[c]
12-step	13 (18)
Source of referral to SMART Recovery, n (%)	
Self	23 (31)
Alcohol and/ or other clinic treatment service	17 (23)
Mental health care service	13 (18)
Legally recommended or mandated	9 (12)
Family member or friend	4 (5)
Other health care provider or service	6 (8)

[a]Missing data for 1 participant.

[b]Missing data for 2 participants.

[c]Three new SMART Recovery participants reported that their first meeting (scheduled for the week before baseline assessment) had been canceled. These participants were due to participate in their first group the week of the baseline assessment.

Feasibility Indicators

Enrollment and Engagement

In total, 216 people expressed interest in participating in the study. Of these 216, 97 (44.9%) were deemed eligible, and 72 (33.3%) went on to enroll. Of the 72 participants enrolled in the study, 68 (94%) created an account, 64 (88%) subsequently used *SMART Track* at least once, and 57 (79%) used *SMART Track* multiple times (mean 16.39, SD 16.10; range 2-83 visits). More than half of the participants (42/72, 58%) used *SMART Track* for ≥5 weeks across the study period (Table 4). *SMART Track* was accessed on 74 unique devices (ie, some participants used the app across multiple devices). Apple iPhone (n=33) and Samsung Galaxy (n=21) smartphones were the primary devices used.

Table 4. Frequency of SMART Track use expressed as the proportion of study participants per time interval across the 20-week study period (n=72).

	Participants, n (%)
Never	8 (11)
1 week	7 (10)
2-4 weeks	15 (21)
5-8 weeks	25 (35)
>8 weeks	17 (24)

Use Patterns

The number of participants using *SMART Track* each week gradually increased across the recruitment period (ie, until the week beginning September 16, 2019), with a gradual decline thereafter (Figure 3). In any one week, the maximum number of study participants using the app was 49% (35/72), and the number of visits to the app ranged from 2.47 to 5.27 (mean 3.39, SD 0.75; Figure 3).

The changes in the number of new, current, resurrected, and dormant users each week (Figure 4) suggest that the participants typically used *SMART Track* intermittently rather than on a regular (weekly) basis.

Figure 3. Total number of study participants using SMART Track each week and the corresponding average number of visits per user.

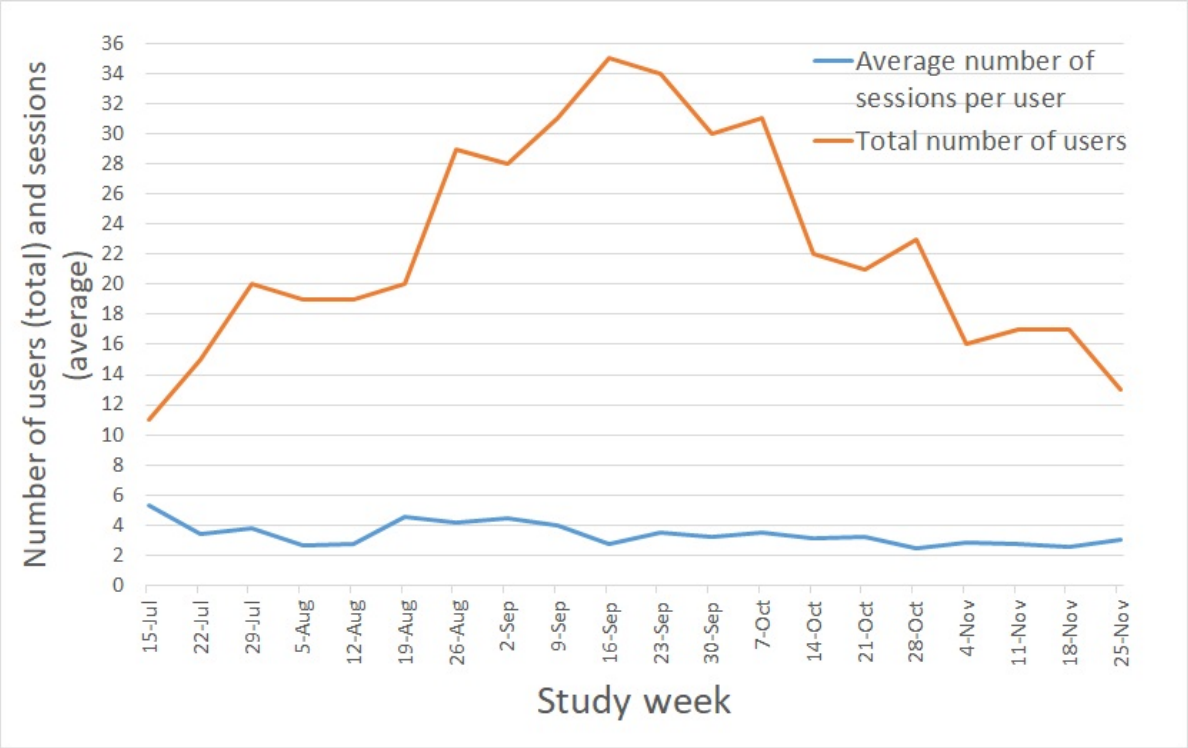
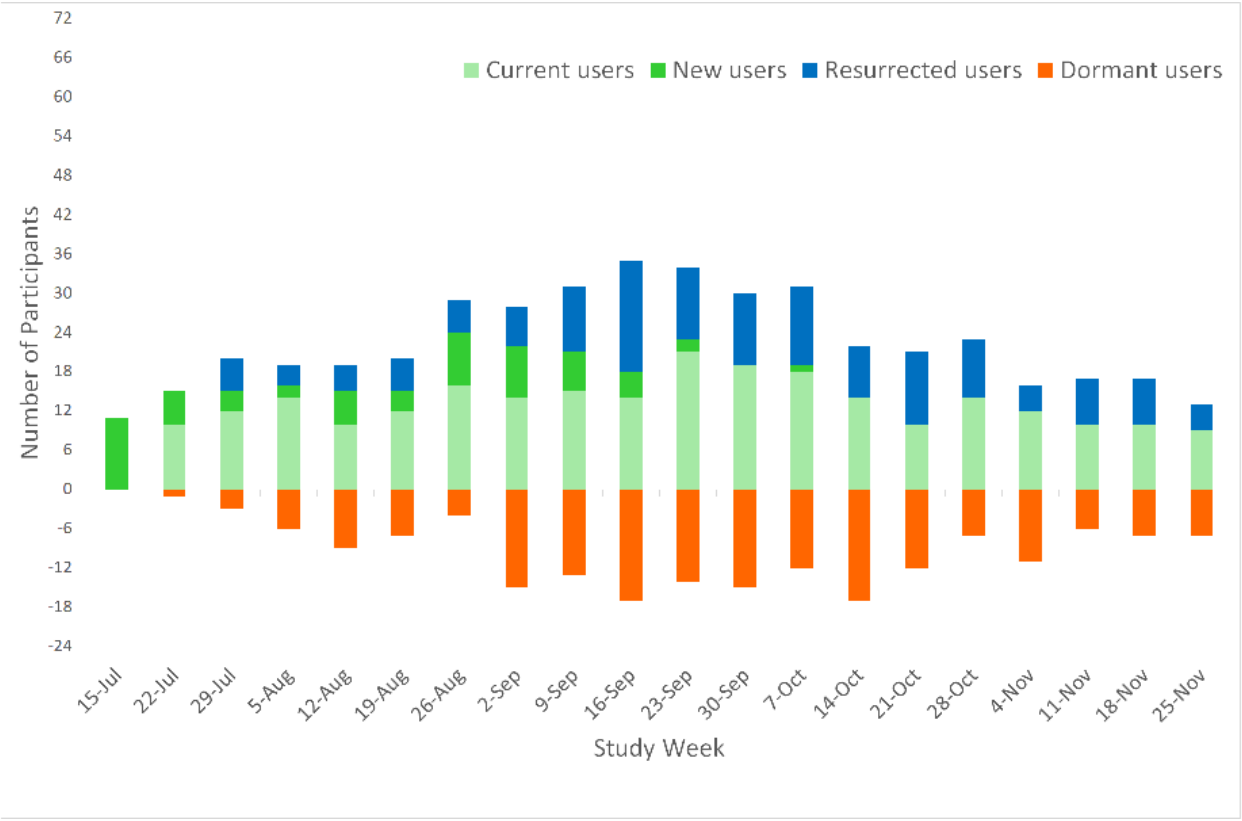


Figure 4. Incoming and outgoing users each week expressed as current versus new versus resurrected versus dormant users.



The two main *SMART Track* screens designed to provide feedback data (*Urges* screen and *Overall Progress* screen) were the most frequently visited sections of the app (Table 5). The participants spent the most time (minutes) using the *Me* screen and the least time viewing the *Overall Progress* screen (Table 5).

Table 5. Use of SMART Track features, expressed as the total number of visits to each of the main screens and the total time spent using each of the main screens.

	Total visits	Total duration (minutes)
Urge button	361	—[a]
Track urges	913	2468.4
Resources	587	443.62
Me	467	3915.94
Overall progress	789	321.33

[a]Duration is not provided for the urge button as use requires a single brief click and is therefore not captured.

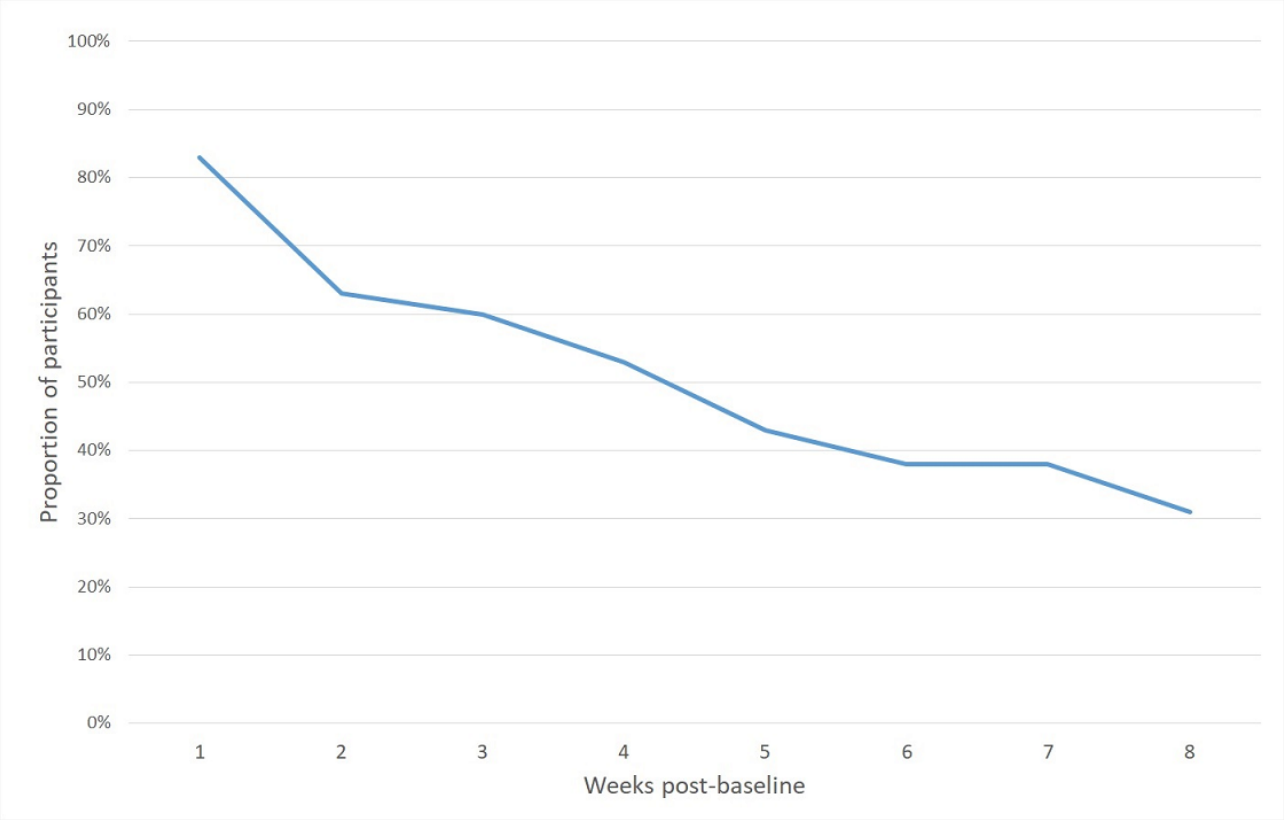
Considerably fewer visits were documented for each of the individual progress screens. These sections of the app were visited, on average, only once or twice per week across the duration of the study by a maximum of 8% (6/72) of the participants (Figure S2 of Multimedia Appendix 1).

The study participants most frequently used the app between 6 PM and 9 PM, with almost a quarter of all visits (500/2166, 23.08%) occurring during this time frame. In the morning, use was greatest between 9 AM and midday (Figure S3 of Multimedia Appendix 1).

Proportion of Missing ROM Data

During the first week of app use, 83% (60/72) of the participants had used *SMART Track* to enter data for at least one ROM instrument (7-day plan, ROM questionnaire, or urge log). There was a reduction across time in the number of participants completing the ROM items. At the end of 8 weeks, almost a third (22/72, 31%) of the participants had provided ROM data, reflecting a 50% reduction compared with week 1 (Figure 5).

Figure 5. Proportion of study participants using SMART Track to enter routine outcome monitoring data.



R&D Costs

To develop the *SMART Track* app, the developer (GHO) received Aus $76,500. However, the true cost to GHO was more than double (Aus $154,034) when the actual time invested by GHO staff (8 staff members; 876 hours) was fully accounted for (Table 4). Furthermore, the research costs to support the development of the app, such as workshops to assess feasibility and usability testing, were estimated at Aus $127,023 (Table 6). This also includes staff time spent by the trial coordinator

and qualitative researcher on development activities. These costs exclude the academic research and evaluation costs that were conducted alongside app development (eg, ethics, recruitment, and economic evaluation; Table 7).

In total, the R&D cost incurred was Aus $203,523. If the true costs to GHO Sydney were included, then the total R&D cost would have been Aus $281,058. Finally, an estimate of the total number of hours invested in the R&D process was estimated at 1485 hours (Table 7). More than three-quarters were in-kind costs and goodwill.

Table 6. Cost of SMART Track development and research costs to support development.

Variable	Hours	Rate	Cost (Aus $)
Development costs			
App developer	452	160	72,320
Account director	48	180	8640
Project manager	73	150	10,875
User experience designer	22	180	3870
User interface designer	209	180	37,530
Strategy director	16	250	4000
Creative director	48	350	16,800
Total	868	1450	154,035
Cost of research to support SMART Track development			
Trial coordinator	983	69	67,813
SMART Recovery technology lead	224	56	12,617
Facilitator support for app	156	58	8992
Qualitative researcher	208	144	29,880
Transcription (20 interviews, focus group)	N/A[a]	N/A	1759
Administrative support	156	38	5962
Total	N/A	N/A	127,023

[a]N/A: not applicable.

Table 7. Total hours spent to support the research and development processa.

	Meetings	Duration	People	Person hours
Expert advisory committee	14	45	10	105
Steering committee	2	30	8	10.7
Trial coordinator and external steering committee members	6	30	7	42
Original development company	4	60	7	56
Interviewing new developers	5	5	3	1.3
GHO: preliminary workshops	3	120	15	1080
GHO meetings: design and development	14	60	8	56
GHO: usability testing sessions	9	30	5	22.5
Qualitative researcher meetings	4	60	6	48
SMART Recovery facilitators: focus groups	8	60	8	64
Total	69	500	77	1485.4

[a]Over three-quarters of time was *in-kind*.

Acceptability Indicators

Qualitative Findings

In total, 28 in-depth qualitative interviews were conducted with 20 app users (group members) and 8 facilitators. The participants tended to perceive and describe their use of *SMART Track* within their broader experiences and competencies related to information and communication technology. Of the 20 app users interviewed, the level of knowledge and prior experience of mHealth apps and other digital resources varied widely and were not related to their age or education level. Moreover, the participants' prior knowledge and experience of apps did not seem to be connected to their use of *SMART Track*. For example, those who demonstrated *high* use of *SMART Track* often described themselves as *new* to apps. Participants with "lots of experience" with apps often described ceasing use after a short time (Alec, group member, low app use).

Compared with another sample of SRAU participants who provided input to inform the development of *SMART Track* [63], the participants in this study expressed less concern about digital support tools replacing face-to-face meetings. These group members tended to perceive the app as complementary to their mutual aid group and described using the resources within it to "stay on track between meetings" (Jasmine, group member, high app use). For these participants, "logging urges and tracking progress" were more desirable than completing routine outcome measures (Campbell, group member, low app use), except when the outcomes data were available to them in "more detail" (Harold, group member, high app use). It is therefore possible that increasing the usability of outcome data tracking activities would increase app users' engagement with outcome measurements. Consistent with other research, end users seem to be more open to completing repeated and routine outcome measures when their understanding of their outcomes is aligned with the outcome measures selected by the program designers [30,74].

The group facilitator seemed to play a key role in implementing the app and collecting routine outcome data. Facilitators who were knowledgeable about the app and purposefully integrated its use in meetings were more likely to report higher app use among the participants. This is consistent with feedback received from client-participants, who described the facilitators' efforts, or lack of efforts, when "inspiring" group members to use the app (Mitchell, group member, high app use). Given the lack of experience that some facilitators have with app use and other information and communication technology, it is possible that basic training would have improved their engagement with implementation. In summary, the posttrial interview findings suggest that *SMART Track* is an engaging platform for collecting routine outcome data, and participant concerns expressed at the pretrial time point were not described after the trial.

Quality Ratings

The uMARS ratings (Table 8) confirmed the acceptability of *SMART Track*. The overall app quality score was *good*, and every domain was rated as either acceptable or good (ie, uMARS rating >3).

Table 8. Quality assessment as indexed by participant responses to the uMARS[a] and DWAI[b].

Variable	Values, mean (SD)	Value, median (range)
uMARS[c]		
Engagement	3.6 (0.5)	3.6 (2.0-5.0)
Functionality	4.1 (0.7)	4.0 (2.5-5.0)
Aesthetics	4.2 (0.6)	4.3 (3.0-5.0)
Information	4.3 (0.5)	4.5 (2.5-5.0)
Overall quality	4.0 (0.5)	4.1 (3.0-4.9)
Subjective quality	3.8 (0.8)	4.0 (1.0-5.0)
Perceived impact	3.7 (0.9)	3.7 (1.7-5.0)
DWAI[d]		
Goals	3.5 (1.0)	3.5 (1.0-5.0)
Tasks	3.7 (1.0)	3.5 (2.0-5.0)
Bond	3.4 (1.1)	3.5 (1.5-5.0)
Overall	3.5 (0.9)	3.7 (1.7-5.0)

[a]uMARS: Mobile App Rating Scale–User Version.

[b]DWAI: Digital Working Alliance Inventory.

[c]All items are rated on a 5-point scale from 1 (inadequate) to 5 (excellent).

[d]All items are rated on a 5-point scale from 1 (seldom) to 5 (always).

Digital Therapeutic Alliance

The DWAI ratings (Table 8) also support the acceptability of the app with the domain scores indicating that, on average, the participants rated the key elements of therapeutic alliance (goals, tasks, and bonds) as being present between *fairly* often and *very* often.

Preliminary Evidence on Outcomes

There was a significant reduction between baseline and 8-week follow-up for the SDS (mean difference 1.93, SD 3.02; 95% CI 1.12 to 2.73) and K-10 scores (mean difference 3.96, SD 8.31; 95% CI 1.75 to 6.17), but there was no change in the SURE scores (mean difference 0.11, SD 7.97; 95% CI –2.02 to 2.24).

ROM Reliability

There was a strong relationship between the clinician-administered SURE [74] (at baseline) and the app-administered SURE (week 1; r=0.89; P<.001). For the quality-of-life item, there was a moderate relationship between baseline clinician administration as part of the EUROHIS-QOL 8-item index [76] and as part of the app-administered measures 2 weeks later (r=0.61; P=.005). For the Kessler Psychological Distress Scale-6 (K-6) [75], there was a moderate relationship between baseline clinician administration as part of the K-10 [97] and app administration as part of the K-6 2 weeks later (r=0.51; P=.02). The internal consistency (Cronbach α) for the SURE and K-6 was high across time points and when collected through clinician-interviewed telephone assessments or within the app (SURE: .86 to .94; K-6: .86 to .90).

Discussion

Principal Findings

This study was designed to assess the feasibility, acceptability, and preliminary outcomes of *SMART Track* for ROM and feedback in SRAU. The qualitative and quantitative findings support the feasibility, acceptability, and utility of *SMART Track* for ROM and feedback in SRAU. The findings also provide insight into avenues for enhancing sustained engagement. SMART Recovery participants were willing to use *SMART Track*, demonstrated sustained use across the 8-week follow-up interval, engaged most with the two main progress screens (*Urges* and *Overall Progress*), and experienced *SMART Track* as useful and consistent with SMART Recovery principles and strategies. Although it is difficult to attribute it directly to the use of *SMART Track* or SMART Recovery, the participants also showed clinical improvement over the 8-week follow-up, specifically reductions in the severity of dependence and psychological distress. Varied rates of ROM completion, minimal use of domain-specific feedback screens, and qualitative feedback suggest that the utility of *SMART Track* would be improved by making minor changes to app functionality and improving attention to implementation strategies.

Engagement With SMART Track

To put engagement with *SMART Track* in perspective, it is helpful to consider the rates of engagement with other mHealth apps. One of the challenges with such comparisons is the considerable variation in the metric used to capture mHealth use (eg, mean number of log-ins, sessions, modules, activities completed [100]) and the degree to which these variables are reported [101,102]. Although several systematic reviews of digital recovery support services [39], digital measurement feedback systems [32], and addiction-related mHealth apps are available [54,103-105], the focus tends to be on content, experience, or outcomes, with little to no data examining participant engagement or use. However, the use of mHealth apps by people in recovery from substance use has been shown to vary from as high as 90% in the first few weeks to as low as 18% after 6 weeks [105]. For people with mental health conditions, engagement varied according to the target mental health condition. The number of *nonusers* (individuals who fail to download or use the intervention) has been calculated as 41% (range 25%-58%) for apps targeting depression and 8% (range 0%-16%) for apps targeting anxiety [100]. Reduced engagement over time was common [100]. Compared with these data, participant engagement with, and sustained use of, *SMART Track* is at least comparable, if not higher than the available evidence.

Another useful point of comparison comes from use trends within the global app marketplace. In 2019, data derived from more than 12,000 apps demonstrated that a quarter of the users will abandon an app after one-time use [106]. In comparison, of the 94% (68/72) of the participants in this study who created a *SMART Track* account, more than one occasion of use was documented for 79% (57/72) of the participants. Benchmarks pertaining specifically to lifestyle-related apps (which include fitness-, health-, and travel-related apps) indicate that the average 2-month retention rate is 36% [107]. In comparison, more than half of the participants in this study used *SMART Track* for between 5 and 8 weeks (25/72, 35%) or longer (17/72, 23%), and week 8 ROM data were provided by 31% (22/72) of the study participants.

Uncertainty exists around what is considered a *good* level of mHealth use. Some addiction-related apps (eg, *In My First Year of Recovery* and *A-CHESS* mHealth interventions) have documented high levels of sustained participant engagement (78% program completion and 4-month retention, respectively [105]). In contrast, *SMART Track* use was intermittent, with a proportion of the participants using the app weekly, whereas others disengaged and re-engaged every few weeks. Emerging evidence suggests that engagement with digital recovery support tools may be influenced by recovery duration [39]. Accordingly, engagement with *SMART Track* is likely to vary widely, given that SRAU caters to people across the spectrum of recovery experiences. It is also possible that users may perceive apps as a *short-term commitment* [108]. Therefore, compared with the use of other digital platforms (eg, the web), app use may be shorter and more sporadic [109]. Moreover, training in *SMART Track* was not extensive. The researchers met with the facilitators at each site to orient them to the features of the app. The participants received an email with brief instructions and an in-app onboarding process, although analytics showed that this was used by less than 45.2% (105/232) of users who downloaded *SMART Track* during the study period. The exemplary use rates documented by *In My First Year of*

Recovery and *A-CHESS* may have been due to the more active role of counselors in encouraging mHealth use [105] and point to the importance of improved integration of *SMART Track* into SMART Recovery groups. Identifying participant and contextual factors that influence engagement represents an important challenge for future research.

Routine Outcome Monitoring

A few studies have examined ROM implementation report data on engagement and attrition rates [8,25,28,110,111]. Compared with recent data from the Netherlands, however, the proportion of participants using *SMART Track* to input ROM data is largely comparable, and in some cases, greater than traditional clinician-completed methods [112]. A 50% response rate has been recently suggested as an acceptable benchmark for ROM data and is likely sufficient to protect against bias and yield valid information about patient progress (see study by de Beurs et al [112] for a discussion). In this study, this benchmark was achieved each week across the first month of data collection, with 83% (60/72), 63% (45/72), 60% (43/72), and 53% (38/72) of the study participants completing at least one of the ROM instruments across the first 4 weeks of app use. ROM completion continued to decline during the second month of data collection, with 31% (22/72) of the sample providing ROM data during week 8. Ongoing efforts are needed to improve ROM completion and understand the participant characteristics associated with drop-off (eg, through attrition analyses). For example, given the voluntary, open-enrolling format of SMART Recovery groups and individual variation in group attendance (0-24 in this study), it would be interesting to examine whether app use varies according to group engagement.

Cost Analysis

The R&D costs of *SMART Track* were estimated at Aus $203,523. However, the developer offered considerable in-kind support, and the overall true cost was Aus $281,058. It is important to reiterate that these are essentially *sunk costs*. If *SMART Track* had been rolled out routinely in SMART Recovery groups, the cost of implementation would have consisted of costs related to app hosting, infrastructure, maintenance, and training. The dynamic nature of the industry makes it difficult to estimate these costs and points to the need for ongoing data collection to understand the long-term real-world feasibility of *SMART Track*.

Opportunities

To maximize participant engagement in ROM, a clear rationale for why the data are being collected and what they will be used for is essential [25]. The measures must be experienced as *relevant* [110] and the process deemed *worthwhile* [113]. Analytics revealed that few people accessed the detailed, personalized feedback provided within each of the domains listed on the summary page. Qualitative feedback highlighted a mismatch between the effort expended and satisfaction with the feedback provided. Given that the participants were asking for feedback that was already provided in the app (but not accessed), we expect that modifications to improve the visibility of these sections of the app will further enhance ROM completion. Importantly, the regular and frequent use of app self-monitoring features has been linked to a longer period of use and reduced likelihood of abandoning apps [114]. Additional features that have been linked to mHealth engagement include an esthetically pleasing interface, ease of use, degree of personalization, reinforcement (eg, rewards and reminders), communication (with peers or professionals), message presentation (including language, tone, and design), and credibility (encompassing trustworthiness and confidentiality [115]). Although these features were considered throughout the design of *SMART Track* (and positively evaluated as part of the quality assessment), an opportunity exists to further enhance the user experience (eg, through improved personalization, greater use of rewards, and the addition of information-sharing or communication capabilities).

Organizational resources are essential for ensuring the sustained implementation of digital interventions to reduce substance use [105], for example, ensuring leadership support; providing adequate training and resources to both staff and service users; leveraging the expertise of service users to contribute to training; and having a process in place to monitor, evaluate, provide feedback, and respond to uptake rates [116]. Training is a particularly important consideration and provides a forum to (1) build clinician knowledge and confidence in app use and features, (2) practice introducing it to participants, and (3) identify and overcome any concerns or perceived barriers to implementation of mHealth to support routine care [117]. The next steps for *SMART Track* include improved facilitator training and support and leveraging of participant and facilitator *champions* (ie, individuals who actively support the use of *SMART Track*).

Strengths

The development of *SMART Track* was grounded in theory [40,57] and user-centered design [57]. Consistent with recommendations for enhancing measurement-based care, *SMART Track* includes both standardized and idiographic outcome assessment and harnesses technology to overcome traditional barriers to ROM (eg, scoring and providing tailored feedback [110]). Quality assessment was conducted using a psychometrically valid tool [88], and *SMART Track* surpassed the minimum acceptable quality benchmark (≥3 [88]) on each of the domains assessed by the uMARS, with an overall quality rating of *good*. This is superior to several published accounts [50]. The observed quality of *SMART Track* likely reflects the user-centric approach to development. However, because the ratings were collected as part of a telephone interview with the researcher, the contribution of response bias cannot be ruled out.

Evidence from mental health settings suggests that mHealth apps that have a clear purpose and simple user interface and are easy and time efficient to navigate and demonstrate were more likely to be used as part of routine practice [117]. Quantitative and qualitative data indicate that *SMART Track* possesses these attributes. Pending minor upgrades and improved training and support, this finding further bolsters our confidence in the routine uptake of *SMART Track* within SMART Recovery groups.

Limitations

This study includes several limitations. In its current format, *SMART Track* is not suitable for people who cannot adequately read and comprehend English. The reliance on written and visual information may also compromise the suitability of *SMART Track* for people with vision impairment. The use of cloud functions for collecting and storing data means that *SMART Track* needs a reliable internet connection to function. The study's approach to assessing engagement is consistent with recommendations for a multidimensional approach using mobile app data analytics (index of *microengagement*), indices of behavior change (macro level of engagement), and participant subjective experience [102]. However, differences in how various studies define engagement and use make it hard to position *SMART Track* within the context of existing studies. Recent guidelines for the measurement and reporting of engagement data in digital interventions may be beneficial in the future [118].

The current findings are derived from a small sample of participants who attended a limited sample of SRAU groups. We did not collect data on the number of participants attending SRAU groups across the study period; therefore, although participant characteristics are comparable with published accounts [99], generalizability is unclear. Furthermore, the short-term nature of the study makes it challenging to position the findings within the often long-term, nonlinear experience of recovery [16]. Finally, because this is a stage 1 feasibility study, our finding that participants reported reduced severity of dependence and psychological distress from baseline to 8 weeks needs to be interpreted cautiously.

Conclusions

The qualitative and quantitative findings support the feasibility and acceptability of *SMART Track* and lend insight into avenues for enhancing sustained engagement. Low rates of engagement and high rates of attrition are known challenges for services working with participants who experience substance use and mental health–related difficulties [119]. Sustained engagement with mHealth apps is notoriously difficult to achieve. In light of these challenges, our findings are promising. *SMART Track* offers a potential solution for ROM and feedback, particularly for people with substance use disorders who attend mutual support groups. Future research should focus on identifying relevant demographic, clinical, and contextual factors that may influence the engagement, attrition, and perceived utility of this promising mHealth app.

Acknowledgments

The authors gratefully acknowledge the creative and technical expertise of the customer experience agency (GHO Sydney). Ryan Chao (executive creative director) provided overall creative direction and led the user experience design process. James Legge (executive strategy director) led the strategy and facilitated the workshops. Marcos Martini (technical lead) was responsible for the overall development of the app across the iOS and Android platforms, Sharon Peng (user experience or user interface designer) designed the user experience and user interface and conducted usability tests, and Phoebe Calcutt (project manager) managed the overall delivery of the app. The authors also gratefully acknowledge the time of, and expert insights from, the members of our steering committee and the valuable support and contributions made by SMART Recovery participants and facilitators to the development of *SMART Track* and the conduct of this research.

Funding for this research was provided by the New South Wales Ministry of Health under the New South Wales Health Alcohol and Other Drugs Early Intervention Innovation Grant Scheme. The funding body did not directly contribute to the design, conduct, analysis, write-up, and submission of this research for publication and does not have the ultimate authority over any of these activities.

Conflicts of Interest

RM is the executive director of SRAU. AA is employed by SMART Recovery as the National Program Manager and Trainer. PJK, FPD, ALB, AS, LH, VM, BL, AKB, JFK, and AA all volunteer as members of the SRAU Research Advisory Committee. The potential perceived conflicts of interest are negligible. The role of study investigators on the research advisory committee or as an employee of SMART Recovery is freely available on the SRAU website (and study participants can be directed to this information as required). Furthermore, the team responsible for informing the study design and overseeing the conduct of the study and data analysis also consists of researchers, clinicians, and statisticians independent of SMART Recovery.

Multimedia Appendix 1
Supplementary figures.
[DOCX File , 142 KB - jmir_v23i10e25217_app1.docx]

Multimedia Appendix 2
Supplementary tables.
[DOCX File , 65 KB - jmir_v23i10e25217_app2.docx]

References

1. National strategic framework for chronic conditions. In: Australian Health Ministers' Advisory Council. Canberra: Australian Government; 2017.
2. National Collaborating Centre for Mental Health (UK). Alcohol-use disorders: diagnosis, assessment and management of harmful drinking (high-risk drinking) and alcohol dependence. In: National Institute for Health and Clinical Excellence: Guidance. Leicester (UK): British Psychological Society; 2011:1-610.
3. National Collaborating Centre for Mental Health (UK). Drug Misuse: Psychosocial interventions. In: National Institute for Health and Clinical Excellence: Guidance. Leicester (UK): British Psychological Society; 2008:1-338.
4. Boswell JF, Constantino MJ, Kraus DR, Bugatti M, Oswald JM. The expanding relevance of routinely collected outcome data for mental health care decision making. Adm Policy Ment Health 2016 Jul 11;43(4):482-491. [doi: 10.1007/s10488-015-0649-6] [Medline: 25861984]
5. Kelly J, Mee-Lee D. Chapter 15 - Quality, accountability, and effectiveness in addiction treatment: The measurement-based practice modela. In: Danovitch I, Mooney L, editors. The Assessment and Treatment of Addiction. Amsterdam: Elsevier; 2019:207-217.
6. Lambert MJ, Harmon C, Slade K, Whipple JL, Hawkins EJ. Providing feedback to psychotherapists on their patients' progress: clinical results and practice suggestions. J Clin Psychol 2005 Feb;61(2):165-174. [doi: 10.1002/jclp.20113] [Medline: 15609358]
7. Carlier IV, Eeden WA. Routine outcome monitoring in mental health care and particularly in addiction treatment: evidence-based clinical and research recommendations. J Addict Res Ther 2017;08(04):1000332. [doi: 10.4172/2155-6105.1000332]
8. Lambert MJ, Whipple JL, Kleinstäuber M. Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring. Psychotherapy (Chic) 2018 Dec;55(4):520-537. [doi: 10.1037/pst0000167] [Medline: 30335463]
9. Fortney JC, Unützer J, Wrenn G, Pyne JM, Smith GR, Schoenbaum M, et al. A tipping point for measurement-based care. Psychiatr Serv 2017 Feb 01;68(2):179-188. [doi: 10.1176/appi.ps.201500439] [Medline: 27582237]
10. Kendrick T, El-Gohary M, Stuart B, Gilbody S, Churchill R, Aiken L, et al. Routine use of patient reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. Cochrane Database Syst Rev 2016 Jul 13;7:CD011119 [FREE Full text] [doi: 10.1002/14651858.CD011119.pub2] [Medline: 27409972]
11. Shimokawa K, Lambert MJ, Smart DW. Enhancing treatment outcome of patients at risk of treatment failure: meta-analytic and mega-analytic review of a psychotherapy quality assurance system. J Consult Clin Psychol 2010 Jun;78(3):298-311. [doi: 10.1037/a0019247] [Medline: 20515206]
12. Goodman JD, McKay JR, DePhilippis D. Progress monitoring in mental health and addiction treatment: A means of improving care. Prof Psychol Res Pract 2013 Aug;44(4):231-246. [doi: 10.1037/a0032605]
13. Thompson C, Sansoni J, Morris D, Cappell J, Williams K. Patient reported outcome measures: An environmental scan of the Australian health care sector. Sydney, NSW: Australian Commission on Safety and Quality in Health Care; 2016:1-89.
14. Kelly PJ, Robinson LD, Baker AL, Deane FP, McKetin R, Hudson S, et al. Polysubstance use in treatment seekers who inject amphetamine: Drug use profiles, injecting practices and quality of life. Addict Behav 2017 Aug;71:25-30. [doi: 10.1016/j.addbeh.2017.02.006] [Medline: 28242532]
15. Kelly P, Deane F, Baker A, Keane C. Client Outcomes Management System (COMS): Data Report 2013. Wollongong, Australia: Illawarra Institute for Mental Health; 2014.
16. Kelly JF, Bergman B, Hoeppner BB, Vilsaint C, White WL. Prevalence and pathways of recovery from drug and alcohol problems in the United States population: Implications for practice, research, and policy. Drug Alcohol Depend 2017 Dec 01;181:162-169 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.09.028] [Medline: 29055821]
17. Manning V, Garfield JB, Best D, Berends L, Room R, Mugavin J, et al. Substance use outcomes following treatment: Findings from the Australian Patient Pathways Study. Aust N Z J Psychiatry 2017 Feb 11;51(2):177-189. [doi: 10.1177/0004867415625815] [Medline: 26769978]
18. Beck AK, Forbes E, Baker AL, Kelly PJ, Deane FP, Shakeshaft A, et al. Systematic review of SMART Recovery: Outcomes, process variables, and implications for research. Psychol Addict Behav 2017 Feb;31(1):1-20. [doi: 10.1037/adb0000237] [Medline: 28165272]
19. Kelly J, Humphreys K, Ferri M. Alcoholics Anonymous and other 12-step programs for alcohol use disorder. Cochrane Database Syst Rev 2020 Mar 11;3(3):CD012880 [FREE Full text] [doi: 10.1002/14651858.CD012880.pub2] [Medline: 32159228]
20. Hennessy EA. Recovery capital: a systematic review of the literature. Addict Res Theory 2017 Mar 08;25(5):349-360. [doi: 10.1080/16066359.2017.1297990]
21. Martinelli TF, van de Mheen D, Best D, Vanderplasschen W, Nagelhout GE. Are members of mutual aid groups better equipped for addiction recovery? European cross-sectional study into recovery capital, social networks, and commitment to sobriety. Drugs Edu Prevent Policy 2020 Nov 09:1-10. [doi: 10.1080/09687637.2020.1844638]
22. Boswell JF, Kraus DR, Miller SD, Lambert MJ. Implementing routine outcome monitoring in clinical practice: benefits, challenges, and solutions. Psychother Res 2015 Jul 26;25(1):6-19. [doi: 10.1080/10503307.2013.817696] [Medline: 23885809]

23. Jensen-Doss A, Haimes EM, Smith AM, Lyon AR, Lewis CC, Stanick CF, et al. Monitoring treatment progress and providing feedback is viewed favorably but rarely used in practice. Adm Policy Ment Health 2018 Jan 8;45(1):48-61 [FREE Full text] [doi: 10.1007/s10488-016-0763-0] [Medline: 27631610]
24. Duncan EA, Murray J. The barriers and facilitators to routine outcome measurement by allied health professionals in practice: a systematic review. BMC Health Serv Res 2012 May 22;12(1):96 [FREE Full text] [doi: 10.1186/1472-6963-12-96] [Medline: 22506982]
25. Solstad SM, Castonguay LG, Moltu C. Patients' experiences with routine outcome monitoring and clinical feedback systems: A systematic review and synthesis of qualitative empirical literature. Psychother Res 2019 Feb 19;29(2):157-170. [doi: 10.1080/10503307.2017.1326645] [Medline: 28523962]
26. Moltu C, Veseth M, Stefansen J, Nøtnes JC, Skjølberg Å, Binder P, et al. This is what I need a clinical feedback system to do for me: A qualitative inquiry into therapists' and patients' perspectives. Psychother Res 2018 Mar 24;28(2):250-263. [doi: 10.1080/10503307.2016.1189619] [Medline: 27219820]
27. Burgess P, Pirkis J, Coombs T. Routine outcome measurement in Australia. Int Rev Psychiatry 2015 Mar 13;27(4):264-275. [doi: 10.3109/09540261.2014.977234] [Medline: 25768326]
28. Gondek D, Edbrooke-Childs J, Fink E, Deighton J, Wolpert M. Feedback from outcome measures and treatment effectiveness, treatment efficiency, and collaborative practice: a systematic review. Adm Policy Ment Health 2016 May 7;43(3):325-343 [FREE Full text] [doi: 10.1007/s10488-015-0710-5] [Medline: 26744316]
29. Davidson K, Perry A, Bell L. Would continuous feedback of patient's clinical outcomes to practitioners improve NHS psychological therapy services? Critical analysis and assessment of quality of existing studies. Psychol Psychother 2015 Mar 21;88(1):21-37. [doi: 10.1111/papt.12032] [Medline: 24850588]
30. NSW Agency for Clinical Innovation. Formative Evaluation of Patient Reported Measures. Chatswood: ACI; 2018.
31. Carlier I, Meuldijk D, Van Vliet IM, Van Fenema E, Van der Wee NJ, Zitman F. Routine outcome monitoring and feedback on physical or mental health status: evidence and theory. J Eval Clin Pract 2012 Feb;18(1):104-110. [doi: 10.1111/j.1365-2753.2010.01543.x] [Medline: 20846319]
32. Lyon AR, Lewis CC, Boyd MR, Hendrix E, Liu F. Capabilities and characteristics of digital measurement feedback systems: results from a comprehensive review. Adm Policy Ment Health 2016 May 10;43(3):441-466 [FREE Full text] [doi: 10.1007/s10488-016-0719-4] [Medline: 26860952]
33. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. Br Med J 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
34. Boudreaux ED, Waring ME, Hayes RB, Sadasivam RS, Mullen S, Pagoto S. Evaluating and selecting mobile health apps: strategies for healthcare providers and healthcare organizations. Transl Behav Med 2014 Dec 24;4(4):363-371 [FREE Full text] [doi: 10.1007/s13142-014-0293-9] [Medline: 25584085]
35. Han M, Lee E. Effectiveness of mobile health application use to improve health behavior changes: a systematic review of randomized controlled trials. Healthc Inform Res 2018 Jul;24(3):207-226 [FREE Full text] [doi: 10.4258/hir.2018.24.3.207] [Medline: 30109154]
36. Milward J, Day E, Wadsworth E, Strang J, Lynskey M. Mobile phone ownership, usage and readiness to use by patients in drug treatment. Drug Alcohol Depend 2015 Jan 01;146:111-115. [doi: 10.1016/j.drugalcdep.2014.11.001] [Medline: 25468818]
37. Drumm J, White N, Swiegers M, Davey M. Mobile Consumer Survey 2017: The Australian cut - Smart everything, everywhere. Deloitte, Sydney. 2017. URL: https://www2.deloitte.com/content/dam/Deloitte/au/Documents/technology-media-telecommunications/deloitte-au-tmt-mobile-consumer-survey-2017-211117-final.pdf [accessed 2021-08-08]
38. Garnett C, Crane D, West R, Brown J, Michie S. Identification of behavior change techniques and engagement strategies to design a smartphone app to reduce alcohol consumption using a formal consensus method. JMIR Mhealth Uhealth 2015 Jun 29;3(2):e73 [FREE Full text] [doi: 10.2196/mhealth.3895] [Medline: 26123578]
39. Ashford RD, Bergman BG, Kelly JF, Curtis B. Systematic review: Digital recovery support services used to support substance use disorder recovery. Human Behav and Emerg Tech 2019 Apr 30;2(1):18-32. [doi: 10.1002/hbe2.148]
40. Jusoh S. A survey on trend, opportunities and challenges of mHealth apps. Int J Interact Mob Technol 2017 Nov 27;11(6):73. [doi: 10.3991/ijim.v11i6.7265]
41. Sama PR, Eapen ZJ, Weinfurt KP, Shah BR, Schulman KA. An evaluation of mobile health application tools. JMIR Mhealth Uhealth 2014 May 01;2(2):e19 [FREE Full text] [doi: 10.2196/mhealth.3088] [Medline: 25099179]
42. Mendiola MF, Kalnicki M, Lindenauer S. Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings. JMIR Mhealth Uhealth 2015 May 13;3(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4283] [Medline: 25972309]
43. Kao C, Liebovitz DM. Consumer Mobile Health Apps: Current state, barriers, and future directions. Phys Med Rehab 2017 May 18;9(5S):106-115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]
44. Carreiro S, Newcomb M, Leach R, Ostrowski S, Boudreaux ED, Amante D. Current reporting of usability and impact of mHealth interventions for substance use disorder: A systematic review. Drug Alcohol Depend 2020 Oct 01;215:108201. [doi: 10.1016/j.drugalcdep.2020.108201] [Medline: 32777691]

45. Staiger PK, O'Donnell R, Liknaitzky P, Bush R, Milward J. Mobile apps to reduce tobacco, alcohol, and illicit drug use: systematic review of the first decade. J Med Internet Res 2020 Nov 24;22(11):e17156 [FREE Full text] [doi: 10.2196/17156] [Medline: 33231555]
46. Crane D, Garnett C, Brown J, West R, Michie S. Behavior change techniques in popular alcohol reduction apps: content analysis. J Med Internet Res 2015 May 14;17(5):e118 [FREE Full text] [doi: 10.2196/jmir.4060] [Medline: 25977135]
47. Hoeppner BB, Schick MR, Kelly LM, Hoeppner SS, Bergman B, Kelly JF. There is an app for that - Or is there? A content analysis of publicly available smartphone apps for managing alcohol use. J Subst Abuse Treat 2017 Nov;82:67-73. [doi: 10.1016/j.jsat.2017.09.006] [Medline: 29021117]
48. Ramo DE, Popova L, Grana R, Zhao S, Chavez K. Cannabis Mobile Apps: A content analysis. JMIR Mhealth Uhealth 2015 Aug 12;3(3):e81 [FREE Full text] [doi: 10.2196/mhealth.4405] [Medline: 26268634]
49. Savic M, Best D, Rodda S, Lubman DI. Exploring the focus and experiences of smartphone applications for addiction recovery. J Addict Dis 2013 Sep 27;32(3):310-319. [doi: 10.1080/10550887.2013.824331] [Medline: 24074196]
50. Chapman C, Champion KE, Birrell L, Deen H, Brierley M, Stapinski LA, et al. Smartphone Apps About Crystal Methamphetamine ("Ice"): Systematic search in app stores and assessment of composition and quality. JMIR Mhealth Uhealth 2018 Nov 21;6(11):e10442 [FREE Full text] [doi: 10.2196/10442] [Medline: 30463834]
51. Ridley K, Wiltshire A, Coleman M. Win Big Fast! An evaluation of mobile applications available in Australia for problem gambling. J Gambl Issues 2020 Sep 04;45:6. [doi: 10.4309/jgi.2020.45.6]
52. Tofighi B, Chemi C, Ruiz-Valcarcel J, Hein P, Hu L. Smartphone apps targeting alcohol and illicit substance use: systematic search in in commercial app stores and critical content analysis. JMIR Mhealth Uhealth 2019 Apr 22;7(4):e11831 [FREE Full text] [doi: 10.2196/11831] [Medline: 31008713]
53. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: A systematic review. J Telemed Telecare 2016 Oct 18;24(1):22-30. [doi: 10.1177/1357633x16673538]
54. Petry N, Meredith S, Alessi S. Smartphone applications to reduce alcohol consumption and help patients with alcohol use disorder: a state-of-the-art review. Adv Health Care Tech 2015 Dec:47. [doi: 10.2147/ahct.s65791]
55. Singh K, Drouin K, Newmark LP, Filkins M, Silvers E, Bain PA, et al. Patient-facing mobile apps to treat high-need, high-cost populations: a scoping review. JMIR Mhealth Uhealth 2016 Dec 19;4(4):e136 [FREE Full text] [doi: 10.2196/mhealth.6445] [Medline: 27993761]
56. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
57. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
58. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for managing health: systematic review and content analysis of findings. J Med Internet Res 2018 Jul 12;20(7):e10480 [FREE Full text] [doi: 10.2196/10480] [Medline: 30001997]
59. Kelly PJ, Beck AK, Baker AL, Deane FP, Hides L, Manning V, et al. Feasibility of a mobile health app for routine outcome monitoring and feedback in mutual support groups coordinated by SMART recovery Australia: protocol for a pilot study. JMIR Res Protoc 2020 Jul 09;9(7):e15113 [FREE Full text] [doi: 10.2196/15113] [Medline: 32673272]
60. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
61. Kelly PJ, Raftery D, Deane FP, Baker AL, Hunt D, Shakeshaft A. From both sides: Participant and facilitator perceptions of SMART Recovery groups. Drug Alcohol Rev 2017 May 01;36(3):325-332. [doi: 10.1111/dar.12416] [Medline: 27246699]
62. Horvath AT, Yeterian J. SMART Recovery: Self-empowering, science-based addiction recovery support. J Group Addict Recovery 2012 Apr;7(2-4):102-117. [doi: 10.1080/1556035x.2012.705651]
63. Gray RM, Kelly PJ, Beck AK, Baker AL, Deane FP, Neale J, et al. A qualitative exploration of SMART Recovery meetings in Australia and the role of a digital platform to support routine outcome monitoring. Addict Behav 2020 Feb;101:106144. [doi: 10.1016/j.addbeh.2019.106144] [Medline: 31648139]
64. Beck AK, Kelly PJ, Deane FP, Baker AL, Hides L, Manning V, et al. Developing a mHealth routine outcome monitoring and feedback app ("Smart Track") to support self-management of addictive behaviours. Front Psychiatry 2021 Jun 18;12:677637 [FREE Full text] [doi: 10.3389/fpsyt.2021.677637] [Medline: 34220583]
65. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 05;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
66. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (Integrate, Design, Assess, and Share): A framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016 Dec 16;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]

67. Deci EL, Ryan RM. The "What" and "Why" of Goal Pursuits: Human needs and the self-determination of behavior. Psychol Inq 2000 Oct;11(4):227-268. [doi: 10.1207/s15327965pli1104_01]
68. Moos RH. Theory-based active ingredients of effective treatments for substance use disorders. Drug Alcohol Depend 2007 May 11;88(2-3):109-121 [FREE Full text] [doi: 10.1016/j.drugalcdep.2006.10.010] [Medline: 17129682]
69. Smart Recovery International. 2019. URL: https://www.smartrecoveryinternational.org/ [accessed 2021-08-08]
70. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug 20;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
71. GHO Sydney. URL: https://ghosydney.com/ [accessed 2021-08-08]
72. SMART Track Google Play. SMART Recovery. 2021. URL: https://play.google.com/store/apps/details?id=au.com.smartrecoveryaustralia [accessed 2021-09-24]
73. SMART Track for iTunes. SMART Recovery. URL: https://apps.apple.com/au/app/smart-track/id1457532295 [accessed 2021-08-08]
74. Neale J, Vitoratou S, Finch E, Lennon P, Mitcheson L, Panebianco D, et al. Development and validation of 'Sure': a Patient reported outcome measure (PROM) for recovery from drug and alcohol dependence. Drug Alcohol Depend 2016 Aug 01;165:159-167 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.06.006] [Medline: 27344196]
75. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SLT, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002 Aug 26;32(6):959-976. [doi: 10.1017/s0033291702006074] [Medline: 12214795]
76. da Rocha NS, Power MJ, Bushnell DM, Fleck MP. The EUROHIS-QOL 8-item index: comparative psychometric properties to its parent WHOQOL-BREF. Value Health 2012 May;15(3):449-457 [FREE Full text] [doi: 10.1016/j.jval.2011.11.035] [Medline: 22583455]
77. Wilson KG, Sandoz EK, Kitchens J, Roberts M. The Valued Living Questionnaire: Defining and measuring valued action within a behavioral framework. Psychol Rec 2017 Jun 30;60(2):249-272. [doi: 10.1007/BF03395706]
78. Miller WR, Rollnick S, McLouth CJ, Walters ST, Hohman M. Motivational Interviewing: Helping People Change. 3rd Ed. New York: Guilford Press; 2012:1-482.
79. Kazantzis N, Deane FP, Ronan KR, L'Abate L. Assessment of homework completion. In: Using Homework Assignments in Cognitive Behavior Therapy. New York: Routledge; 2005:50-60.
80. NGO Drug and alcohol and mental health information management project: determining the treatment outcomes data collection set. NADA. 2009. URL: https://www.nada.org.au/wp-content/uploads/2018/06/determining_the_data_collection_set_dec09.pdf [accessed 2021-08-31]
81. Schluter MG, Hodgins DC, Wolfe J, Wild TC. Can one simple questionnaire assess substance-related and behavioural addiction problems? Results of a proposed new screener for community epidemiology. Addiction 2018 Aug;113(8):1528-1537. [doi: 10.1111/add.14166] [Medline: 29357188]
82. Client Outcome Measures (COMS). Network of Alcohol & Other Drugs Agencies and NSW Ministry of Health, Sydney. 2012. URL: http://tutorial.nada.org.au/client-outcome-measures/ [accessed 2021-08-31]
83. Lawrinson P, Copeland J, Indig D. The Brief Treatment Outcome Measure: Opioid Maintenance Pharmacotherapy (BTOM) Manual. University of New South Wales, Sydney: National Drug and Alcohol Research Centre; 2003.
84. Lives of Substance: Curtin University. National Drug Research Institute. 2016. URL: https://www.livesofsubstance.org/resources-information/ [accessed 2021-08-08]
85. SMART Recovery Facilitator Training Manual: Practical information and tools to help you facilitate a SMART recovery group. SMART Recovery Australia. 2015. URL: https://smartrecoveryaustralia.com.au/ [accessed 2021-08-08]
86. SMART Recovery Australia Participants' Manual: Tools and strategies to help you manage addictive behaviours. SMART Recovery Australia. URL: https://smartrecoveryaustralia.com.au/ [accessed 2021-08-08]
87. Amplitude Analytics. URL: https://amplitude.com/ [accessed 2021-08-08]
88. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
89. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
90. Gossop M, Darke S, Griffiths P, Hando J, Powis B, Hall W, et al. The Severity of Dependence Scale (SDS): psychometric properties of the SDS in English and Australian samples of heroin, cocaine and amphetamine users. Addiction 1995 May;90(5):607-614. [doi: 10.1046/j.1360-0443.1995.9056072.x] [Medline: 7795497]
91. Kessler RC, Barker PR, Colpe LJ, Epstein JF, Gfroerer JC, Hiripi E, et al. Screening for serious mental illness in the general population. Arch Gen Psychiatry 2003 Feb 01;60(2):184-189. [doi: 10.1001/archpsyc.60.2.184] [Medline: 12578436]
92. Henson P, Wisniewski H, Hollis C, Keshavan M, Torous J. Digital mental health apps and the therapeutic alliance: initial review. BJPsych Open 2019 Jan 29;5(1):e15 [FREE Full text] [doi: 10.1192/bjo.2018.86] [Medline: 30762511]

93. Major G, Ling R, Searles A, Niddrie F, Kelly A, Holliday E, et al. The Costs of Confronting Osteoporosis: Cost study of an Australian fracture liaison service. JBMR Plus 2019 Jan 18;3(1):56-63 [FREE Full text] [doi: 10.1002/jbm4.10046] [Medline: 30680364]
94. Ling R, Rush A, Carter C, Carpenter J, Watson PH, Byrne JA, et al. An Australian Biobank Certification Scheme: A study of economic costs to participating biobanks. Biopreserv Biobank 2018 Feb;16(1):53-58. [doi: 10.1089/bio.2017.0095] [Medline: 29215912]
95. Yoong SL, Grady A, Wiggers J, Flood V, Rissel C, Finch M, et al. A randomised controlled trial of an online menu planning intervention to improve childcare service adherence to dietary guidelines: a study protocol. BMJ Open 2017 Sep 11;7(9):e017498 [FREE Full text] [doi: 10.1136/bmjopen-2017-017498] [Medline: 28893755]
96. Beecham J, Knapp M. Costing psychiatric interventions. In: Thornicroft G, editor. Measuring Mental Health Needs (Second Edition). London, U.K: Royal College of Psychiatrists; 2001:200-224.
97. Andrews G, Slade T. Interpreting scores on the Kessler Psychological Distress Scale (K10). Aust N Z J Public Health 2001 Dec;25(6):494-497. [doi: 10.1111/j.1467-842x.2001.tb00310.x] [Medline: 11824981]
98. Lieberman ES. Nested analysis as a mixed-method strategy for comparative research. Am Polit Sci Rev 2005 Sep 02;99(3):435-452. [doi: 10.1017/s0003055405051762]
99. Kelly PJ, Deane FP, Baker AL. Group cohesion and between session homework activities predict self-reported cognitive-behavioral skill use amongst participants of SMART Recovery groups. J Subst Abuse Treat 2015 Apr;51:53-58. [doi: 10.1016/j.jsat.2014.10.008] [Medline: 25535099]
100. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: A systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
101. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of ehealth interventions: systematic review of the literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
102. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mhealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941 [FREE Full text] [doi: 10.2196/11941] [Medline: 30664463]
103. Weaver ER, Horyniak DR, Jenkinson R, Dietze P, Lim MS. "Let's get Wasted!" and Other Apps: Characteristics, acceptability, and use of alcohol-related smartphone applications. JMIR Mhealth Uhealth 2013 Jun 25;1(1):e9 [FREE Full text] [doi: 10.2196/mhealth.2709] [Medline: 25100681]
104. Song T, Qian S, Yu P. Mobile health interventions for self-control of unhealthy alcohol use: systematic review. JMIR Mhealth Uhealth 2019 Jan 29;7(1):e10899 [FREE Full text] [doi: 10.2196/10899] [Medline: 30694200]
105. Nesvåg S, McKay JR. Feasibility and effects of digital interventions to support people in recovery from substance use disorders: systematic review. J Med Internet Res 2018 Aug 23;20(8):e255 [FREE Full text] [doi: 10.2196/jmir.9873] [Medline: 30139724]
106. Rodde T. 21% of users abandon an app after one use. Localytics. 2018. URL: http://info.localytics.com/blog/25-of-users-abandon-apps-after-one-use [accessed 2021-07-30]
107. Rodde T, Cronin K, Noone S. The 2018 Mobile Benchmark Report. Localytics. 2018. URL: https://uplandsoftware.com/localytics/resources/ebook/2018-mobile-benchmark-report/ [accessed 2021-08-08]
108. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
109. Morrison LG, Hargood C, Lin SX, Dennison L, Joseph J, Hughes S, et al. Understanding usage of a hybrid website and smartphone app for weight management: a mixed-methods study. J Med Internet Res 2014 Oct 22;16(10):e201 [FREE Full text] [doi: 10.2196/jmir.3579] [Medline: 25355131]
110. Lewis CC, Boyd M, Puspitasari A, Navarro E, Howard J, Kassab H, et al. Implementing measurement-based care in behavioral health: a review. JAMA Psychiatry 2019 Mar 01;76(3):324-335 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.3329] [Medline: 30566197]
111. Delgadillo J, Overend K, Lucock M, Groom M, Kirby N, McMillan D, et al. Improving the efficiency of psychological treatment using outcome feedback technology. Behav Res Ther 2017 Dec;99:89-97. [doi: 10.1016/j.brat.2017.09.011] [Medline: 29024821]
112. de Beurs E, Warmerdam L, Twisk J. Bias through selective inclusion and attrition: Representativeness when comparing provider performance with routine outcome monitoring data. Clin Psychol Psychother 2019 Jul 23;26(4):430-439 [FREE Full text] [doi: 10.1002/cpp.2364] [Medline: 30882974]
113. Foster A, Croot L, Brazier J, Harris J, O'Cathain A. The facilitators and barriers to implementing patient reported outcome measures in organisations delivering health related services: a systematic review of reviews. J Patient Rep Outcomes 2018 Dec 3;2(1):46 [FREE Full text] [doi: 10.1186/s41687-018-0072-3] [Medline: 30363333]
114. Lee K, Kwon H, Lee B, Lee G, Lee JH, Park YR, et al. Effect of self-monitoring on long-term patient engagement with mobile health applications. PLoS One 2018 Jul 26;13(7):e0201166 [FREE Full text] [doi: 10.1371/journal.pone.0201166] [Medline: 30048546]

115. Wei Y, Zheng P, Deng H, Wang X, Li X, Fu H. Design features for improving mobile health intervention user engagement: systematic review and thematic analysis. J Med Internet Res 2020 Dec 09;22(12):e21687 [FREE Full text] [doi: 10.2196/21687] [Medline: 33295292]
116. Ford JH, Alagoz E, Dinauer S, Johnson KA, Pe-Romashko K, Gustafson DH. Successful Organizational Strategies to Sustain Use of A-CHESS: A mobile intervention for individuals with alcohol use disorders. J Med Internet Res 2015 Aug 18;17(8):e201 [FREE Full text] [doi: 10.2196/jmir.3965] [Medline: 26286257]
117. Hoffman L, Benedetto E, Huang H, Grossman E, Kaluma D, Mann Z, et al. Augmenting mental health in primary care: A 1-year study of deploying smartphone apps in a multi-site primary care/behavioral health integration program. Front Psychiatry 2019 Feb 28;10:94 [FREE Full text] [doi: 10.3389/fpsyt.2019.00094] [Medline: 30873053]
118. Miller S, Ainsworth B, Yardley L, Milton A, Weal M, Smith P, et al. A framework for Analyzing and Measuring Usage and Engagement Data (AMUsED) in digital interventions: Viewpoint. J Med Internet Res 2019 Feb 15;21(2):e10966 [FREE Full text] [doi: 10.2196/10966] [Medline: 30767905]
119. Perkins KS, Tharp BE, Ramsey AT, Silver-Wolf DP. Mapping the evidence to improve retention rates in addiction services. J Soc Work Pract Addict 2016 Jul 29;16(3):233-251. [doi: 10.1080/1533256x.2016.1200055]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
DWAI: Digital Working Alliance Inventory
K-10: Kessler Psychological Distress Scale-10
K-6: Kessler Psychological Distress Scale-6
MARS: Mobile App Rating Scale
mHealth: mobile health
R&D: research and development
ROM: routine outcome monitoring
SDS: Severity of Dependence Scale
SRAU: SMART Recovery Australia
SURE: Substance Use Recovery Evaluator
uMARS: Mobile App Rating Scale–User Version

Edited by R Kukafka; submitted 22.11.20; peer-reviewed by J Bjornestad, J Debenham, C Rauschenberg, S Kershaw; comments to author 30.01.21; revised version received 25.03.21; accepted 25.04.21; published 06.10.21.

Please cite as:
Kelly PJ, Beck AK, Deane FP, Larance B, Baker AL, Hides L, Manning V, Shakeshaft A, Neale J, Kelly JF, Oldmeadow C, Searles A, Palazzi K, Lawson K, Treloar C, Gray RM, Argent A, McGlaughlin R
Feasibility of a Mobile Health App for Routine Outcome Monitoring and Feedback in SMART Recovery Mutual Support Groups: Stage 1 Mixed Methods Pilot Study
J Med Internet Res 2021;23(10):e25217
URL: https://www.jmir.org/2021/10/e25217
doi:10.2196/25217
PMID:34612829

Review

Mobile Safety Alarms Based on GPS Technology in the Care of Older Adults: Systematic Review of Evidence Based on a General Evidence Framework for Digital Health Technologies

Maria Ehn[1], MSc, PhD; Matt X Richardson[2], PhD; Sara Landerdahl Stridsberg[3], MA; Ken Redekop[4], PhD; Sarah Wamala-Andersson[2], MSc, PhD

[1]School of Innovation, Design and Engineering, Mälardalen University, Västerås, Sweden

[2]School of Health, Care and Social Welfare, Mälardalen University, Eskilstuna, Sweden

[3]University Library, Mälardalen University, Västerås, Sweden

[4]Erasmus School of Health Policy and Management, Erasmus Universiteit Rotterdam, Rotterdam, Netherlands

Corresponding Author:
Maria Ehn, MSc, PhD
School of Innovation, Design and Engineering
Mälardalen University
Box 883
Västerås, SE-721 23
Sweden
Phone: 46 21 107093
Email: maria.ehn@mdh.se

Abstract

Background: GPS alarms aim to support users in independent activities. Previous systematic reviews have reported a lack of clear evidence of the effectiveness of GPS alarms for the health and welfare of users and their families and for social care provision. As GPS devices are currently being implemented in social care, it is important to investigate whether the evidence of their clinical effectiveness remains insufficient. Standardized evidence frameworks have been developed to ensure that new technologies are clinically effective and offer economic value. The frameworks for analyzing existing evidence of the clinical effectiveness of GPS devices can be used to identify the risks associated with their implementation and demonstrate key aspects of successful piloting or implementation.

Objective: The principal aim of this study is to provide an up-to-date systematic review of evidence based on existing studies of the effects of GPS alarms on health, welfare, and social provision in the care of older adults compared with non–GPS-based standard care. In addition, the study findings were assessed by using the evidence standards framework for digital health technologies (DHTs) established by the National Institute for Health and Care Excellence (NICE) in the United Kingdom.

Methods: This review was conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. Primary studies published in peer-reviewed journals and gray literature from January 2005 to August 2020 were identified through searches in 13 databases and several sources of gray literature. Included studies had individuals (aged ≥50 years) who were receiving social care for older adults or for persons with dementia; used GPS devices as an intervention; were performed in Canada, the United States, European Union, Singapore, Australia, New Zealand, Hong Kong, South Korea, or Japan; and addressed quantitative outcomes related to health, welfare, and social care. The study findings were analyzed by using the NICE framework requirements for *active monitoring* DHTs.

Results: Of the screened records, 1.6% (16/986) were included. Following the standards of the NICE framework, practice evidence was identified for the tier 1 categories *Relevance to current pathways in health/social care system* and *Acceptability with users*, and minimum evidence was identified for the tier 1 category *Credibility with health, social care professionals*. However, several evidence categories for tiers 1 and 2 could not be assessed, and no clear evidence demonstrating effectiveness could be identified. Thus, the evidence required for using DHTs to track patient location according to the NICE framework was insufficient.

Conclusions: Evidence of the beneficial effects of GPS alarms on the health and welfare of older adults and social care provision remains insufficient. This review illustrated the application of the NICE framework in analyses of evidence, demonstrated successful piloting and acceptability with users of GPS devices, and identified implications for future research.

(J Med Internet Res 2021;23(10):e27267) doi:10.2196/27267

KEYWORDS

GPS; older adults; dementia; evidence; digital health; NICE; mobile phone

Introduction

Background

Health and welfare technologies (HWTs) are "technology-based interventions that aim at maintaining or promoting health, well-being, quality of life (QoL) and/or increasing efficiency in the service delivery system of welfare, social and health care services, while improving working conditions of the staff" [1]. This definition can be regarded as an integration of the Scandinavian term *welfare technology* [2] and the global concept *digital health* [3], including a broader definition of health according to the World Health Organization: "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity." GPS alarms are examples of HWTs that aim to support users in independent activities of daily living, both indoors and outdoors, something that is needed by large groups of older adults who want to maintain independent living in spite of cognitive impairments, for example, persons with dementia in the early or middle stage. Globally, approximately 50 million people have dementia, and nearly 10 million new persons develop dementia every year [4]. The global population of persons with dementia is estimated to increase to 82 million in 2030 and to 152 million in 2050 [4].

GPS solutions enable the user to send an alarm to home care, family members, or both, either at home or outdoors. The user's position can be localized through GPS coordinates when an alarm is sent. GPS alarms may also include a geofencing function that automatically generates an alarm if the user leaves a predefined geographical area. A European overview has identified ongoing implementation of GPS alarms in the municipal care of older adults in Sweden and Norway [5].

Research on the implementation of welfare technology in Swedish municipalities has identified several barriers to its use, including lack of supporting evidence of the benefits and positive outcomes [6]. Hence, evidence of positive outcomes of GPS alarms can justify large-scale implementation and confidence in their use.

When making implementation decisions, it is important to consider what kind of evidence is needed to make an informed decision. To date, there is no national Swedish evidence-based framework to support decision-making processes related to HWTs [7]. However, the National Institute for Health and Care Excellence (NICE) in the United Kingdom, together with relevant stakeholders, has produced an evidence standards framework for digital health technologies (DHTs) to ensure that new technologies can demonstrate clinically effective and economic value in a structured and transparent manner [8]. The framework classifies DHTs by function and thereby allows them to be stratified into evidence tiers based on the potential risk to users. For example, DHTs that track patient location can be classified as *active monitoring* and thereby stratified into evidence tier 3b in the framework. Good levels of evidence for tier 3 DHTs include evidence of credibility with health and social care professionals, relevance to current care pathways in health care systems, acceptability with users, equalities considerations in provision and use, accurate and reliable measurements and transmission of data, reliable information content, ongoing and planned continuous data collection to follow up on the use and value of the DHT, quality and safeguarding, and demonstration of its effectiveness according to intended outcomes. The framework has been applied in examples of case studies that demonstrate evidence of effectiveness and economic value of a number of DHTs [9]. These studies are based largely on information provided by the developers that has not been independently verified. Information from studies published in peer-reviewed journals and in gray literature can therefore strengthen the validity of evidence of DHTs.

Two previous scoping reviews of tracking technology in the care of older adults conducted by Røhne et al [10] and Neubauer et al [11] identified limited evidence in peer-reviewed scientific studies for the effects of GPS use on health, quality, and cost-efficiency among users and spouses and in the health and welfare sector [10] as well as pointed out that further research was needed to identify technologies with high levels of evidence for effectiveness and usability [11]. Moreover, a synthesizing review of empirical evidence on the broader use of GPS technologies by home-dwelling persons with dementia and their family caregivers (CGs), conducted by Bartlett et al [12], found only nontrial evidence and demonstrated the lack of large-scale studies. As GPS devices are currently being implemented in social care, it is important to investigate whether the evidence of their clinical effectiveness remains insufficient. By analyzing existing evidence of GPS devices using an established framework, the risks associated with their implementation and conclusions from successful piloting or implementations can be identified.

Objectives

The aims of this review are to (1) systematically update evidence from existing studies of the effects of GPS alarms on the health, welfare, and social provision in receivers of care for older adults compared with non–GPS-based standard care and (2) review the findings using an established evidence framework for DHTs. Studies from both peer-reviewed journals and gray literature were included, and the findings were compared with the NICE evidence standards framework for DHTs [8]. The goal is to assess the available evidence according to the desired evidence standards of an established framework to support decision-making in future implementations of GPS-based alarms.

Methods

Protocol and Registration

This systematic literature review was reported according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [13].

Eligibility Criteria

The elements of the review (aim, eligibility criteria, and outcomes) were defined as shown in Textbox 1.

Textbox 1. Review elements.

Aim

- To conduct a systematic review of existing evidence of the effects of GPS alarms on the health, welfare, and social provision in the care of older adults compared with standard non–GPS-based care

Inclusion criteria

- Original, peer-reviewed publications published between January 2005 and August 2020 in the English language
- Studies performed in Canada, the United States, European Union, Australia, New Zealand, Singapore, Hong Kong, South Korea, or Japan
- Population aged ≥50 years
- GPS-based alarm interventions enabled the users to initiate alarms, with or without localization and with or without geofencing functions

Exclusion criteria

- Qualitative studies, technical validations, proof-of-concept studies, system descriptions, reviews, and editorials

Outcomes

- Outcomes related to the health and welfare of users and their informal caregivers, and outcomes related to provision in the care of older adults

Search Strategy

A three-step search strategy was used: (1) an initial search of electronic peer-reviewed scientific publication databases; (2) a search of gray literature databases, trial registers, and Google Scholar; and (3) a search of gray literature from the Nordic countries in targeted websites set up by authorities and organizations, academic publication databases, and Google Scholar. All three steps used a snowballing approach [14,15] in which the articles that reached title-, abstract-, or full-text screening of an initial database search were used in iterations of both backward (reference search) and forward (citation search) snowballing.

Information Sources and Searches

Summary and Complete Strategies

A summary of information sources and retrieved records for all the search steps is presented in Multimedia Appendix 1 [10,16-19]. The complete database-specific search strategies for the initial searches of scientific and gray literature are presented in Multimedia Appendix 2.

Scientific Literature

The literature search was completed on August 20, 2020, in the following electronic databases: Academic Search Elite, APA PsycINFO, Applied Social Sciences Index & Abstracts, CINAHL Plus, Cochrane Library, International Bibliography of the Social Sciences, IEEE Xplore, PubMed, Scopus, SocINDEX, Social Services Abstracts, Sociological Abstracts, and Web of Science Core Collection. The search strategy included both free-text and controlled vocabulary. The search was limited to articles published in 2005 or later.

Gray Literature

A primary search in gray literature databases was conducted on September 8 and 9, 2020. The databases searched included Base, OpenGrey, OAIster, DART-Europe, ProQuest Dissertations & Theses Global, WHO ICTRP, ClinicalTrials.gov, International HTA Database, and Google Scholar. The search strategy was simplified and adapted to the search interface of gray literature databases. Finally, a larger search for gray literature from the Nordic countries was conducted by searching the websites of government agencies and organizations working with health and welfare issues. Publication databases from universities in the Nordic countries were also searched including the databases and websites of associations of local authorities and regions in the Nordic countries and Google Scholar. The searches were conducted from September 22, 2020, to September 29, 2020, and involved searching for terms in Swedish, Norwegian, Danish, Finnish, and Icelandic.

Study Selection

The relevant records were downloaded to the reference management software EndNote (Clarivate Analytics). In the search for Nordic gray literature, the retrieved records were screened to determine their relevance related to the aim of the review before downloading the publications to EndNote. In the other searches, all retrieved records were directly downloaded to EndNote. After the removal of duplicates, the records were transferred to the systematic review software Covidence (Veritas Health Innovation Ltd). All four steps were performed by an information science specialist or librarian. The Covidence software, which automatically detected and removed any remaining duplicates, was used for title and abstract screening, full-text review, and data extraction. The screening of titles or abstracts, eligibility assessment of full-text articles, and full-text

screening were performed independently by 2 reviewers; any conflicts were resolved by a third reviewer.

Data Extraction

Information on the included studies' aims, design, conduct, population, intervention, and outcomes as well as the results for relevant outcomes were extracted from the publications by 2 reviewers independently using a predefined template. Any conflicts in eligibility assessment were resolved through discussion between the reviewers. Information regarding the included studies is presented in Table 1.

Risk-of-Bias Assessment

Publications that aimed to demonstrate evidence of effectiveness (for tier 3a) were assessed for risk of bias at the study level by 2 researchers independently. The criteria assessed for the randomized study followed the Cochrane Risk-of-Bias Tool 2.0 guidelines [20] and included the method used for random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, the completeness of outcome data, the possibility of selective outcome reporting, and other sources of bias. For nonrandomized studies, the Risk of Bias in Non-randomized Studies of Interventions tool [21] was used to assess bias due to confounding and missing data as well as selective reporting and in the selection of participants and classification of, and deviations from, interventions. Consensus resulted in inclusion of the risk assessment in this review's summary of findings. Any conflicts in risk assessment were discussed by the 2 researchers (ME and MR) and any conflicts remaining were resolved by the third reviewer (SLS).

Analysis of Findings From the Systematic Review via the Application of the Evidence Framework

The extracted data for relevant outcomes were categorized and summarized according to the tiers 1-3b evidence categories in the NICE evidence standards framework for DHTs described below. An overview of each study's contribution to the respective evidence categories is presented in Table 2, and a more extensive version of the table, including the criteria for minimum evidence and best practice standards in all evidence categories, is presented in Multimedia Appendix 3 [22-37]. The table was prepared by one of the researchers and reviewed by a second reviewer. For each evidence category, data extracted from all included studies were compared with the defined minimum evidence and best practice standards of evidence. The extracted data were summarized for each evidence category, and the extent to which the results met the requirements for minimum evidence and best practice standards in each category was assessed. The process was carried out by one of the researchers and reviewed by a second reviewer. Any assessment conflicts were resolved through discussion between the reviewers. The summarized data according to the evidence framework are presented in this review's summary of the findings (Table 2).

The Standard Evidence Framework for DHTs

The framework classifies DHTs by function and thereby allows them to be stratified into evidence tiers based on the potential risk to users [8]. Examples of effectiveness and economic value of digital health case studies have been provided to illustrate how the framework can be used [9]. DHTs that track patient location were classified as *active monitoring* and thereby stratified into evidence tier 3b. Moreover, the NICE guidelines state that "best practice evidence standards in each relevant evidence tier should be used for DHTs that present a potential high risk" [8]. GPS alarms can be identified as high-risk DHTs because the intended user group is a vulnerable group and GPS alarm failure could have serious consequences for the user, and in some countries GPS alarms might be used without regular support from social care professionals.

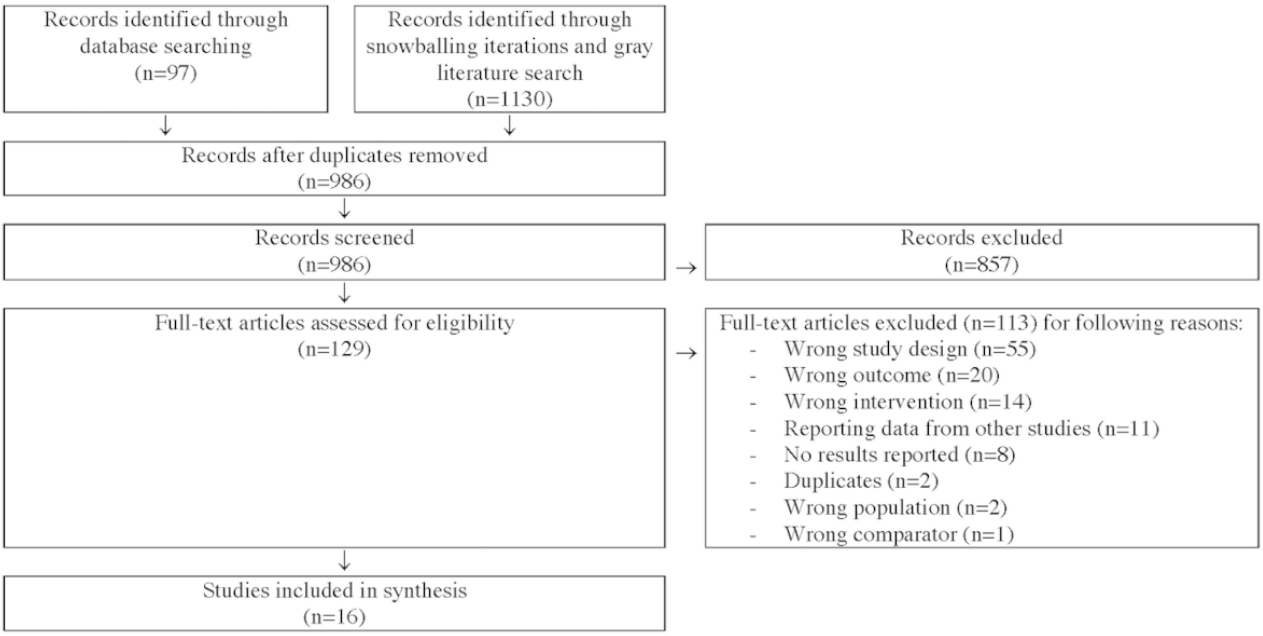
Good levels of evidence for tier 3 DHTs are presented in Multimedia Appendix 3 [22-37]. Evidence standards of the tier 1 and tier 2 categories require documentation that demonstrate that specific aspects have been considered and that the procedures have been completed with a certain level of quality control. Therefore, both the results and aims of the included studies were considered for tier 1 evidence categories. In contrast, tiers 3a and 3b require evidence demonstrating effectiveness.

Results

Study Selection

The literature searches identified 1227 records. After the removal of 240 (240/1227, 19.56%) duplicates, the titles or abstracts of 80.36% (986/1227) of the publications were screened for relevance; 69.85%, (857/1227) of articles were excluded in the screening, and 10.51% (129/1227) full-text publications were assessed for eligibility according to the inclusion and exclusion criteria presented in Textbox 1. Of the 129 full-text publications, 113 (87.6%) were excluded, and the remaining 16 (12.4%) articles were included in the final assessment. Figure 1 provides an overview of the publication selection process.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of publications assessed in each step of the review process.



Study Characteristics

The 16 included studies were published as original articles in scientific journals (n=6), conference papers (n=2), a master's thesis (n=1), and project reports (n=7). A summary of the characteristics of each publication is presented in Table 1.

Of the 6 peer-reviewed journal studies, 1 was a randomized controlled trial (RCT), and the others were nonrandomized studies with either an experimental (n=4) or an observational design (n=1). The other 10 studies had the following study designs: observational design (n=5, of which 3 used mixed methods and two used quantitative methods), pre- and postintervention design using mixed methods (n=2), economic evaluation (n=2), and within-subjects design using mixed methods (n=1). Of these 10 studies, 5 were part of larger innovation or pilot implementation projects.

The study populations were older adults in general (n=4), older adults with dementia and their family CGs (n=11), and municipal employees working with GPS trackers for persons with dementia (n=1).

Table 1. Characteristics of the included studies (N=16).

Study, country	Aim	Intervention	Duration	Design	Participants
Magnusson et al [22], Sweden[a,b]	To investigate views and experiences of persons with dementia living at home, their CGs[c], and the staff involved in use and implementation of advanced electronic tracking; to analyze results in relation to ethics; and to discuss methodological aspects of research involving persons with dementia	Extended safety and support system Posifon in mobile phone	9 months	Nonrandomized experimental study with pre- and postintervention measurements	Older persons with dementia living at home and their family CGs. Recruited: 76 persons with dementia and 76 CGs; final sample: 20 persons with dementia and 36 CGs
Megges et al [23], Germany[a,b]	To perform an in-depth comparison of the user experiences of persons with dementia and their informal CGs with 2 similar commercial GPS watches in home dementia care; to study the products' clinical effectiveness in persons with dementia and CGs	2 GPS watches for people with orientation impairment (products A and B)	4 weeks for each product	Nonrandomized experimental study, 2×2 crossover	Dyads: persons with dementia and CGs. Recruited: 12 persons with dementia and 12 CGs; final sample: 4 persons with dementia and 8 CGs
Milne et al [24], Scotland[a,b]	To determine the feasibility of a prospective randomized controlled trial	GPS in watch or pendant, some with a geofencing option. CGs track persons with dementia and are contacted by monitoring agency if persons with dementia report that they are lost or have reached the geofence	4 months (mean and median), range 1-7 months	Mixed methods observational design	20 dyads: persons with dementia with a history of wandering and CGs
Olsson et al [25], Sweden[a,b]	To investigate the effects of using tracking technology on independent outdoor activities and psychological well-being in 3 persons with dementia and their spouses	Passive positioning alarm (transmitter based on GPS, cell phone, and support person)	Intervention phase B1: 5-7 weeks; intervention phase B2: 4-5 weeks	Nonrandomized experimental study with crossover design	3 dyads: persons with dementia and CGs
Pot et al [26], The Netherlands[a,b]	To investigate feasibility, acceptability, and effectiveness of 3-month use of GPS by persons with dementia and CGs	Tracking device (GPS and general packet radio service) worn on belt, including track and trace function, telephone contact, and loudspeaker function	3 months	Nonrandomized experimental study	Dyads: persons with dementia and CGs; persons with dementia in early stage of dementia. Recruited: 33; included: 28
Scheffer et al [27], The Netherlands[a,b]	To investigate the effects of a mobile safety alarm on frequency of going outside and experiences regarding fear of falling, feelings of being unsafe, and quality of life in older adults	Mobile safety alarm with a built-in drop sensor using a positioning system over a mobile phone network	6 months	Randomized controlled trial	Older adults using indoor alarm. Included: 203 (100 in intervention and 103 in control); final sample: 135 (58 in intervention and 77 in control)

Study, country	Aim	Intervention	Duration	Design	Participants
Ribas Miquel et al [37], Spain[d]	To describe the perceptions and experiences of professionals and family members of users and nonusers of GPS tracking devices	GPS tracking device	N/A[e]	Mixed methods observational design	30 care professionals, 7 family members of care receivers who use a GPS tracking device, and 7 family members of care receivers not using a tracking device
Røhne et al [28], Norway[b,f]	To verify if or how mobile safety alarms make older adults more independent, increase their mobility and physical activity, engage relatives and lead to reduced need for care, and increase the ability of older adults to live longer at home	Global System for Mobile communication or GPS-based alarm unit (hanging around the user's neck) with geofencing, voice connection, and tracking ability	2-9 months	Mixed methods with pre- and postintervention design[f]	Populations of older adults in Bærum, Skien, and Stavanger municipalities. Included in pilot: 71; included in evaluation: 46
Sørli [29], Norway[g]	To investigate the effects and the experiences of municipal employees of GPS use among persons with dementia on the quality of the municipal services	GPS tracking device	8 months	Observational design	19 municipal employees working with GPS for persons with dementia
Ausen et al [30], Norway[h]	To establish knowledge about users' and employees' experience of piloted GPS; to establish knowledge of the effects of use of the technology on users, employees, and the service; to identify potential gains; to describe service models for the use of various security and coping technologies in the municipality	GPS unit (possible to attach to key chain) for localization with geofencing alarm	1 year (mean)	Observational design[f]	Older adults living at home or in nursing homes in Larvik. 47 care receivers; number of employees unclear
Boysen and Støle [31], Norway[h]	To establish knowledge of the users' and employees' experience of the technologies being piloted; to identify potential gains; and to propose measures to realize the benefits	Mobile safety alarm with GPS tracking and geofencing (Safemate)	6 months	Mixed methods with pre- and postintervention design[f]	9 persons with dementia and with affiliation with either housing associations, nursing homes, or activity centers; 13 home users with follow-up of relatives

Study, country	Aim	Intervention	Duration	Design	Participants
Dahlberg [32], Sweden[h]	To perform a socioeconomic analysis in the form of a cost-benefit assessment of a mobile security alarm	Mobile safety alarm with GPS tracking, geofencing, and voice communication (Posifon, same as [22])	9-10 months	Economic evaluation	Persons with mild and more advanced dementia; persons with more advanced dementia with municipal care; all living in a home setting with some form of care from relatives. Approximately 80 (same as [22])
Malmquist [33], Sweden[h]	To assess the costs and benefits of passive alarms for users, their relatives, the municipality, and society; to develop a decision-making basis for prescription of different passive alarms; to increase knowledge regarding passive alarms and their costs and benefits for stakeholders	GPS or Global System for Mobile communication passive position alarm (bracelet) with geofencing function	37-260 days	Economic evaluation	8 persons with dementia living in home settings in Östersund municipality and their relatives or informal CGs. Prescribers of alarms within the municipality
Øderud et al [34], Norway[h]	To investigate how location technology can be organized and integrated as part of the municipalities' operational health and care services and to develop service models for interaction among public, private, and volunteer service providers to help persons with dementia and their CGs	GPS devices (1 of 3) with alarm, tracking, and voice communication features	Up to 1 year (58% of the users); 1-2 years (35% of the users), more than 2 years (7% of the users)	Mixed methods observational design[i]	Oslo inhabitants with dementia or dementia-like conditions, 94% living in their own home. 109 users; 216 in total (users and their relatives, employees in care services and at alarm center, and localization technology providers)
Røhne et al [35], Norway[h]	To investigate how mobile safety alarms can contribute toward enabling older adults to reside at home for as long as possible and to establish knowledge of how today's mobile safety alarms can be developed	Mobile safety alarm with tracking and voice communication	8-10 months	Mixed methods observational design[j]	Older adults living independently Recruited: 12; final sample: 10

Study, country	Aim	Intervention	Duration	Design	Participants
Vidensförmidling and Syd [36], Denmark[h]	To investigate whether early allocation of GPS alarms can provide greater security and quality of life for persons with dementia and their relatives and thereby reduce the need for help from the municipality	A combined GPS or Global System for Mobile communication device carried in a pocket or worn on the belt, individually adapted to the individual. Unit equipped with a call button, with which the person with dementia can call for help in the event of a fall or similar incident	At least three months	Mixed methods, within-subjects design; economic evaluation[k]	180 persons with dementia and their informal CGs living in the home setting in 5 municipalities

[a]Journal article.

[b]Peer-reviewed.

[c]CG: caregiver.

[d]Conference paper.

[e]N/A: not applicable.

[f]Part of a larger project that, according to the publication, used methods from research-supported, demand-driven innovation.

[g]Master's thesis.

[h]Project report.

[i]Part of a larger project which, according to the publication, used methods from demand-driven innovation and service design to go from pilot to operation.

[j]Part of pilot implementation study.

[k]Part of demonstration project.

Summary of Findings

Table 2 presents an overview of the study findings from individual studies assessed using the NICE evidence standards framework for DHTs [8]. A more extensive version of the table, including the definition of minimum evidence and best practice standards for each category, is presented in Multimedia Appendix 3 [22-37]. As can be seen in Table 2, the included studies provide best practice evidence according to the standards of two of the five tier 1 evidence categories, that is, *Relevance to current pathways in health/social care system* and *Acceptability with users*. In more than 60% of the included studies, GPS devices had been successfully piloted or implemented in social care systems. All these studies had been performed in Nordic countries as part of larger projects supporting the development of products, services, and decision-making processes. Most of these projects were part of national government programs that aimed to stimulate the use of welfare technology. Furthermore, best practice evidence showing that representatives from the intended user groups (older adults and persons with dementia) were involved in the testing of the GPS alarms and that the users were satisfied with them was identified in 38% (6/16) of the included studies. Table 2 also shows that minimum standard evidence was identified for the tier 1 category *Credibility with health, social care professionals* because relevant social care professionals had been involved in 75% (12/16) of the included studies.

However, the evidence according to the standards for tier 2 evidence categories could not be assessed from the included studies. For example, because the alarm systems do not provide general information or advice to users concerning health, healthy living, lifestyle, diseases, illnesses, or conditions, the minimum and best practice standards were not relevant for this assessment. Nevertheless, the information that the alarm systems provides to formal and informal CGs and security service providers about user position and emergency situations must be accurate. None of the evaluated studies investigated the reliability of this information. However, one study elaborated on alarm testing and the timeliness of the transmission of information in case of an alarm during the development of test routines, and two studies investigated the CGs' views on the accuracy of the information regarding the user position and the user-friendly aspects of the interfaces. The two latter studies identified that some situations (eg, when users reach places with poor mobile coverage) can limit the updating of user position and thereby reduce the reliability of the system's information content.

Furthermore, evidence of ongoing data collection to show use and value could not be identified from the included studies. Indeed, several studies demonstrated use and value on study follow-up occasions. Use was reported both for specific system functions and on a system level with varied levels of detail. Although some studies collected use data from system logs, reporting by CGs or users was more common. The measured values for users (persons with dementia or older adults) related to improved outdoor activity (increased independence, fewer worries, increased frequency with regard to visiting new places and making longer trips, and increased or maintained physical activity level); improved relationship with CGs (fewer conflicts with CGs and more freedom); increased security and safety (increased security in daily life and prolonged period living independently); and increased QoL. The measured values for CGs of persons with dementia related to improved well-being, QoL, enhanced possibility of giving more freedom to persons with dementia, and improved security and safety. Finally, examples of measures for safeguarding (service models, test routines, and role of relatives or alarm center) in the use of GPS

alarms were described in three of the included studies. However, evidence according to the standards could not be identified.

Moreover, the included studies provide clear evidence of effectiveness in outcomes or improvements in outcomes relevant to tiers 3a and 3b. Of the 16 included studies, 10 [22-27,32,33,36,37] investigated the outcomes that could be related to effectiveness. The findings of the included RCT [27] were compared with the framework's standards for best practice evidence to demonstrate effectiveness in outcomes or improvements in outcomes. However, no increase in the frequency of older adults going outside was found in the intervention group, and no significant differences in secondary outcomes, including fear of falling, feelings of unsafety, or QoL, were identified [27]. The other nine studies that used observational or quasi-experimental designs were assessed with regard to requirements for minimum evidence standards (ie, demonstrating effectiveness in outcomes or improvements in outcomes). Improvements were indicated with regard to an increase in the percentage of days that persons with dementia were engaged in independent outdoor activity (there were indications of an increase in three cases based on CG reports; no statistical data were available) [25] and to a decrease in role-overload and feelings of worry for CGs of persons with dementia ($P>.05$) [26]. Furthermore, GPS tracker use was associated with an important decrease in time spent searching for persons with dementia who were lost [24]. However, the data were based on CG recall and could not be objectively verified. In addition, this outcome is more related to efficiency than effectiveness. Interestingly, one study found decreased activity among persons with dementia because of disease progression [22], and another study identified no significant changes in burden or QoL for CGs of persons with dementia [23]. In contrast, economic evaluations indicated reduced costs for the care of persons with dementia because of prolonged periods living independently instead of special housing (up to 3 months) [32,33,36]. As a crossover design was used, the difference in the mean CG burden between relatives of persons with dementia using or not using GPS trackers was indicated in small samples [37].

Table 2. Summary of findings in relation to the evidence categories of the evidence standards framework for digital health technologies (N=16).

Tiers	Findings	Studies, n (%)	Risk of bias[a]
Tier 1			
Credibility with health and social care professionals	• Minimum evidence standards show that relevant social care professionals were involved in the design, development, or testing of the GPS devices • In 12 (75%) [22,24,25,28-36] of the included studies, social care professionals were involved in the testing of the GPS devices to a varied extent	• 12 (75)	—[b]
Relevance to current pathways in health or social care system	• Minimum and best practice evidence standards show that GPS devices have been successfully piloted or implemented in social care systems. This was described in 10 (63%) of the included studies. Of these 10, 3 were performed in Sweden [22,32,33], 6 in Norway [28-31,34,35], and 1 in Denmark [36]. All of them were part of larger projects supporting development of products, services, and decision-making processes to support OAsc and their families in their homes. Most of those projects were part of national government programs that aimed to stimulate the use of welfare technology	• 10 (63)	—
Acceptability with users	• Best practice evidence shows that representatives from the intended user groups (persons with dementia and OAs) were involved in the design, development, or testing of the DHTd and to show that users were satisfied with the DHT • Representatives from the intended user groups (OAs in general or persons with dementia) were involved in testing of the GPS alarms in 15 (94%) of the studies (ie, all the included studies except [29]) • Six (38%) of these studies showed that the users were satisfied with the alarms: 77% of the CGse of persons with dementia stated that they would recommend the use of GPS alarms in the Pot et al study [26]; 97% of the OAs who participated in the Røhne et al study [28] and 90% of the OAs in the Røhne et al study [35] stated that they were satisfied with the alarm • All older users in the Ausen et al study [30] would recommend others in similar situations to use the GPS alarm • User satisfaction was confirmed in the interviews in the Milne et al study [24] and in the values identified in the Boysen et al study [31]	• 6 (38)	—
Equalities considerations	• No information retrieved from included studies. Socioeconomic aspects were not addressed • However, persons with dementia and OAs may be considered vulnerable groups	• 0 (0)	—
Accurate and reliable measurements (if relevant)	• No information retrieved from included studies. Technical validations were not included in the review. All included studies used commercial products	• 0 (0)	—
Accurate and reliable transmission of data (if relevant)	• No information retrieved from included studies. Technical validations were not included in the review. All included studies used commercial products	• 0 (0)	—
Tier 2			
Reliable information content	• Minimum and best practice standards category is not relevant for GPS devices because they do not provide general information or advice to users concerning health, healthy living, lifestyle, diseases, illnesses, or conditions • However, for reliability of information on user position and emergency situations, data on user testing were provided by 1 study (6%) [28], and data on CGs' perceptions of the accuracy of the GPS information were provided by 2 (13%) studies. For example, relatives and staff in the Øderud et al study [34] had experienced slow or unreliable information on the user's position. Moreover, [30] reported cases of poor mobile coverage that had resulted in failures in updating user position	• 0 (0)	—

Tiers	Findings	Studies, n (%)	Risk of bias[a]
Ongoing data collection to show use	• Cannot be assessed from the included studies. Evidence of ongoing data collection (required according to evidence standards for the category) was not reported in the included studies • However, 10 (63%) of the included studies presented data on use on specific occasions related to the interventions • In all, 3 studies (19%) reported quantitative data on usage period: [30] and [34] presented the number of participants who had used GPS trackers for up to 1 year and between 1 and 2 years, respectively. [33] reported the number of days that each participant had used the GPS trackers (mean 158 days, median 210 days, and range 37-260 days) • A total of 3 (19%) studies [22,25,28] included system logs in the collection of data to investigate use. Interestingly, [22] saw that the extent to which persons with dementia used mobile phone–based GPS varied widely among the participants. Moreover, [28] described that the logs from the technical systems were thoroughly analyzed to understand the role and function of users, alarm units, response center, CGs, and relatives • In all, 4 (25%) studies [23,24,26,27] based the data collection of use on the recall of the users or their CGs, and 2 of these [26,27] reported that the persons with dementia did not always take along the GPS devices (mobile phone or tracker worn on the belt) when going out and that the devices were not always switched on	• 0 (0)	—
Ongoing data collection to show value	• Cannot be assessed from the included studies. Evidence of ongoing data collection to show value (required according to evidence standards for the category) was not reported • However, 11 (69%) of the included studies presented data on use for values related to the health and welfare outcomes of users (OAs, persons with dementia, and CGs of persons with dementia) on specific occasions related to the interventions; one (6%) study [22] identified that CGs experienced that the persons with dementia had become more independent in outdoor activity; 1 (6%) study [24] identified that CGs and staff saw that GPS trackers could give persons with dementia in milder stages of dementia and their CGs increased freedom and decreased stress and anxiety; and 1 (6%) study [26] identified values perceived by some of the CRsf, including increased freedom and decreased worries and fewer conflicts with CGs when going outside alone. Moreover, the CGs experienced that they gave more freedom to the CR and some experienced fewer conflicts with the CR • Another study (6%) [28] identified that more than 50% of the users thought that the GPS alarm helped to increase their freedom • One study (6%) [29] noted that more than 50% of the participating staff perceived that GPS trackers for persons with dementia could, to some degree, free up time for service providers by reducing the number of inspections they carried out to see if the person is well, driving to and from the user and following the user on walks; [30] identified that all persons with dementia thought that GPS trackers enabled them to increase or maintain physical activity, to increase freedom in outdoor activities, and that all relatives experienced that the GPS trackers increased their feelings of safety when leaving the person with dementia by themselves; and [31] identified positive values of GPS trackers both in shared housing for persons with dementia, including freedom for persons with dementia, decreased stress and anxiety for employees, time savings for staff and cost reduction, and for home users, including increased security, with, in some cases, increased outdoor activity and CG relief • Another study [32] identified that 5 of the 8 GPS tracker users experienced increased security and could continue to live at home for a longer periods. In addition, 5 of the 8 relatives experienced fewer concerns and worries; [34] noted that most of the users perceived that the GPS trackers provide security (for the user, CG, and staff), increase freedom for the user and sometimes also the CG, as well as help the user to be physically active and maintain their activity level; [35] identified that most of the GPS alarm users experienced that it increased their safety and freedom in daily life; and [36] identified that the GPS tracker increased the security and quality of life of persons with dementia and their CGs	• 0 (0)	—
Quality and safeguarding		• 0 (0)	—

Tiers	Findings	Studies, n (%)	Risk of bias[a]
	• Cannot be assessed from the included studies • However, 3 (19%) of the included studies had a study aim or presented data related to system-level quality and safeguarding: [30] and [34] presented the service model for implementation of GPS trackers in the homes of older adults, which included safeguarding measures taken by the municipality. Moreover, [34] presented data on the roles of different actors (users, relatives, and alarm centers) in charging and administration of the alarm as well as locating and retrieving the user, if necessary • One study (6%) [28] described the establishment of an initial test routine to encourage users to regularly trigger the alarm when out walking		
Tier 3a			

Demonstrating effectiveness in outcomes or improvements in outcomes	• Effectiveness is not demonstrated in outcomes or improvements in outcomes according to best practice standards: no increase in the frequency of OAs going outside; no significant differences in changes in fear of falling, feelings of unsafety, or quality of life [27]
	• Effectiveness is not demonstrated in outcomes or improvements in outcomes according to minimum evidence standards. Indications were identified for the following:
	  • Decrease in time searching for person with dementia (from a mean of 3-4 hours per event to 40 minutes) [24]
	  • Increase in the number of days that person with dementia was engaged in independent outdoor activity (three cases, no statistics available) [25]
	  • Decrease in role-overload of CGs of persons with dementia (P=.126; d=–0.25 for all CGs, and P=.119; d=–0.34 for CGs who could reach CR with the mobile alarm) and in feelings of worry (P=.08; d=–0.32 for all CGs, and P=.057; d=–0.46 for CGs who could reach CR with the alarm) [26]
	  • Reduction in costs for care of persons with dementia because of prolonged time that the person could live independently instead of in special housing (up to 3 months) [32,33,36]
	  • Difference in mean CG burden between relatives of persons with dementia using and not using GPS (P=.04) was indicated in small samples because a crossover design was used [37]
	• Outcomes investigated with negative results:
	  • Activity of person with dementia: reduced because of disease progression [22]
	  • Burden and quality of life for CGs of persons with dementia: no significant changes [23]
		• 1 (6)
		• 7 (44)
		• 2 (12)
			• High
			• High
			• High

Use of appropriate behavior change techniques (if relevant)	• Not relevant for GPS devices: no behavior change techniques used	• 0 (0)	—
Tier 3b			
Demonstrating effectiveness: improvements in outcomes	• Effectiveness is not demonstrated in improvements in outcomes according to minimum evidence or best practice standards (see above)	• 0 (0)	—

[a]Risk-of-bias assessment: a “high” risk of bias is equivalent to “high” for the Cochrane Risk-of-Bias Tool 2.0 (randomized studies) or “serious/critical” for the Risk of Bias in Non-randomized Studies of Interventions Tool (nonrandomized studies), and a “low” risk of bias is equivalent to “low” for the Cochrane Risk-of-Bias Tool 2.0 (randomized studies) or “low/moderate” for the Risk of Bias in Non-randomized Studies of Interventions Tool (nonrandomized studies).

[b]Not available.

[c]OA: older adult.

[d]DHT: digital health technology.

[e]CG: caregiver.

[f]CR: care receiver.

Risk of Bias

The summaries of the assessed risk of bias for individual studies that aimed to demonstrate evidence of effectiveness are presented in Table 3 (nonrandomized studies) and Table 4 (randomized study). The overall risk of bias was assessed to be serious or critical in all nonrandomized studies, particularly bias due to confounding and in measuring outcomes

[22-26,32,33,36,37]. The RCT was assessed as having a high risk of bias, in particular with regard to the blinding of patients or personnel and incomplete outcome data [27].

Table 3. Risk-of-bias assessments for individual studies investigating effectiveness according to standardized criteria for nonrandomized studies using the Risk of Bias in Non-randomized Studies of Interventions Tool. The tool's scale for ascending risk is low, moderate, serious, and critical.

Domains of bias	Study								
	Magnusson et al [22]	Megges et al [23]	Milne et al [24]	Olsson et al [25]	Pot et al [26]	Ribas Miquel et al [37]	Dahlberg [32]	Malmquist [33]	Vidensförmidling and Syd [36]
Bias due to confounding	Serious	Serious	Critical	Moderate	Serious	Serious	Critical	Critical	Critical
Bias in selection of participants	Low	Low	Critical	Serious	Low	Serious	No information	Serious	No information
Bias in classification of interventions	Low	Low	Low	Low	Low	Moderate	No information	Moderate	Moderate
Bias in deviations from intervention	Low	No information	Low	Low	Serious	No information	Serious	No information	No information
Bias due to missing data	Critical	Moderate	Serious	Low	Moderate	No information	No information	Low	No information
Bias in measurement of outcomes	Serious	Serious	Serious	Serious	Serious	Serious	Moderate	Critical	Serious
Bias in selective reporting	Low	Moderate	Serious	Low	Low	Low	Serious	Serious	Serious
Overall bias	Critical	Serious	Critical	Serious	Serious	Serious	Critical	Critical	Critical
Comments or direction of bias	Unpredictable	Unpredictable	Unpredictable	Unpredictable	Unpredictable	Unpredictable	Unpredictable	Unpredictable	Unpredictable

Table 4. Risk-of-bias assessments for individual studies investigating effectiveness according to standardized criteria for randomized studies using the Cochrane Risk-of-Bias Tool 2.0. Bias is assessed as low, high, or unsure (when sufficient information is not available to allow assessment).

Risk of bias domain	Study (Scheffer et al [27])
Random sequence generation	Low
Allocation concealment	Low
Blinding of participants or personnel	High
Blinding of outcome assessment	Unsure
Incomplete outcome data	High
Selective reporting	Low
Other bias	Unsure
Comments	High dropout, especially in the intervention group where the provided foremost reason was that the participants found the alarm too big and heavy

Discussion

Principal Findings

GPS alarms are implemented in social care with the aim of supporting users in independent activities of daily living, particularly outdoors. This systematic review included 16 studies investigating the effects of GPS alarms on health, welfare, and social provision in older adult care. The review demonstrates best practice evidence in peer-reviewed and gray literature for two of the tier 1 evidence categories (*Relevance to current pathways in health/social care system* and *Acceptability with users*) of the NICE evidence standards framework for DHTs [8]. The approach of using an existing framework to assess the quality of studies is relevant for all types of DHTs. This review also presents evidence of the minimum standard for the tier 1 category *Credibility with health, social care professionals*, although the minimum standard is insufficient for high-risk DHTs such as technologies that track patient location. More specifically, the studies in this review showed that social care professionals have been involved in the testing of alarm systems; that GPS trackers have been successfully piloted or implemented in the Swedish, Norwegian, and Danish social care systems; that representatives from the intended user groups (persons with dementia and older adults) have been involved in testing the alarm systems; and that users were satisfied with the alarm systems. Although a number of studies reported findings regarding use, value, and measures for safeguarding at specific time points, our review concluded that evidence categories for tier 2 could not be assessed from the included studies. Finally,

this review identified a lack of clear evidence for effectiveness according to the standards of evidence categories in tiers 3a and 3b in the NICE framework, which is required for DHTs that track patient location [8]. Moreover, the overall risk of bias of the included studies that evaluated effectiveness was assessed to be high. Therefore, the study findings should be interpreted with caution.

Some of the included studies demonstrated negative or absent effects on healthy behaviors and on the users' health and welfare. For example, the only RCT included in our review demonstrated no effect of GPS trackers on older adults' frequency of going out, feelings of unsafety, or fear of falling, and the authors stated that "some of the participants did not take the mobile alarm outside with them at all times. This might have been caused by perceived user - unfriendliness of the alarm [27].

Moreover, a nonrandomized intervention study demonstrated that activity among persons with dementia was reduced during the intervention and concluded that this was due most likely to disease progression [22]. The need to identify primary user groups has been addressed in previous research on GPS alarms for older adults [10]. Both examples illustrate the complexity of this type of intervention in populations of persons with dementia, which might increase the risk of bias due to deviations from the interventions and to missing data, respectively. Additional potential challenges for RCTs investigating the use of GPS alarms in persons with dementia were identified in a feasibility study [24] and included challenges in finding social care staff willing to recruit participants, randomizing the participants, and finding participants or CGs willing to participate as controls [24].

It should be pointed out that GPS trackers and the context of use varied among the studies included in this review; simply put, these studies evaluated different wearables, technical infrastructures, and supporting services. For example, some studies reported on insufficient usability of the trackers, which might have limited their use. However, the rapid development of mobile health systems has enabled the incorporation of GPS alarms in discrete wearables such as bracelets and pendants or consumer products that older adults are already using in daily life. These more sophisticated wearables might help to overcome the obstacles of poor usability and the potential stigma of older GPS devices such as mobile trackers attached to the belt or additional mobile phones. Hence, the usability, acceptability, and desirability of GPS wearables need to be addressed in future research [38]. Other aspects that are relevant to investigate in future studies of GPS use include the system's ease of use (including usability and learnability) and the users' readiness (including expectations [39] and eHealth literacy [40]). In addition, involving user groups in the design of user interfaces and supporting services can be effective in the development of new systems that meet the needs of user groups and the context of use [41,42].

The alarm systems studied have been implemented to varying degrees in the social care systems of different countries. User and CG support services also varied across the studies: in some cases, CGs were the sole alarm receivers, whereas in others, resources from a security company or the municipality have supported the CGs.

Interestingly, economic effects are likely to be affected by the value models and organization of national social care systems. In this review, the heterogeneity of the studies in terms of population, context, and systems was high; therefore, the generalizability of the results is uncertain.

Further research on service development, implementation of information and communication technologies in the public sector, profit realization, processes for change, contribution to political goals, business models, and health and welfare has been recommended to support the implementation of GPS alarms in older adult care [10]. Studies in this review that were found in gray literature have demonstrated progress in these areas (eg, usable and acceptable systems that were properly integrated in the social care services of Nordic countries). These studies were conducted as a part of larger projects—often within national programs to promote the use of HWTs among older adults—that aimed to study and develop products, services, and methods for implementation.

Notably, the studies included in this review identified outcomes that might be relevant for future studies investigating the effectiveness of GPS-based mobile alarms on the health and welfare of older adults and social care provision. Furthermore, outcomes related to the health and welfare of CGs of persons with dementia might be relevant for further investigation [23,26]. Other outcomes related more to efficiency, such as time spent searching for persons with dementia who were lost [24], may be less useful in assessing effectiveness and should therefore be considered as complementary or secondary.

This review identified a need to establish evidence in several evidence categories in all tiers of the NICE evidence framework. Evidence for tiers 1 and 2 might be identified from several sources, including product documentation from suppliers of HWT products; expert authorities; and initiatives for service development, piloting, and implementation of HWT products. However, for tiers 3a and 3b, further research on the effectiveness of GPS-based mobile alarms is still needed.

Limitations

A potential limitation of this review is the exclusion of studies with a qualitative design as well as technical validation studies. The findings from these types of studies might have been relevant for providing evidence for tiers 1 and 2. The study findings were assessed using the NICE evidence standards framework for DHTs [8], which includes evidence categories (especially in tiers 3a and 3b) that do not support qualitative studies. However, other evaluation strategies exist [43]. In addition, the searches were restricted to the English and Nordic languages, thus excluding studies published in other languages.

Comparison With Prior Work

The findings of this review are in line with previous reviews of GPS tracker use among older adults (including persons with dementia), which identified limited evidence on the effectiveness of GPS tracker use on health-related outcomes, for example [10-12]. The previous reviews provided a broader picture of

state-of-the-art research on GPS tracker use in older adult care [10] and for managing wandering behavior among persons with dementia [11]. For example, the study by Røhne et al [10] categorized findings in different research areas from 74 articles published from 1998 to 2016. The study by Røhne et al [10] concluded that research on GPS had increased significantly by 2014-2015, most likely because of the increased use of smartphones, wearables, and other health technologies, and had concentrated, up to that time, on identifying the primary user groups and their needs and experiences. Moreover, the study by Røhne et al [10] identified knowledge gaps related to service development, implementation of information and communication technologies in the public sector, profit realization, change processes, contributions to political objectives, business models, and health and well-being that need to be filled for location services to be used as ordinary services in Norwegian municipalities. This systematic review could not identify a large increase in published studies of the effectiveness of GPS alarms on health, welfare, and social care outcomes after 2014. On the contrary, the publication years of the studies included in this review are rather evenly distributed over the search period (two each in 2012 and in the years 2014-2017, one each in 2011 and 2018, and four in 2013). It is worth mentioning that the included economic evaluations were published in 2011 and 2013.

Moreover, the scoping review by Neubauer et al [11], which included studies that used any type of study design or methodology with positive or negative results, identified that the most commonly used wander management technology—from scholarly and gray literature—was GPS, followed by alarms and sensors, with only 22% of the devices clinically tested in home or institution settings. The review by Neubauer et al [11] concluded that further research is needed to identify technologies with high levels of evidence for effectiveness and usability. The seven studies on mobile locators included in the review by Neubauer et al [11] were published in the period 2000-2012. Interestingly, no overlap in the included studies was identified between the review by Neubauer et al [11] and this review. However, this review had a small overlap of four studies with a synthesizing review by Bartlett et al [12], which included empirical studies of persons with dementia or their family CGs or both using GPS. A total of 23 studies published in the period 2007-2016 were included, and data were synthesized across three identified themes: using GPS to prevent harm and promote well-being; taking control; and value of GPS data [12]. The review by Bartlett et al [12] found only nontrial evidence and identified a lack of large-scale studies. This review therefore complements the previous reviews in several aspects: first, by analyzing data from different studies, mainly because of the inclusion of only quantitative results; second, by contributing results published after 2016; and third, by reviewing the current evidence on GPS trackers according to the requirements of the NICE framework. So far, the NICE framework has been used for examples of case studies that demonstrate evidence of the effectiveness and economic value of a number of DHTs [9]. These studies are based largely on information provided by the developers that has not been independently verified. Study results published in peer-reviewed and gray literature can therefore strengthen the validity of evidence of DHTs. We have not been able to identify prior systematic reviews of evidence related to HWTs that have applied the NICE framework to the study findings. However, one study has used the framework for classifying medical mobile apps [44]. We envision that the number of studies that apply the NICE framework for various purposes will increase in the future and contribute to demonstrating evidence related to HWTs.

Implications for Clinical Practice and Future Research

This review has identified a need for further research to provide the required evidence for the effects of the use of GPS-based mobile alarms on the health and welfare of older adults and social care provision. On the basis of the NICE evidence standards framework, examples of successful piloting of GPS alarms in social care systems and of testing that involved older adults and CGs have been identified.

The results can be beneficial to the social care organizations that see the potential in GPS alarms to support older citizens' independence in daily living activities. First, we demonstrated that there is a lack of evidence of the clinical effectiveness of GPS trackers in the care of older adults. Hence, the ongoing implementations of GPS trackers have not been based on evidence of their clinical effectiveness. Second, we clarified the risks of implementing patient-locating systems such as GPS with insufficient evidence based on the NICE evidence framework. On the basis of the framework, the types of potential risks associated with the use of HWTs with insufficient evidence (eg, safety risks for the user and inefficient use of resources) can be identified. Social care organizations can use this knowledge to make informed decisions on whether they should wait for, or demand, more evidence before they start using new HWTs. Hence, increased awareness of what can be gained from better evidence is important. Third, we presented experiences from successful piloting of GPS trackers in social care systems, for example, with regard to suitable user groups, usable and acceptable GPS solutions, customization and development of products and services, processes for decision-making, and implementation. The level of evidence regarding effectiveness and economic value needs to be considered in decision-making processes on the implementation of HWTs when potential benefits and risks need to be balanced. After all, money spent on improving the health of older adults needs to be spent wisely and efficiently.

This review has identified critical knowledge gaps that need to be addressed in future research, most importantly with regard to clinical effectiveness. Open questions include the clinical effectiveness of specific GPS trackers for certain user groups with defined supporting services. The included studies can contribute with positive examples of implementation on which to base future research (eg, products, user groups, and service models) and with challenges that have been experienced in previous research. Potential evolutions of this review are future studies on the clinical effectiveness of already implemented or piloted GPS trackers in their *real-world* implementation environment. Thus, the interventions to be evaluated should include user groups and service models that have been identified and developed in the previous implementation or pilot projects.

In addition, generating real-world evidence during the implementation of GPS trackers in social care systems seems to be of great importance in addressing the evidence gaps.

Conclusions

There is insufficient evidence for the effects of the use of GPS-based mobile alarms on the health and welfare of older adults, as well as social care provision. Best practice evidence for two of the tier 1 evidence categories and evidence of minimum standard for the tier 1 category of the NICE evidence standards framework for DHTs [8] were identified. This review identified that social care professionals have been involved in the testing of GPS trackers, that GPS trackers have been successfully piloted or implemented in the Nordic social care systems, that representatives from the intended user groups (persons with dementia and older adults) have been involved in testing the devices, and that users were satisfied with them. Although a number of studies have contributed data demonstrating use, value, and measures for safeguarding at specific time points, our review concluded that the evidence categories for tier 2 could not be assessed from the included studies. Finally, this review demonstrated a lack of clear evidence of effectiveness according to the standards of evidence categories in tiers 3a and 3b in the NICE framework, which is required for DHTs that track patient location [8]. Future research needs to address clinical effectiveness broadly and incorporate aspects related to products, user groups, service models, and challenges in social care systems in the *real world*.

Acknowledgments

This research is funded by Mälardalen University and the Sörmlandskontraktet regional funding partnership.

Authors' Contributions

ME conducted the literature assessment, data extraction, and risk-of-bias analysis together with MR; summarized the results and assessed them by using the NICE framework; and wrote the manuscript with input from all coauthors. MR conducted the literature assessment, data extraction, and risk-of-bias analysis together with ME and reviewed the study findings and their application by using the NICE framework. SLS contributed search strategies, performed literature searches, documented the search results, prepared the references for screening, and resolved conflicts that arose during the screening. SWA conceptualized the research questions. KR made contributions related to the NICE framework. All authors contributed to the study conceptualization, scientific discussions, and the manuscript review, and all have accepted the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of the search process and the results.
[DOCX File , 19 KB - jmir_v23i10e27267_app1.docx]

Multimedia Appendix 2
Complete database-specific search strategies for the initial searches of scientific and gray literature.
[DOCX File , 22 KB - jmir_v23i10e27267_app2.docx]

Multimedia Appendix 3
Study findings related to the requirements of minimum evidence and best practice standards of each evidence category in the NICE evidence standards framework. NICE: National Institute for Health and Care Excellence.
[DOCX File , 34 KB - jmir_v23i10e27267_app3.docx]

References

1. Andersson WS. International health and welfare technology webinar series. Malardalen University Sweden. 2020. URL: https://www.mdh.se/en/malardalen-university/research/health-and-welfare-/hwt-research-webinar-series [accessed 2021-06-15]
2. Centre N. Welfare technology. Nordic Welfare Centre. URL: https://nordicwelfare.org/en/welfare-policy/welfare-technology/ [accessed 2021-06-15]
3. WHO Guideline: Recommendations on Digital Interventions for Health System Strengthening. World Health Organization. 2019. URL: https://www.who.int/reproductivehealth/publications/digital-interventions-health-system-strengthening/en/ [accessed 2021-06-15]
4. The global dementia observatory. World Health Organization. 2018. URL: https://www.who.int/publications/i/item/who-msd-mer-18.1 [accessed 2021-06-15]
5. Technologies in care for older people. European Parliamentary Technology Assesment. 2019. URL: https://eptanetwork.org/images/documents/minutes/EPTA_report_2019.pdf [accessed 2021-06-15]

6. Frennert S, Baudin K. The concept of welfare technology in Swedish municipal eldercare. Disabil Rehabil 2019 Sep 10;43(9):1220-1227. [doi: 10.1080/09638288.2019.1661035] [Medline: 31503509]
7. Professionalisera införandet av ny hälso- och välfärdsteknik. Dagens Medicin. 2020. URL: https://www.dagensmedicin.se/opinion/debatt/professionalisera-inforandet-av-ny-halso-och-valfardsteknik/ [accessed 2021-06-15]
8. Evidence standards framework for digital health technologies. National Institute for Health and Care Excellence. 2018 Dec. URL: https://www.nice.org.uk/corporate/ecd7 [accessed 2021-06-15]
9. Craig JS, Russell J. Examples of effectiveness and economical digital health case studies. York Health Economics Consortium (YHEC). 2019. URL: https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/evidence-case-studies.pdf [accessed 2021-06-15]
10. Røhne M, Tone D, Øystein D. GPS som varslings-og lokaliseringsteknologi i helse og omsorg. SINTEF. 2017. URL: https://sintef.brage.unit.no/sintef-xmlui/handle/11250/2478904 [accessed 2021-06-15]
11. Neubauer NA, Lapierre N, Ríos-Rincón A, Miguel-Cruz A, Rousseau J, Liu L. What do we know about technologies for dementia-related wandering? A scoping review: examen de la portée : que savons-nous à propos des technologies de gestion de l'errance liée à la démence? Can J Occup Ther 2018 Jun;85(3):196-208. [doi: 10.1177/0008417418777530] [Medline: 29972049]
12. Bartlett R, Brannelly T, Topo P. Using gps technologies with people with dementia: a synthesising review and recommendations for future practice. Tidsskrift for omsorgsforskning 2019 Dec;5(3):84-98. [doi: 10.18261/issn.2387-5984-2019-03-08]
13. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
14. Jalali S, Wohlin C. Systematic literature studies: database searches vs. backward snowballing. In: Proceedings of the ACM-IEEE international symposium on Empirical software engineering and measurement. 2012 Presented at: ACM-IEEE international symposium on Empirical software engineering and measurement; Sep 19-20, 2012; Lund, Sweden. [doi: 10.1145/2372251.2372257]
15. Wohlin C, Mendes E, Felizardo K, Kalinowski M. Guidelines for the search strategy to update systematic literature reviews in software engineering. Inf Softw Technol 2020 Nov;127:106366 [FREE Full text] [doi: 10.1016/j.infsof.2020.106366]
16. Carswell W, McCullagh PJ, Augusto JC, Martin S, Mulvenna MD, Zheng H, et al. A review of the role of assistive technology for people with dementia in the hours of darkness. Technol Health Care 2009;17(4):281-304. [doi: 10.3233/THC-2009-0553] [Medline: 19822946]
17. Hall A, Wilson CB, Stanmore E, Todd C. Implementing monitoring technologies in care homes for people with dementia: A qualitative exploration using Normalization Process Theory. Int J Nurs Stud 2017 Jul;72:60-70 [FREE Full text] [doi: 10.1016/j.ijnurstu.2017.04.008] [Medline: 28494333]
18. Landau R, Auslander GK, Werner S, Shoval N, Heinik J. Who should make the decision on the use of GPS for people with dementia? Aging Ment Health 2011 Jan;15(1):78-84. [doi: 10.1080/13607861003713166] [Medline: 20924823]
19. Larsson P, Gill M, Lind MS. Framtidens teknik i omsorgens tjänst. SOU. 2020. URL: https://www.regeringen.se/494156/contentassets/576aa4588db340b0ad052537ae90511d/framtidens-teknik-i-omsorgens-tjanst-sou-2020_14.pdf [accessed 2021-10-07]
20. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
21. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016 Oct 12;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
22. Magnusson L, Sandman L, Gustav RK, Hanson E. Extended safety and support systems for people with dementia living at home. J Assist Technol 2014 Dec 9;8(4):188-206.
23. Megges H, Freiesleben SD, Rösch C, Knoll N, Wessel L, Peters O. User experience and clinical effectiveness with two wearable global positioning system devices in home dementia care. Alzheimers Dement (N Y) 2018 Nov 9;4:636-644 [FREE Full text] [doi: 10.1016/j.trci.2018.10.002] [Medline: 30519629]
24. Milne H, van der Pol M, McCloughan L, Hanley J, Mead G, Starr J, et al. The use of global positional satellite location in dementia: a feasibility study for a randomised controlled trial. BMC Psychiatry 2014 May 30;14:160 [FREE Full text] [doi: 10.1186/1471-244X-14-160] [Medline: 24885489]
25. Olsson A, Engström M, Åsenlöf P, Skovdahl K, Lampic C. Effects of tracking technology on daily life of persons with dementia: three experimental single-case studies. Am J Alzheimers Dis Other Demen 2015 Feb;30(1):29-40 [FREE Full text] [doi: 10.1177/1533317514531441] [Medline: 24771764]
26. Pot AM, Willemse BM, Horjus S. A pilot study on the use of tracking technology: feasibility, acceptability, and benefits for people in early stages of dementia and their informal caregivers. Aging Ment Health 2012;16(1):127-134. [doi: 10.1080/13607863.2011.596810] [Medline: 21780960]

27. Scheffer AC, Scholte op Reimer WJ, van Dijk N, van Munster BC, Abu-Hanna A, Levi M, et al. Effect of a mobile safety alarm on going outside, feeling safe, fear of falling, and quality of life in community-living older persons: a randomized controlled trial. J Am Geriatr Soc 2012 May;60(5):987-989. [doi: 10.1111/j.1532-5415.2012.03932.x] [Medline: 22587862]
28. Røhne M, Boysen ES, Ausen D. Wearable and mobile technology for safe and active living. Stud Health Technol Inform 2017;237:133-139. [Medline: 28479556]
29. Sørli A. Effektstudie ved bruk av GPS til personer med demens. NTNU Open. 2013. URL: https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/143676 [accessed 2021-06-15]
30. Ausen D, Svagård I, Øderud T, Sørensen E, Stanarevic M. Selvstendig, trygg og aktiv i Larvik. Erfaringer fra velferdsteknologiprosjektet i Larvik kommune. SINTEF. 2016. URL: https://sintef.brage.unit.no/sintef-xmlui/handle/11250/2421428 [accessed 2021-06-15]
31. Boysen E, Støle N. Erfaringer og funn fra pilotering av velferdsteknologi i Stavanger kommune. SINTEF. 2016. URL: https://sintef.brage.unit.no/sintef-xmlui/handle/11250/2421431 [accessed 2021-06-15]
32. Dahlberg Å. Nyttokostnadsanalys vid införande av välfärdsteknologi -exemplet Posifon. Hjälpmedelsinstitutet. 2013. URL: https://posifon.se/wp-content/uploads/2016/12/Nyttokostnadsanalys-inforande-valfardsteknologi.pdf [accessed 2021-06-15]
33. Malmquist C. Utvärdering av projekt Förskrivning av larm i Östersund-ekonomi och proces. Östersund: payoff AB. 2013. URL: https://docplayer.se/3797329-Utvardering-av-projekt-forskrivning-av-larm-i-ostersund-ekonomi-och-process-ostersund-2013-01-03.html [accessed 2021-06-15]
34. Øderud T, Ausen D, Aketun S, Thorgersen M. GPS for trygghet, frihet og mestring. Fra prosjekt til drift-Bruk av GPS for lokalisering av personer med demens/kognitiv svikt. Sintef Rapport. 2017. URL: https://sintef.brage.unit.no/sintef-xmlui/handle/11250/2489904 [accessed 2021-06-15]
35. Røhne M, Svagård I, Ausen D, Fossberg AB, Husebø I, Øverli T. Bo lenger hjemme med mobil trygghetsalarm? In: Erfaringer med mobil trygghetsalarm i Bærum kommune. Norway: SINTEF; 2015.
36. Vidensförmidling and Syd. Demonstrationsprojekt med brug af GPS system i eget hjem. 2011. URL: https://docplayer.dk/10583158-Demonstrationsprojekt-med-brug-af-gps-system-i-eget-hjem.html [accessed 2021-06-15]
37. Ribas Miquel G, Pérez González A, Beltran Vilella M, Boix Roqueta E, Ferré Munté M, Reig Garcia G, et al. Uso de localizadores para personas mayores. Ariadna 2013;1(1):68-72. [doi: 10.6035/ariadna.2013.1.11]
38. Bichard J, Coleman R, Langdon P. Does my stigma look big in this? Considering acceptability and desirability in the inclusive design of technology products. In: Universal Acess in Human Computer Interaction. Berlin, Heidelberg: Springer; 2007.
39. Venkatesh V, Morris M, Davis G, Davis F. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425-478 [FREE Full text] [doi: 10.2307/30036540]
40. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan 27;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
41. Abras CM, Preece J. User-centered design. In: Encyclopedia of Human-Computer Interaction. Thousand Oaks, CA, US: Sage Publications; 2004.
42. Norman D, Draper. In: Norman D, Draper S, editors. User-Centered System Design: New Perspectives on Human-computer Interaction. Hillsdale, NJ, US: CRC Press; 2018.
43. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. NPJ Digit Med 2019 May 13;2:38 [FREE Full text] [doi: 10.1038/s41746-019-0111-3] [Medline: 31304384]
44. Nwe K, Larsen ME, Nelissen N, Wong DC. Medical mobile app classification using the national institute for health and care excellence evidence standards framework for digital health technologies: interrater reliability study. J Med Internet Res 2020 Jun 05;22(6):e17457 [FREE Full text] [doi: 10.2196/17457] [Medline: 32501271]

Abbreviations

CG: caregiver
DHT: digital health technology
HWT: health and welfare technology
NICE: National Institute for Health and Care Excellence
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QoL: quality of life
RCT: randomized controlled trial

Edited by R Kukafka; submitted 19.01.21; peer-reviewed by S Frennert, V Stara; comments to author 30.03.21; revised version received 21.04.21; accepted 14.06.21; published 11.10.21.

Please cite as:
Ehn M, Richardson MX, Landerdahl Stridsberg S, Redekop K, Wamala-Andersson S
Mobile Safety Alarms Based on GPS Technology in the Care of Older Adults: Systematic Review of Evidence Based on a General Evidence Framework for Digital Health Technologies
J Med Internet Res 2021;23(10):e27267
URL: https://www.jmir.org/2021/10/e27267
doi:10.2196/27267
PMID:34633291

Original Paper

Accuracy and Diversity of Wearable Device–Based Gait Speed Measurement Among Older Men: Observational Study

Min-gu Kang[1], MSc, MD; Seong-Ji Kang[2,3], MSc, MD; Hye-Kang Roh[3], MSc; Hwa-Young Jung[3], MSc; Sun-wook Kim[4], MSc, MD; Jung-Yeon Choi[4], MD, PhD; Kwang-il Kim[4,5], MD, PhD

[1]Department of Internal Medicine, Chonnam National University Bitgoeul Hospital, Gwangju, Republic of Korea

[2]Graduate School of Health Science and Management, Yonsei University, Seoul, Republic of Korea

[3]WELT Corp, Ltd, Seoul, Korea, Seoul, Republic of Korea

[4]Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam, Republic of Korea

[5]Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea

Corresponding Author:
Kwang-il Kim, MD, PhD
Department of Internal Medicine
Seoul National University Bundang Hospital
82 Gumi-Ro 173 Beon-Gil, Seongnamsi Gyeonggi-Do
Seongnam, 13620
Republic of Korea
Phone: 82 317877032
Email: kikim907@snu.ac.kr

Abstract

Background: Gait speed measurements are widely used in clinical practice, as slow gait is a major predictor of frailty and a diagnostic criterion for sarcopenia. With the development of wearable devices, it is possible to estimate the gait speed in daily life by simply wearing the device.

Objective: This study aims to accurately determine the characteristics of daily life gait speed and analyze their association with sarcopenia.

Methods: We invited community-dwelling men aged >50 years who had visited the outpatient clinic at a tertiary university hospital to participate in the study. Daily life gait speed was assessed using a wearable smart belt (WELT) for a period of 4 weeks. Data from participants who wore the smart belt for at least 10 days during this period were included. After 4 weeks, data from a survey about medical and social history, *usual gait speed* measurements, handgrip strength measurements, and dual-energy x-ray absorptiometry were analyzed.

Results: A total of 217,578 daily life gait speed measurements from 106 participants (mean age 71.1, SD 7.6 years) were analyzed. The mean daily life gait speed was 1.23 (SD 0.26) m/s. The daily life gait speed of the participants varied according to the time of the day and day of the week. Daily life gait speed significantly slowed down with age (P<.001). Participants with sarcopenia had significantly lower mean daily life gait speed (mean 1.12, SD 0.11 m/s) than participants without sarcopenia (mean 1.23, SD 0.08 m/s; P<.001). Analysis of factors related to mean daily life gait speed showed that age and skeletal muscle mass of the lower limbs were significantly associated characteristics.

Conclusions: More diverse and accurate information about gait speed can be obtained by measuring daily life gait speed using a wearable device over an appropriate period, compared with one-time measurements performed in a laboratory setting. Importantly, in addition to age, daily life gait speed is significantly associated with skeletal muscle mass of the lower limbs.

(J Med Internet Res 2021;23(10):e29884) doi:10.2196/29884

KEYWORDS

gait speed; sarcopenia; skeletal muscle mass; wearable device

Introduction

Background

Globally, populations are aging, including in Korea, where the population is aging rapidly because of extended life expectancy [1] and a low fertility rate. In 2017, more than 14% of the total population in Korea was aged ≥65 years.

Frailty, as a reflection of decreased physiological reserve, is closely associated with increased biological age [2], concurrent medical conditions, morbidity, and decreased survival in older individuals [3]. Frailty assessments are clinically useful for determining the heterogeneous health status of older individuals [4]. Sarcopenia, defined as low muscle mass and low muscle strength, is a key characteristic of frailty [5-7]. Recently, the treatment of sarcopenia has become a key strategy for preventing and overcoming frailty [8].

Slow gait speed is a major feature of frailty [9] and a diagnostic criterion for sarcopenia [6,7]; therefore, gait speed measurement is widely used to assess frailty. However, gait speed measurement methods are not completely standardized, and the term *usual gait speed* refers to a measure that has been commonly used as part of this fast, safe, and inexpensive assessment [10,11]. *Usual gait speed* is calculated by instructing the individual to walk a certain distance at their usual pace in the laboratory and then measuring the time it takes to cover that distance.

Currently, technology is advancing at a rapid pace in the area of wearable devices, which can be used to measure a variety of attributes. Wearable sensors allow frequent and continuous measurements of body functions, including various physical activities, such as walking, running, and biking [12]. With the availability of these wearable devices, it is possible to measure and track daily life gait speed in the real world without significant additional effort by asking the patient to wear the device [13,14]. Moreover, it is questionable whether the *usual gait speed* measured in the laboratory can represent daily gait speed in an individual's real world [15]. Few studies have compared *usual gait speed* and real-world gait speed in analyses of the associations between muscle mass or muscle strength and gait speed.

Objectives

The primary aim of this study was to identify the characteristics of daily life gait speed in community-dwelling older male adults and to analyze the association of these characteristics with sarcopenia. The secondary objective was to compare the results of daily life gait speed obtained using a wearable device with *usual gait speed* measured in the laboratory by analyzing the associations of muscle mass and muscle strength with gait speed.

Methods

Study Population

This observational study was conducted at Seoul National University Bundang Hospital. Men aged >50 years who could walk unassisted were recruited consecutively from November 16, 2018, to April 12, 2019.

Study Protocol

The WELT (WELT Corp, Ltd) is a belt-type wearable device that uses a triaxial accelerometer to continuously measure the wearer's gait speed while walking. The smart belt measures the walking speed using the step interval time and stride length. An algorithm for detecting and analyzing peaks is used to identify the steps. Two consecutive peaks are detected as steps occur, and the step interval time between the peaks is measured in 0.1-second increments. There are several ways to estimate stride length, and WELT adopts a method that uses a sex-based constant and height [16,17]. As a result, it is possible to continuously measure gait speed that changes with the step interval time.

Participants were asked to wear the WELT for 4 weeks as they went about their daily lives, and their daily life gait speed was recorded. After 4 weeks, a survey that included questions about social history and past medical history was administered, *usual gait speed* and handgrip strength measurements were obtained, and a dual-energy x-ray absorptiometry (DEXA) was performed.

Usual gait speed was calculated for each participant using distance in meters and time in seconds. A marked walkway and an automated laser-gated chronometer attached to the wall were used to calculate gait speed [18]. The walkway consisted of a 1-m acceleration phase, 4.5-m timed section, and a 1-m deceleration phase. The chronometer started and stopped automatically as each participant crossed into and out of the timed section. The participants were instructed to walk at their usual pace. Handgrip strength was measured using a Jamar Plus digital hand dynamometer (Patterson Medical). Participants were seated in a chair with their shoulder adducted and neutrally rotated, their elbow flexed to 90°, and their forearm and wrist neutrally positioned. A total of three consecutive measurements of the dominant hand were taken with a brief rest between measurements, and the average scores were obtained and recorded in kilograms. DEXA was used to measure the limb skeletal muscle mass of the participants. It is very useful for body composition analysis because fat mass, nonbone lean mass, and bone mineral content can be differentiated at both the regional and whole-body levels [19]. A Horizon W machine (Hologic Inc) was used to measure appendicular skeletal muscle mass [20]. Before the DEXA scan, participants were asked to remove all metal objects and change into hospital gowns. Scanning was performed by an imaging technician in the laboratory. The scan time was approximately 10 minutes. Appendicular skeletal muscle mass was calculated as the sum of the muscle mass of both arms and legs.

Sarcopenia was defined based on the diagnostic criteria of the Asian Working Group for Sarcopenia [6]. Low muscle strength was defined as a handgrip measurement of <28 kg. Low physical performance was defined as a *usual gait speed* of <1.0 m/s. Sarcopenia was diagnosed when the individual demonstrated a low appendicular skeletal muscle mass and low muscle strength or low physical performance.

Daily life gait speed was analyzed using data from participants who wore the wearable device for at least 10 days during the 4-week study period. The characteristics of daily life gait speed in the real world were identified, and the association between

daily life gait speed and sarcopenia was analyzed. In addition, *usual gait speed* and daily life gait speed were compared by an analysis of the association between muscle mass or muscle strength and gait speed.

Statistical Analysis

All statistical analyses were performed using SPSS version 21.0 (IBM Corporation). Continuous variables were expressed as mean (SD) and were compared using either a two-tailed unpaired *t* test or one-way analysis of variance. Discrete variables were expressed as counts and percentages, and the proportions were compared using the chi-square test or Fisher exact test. We used correlation analysis and linear regression analysis to identify factors related to daily life gait speed. All statistical analyses were two tailed, and $P<.05$ were considered statistically significant.

Ethical Standards

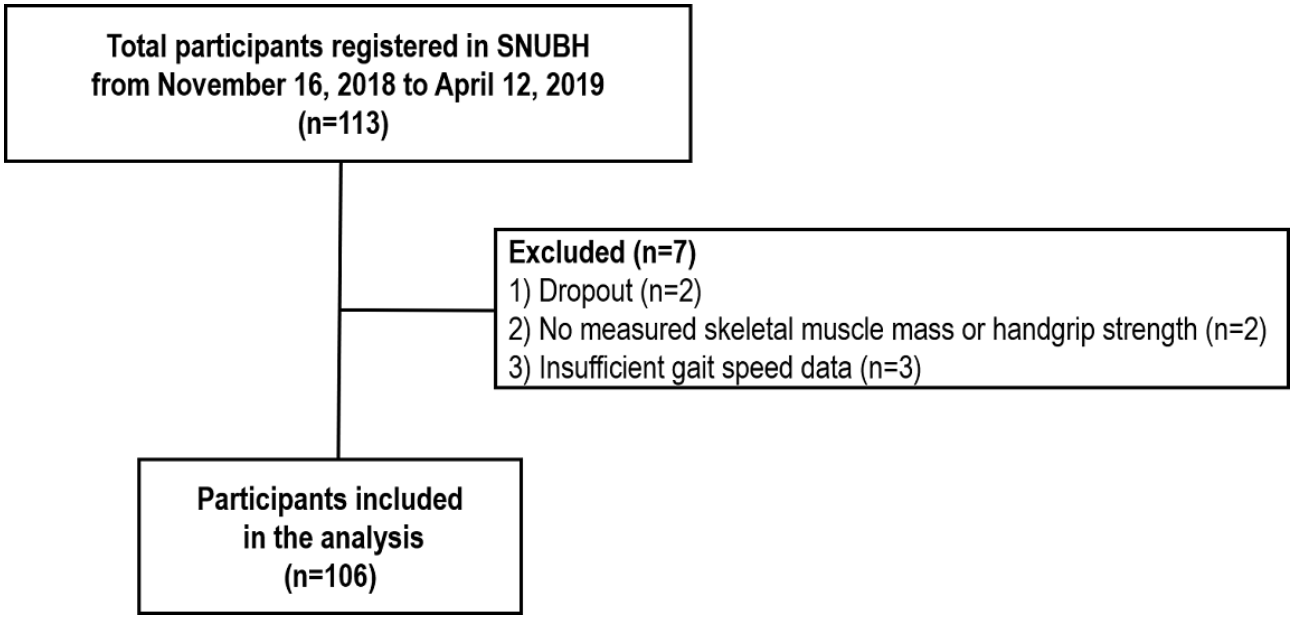
This study was approved by the institutional review board of Seoul National University Bundang Hospital (Institutional Review Board No. B-1808/486-002). All participants provided written informed consent before participation. The study was performed in accordance with the principles of the Declaration of Helsinki.

Results

Characteristics of Study Participants

A total of 217,578 daily life gait speed measurements were analyzed from 106 participants who completed the study (Figure 1). The average age of the participants was 71.1 (SD 7.6; range 52-90) years, and the average number of gait speed measurements per participant was 2052.6 (SD 1022.3; range 618-4783). The mean daily life gait speed of the participants was 1.23 (SD 0.26; range 0.94-1.39) m/s, and the average walking time per day was 88.0 (SD 40.2; range 30.5-176.5) minutes.

Figure 1. Flow diagram of the study population. SNUBH: Seoul National University Bundang Hospital.



Characteristics of Daily Life Gait Speed Measured Using a Wearable Device

The characteristics of daily life gait speed were presented in Figure 2. The daily life gait speed was significantly lower in older individuals ($P<.001$). Participants walked the fastest between 5 AM and 7 AM and the slowest at night ($P<.001$). There was a statistically significant difference in daily life gait speed according to the day of the week ($P<.001$), and weekday gait speed (mean 1.23, SD 0.26 m/s) was significantly faster than weekend gait speed (mean 1.22, SD 0.26 m/s; $P<.001$). Walking time per day differed significantly on weekdays (mean 93. 4, SD 47.0 minutes) and weekends (mean 74.5, SD 41.6 minutes; $P<.001$).

To compare daily life gait speed with *usual gait speed*, the daily life gait speed for each participant was calculated as a percentile compared with their own *usual gait speed*. The participants' daily life gait speed varied widely. The corresponding percentiles ranged from 30.3 percentile to 181.0 percentile, with a median value of 102.9 percentile. In their daily lives, a walking speed >25% faster than their own *usual gait speed* was observed in 13.46% (29,279/217,578) of the total number of measurements, and a walking speed >25% slower than their own *usual gait speed* was observed in 14.58% (31,728/217,578) of the total number of measurements. In the analysis of the correlation between daily life gait speed and *usual gait speed*, only a negligible correlation was confirmed (Pearson $r=0.155$; $P<.001$; Figure 3). However, in the analysis of the correlation between mean daily life gait speed and *usual gait speed*, a moderate positive correlation was confirmed (Pearson $r=0.504$; $P<.001$; Figure 3). In the analysis of the correlation between the SD of daily life gait speed and *usual gait speed*, only a negligible correlation was confirmed (Pearson $r=0.195$; $P=.046$; Figure 3). In addition, no significant correlation was found

between the coefficient of variation of daily life gait speed and *usual gait speed* (Pearson r=−0.139; P=.16; Figure 3).

Figure 2. The characteristics of daily life gait speed measured using a wearable device. Comparison of daily life gait speed (A) by age group; (B) by time of day; (C) by day of the week; (D) between weekdays and weekends.

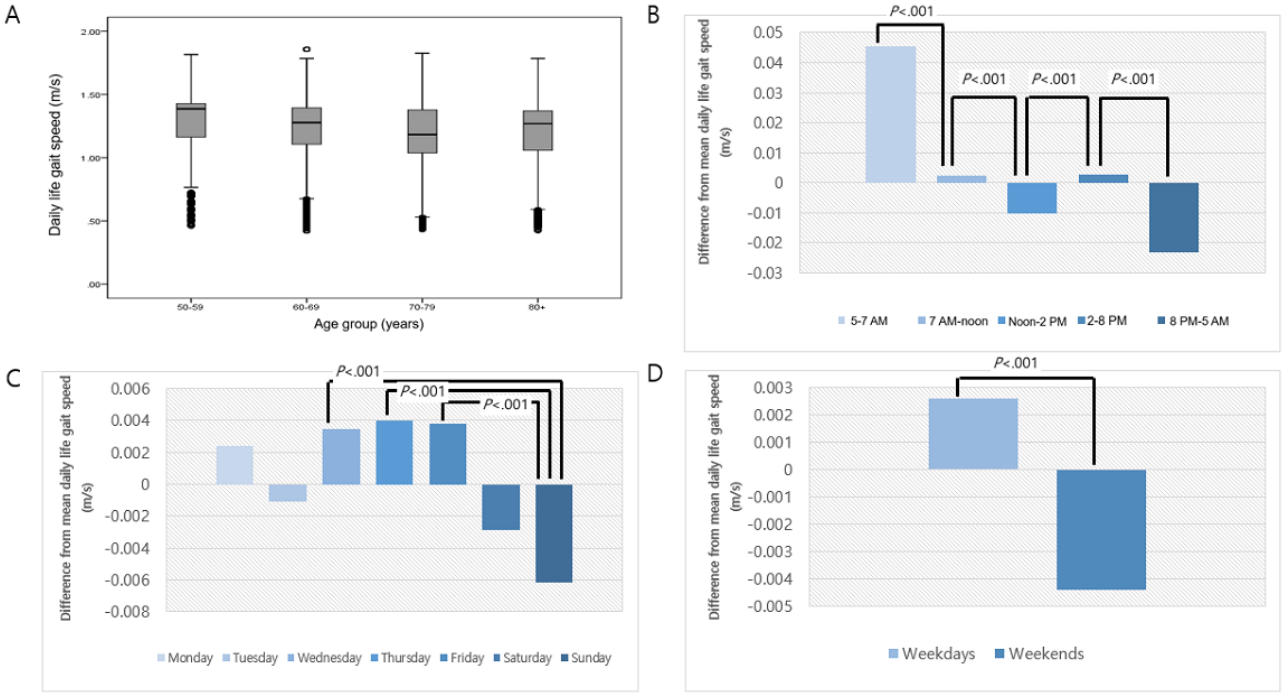
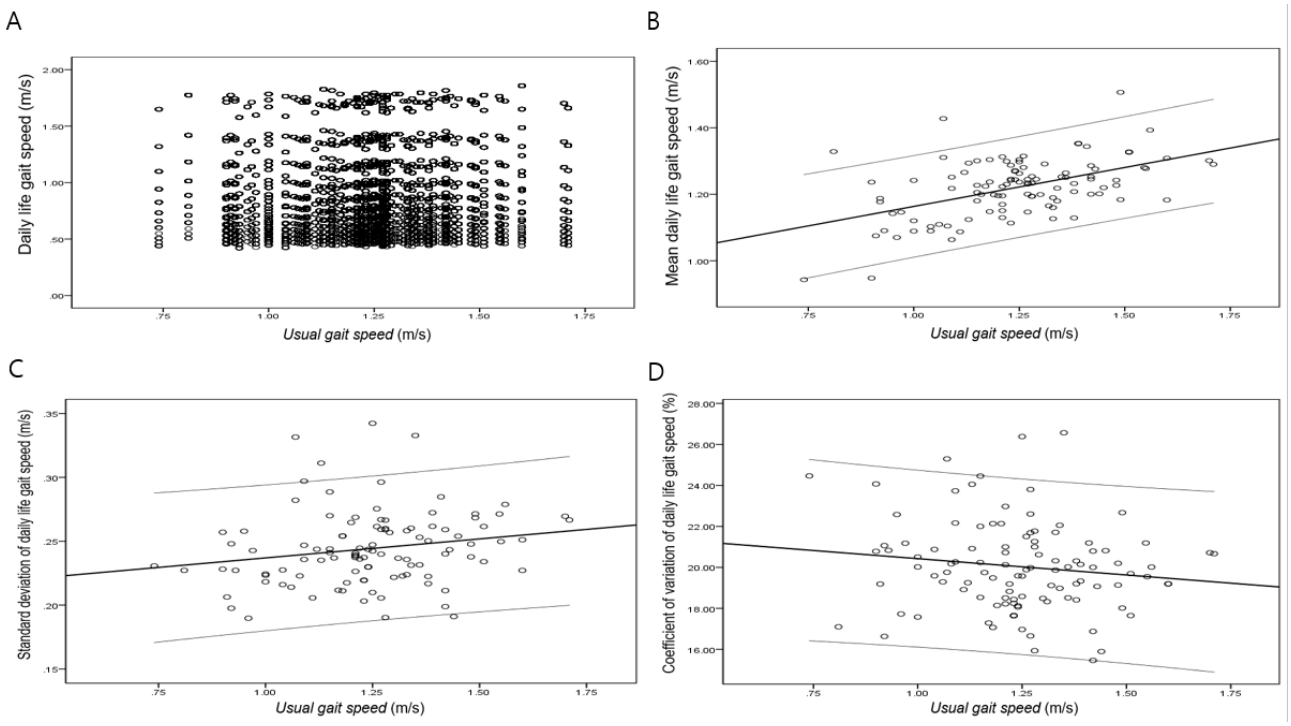


Figure 3. Scatterplot of daily life gait speed and usual gait speed. (A) Scatterplot of daily life gait speed and usual gait speed (Pearson correlation coefficient = 0.155; P<.001). (B) Scatterplot of mean daily life gait speed and usual gait speed (Pearson correlation coefficient = 0.504; P<.001). (C) Scatterplot of SD of daily life gait speed and usual gait speed (Pearson correlation coefficient = 0.195; P=.046). (D) Scatterplot of coefficient of variation of daily life gait speed and usual gait speed (Pearson correlation coefficient = −0.139; P=.16).



Comparison Between Study Participants With and Without Sarcopenia

Participants with low muscle strength (handgrip strength <28 kg) had significantly lower daily life gait speed than those with normal muscle strength (mean 1.15, SD 0.25 m/s vs mean 1.23, SD 0.26 m/s; P<.001). There were also significant differences in daily life gait speed between participants with low muscle mass (appendicular skeletal muscle mass <7.0 kg/m^2) and those

with normal muscle mass (mean 1.22, SD 0.26 m/s vs mean 1.25, SD 0.26 m/s; *P*<.001).

A total of 13 participants (13/106, 12.3%) were diagnosed with sarcopenia based on the diagnostic criteria of the Asian Working Group for Sarcopenia. Participants with sarcopenia (mean 76.0, SD 6.2 years) were older than participants without sarcopenia (mean 70.5, SD 7.6 years; *P*=.01) and had significantly lower mean daily life gait speed than normal participants (mean 1.12, SD 0.11 m/s vs mean 1.23, SD 0.08 m/s; *P*<.001). Sarcopenic participants had significantly lower handgrip strength and lower appendicular skeletal muscle mass. In addition, among participants with sarcopenia, the proportion of patients who had angina and were current smokers was significantly higher (Table 1).

Table 1. Comparison between study participants with and without sarcopenia.

Characteristics	Total (N=106)	Sarcopenia (n=13)	Normal (n=93)	*P* value
Age (years)				.01
Mean (SD)	71.1 (7.6)	76.0 (6.2)	70.5 (7.6)	
Range	52-90	67-88	52-90	
BMI (kg/m^2)				.54
Mean (SD)	24.6 (2.5)	25.0 (2.5)	24.5 (2.5)	
Range	17.0-30.8	20.7-29.2	17.0-30.8	
Education period (years)				.38
Mean (SD)	14.3 (3.2)	13.5 (3.3)	14.4 (3.2)	
Range	6-20	8-20	6-20	
Job status				.79
Incumbent	44 (41.5)	5 (38.5)	39 (41.9)	
Retired	61 (57.5)	8 (61.5)	53 (57)	
Hypertension, n (%)	67 (63.2)	11 (84.6)	56 (60.2)	.13
Diabetes mellitus, n (%)	28 (26.4)	3 (23.1)	25 (26.9)	.99
Angina, n (%)	22 (20.8)	6 (46.2)	16 (17.2)	.03
Arthritis, n (%)	18 (17)	2 (15.4)	16 (17.2)	.99
Smoking, n (%)				.004
Current smoker	16 (15.1)	6 (46.2)	10 (10.8)	
Ex-smoker	76 (71.7)	6 (46.2)	70 (75.3)	
Never smoker	14 (13.2)	1 (7.7)	13 (14)	
Usual gait speed (m/s)				<.001
Mean (SD)	1.24 (0.19)	0.97 (0.12)	1.28 (0.16)	
Range	0.74-1.71	0.74-1.20	0.91-1.71	
Daily life gait speed (m/s)				<.001
Mean (SD)	1.22 (0.09)	1.12 (0.11)	1.23 (0.08)	
Range	0.94-1.39	0.94-1.33	1.06-1.39	
Handgrip strength[a] (kg)				<.001
Mean (SD)	36.0 (6.6)	27.4 (5.4)	37.2 (5.8)	
Range	22.1-50.3	22.1-38.0	24.5-50.3	
Appendicular skeletal muscle mass/height^2 (kg/m^2)				.003
Mean (SD)	6.74 (0.74)	6.39 (0.34)	6.79 (0.77)	
Range	5.21-8.87	5.68-6.97	5.21-8.87	

[a]Average grip strength measured three times in the dominant hand.

Participants with sarcopenia walked faster from 5 AM to 7 AM than at other times of the day (*P*<.001); however, there were no significant differences between weekday and weekend gait speeds (*P*=.64). Participants with sarcopenia (mean 78.2, SD 45.4 minutes) had significantly less walking time per day than participants without sarcopenia (mean 89.3, SD 46.3 minutes;

P=.03). In addition, there was no significant difference in walking time per day between weekdays (mean 82.4, SD 46.3 minutes) and weekends (mean 67.8, SD 42.1 minutes) in the participants with sarcopenia (*P*=.17). In the analysis of the correlation between daily walking time and sarcopenia-related factors, it was observed that grip strength, skeletal muscle mass of the lower limbs, and *usual gait speed* had a negligible correlation with walking time (walking time and grip strength: Pearson *r*=0.111; *P*=.003; walking time and skeletal muscle mass of the lower limbs: Pearson *r*= 0.152; *P*<.001; and walking time and *usual gait speed*: Pearson *r*=0.112; *P*=.002). Analysis of the correlation between daily walking time and mean daily life gait speed revealed a low positive correlation (Pearson *r*=0.315; *P*<.001).

There was no significant difference between the mean daily life gait speed (mean 1.12, SD 0.11 m/s) and the median value of daily life gait speed (mean 1.11, SD 0.12 m/s; *P*=.45) in participants with sarcopenia, but those without sarcopenia had median values (mean 1.27, SD 0.13 m/s) significantly higher than the mean values (mean 1.23, SD 0.08 m/s; *P*<.001). There was no significant difference in the SD of daily life gait speed in participants with sarcopenia (mean 0.23, SD 0.02 m/s) and those without sarcopenia (mean 0.25, SD 0.03 m/s; *P*=.06). There was also no significant difference in the coefficient of variation in daily life gait speed between participants with sarcopenia (mean 20.58%, SD 2.37%) and those without sarcopenia (mean 19.96%, SD 2.13%; *P*=.34).

Factors Associated With Mean Daily Life Gait Speed

The correlation analysis of the potential factors associated with mean daily life gait speed showed that age, height, and skeletal muscle mass of the lower limbs were significantly correlated with mean daily life gait speed (Table 2). In a linear regression analysis with correlated factors, age and lower skeletal muscle mass were significantly associated with mean daily life gait speed. Participants who were younger and had more skeletal muscle mass in their lower limbs walked faster (Table 3).

Table 2. Correlation analysis (Pearson *r* and one/two-tailed *P* values) with mean daily life gait speed.[a]

Variable	Mean daily life gait speed	Age	Height	Body weight	BMI	Skeletal muscle mass of the lower limbs
Mean daily life gait speed						
r	1	−0.357	0.333	0.108	−0.074	0.383
P value	—[b]	<.001	<.001	.27	.45	<.001
Age						
r	−0.357	1	−0.274	−0.237	−0.073	−0.338
P value	<.001	—	.004	.01	.46	<.001
Height						
r	0.333	−0.274	1	0.466	−0.062	0.523
P value	<.001	.004	—	<.001	.53	<.001
Body weight						
r	0.108	−0.237	0.466	1	0.824	0.797
P value	.27	.01	<.001	—	<.001	<.001
BMI						
r	−0.074	−0.073	−0.062	0.824	1	0.596
P value	.45	.46	.53	<.001	—	<.001
Skeletal muscle mass of the lower limbs						
r	0.383	−0.338	0.523	0.797	0.596	1
P value	<.001	<.001	<.001	<.001	<.001	—

[a]The correlation is significant at a level of .001 (one/two-tailed).

[b]Not applicable.

Table 3. Linear regression analysis.[a]

Independent variable	B	β	*t* statistic	*P* value	Variance inflation factor
Constant	0.837	N/A[b]	N/A	N/A	N/A
Age (years)	−0.003	−.240	−2.566	.01	1.146
Height (m)	0.259	.150	1.454	.15	1.397
Skeletal muscle mass of the lower limbs (kg)	0.011	.223	2.116	.04	1.459

[a]Dependent variable is mean daily life gait speed.

[b]N/A: not applicable.

Comparison Between Daily Life Gait Speed and Usual Gait Speed in the Analysis of Associations With Muscle Mass and Muscle Strength

A correlation analysis of gait speed, handgrip strength, and appendicular skeletal muscle mass showed that mean daily life gait speed was significantly and positively correlated with handgrip strength (Pearson r=0.380; P<.001) and appendicular skeletal muscle mass (Pearson r=0.355; P<.001). *Usual gait speed* had a significant positive correlation with handgrip strength (Pearson r=0.501; P<.001) but was less correlated with muscle mass (Pearson r=0.227; P=.02).

Discussion

Principal Findings

In this study, we showed that daily life gait speeds varied depending on the time of day and day of the week. Daily life gait speed showed greater variability than *usual gait speed*. The mean daily life gait speed was lower in participants with sarcopenia. When analyzing factors related to daily life gait speed, age and skeletal muscle mass of the lower limbs were significantly associated with mean daily life gait speed. The mean daily life gait speed was significantly and positively correlated with handgrip strength and appendicular skeletal muscle mass.

Walking speed is a valid and reliable measure for assessing the functional status of older adults [21], and identification of a slow walking speed is a simple approach to the diagnosis of frailty [22]. Gait speed at an individual's usual pace is known to be associated with survival in older adults [23] and has been widely used as a tool to predict adverse outcomes in community-dwelling older individuals, including falls, disability, institutionalization, and mortality [9,24]. However, one study demonstrated that *usual gait speed* does not represent daily life gait speed because it is not strongly related to daily gait speed [15]. Assessment of daily life gait speed showed that the men in our study walked significantly faster in the morning than during the rest of the day, and men with sarcopenia also walked faster in the morning. People may walk faster in the early morning because they usually have to go to work or attend social activities at a specified time. Participants in our study also walked faster on weekdays than on weekends. However, this difference was not significant in individuals with sarcopenia. This may be related to a decrease in the daily life gait speed of participants with sarcopenia, especially a decrease in the use of a fast gait speed. This is supported by the fact that men without sarcopenia often walked faster than their mean daily life gait speed, whereas participants with sarcopenia did not. As gait speed varied depending on the time of day and day of the week and the distribution of gait speed was very diverse, it would be difficult to accurately reflect the actual gait speed of individuals using one or two *usual gait speed* measurements obtained in the clinical setting. In fact, we found that the Pearson correlation coefficient between daily life gait speed and *usual gait speed* was negligible [25]. In addition, although the direct correlation between walking time and sarcopenia-related factors is not clear, people who spend more time walking tend to walk faster in their daily lives. An increase in walking-related exercise may be associated with a decrease in sarcopenia [26].

In this study, the mean value was used to represent the daily life gait speed to identify factors related to gait speed and to analyze the associations between gait speed and sarcopenia. Age and skeletal muscle mass of the lower limbs were significantly associated with the mean daily life gait speed. Reduced walking speed is associated with increasing age [27,28]. A decrease in the muscle mass of the lower limbs is also associated with a decrease in gait speed [29].

The participants in our study who had sarcopenia walked significantly more slowly in their daily lives than those without sarcopenia. This result was expected considering that patients with sarcopenia demonstrate various deteriorations in their physical performance [30,31]. Gait speed is a representative indicator of physical performance and is included in the diagnostic criteria for sarcopenia [6,7]. Diagnostic criteria commonly use the results of a 4- or 6-m *usual gait speed* test. Analyzing associations between daily life gait speed, muscle strength, and muscle mass, daily life gait speed also accurately reflects the sarcopenic state.

Advances in wearable devices have made it possible to continuously measure daily gait speed in the real world. As men usually wear belts in their daily lives, belt-type wearable devices can be used to measure daily life gait speed as naturally as possible over a long period. In fact, smart belt wearers can check their average walking speed on a daily, weekly, or monthly basis using a specific application. Therefore, older people would be able to identify a decrease in their physical performance in real life and, if so, would be able to visit health care providers for active intervention.

Strengths and Limitations

This study had several strengths. First, the analysis was performed using gait speed data measured more than 200,000

times over a period of ≥10 days. Therefore, the results reflect the actual gait speed of the participants more accurately than the current common practice of *usual gait speed* measured in a laboratory at only one point in time. Second, the skeletal muscle mass of participants was accurately measured by the standard method using DEXA, and it was confirmed that the skeletal muscle mass of the lower limbs had a statistically significant association with gait speed. Third, we found that walking speed could be measured continuously over a long period using a belt-type wearable device. On the basis of these results, it is feasible to carry out more precise longitudinal studies in the future, including studies focused on the prediction of a negative prognosis related to a change in walking speed.

This study has some limitations. First, we could not determine a causal relationship between skeletal muscle mass and gait speed because of the limitations inherent to the cross-sectional study design. A second limitation was that the study included only male participants. We chose male participants as Korean women who met our age criterion of >50 years do not usually wear belts, and we wanted to be certain that the participants could wear the device with ease. Accordingly, the findings observed in this study cannot be applied to women who usually do not use belts. To overcome this problem, we considered two solutions to this limitation. One was to improve the esthetic look of the belt so that women could wear it as a fashion accessory. The other was to place a plastic case containing the sensor, circuit, and battery in the same position as the buckle of the belt in the form of a clip.

Finally, we developed an algorithm for detecting and analyzing gait speed, but this has not been fully validated. However, a previous study reported that a method based on the detection of heel-strike events using a triaxial accelerometer to confirm that each step is suitable for measuring gait speed [32]. In our unpublished data, when comparing the gait data measured using WELT with that measured using video readings in 10 healthy adults, the mean step count was 95.9% (SD 4.2%) consistent. The mean sensitivity of the WELT step detection algorithm was 87.9% (SD 2.8%), and the mean positive predictive value was 92.4% (SD 3.3%). However, there were no data regarding gait speed in patients with pathological gait patterns such as neurologic (Parkinson disease) or orthopedic conditions. Further studies are required to confirm the validity of wearable device–based gait speed measurements.

Conclusions

Diverse and accurate information can be obtained by measuring daily life gait speed using a wearable device. Daily life gait speed is significantly associated with age and skeletal muscle mass of the lower limbs. As gait speed is a representative indicator of physical performance, older individuals would be able to detect a decrease in physical performance in real life by checking their walking speed. In addition, as it is possible to measure daily gait speed continuously over a long period using a wearable device, daily life gait speed data will be available for use in future longitudinal studies. Further studies are needed, especially studies aimed at developing a method to accurately measure women's daily life gait speed as naturally as possible.

Acknowledgments

This study was supported by Seoul National University Bundang Hospital (SNUBH 14-2018-001). The funder of this study was not involved in study design, collection of data, interpretation of data, or writing of the manuscript.

Conflicts of Interest

SJK and HKR are cofounders of WELT Corp, Ltd. HYJ is a full-time employee of WELT Corp, Ltd. The other authors declare no conflicts of interest.

References

1. Kontis V, Bennett JE, Mathers CD, Li G, Foreman K, Ezzati M. Future life expectancy in 35 industrialised countries: projections with a Bayesian model ensemble. Lancet 2017 Apr;389(10076):1323-1335. [doi: 10.1016/s0140-6736(16)32381-9] [Medline: 28236464]
2. Mitnitski A, Collerton J, Martin-Ruiz C, Jagger C, von Zglinicki T, Rockwood K, et al. Age-related frailty and its association with biological markers of ageing. BMC Med 2015 Jul 13;13:161 [FREE Full text] [doi: 10.1186/s12916-015-0400-x] [Medline: 26166298]
3. Klein BE, Klein R, Knudtson MD, Lee KE. Frailty, morbidity and survival. Arch Gerontol Geriatr 2005;41(2):141-149. [doi: 10.1016/j.archger.2005.01.002] [Medline: 16085065]
4. Mach M, Watzal V, Hasan W, Andreas M, Winkler B, Weiss G, et al. Fitness-tracker assisted frailty-assessment before transcatheter aortic valve implantation: proof-of-concept study. JMIR Mhealth Uhealth 2020 Oct 15;8(10):e19227 [FREE Full text] [doi: 10.2196/19227] [Medline: 33055057]
5. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, Cardiovascular Health Study Collaborative Research Group. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001 Mar;56(3):146-156. [doi: 10.1093/gerona/56.3.m146] [Medline: 11253156]
6. Chen L, Woo J, Assantachai P, Auyeung T, Chou M, Iijima K, et al. Asian Working Group for Sarcopenia: 2019 consensus update on sarcopenia diagnosis and treatment. J Am Med Dir Assoc 2020 Mar;21(3):300-307. [doi: 10.1016/j.jamda.2019.12.012] [Medline: 32033882]

7. Cruz-Jentoft AJ, Bahat G, Bauer J, Boirie Y, Bruyère O, Cederholm T, Writing Group for the European Working Group on Sarcopenia in Older People 2 (EWGSOP2),the Extended Group for EWGSOP2. Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing 2019 Jan 01;48(1):16-31 [FREE Full text] [doi: 10.1093/ageing/afy169] [Medline: 30312372]
8. Hoogendijk EO, Afilalo J, Ensrud KE, Kowal P, Onder G, Fried LP. Frailty: implications for clinical practice and public health. Lancet 2019 Oct 12;394(10206):1365-1375. [doi: 10.1016/S0140-6736(19)31786-6] [Medline: 31609228]
9. van Kan GA, Rolland Y, Andrieu S, Bauer J, Beauchet O, Bonnefoy M, et al. Gait speed at usual pace as a predictor of adverse outcomes in community-dwelling older people an International Academy on Nutrition and Aging (IANA) Task Force. J Nutr Health Aging 2009 Dec;13(10):881-889. [doi: 10.1007/s12603-009-0246-z] [Medline: 19924348]
10. Cesari M, Kritchevsky SB, Penninx BW, Nicklas BJ, Simonsick EM, Newman AB. Prognostic value of usual gait speed in well-functioning older people--results from the Health, Aging and Body Composition Study. J Am Geriatr Soc Oct 2005 Oct;53(10):1675-1680. [doi: 10.1111/j.1532-5415.2005.53501.x] [Medline: 16181165]
11. Hardy SE, Perera S, Roumani YF, Chandler JM, Studenski SA. Improvement in usual gait speed predicts better survival in older adults. J Am Geriatr Soc 2007 Nov;55(11):1727-1734. [doi: 10.1111/j.1532-5415.2007.01413.x] [Medline: 17916121]
12. Li X, Dunn J, Salins D, Zhou G, Zhou W, Schüssler-Fiorenza RS, et al. Digital Health: Tracking physiomes and activity using wearable biosensors reveals useful health-related information. PLoS Biol 2017 Jan;15(1):e2001402 [FREE Full text] [doi: 10.1371/journal.pbio.2001402] [Medline: 28081144]
13. Kim K, Gollamudi SS, Steinhubl S. Digital technology to enable aging in place. Exp Gerontol 2017 Feb;88:25-31. [doi: 10.1016/j.exger.2016.11.013] [Medline: 28025126]
14. Mueller A, Hoefling HA, Muaremi A, Praestgaard J, Walsh LC, Bunte O, et al. Continuous digital monitoring of walking speed in frail elderly patients: noninterventional validation study and longitudinal clinical trial. JMIR Mhealth Uhealth 2019 Nov 27;7(11):e15191 [FREE Full text] [doi: 10.2196/15191] [Medline: 31774406]
15. Van Ancum JM, van Schooten KS, Jonkman NH, Huijben B, van Lummel RC, Meskers CG, et al. Gait speed assessed by a 4-m walk test is not representative of daily-life gait speed in community-dwelling adults. Maturitas 2019 Mar;121:28-34. [doi: 10.1016/j.maturitas.2018.12.008] [Medline: 30704562]
16. Grieve DW, Gear RJ. The relationships between length of stride, step frequency, time of swing and speed of walking for children and adults. Ergonomics 1966 Sep;9(5):379-399. [doi: 10.1080/00140136608964399] [Medline: 5976536]
17. Guest R, Miguel-Hurtado O, Stevenage S, Black S. Exploring the relationship between stride, stature and hand size for forensic assessment. J Forensic Leg Med 2017 Nov;52:46-55. [doi: 10.1016/j.jflm.2017.08.006] [Medline: 28865387]
18. Choi J, Kim K, Choi Y, Ahn S, Kang E, Oh H, et al. Comparison of multidimensional frailty score, grip strength, and gait speed in older surgical patients. J Cachexia Sarcopenia Muscle 2020 Apr;11(2):432-440 [FREE Full text] [doi: 10.1002/jcsm.12509] [Medline: 31912668]
19. Guerri S, Mercatelli D, Gómez MP, Napoli A, Battista G, Guglielmi G, et al. Quantitative imaging techniques for the assessment of osteoporosis and sarcopenia. Quant Imaging Med Surg 2018 Feb;8(1):60-85 [FREE Full text] [doi: 10.21037/qims.2018.01.05] [Medline: 29541624]
20. Lee SY, Ahn S, Kim YJ, Ji MJ, Kim KM, Choi SH, et al. Comparison between dual-energy x-ray absorptiometry and bioelectrical impedance analyses for accuracy in measuring whole body muscle mass and appendicular skeletal muscle mass. Nutrients 2018 Jun 07;10(6):1-14 [FREE Full text] [doi: 10.3390/nu10060738] [Medline: 29880741]
21. Middleton A, Fritz SL, Lusardi M. Walking speed: the functional vital sign. J Aging Phys Act 2015 Apr;23(2):314-322 [FREE Full text] [doi: 10.1123/japa.2013-0236] [Medline: 24812254]
22. Castell M, Sánchez M, Julián R, Queipo R, Martín S, Otero A. Frailty prevalence and slow walking speed in persons age 65 and older: implications for primary care. BMC Fam Pract 2013;14:86 [FREE Full text] [doi: 10.1186/1471-2296-14-86] [Medline: 23782891]
23. Studenski S, Perera S, Patel K, Rosano C, Faulkner K, Inzitari M, et al. Gait speed and survival in older adults. J Am Med Assoc 2011 Jan 5;305(1):50-58 [FREE Full text] [doi: 10.1001/jama.2010.1923] [Medline: 21205966]
24. Verghese J, Holtzer R, Lipton RB, Wang C. Quantitative gait markers and incident fall risk in older adults. J Gerontol A Biol Sci Med Sci 2009 Aug;64(8):896-901 [FREE Full text] [doi: 10.1093/gerona/glp033] [Medline: 19349593]
25. Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. Malawi Med J 2012 Sep;24(3):69-71 [FREE Full text] [Medline: 23638278]
26. Yoo S, No M, Heo J, Park D, Kang J, Kim SH, et al. Role of exercise in age-related sarcopenia. J Exerc Rehabil 2018 Aug;14(4):551-558 [FREE Full text] [doi: 10.12965/jer.1836268.134] [Medline: 30276173]
27. Himann JE, Cunningham DA, Rechnitzer PA, Paterson DH. Age-related changes in speed of walking. Med Sci Sports Exerc 1988 Apr;20(2):161-166. [doi: 10.1249/00005768-198820020-00010] [Medline: 3367751]
28. Busch TD, Duarte YA, Nunes D, Lebrão ML, Naslavsky MS, dos Santos Rodrigues A, et al. Factors associated with lower gait speed among the elderly living in a developing country: a cross-sectional population-based study. BMC Geriatr 2015 Apr 01;15:35 [FREE Full text] [doi: 10.1186/s12877-015-0031-2] [Medline: 25880124]

29. Beavers KM, Beavers DP, Houston DK, Harris TB, Hue TF, Koster A, et al. Associations between body composition and gait-speed decline: results from the Health, Aging, and Body Composition study. Am J Clin Nutr 2013 Mar;97(3):552-560 [FREE Full text] [doi: 10.3945/ajcn.112.047860] [Medline: 23364001]
30. Cesari M, Landi F, Vellas B, Bernabei R, Marzetti E. Sarcopenia and physical frailty: two sides of the same coin. Front Aging Neurosci 2014;6:192 [FREE Full text] [doi: 10.3389/fnagi.2014.00192] [Medline: 25120482]
31. Dutta C, Hadley EC, Lexell J. Sarcopenia and physical performance in old age: overview. Muscle Nerve Suppl 1997;5:5-9. [Medline: 9331374]
32. Fortune E, Lugade V, Morrow M, Kaufman K. Validity of using tri-axial accelerometers to measure human movement - Part II: Step counts at a wide range of gait velocities. Med Eng Phys 2014 Jun;36(6):659-669 [FREE Full text] [doi: 10.1016/j.medengphy.2014.02.006] [Medline: 24656871]

Abbreviations

DEXA: dual-energy x-ray absorptiometry
WELT: wearable smart belt

Edited by R Kukafka; submitted 24.04.21; peer-reviewed by V Watzal, M Perez-Zepeda, E Sadeghi-Demneh; comments to author 28.05.21; revised version received 21.07.21; accepted 12.08.21; published 11.10.21.

Please cite as:
Kang MG, Kang SJ, Roh HK, Jung HY, Kim S, Choi JY, Kim KI
Accuracy and Diversity of Wearable Device–Based Gait Speed Measurement Among Older Men: Observational Study
J Med Internet Res 2021;23(10):e29884
URL: https://www.jmir.org/2021/10/e29884
doi:10.2196/29884
PMID:34633293

Original Paper

A Mobile App (AMOR Mama) for Women With Breast Cancer Undergoing Radiation Therapy: Functionality and Usability Study

Flávia Oliveira de Almeida Marques da Cruz[1], PhD; Edison Tostes Faria[2], MD; Pabblo Cardelino Ghobad[3], MSc; Leandro Yukio Mano Alves[4], PhD; Paula Elaine Diniz dos Reis[5], PhD

[1]Interdisciplinary Laboratory of Research Applied to Clinical Practice in Oncology, School of Health Sciences, University of Brasilia (UnB), Brasilia, Brazil

[2]Faculty of Medicine, University of Brasilia (UnB), Brasilia, Brazil

[3]Federal Senate of Brazil, Brasilia, Brazil

[4]State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil

[5]Department of Nursing, School of Health Sciences, University of Brasilia (UnB), Brasilia, Brazil

Corresponding Author:
Flávia Oliveira de Almeida Marques da Cruz, PhD
Interdisciplinary Laboratory of Research Applied to Clinical Practice in Oncology
School of Health Sciences
University of Brasilia (UnB)
Darcy Ribeiro s/n - Asa Norte
Brasilia, 70910-900
Brazil
Phone: 55 61998552290
Email: flavinhaoliveira5@hotmail.com

Abstract

Background: Mobile apps targeting women with breast cancer can facilitate access to information, improve well-being, and record reports of treatment-related symptoms. However, it is important to confirm the benefits of these apps before they are used as a tool in clinical care.

Objective: The aim of this study was to evaluate the functionality and the usability of a mobile app created to guide and monitor patients with breast cancer undergoing radiation therapy.

Methods: The evaluation process of the mobile app was performed in 2 steps with 8 professionals, including nurses, physician, medical physicists, and communication networks engineer. The first step was the focus group, which allowed obtaining suggestions proposed by the participants regarding the improvement of the mobile app. The second step was the individual filling in of an evaluation tool to obtain objective measures about the mobile app. A minimum concordance index of 80% was considered to ensure the adequacy of the material.

Results: After the mobile app was evaluated by 8 professionals, only 1 item of the evaluation tool, that is, concerning the potentiality of the app to be used by users of different educational levels, obtained a concordance index <80%.

Conclusions: The mobile app titled “AMOR Mama” was considered suitable, which suggests its contribution to an educational health technology to guide and monitor patients with breast cancer undergoing radiation therapy. More studies with this target population should be carried out to assess the performance and quality of the mobile app during its use.

(J Med Internet Res 2021;23(10):e24865) doi:10.2196/24865

KEYWORDS

mobile applications; health education; nursing care; oncology nursing; educational technology; breast neoplasms; radiation therapy

Introduction

Cancer is a leading cause of disability and mortality worldwide, affecting more than 18 million people each year [1]. Breast cancer is the most frequent among women, with approximately 2.1 million new cases diagnosed in 2018 [2]. Radiation therapy is a treatment modality that uses ionizing radiation to prevent the multiplication of tumor cells and to determine their death by delivering radiation to the tumor with the least possible damage to the healthy surrounding tissues [3]. When the breast

region is exposed to radiation therapy, the most common side-effects are pain, radiation dermatitis, restricted mobility, local sensory alteration, and fatigue [4]. In addition to the physical factors related to the disease and treatment, it is worth mentioning the negative psychological impacts caused by breast cancer, mainly on the perception of body image, sexuality, and femininity [5]. Therefore, nurses need to perform their care role related to cancer and therapy effects as a facilitator in the coping process when providing comprehensive and individualized care to patients.

Mobile apps can help individuals in managing their own health and well-being as well as in promoting healthy lifestyles and obtaining quick access to important information. The use of mobile technologies has brought innovative possibilities for improving health care provision [6]. A study reported the development of a mobile app prototype to assist health care professionals in the prevention and classification of pressure ulcers. The app was evaluated for technical and functional quality by experts in the field of computing and nursing, and it was considered adequate with respect to reliability, usability, efficiency, functionality, and portability [7]. Another study described the development and evaluation of an app to support decision-making in the process of early mobilization of critical patients admitted to an intensive care unit. The app was evaluated by 58 physiotherapy undergraduate students, and it was considered practical and easy to understand and manipulate [8]. A separate study focused on the development and testing of a tablet app to collect computerized data on tobacco use among psychiatric patients and the general population, thereby replacing paper surveys. This app prevented human errors, allowed automatic tabulation, and made the interviews less tiring [9]. A recent systematic review suggests that apps for women with breast cancer can facilitate access to information and improve patients' well-being as well as improve the report of symptoms and adverse treatment-related effects. However, the real benefit of using apps is still uncertain since the apps are often introduced into clinical care before necessary research is carried out to confirm their benefits [10]. In this context, this study aimed to evaluate the functionality and the usability of a mobile app created to guide and monitor patients with breast cancer undergoing radiation therapy.

Methods

Study Design

This is a methodological research focusing on the evaluation of the functionality and usability of a mobile app prototype for patients with breast cancer undergoing radiation therapy. The app was entitled "AMOR Mama" and features content of an educational manual developed and validated in a previous study called as "Orientations Manual: breast radiation therapy" [11]. The manual contains 36 pages, divided into pretextual items (cover, back cover, cataloguing sheet, index, presentation, and registration card), textual items (chapters on radiation therapy, stages of treatment, adverse effects of radiation therapy, and how to prevent them), and posttextual items (latest information, weekly diary, and bibliographical references).

The features and functionality of the app were then defined and associated with the textual content of the validated manual. Among them, we can mention user registration; registration of events in calendar and medications in use, including the possibility of generating automatic reminders through notifications; calculation of the water intake target according to the user's weight; recording and monitoring of signs and symptoms related to breast radiation therapy, as well as recommendations and necessary care. The data registration was planned to be used in the preparation of reports that can be forwarded to the health team involved in the user's care, thus facilitating their monitoring.

We aimed to develop a prototype of an app with relevant information and useful resources to encourage interaction between the user and the app. It has 5 menus that list the main sections: Home Page, About Radiation Therapy, Diary, Schedule, and Settings. One of the features of "Settings" is the tutorial that was designed to be the user's initial experience during their first access.

Methodological Reference

We chose to use the standard published by the International Organization of Standardization and the International Electrotechnical Commission, revised and translated by the Brazilian Association of Technical Standards, named as NBR ISO/IEC 9126-1. This standard describes a quality model for software products, which is composed of 2 parts: internal and external quality and quality in use [12]. The internal and external quality of the software refer to a group of characteristics that the model must present to be able to meet the user's needs, such as functionality, usability, reliability, efficiency, maintainability, and portability [12]. Quality in use involves 4 characteristics: effectiveness, productivity, safety, and satisfaction. It is related to the evaluation of the final product from the user's point of view, which is measured according to the performance obtained by the target audience when using the software. It is necessary to guarantee the internal and external quality of software to obtain the quality in use [12].

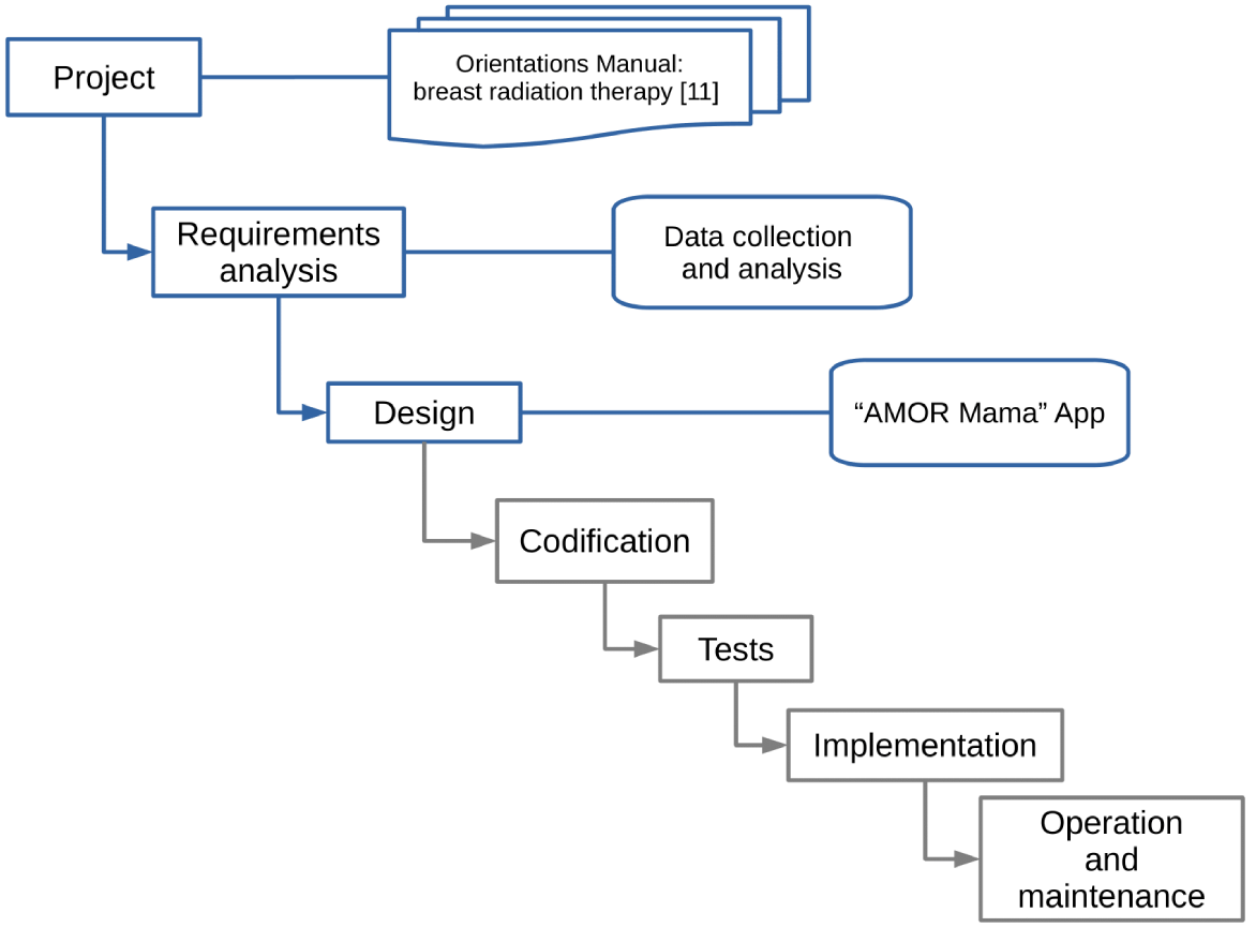
This study carried out the internal and external quality of the software, and the characteristics of functionality and usability were investigated since what is being evaluated is a prototype of an app, that is, a set of screens that present the interface and the proposed functions for the app and not the software that is already developed. The functionality is related to the potentiality of the app to offer the functions and resources that meet the needs of users, that is, the degree to which the software is able to provide functions that satisfy such needs [12,13]. Usability is achieved when the software has intelligibility, apprehensibility, and operability, that is, when it is easy to understand, to learn, and to operate [12]. Good usability is directly related to the success of the software in which the user is able to employ the functions and resources offered by the app with ease and efficiency [13].

Process of the Mobile App Design Assessment

This study carried out actions that involve the development of software up to the modeling phase, that is, the elaboration and analysis of a project. A project is the representation of the

aspects of software that are visible to the user such as the interface layout and the display formats on the screen [13]. Figure 1 shows the software development cycle: the blue boxes show the steps that were developed in this study while the gray boxes show the steps that will be developed in a future study.

Figure 1. Phases of software development. The blue boxes show the steps that were developed in this study while the gray boxes show the steps that will be developed in a future study.



The evaluation process of the app prototype was performed in 2 steps. The first one was the focus group, which allowed obtaining the suggestions proposed by the participants regarding the improvement of the app. The second step was the individual filling in of an evaluation tool to obtain objective measures about the mobile app. The researcher presented the display screens of the mobile app prototype during the meeting of the focus groups, with the aid of a multimedia projector. The interface was exhibited and its functionality demonstrated.

Research Participants

Intentional nonprobabilistic sampling was used, considering that the sample must be defined according to the study objectives. The factors used to choose these professionals were then determined according to their academic degree, specialization, scientific production, knowledge, and time of work, according to the adaptation of the evaluation system defined in the Fehring model (1987) [14]. The recommendation is that at least 8 participants perform the design assessment of the mobile app [15]. The invitation was sent by email to 20 professionals qualified to participate in the study. Six professionals expressed interest in participating; however, they were not available during the days and times foreseen for the focus group, while 6 other professionals did not respond to the invitation even after 3 attempts to contact them. Therefore, 8 professionals participated in this research who attended the focus group at an agreed date and time.

Data Collection and Analysis

The data were collected from November to December 2019. There were difficulties in scheduling a single meeting with all the professionals, given the specificities and individual demands. Therefore, 2 focus groups were held at different dates and times, previously agreed with the participants. The suggestions obtained through the focus groups were transcribed and analyzed by the researchers both within each focus group and comparatively between the 2 focus groups carried out to identify suggestion patterns. The quantitative data were collected using a Likert scale evaluation tool, which allowed objective measures to be obtained in relation to participants' opinions about the app prototype. This tool was adapted from a preexisting questionnaire [16], with the necessary changes and adjustments in relation to the theme addressed in this study. It has 5 levels of judgment regarding the items: no, in parts, not sure, yes, and certainly. The options yes and certainly were grouped and represent the adequacy of the item. To determine the adequacy

of the app prototype, a minimum concordance index of 80% was determined among the participants [17]. The set composed of options yes (Y) and certainly (C) should comprise at least 80% of professionals' responses to determine the quality of the app. Thus, concordance index was calculated by the mathematical formula: (Y+C)×100/total responses. Each item that obtained concordance index below 80% was analyzed and the app was improved according to the participants' suggestions, scientific literature, and clinical evidence.

Ethical Aspects of the Research

This research project was sent to the Research Ethics Committee of the College of Health Sciences of the University of Brasilia (CEP/FS-UnB) and approved by Opinion 2.608.031, CAAE: 71348417.7.0000.0030.

Results

Profile of the Research Participants

For the mobile app evaluation, the sample consisted of 8 professionals, including nurses, doctor, medical physicists, and communication network engineer. Regarding gender, 5 were women and 3 were men. The age of the professionals ranged from 24 to 43 years (mean 32 [SD 6] years), while educational background ranged from 2 to 20 years (mean 8 [SD 6] years) and that of working in the subject area from 1 to 20 years (mean 7 [SD 6] years). Regarding the highest academic degree, 6 participants had master's degrees and 2 were specialists.

Process of the Mobile App Design Assessment

The professionals' evaluations (N=8) were analyzed quantitatively. In the evaluation tool, the options yes and certainly were grouped to represent the adequacy of the item. Table 1 shows the judgment of the experts and the concordance index for each item of the evaluation tool. The focus was to investigate the participants' opinion regarding the usability of the mobile app and the possibility of it being recommended for wide use by patients with breast cancer undergoing radiotherapy. Only item 19 did not reach the minimum concordance index established, reaching 75% of agreement among the experts. All other items reached a minimum concordance index of 80% (range 88%-100%).

In item 19, the professional who marked the option "in parts" gave the following justification for their choice: "I believe that the initial guidelines for tomography and symptoms—as presented in the text—are not simple for patients with different levels of education (E4)."

Table 1. Professionals' assessment regarding the usability and functionality of the mobile app prototype.

Evaluation items	Experts' levels of judgement (N=8)					Concordance index (%)[a]
	No	In parts	Not sure	Yes	Certainly	
1. It seems to be easy to understand the concept and manipulation of the app.	0	0	0	1	7	100
2. The app appears to be self-explanatory.	0	0	0	0	8	100
3. The tutorial is useful to optimize the use of the app.	0	0	0	1	7	100
4. The organization of the menu is logical.	0	0	0	1	7	100
5. Icons are clear and intuitive.	0	0	0	1	7	100
6. The look of the app (design and layout) is attractive.	0	0	0	2	6	100
7. It appears to be easy to navigate the app screens.	0	0	0	1	7	100
8. The sequence of actions in the app is consistent.	0	0	0	1	7	100
9. The app has useful functions.	0	0	0	2	6	100
10. The app appears to be accurate in performing its functions.	0	0	0	1	7	100
11. It seems easy to find information in the app.	0	0	0	2	6	100
12. It seems easy to insert personal information in the app.	0	0	0	2	6	100
13. It seems easy to insert events and medications into the app.	0	0	0	3	5	100
14. The texts in the app are easy to read.	0	0	0	2	6	100
15. The illustrations are important to complement the texts and facilitate the use of the app.	0	0	0	0	8	100
16. The screens are visually pleasing and seem to encourage the use of the app.	0	0	0	1	7	100
17. Using the app seems simple and accessible.	0	0	0	2	6	100
18. It seems possible to maintain interaction between the user and the app.	0	0	0	2	6	100
19. The app can be used by users of different educational levels.	0	1	1	4	2	75
20. I believe that most users would learn how to use the app.	0	0	1	2	5	88
21. The app can be used anywhere, such as at home, hospital, or street.	0	0	1	1	6	88
22. I would recommend the app to be used by women with breast cancer undergoing radiation therapy.	0	0	0	0	8	100

[a]Concordance index was calculated by the sum of yes and certainly judgments: (Yes+Certainly)×100/total of responses.

Research Participants' Suggestions

After analyzing the suggestions of the focus groups, the changes proposed by the participants and the researchers' decisions regarding whether to accept such changes in the final version of the app are described in Table 2. It is worth mentioning that the communication networks engineer was the professional that suggested the most number of changes, which are mainly related to the functions and resources of the app.

Table 2. Summary of the participants' proposed changes and the researchers' decision.

Items (list of changes)	Participants' changes proposed	Decision
Icon	Change the app title "A.M.O.R. Mama" to "AMOR Mama"	Accepted
Tutorial		
1	Replace "About the app" (from Portuguese *Sobre o app*) with "Tutorial"	Accepted
2	Add an introduction screen to the tutorial	Accepted
3	Add information about the home page	Accepted
4	Replace "Clicking on this icon at the navigation bar (…)" (from Portuguese *Clicando neste ícone na barra de navegação(...)*) with "Here you will find your diary, where you will be able to insert (…)" (from Portuguese *Aqui você encontrará o seu Diário, onde poderá registrar (...)*)	Accepted
5	Replace "Clicking on this icon at the navigation bar (…)" with "On this icon you will access your calendar, where you will be able to insert (…)" (from Portuguese *Neste ícone você terá acesso a sua Agenda, onde poderá cadastrar (...)*)	Accepted
6	Replace "Clicking on this icon at the navigation bar (…)" with "Here you will be able to change your personal data and obtain information about (…)" (from Portuguese *Aqui você poderá alterar seus dados pessoais e obter informações sobre (...)*)	Accepted
7	Add a tutorial completion screen	Accepted
Home page	Add a home page that includes the information recorded daily in the schedule	Accepted
About radiation therapy		
1	Add explanatory videos recorded by the team and in the real treatment environment	Accepted
2	Add links to informational content available on the web	Not accepted
3	In "Steps of the Treatment" (from Portuguese *Etapas do Tratamento*), include the possibility of adding the date and time of the tomography	Accepted
4	In "Adverse Effects" (from Portuguese *Efeitos Adversos*), add a direct link to the schedule, allowing the insertion of appointments already scheduled at the institution	Accepted
Diary		
1	Add "Export data" (from Portuguese *Exportar dados*) next to the download icon	Accepted
2	Add the possibility to select the period and data to download	Accepted
3	Replace "Symptoms" (from Portuguese *Sintomas*) with "Signs and Symptoms" (from Portuguese *Sinais e Sintomas*)	Accepted
4	In "Signs and Symptoms," include a postmastectomy chest wall image	Not accepted
5	In "Care" (from Portuguese *Cuidados*), replace the checkboxes with radio buttons	Accepted
6	In "Care," include information about the water intake goal	Accepted
7	In "Notes" (from Portuguese *Anotações*), include the possibility of adding photos next to the user's comments	Accepted
Schedule		
1	Add a screen that highlights the 2 features of this topic ("Events" from Portuguese *Eventos* and "Medicines" from Portuguese *Medicamentos*)	Accepted
2	In "Events," allow the exclusion of an event previously included (add an "X" next to the notification symbol)	Accepted
3	In "Medicines," allow the exclusion of a drug previously included (add an "X" next to the notification symbol)	Accepted
4	In "Medicines," add mechanisms that facilitate their insertion	Accepted
5	In "Medicines," add a list of possible drugs to be added	Not accepted
Settings		
1	Add a screen that highlights the 3 features of this topic ("Registration" from Portuguese *Cadastro*, "Tutorial," and "About the App")	Accepted
2	In "About the app," add "Contact Info" (from Portuguese *Contato*), "Terms of Use" (from Portuguese *Termos de Uso*), and "Privacy Policy" (from Portuguese *Política de Privacidade*)	Accepted

Discussion

Main Results

The design of the mobile app "AMOR Mama" was based on an educational manual developed and validated in a previous study named as "Orientations Manual: breast radiation therapy" [11]. The features and functionality of the app were then defined and associated with the textual content of the validated manual. The participants' judgments were generally consistent (Table 1). Only 1 item of the evaluation tool, which was related to the potentiality of the app to be used by users of different educational levels, obtained a concordance index <80%, considering that 1 expert chose the option "not sure" and another chose the option "in parts" for that item. One of the experts believed that the guidelines presented in the text are not accessible for people with different levels of education. One way to reduce the difficulties in reading and facilitate communication is by using images that are associated with the textual content. Thus, to create a visual representation and illustrate the app, figures and photos obtained in the real treatment environment of patients undergoing radiation therapy were included.

Remote access to health information through mobile technologies contributes to the solution of the patient's needs without restriction of time and space, thereby promoting the overcoming of barriers and both, the extension of assistance, and the reach of health benefits [18]. In this way, caregivers and family members who are often unable to be present in the treatment environment of patients can help those who have a low level of education, especially in relation to reading and interpreting the information offered by an app that can be accessed freely in any place at any time of the day. Furthermore, a suggestion that emerged during a discussion in one of the focus groups and which was accepted to be implemented in the app was to add explanatory videos related to the stages of radiation therapy, with the health team as actors in a real environment of treatment. This suggestion is very pertinent and useful to complement the textual content and facilitate the understanding for users, especially for those who have a low level of education. In the following study of coding and evaluation of the software, the stages of preparing the scripts, training the team, and recording and editing the explanatory videos to be inserted in the app will also be carried out. Therefore, the content offered by the app must be intelligible, informative, succinct, objective, and delivered to the user in an appropriate manner, that is, through correct texts in relation to spelling and concordance and associated with representative images, which assist in the interpretation of the information provided. The aesthetics must be pleasant, reducing distractions and favoring the interaction between the user and the software [13]. The content must also be constructed consistently. The formatting of the text and its font style, size, and color must be the same throughout all the screens developed, thereby maintaining a standard. In any software design model, consistency is an attribute that must be considered. The graphic design must consist of a pattern of style, aspects, and colors, which provides a consistent interface in relation to the interaction, navigation, and display of content and functionalities [13]. Therefore, a graphic design with well-planned mechanisms provides spontaneous and intuitive interaction with the user. In this study, the interface of the app was designed to be simple, easy to understand, visibly pleasant and attractive, as proposed by a study that focused on the development of an app for teaching the International Classification for Nursing Practice. The development of this app also took into account the importance of attractiveness and objectivity so that some aspects such as the choice of textual content and its font and size as well as the colors of the design provided visual and dynamic comfort in the delivery of information [19]. However, in addition to consistency, the app must have an identity. The aesthetics of the interface must provide affinity with the user to whom the software will be directed [13]. In the elaboration of the app prototype proposed in this study, female aspects were taken into consideration, such as the choice of the predominant color and the images related to the female universe (clothing being illustrated by images of dress and bra and skin care being illustrated by photographs of a woman and female breasts) in addition to the communication being directed to the female gender. Regardless of the information or functionality provided by the app, the screen layout is responsible for shaping the user's perception of the software, creating a medium of communication between them. In this context, a software user expects to find content and functions relevant to their needs in addition to intuitive and predictable navigation. Without adequate graphic design, an app for mobile devices may be quite functional, but it would not be attractive and, in turn, would not have good usability [13].

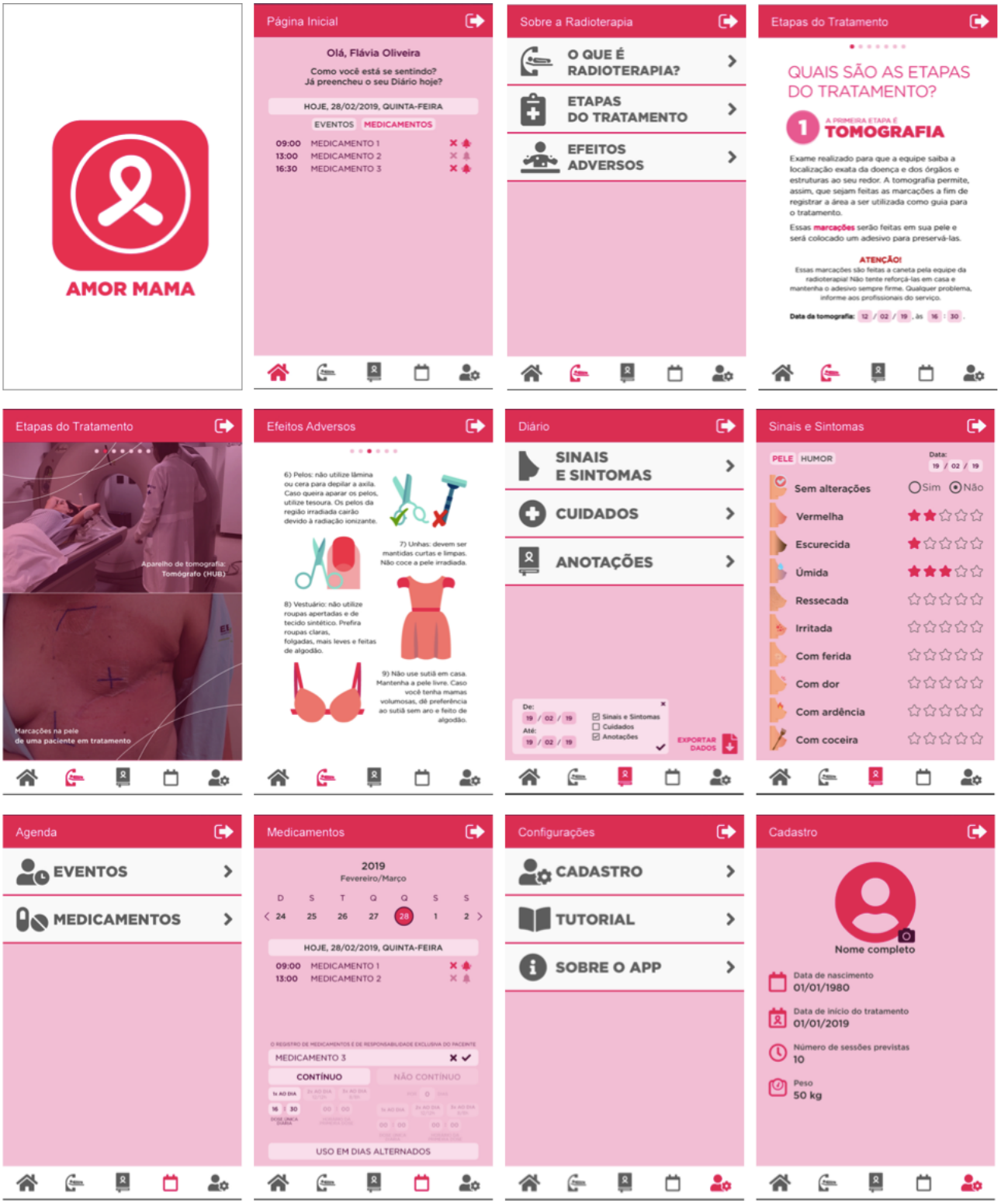
Among the features and functionality available in the prototype of the app proposed in this study, we can mention the registration of events in a calendar, with the possibility of generating automatic reminders through notifications with 24 hours in advance. The user's freedom of choice in relation to enabling or disabling notifications is an important aspect that was already foreseen in the project and was reinforced during the discussion of one of the focus groups. This possibility of adapting to different users is very relevant, considering that an interface must be able to provide flexible interaction since different users may have different preferences for interacting with the app [13]. This customization is also found in the registration where the user can insert a photo and her personal information such as name, weight, date of birth, start date of treatment, and number of sessions planned. Body weight is initially used to calculate the goal of daily water intake, which should be 30 to 35 ml/kg for adults and 25 mL/kg for the elderly, considering the variations according to the symptoms and tolerance presented by each patient [20].

A study also included the possibility of personalization in the development of the proposed app. This app can be useful for recording and monitoring various personal health data such as results of laboratory tests, medications in use, immunization, allergies, body weight, and blood pressure in addition to information related to physical activity, nutrition, and sleep. The modules are customizable and flexible since users can make choices according to their own needs, thereby hiding or showing different information [21]. The suggestion to include links in the app that direct to informative content available on the web

was not accepted. The researchers believe that the information already validated by specialists and available on the app is sufficient for the adequate guidance of users. An exaggerated amount of information can be harmful, as suggested by a study that evaluated the effects of an app on the anxiety levels of patients undergoing breast cancer surgery [22].

The authors of this study also have no control over external content inserted in the app through links and are not responsible for its availability and the updates. In addition, the direction and access to the content available on the web depends on the internet connection, which hurts the initial proposal of offline browsing for the app proposed in this study. Therefore, after making the changes suggested by the participants, the final product of the elaboration and improvement of the app prototype resulted in a model ready to be coded as software, with essential content for patients with breast cancer undergoing radiation therapy and with an interface that favors communication, interactivity, and its use, as shown in Figure 2.

Figure 2. Screenshots of the AMOR Mama app.



Limitations

The limitation in this study was related to the difficulty in conducting only 1 focus group, including all participants at once in the design assessment of the mobile app, given the specificities and individual demands of each professional. However, this limitation was overcome by the fact that the focus group was used as a complementary technique to the quantitative assessment, which occurred individually and allowed to obtain objective measures about the evaluation. Another limitation is related to the number of participants included. Despite the recommended minimum of 8 participants, it is believed that a larger sample could have improved the design assessment of the mobile app.

Conclusions

The mobile app prototype entitled "AMOR Mama," after having been improved based on the participants' suggestions, was considered adequate. This indicates its contribution as a technology for health education aimed at the dissemination of guidelines and monitoring of women with breast cancer undergoing radiation therapy. This app will be developed through software codification, and studies with the target population must be carried out to assess the performance and quality of the app during its use.

Conflicts of Interest

None declared.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394-424 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
2. Brasil, Ministério da Saúde. Estimativa 2020: incidência de câncer no Brasil. Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA). URL: https://www.inca.gov.br/sites/ufu.sti.inca.local/files/media/document/estimativa-2020-incidencia-de-cancer-no-brasil.pdf [accessed 2021-09-22]
3. Silva LFO, Santos LB. Física médica aplicada à radioterapia. Diretrizes Oncológicas 2. URL: https://diretrizesoncologicas.com.br/wp-content/uploads/2018/10/Diretrizes-oncol%C3%B3gicas-2_Parte37.pdf [accessed 2021-09-22]
4. Reis PED, Ferreira EB, Bontempo PSM. Radiodermatites: prevenção e tratamento. Diretrizes Oncológicas 2. URL: https://diretrizesoncologicas.com.br/wp-content/uploads/2018/10/Diretrizes-oncol%C3%B3gicas-2_Parte41.pdf [accessed 2021-09-22]
5. Varela AIS, Rosa LM, Sebold N, Iaverde AG, Maçaneiro A, Erdmann AL. Commitment to sexuality of women with breast cancer. Revista Cofen. 2017. URL: http://revista.cofen.gov.br/index.php/enfermagem/article/view/764/359 [accessed 2020-03-04]
6. Mobile medical applications: guidance for industry and food and drug administration staff. Federal Register. URL: https://www.federalregister.gov/documents/2013/09/25/2013-23293/mobile-medical-applications-guidance-for-industry-and-food-and-drug-administration-staff [accessed 2020-03-14]
7. Tibes CMS. Aplicativo móvel para prevenção e classificação de úlceras por pressão. Repositorio Ufscar. URL: https://repositorio.ufscar.br/bitstream/handle/ufscar/3287/6796.pdf?sequence=1 [accessed 2021-09-22]
8. Souza JF, Gonçalves FB, Queiroz VAR, Queiroz RS. Evaluation of an application for aid raising a decision-making critical patients. Revista Saúde.Com. URL: https://periodicos2.uesb.br/index.php/rsc/article/view/344/277 [accessed 2020-05-23]
9. Oliveira RMD, Duarte AF, Alves D, Furegato ARF. Development of the TabacoQuest app for computerization of data collection on smoking in psychiatric nursing. Rev Lat Am Enfermagem 2016 Aug 29;24:e2726 [FREE Full text] [doi: 10.1590/1518-8345.0661.2726] [Medline: 27579930]
10. Cruz FOAM, Vilela RA, Ferreira EB, Melo NS, Reis PEDD. Evidence on the Use of Mobile Apps During the Treatment of Breast Cancer: Systematic Review. JMIR Mhealth Uhealth 2019 Aug 27;7(8):e13245 [FREE Full text] [doi: 10.2196/13245] [Medline: 31456578]
11. Cruz FOAM, Faria ET, Reis PED. Validation of an educational manual for breast cancer patients undergoing radiotherapy. Rev Lat Am Enfermagem 2020;28:e3384 [FREE Full text] [doi: 10.1590/1518-8345.3197.3384] [Medline: 33084780]
12. Associação Brasileira de Normas Técnicas. Engenharia de software-qualidade de produto. parte 1: modelo de qualidade. ABNT-Associação Brasileira de Normas Técnicas NBR ISO/IEC 9126-1. URL: https://www.academia.edu/26260622/ABNT_Associa%C3%A7%C3%A3o_Brasileira_de_Normas_T%C3%A9cnicas_NBR_ISO_IEC_9126_1_Engenharia_de_software_Qualidade_de_produto_Parte_1_Modelo_de_qualidade [accessed 2021-09-22]
13. Pressman RS. Engenharia de Software: Uma Abordagem Profissional. URL: https://www.docsity.com/pt/pressman-r-s-maxim-b-r-engenharia-de-software-uma-abordagem-profissional-8-ed-porto-alegre-amgh-2016-e-book/5818987/ [accessed 2021-09-22]
14. Melo RP, Moreira RP, Fontenele FC, Aguiar SCA, Joventino ES, Carvalho EC. Critérios de seleção de experts para estudos de validação de fenômenos de enfermagem. Revista da Rede de Enfermagem do Nordeste. URL: https://www.redalyc.org/articulo.oa?id=324027975020 [accessed 2020-05-28]

15. Engenharia de software - Requisitos e avaliação da qualidade de produto de software (SQuaRE). Formato comum da indústria (FCI) para relatórios de teste de usabilidade. Associação Brasileira de Normas Técnicas ISO 25062. URL: https://www.abntcatalogo.com.br/norma.aspx?ID=086972 [accessed 2021-09-22]
16. Gloria HS. Avaliação de um conjunto de heurísticas de usabilidade para aplicativos de smartphones na área da saúde por meio de testes de usabilidade. Universidade Federal de Santa Catarina. URL: http://www.gqs.ufsc.br/files/2020/02/TCC2_Heloisa_Gloria.pdf [accessed 2021-09-22]
17. Pasquali L. Psicometria: Teoria e Aplicações. URL: https://www.scielo.br/j/reeusp/a/Bbp7hnp8TNmBCWhc7vjbXgm/?lang=pt&format=pdf [accessed 2021-09-22]
18. Barra DCC, Paim SMS, Sasso GTMD, Colla GW. Métodos para desenvolvimento de aplicativos móveis em saúde: revisão integrativa da literatura. Texto contexto-enferm 2018 Jan 08;26(4):1-12. [doi: 10.1590/0104-07072017002260017]
19. Mota NP, Vieira CMDA, Nascimento MNR, Bezerra AM, Quirino GDS, Félix NDDC. Mobile application for the teaching of the International Classification for Nursing Practice. Rev. Bras. Enferm 2019 Aug;72(4):1020-1027. [doi: 10.1590/0034-7167-2018-0751]
20. Brasil, Ministério da Saúde. Consenso nacional de nutrição oncológica. Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA). URL: https://bvsms.saude.gov.br/bvs/publicacoes/consenso_nacional_nutricao_oncologico.pdf [accessed 2021-09-22]
21. Zhou L, DeAlmeida D, Parmanto B. Applying a User-Centered Approach to Building a Mobile Personal Health Record App: Development and Usability Study. JMIR Mhealth Uhealth 2019 Jul 05;7(7):e13194 [FREE Full text] [doi: 10.2196/13194] [Medline: 31278732]
22. Foley NM, O'Connell EP, Lehane EA, Livingstone V, Maher B, Kaimkhani S, et al. PATI: Patient accessed tailored information: A pilot study to evaluate the effect on preoperative breast cancer patients of information delivered via a mobile application. Breast 2016 Dec;30:54-58. [doi: 10.1016/j.breast.2016.08.012] [Medline: 27611236]

Edited by R Kukafka; submitted 26.10.20; peer-reviewed by R Vilela, SS Funghetto, A Teles; comments to author 06.12.20; revised version received 25.01.21; accepted 16.03.21; published 13.10.21.

Please cite as:
Cruz FODAMD, Faria ET, Ghobad PC, Alves LYM, Reis PEDD
A Mobile App (AMOR Mama) for Women With Breast Cancer Undergoing Radiation Therapy: Functionality and Usability Study
J Med Internet Res 2021;23(10):e24865
URL: https://www.jmir.org/2021/10/e24865
doi:10.2196/24865
PMID:34643531

Viewpoint

Building the Digital Mental Health Ecosystem: Opportunities and Challenges for Mobile Health Innovators

Benedetta Spadaro[1*], MPhil; Nayra A Martin-Key[1*], PhD; Sabine Bahn[1,2], MD, MRCPsych, PhD

[1]Cambridge Centre for Neuropsychiatric Research, Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, United Kingdom

[2]Psyomics Ltd, Cambridge, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Sabine Bahn, MD, MRCPsych, PhD
Cambridge Centre for Neuropsychiatric Research
Department of Chemical Engineering and Biotechnology
University of Cambridge
Philippa Fawcett Drive
Cambridge, CB3 0AS
United Kingdom
Phone: 44 1223 334151
Email: sb209@cam.ac.uk

Abstract

Digital mental health technologies such as mobile health (mHealth) tools can offer innovative ways to help develop and facilitate mental health care provision, with the COVID-19 pandemic acting as a pivot point for digital health implementation. This viewpoint offers an overview of the opportunities and challenges mHealth innovators must navigate to create an integrated digital ecosystem for mental health care moving forward. Opportunities exist for innovators to develop tools that can collect a vast range of active and passive patient and transdiagnostic symptom data. Moving away from a symptom-count approach to a transdiagnostic view of psychopathology has the potential to facilitate early and accurate diagnosis, and can further enable personalized treatment strategies. However, the uptake of these technologies critically depends on the perceived relevance and engagement of end users. To this end, behavior theories and codesigning approaches offer opportunities to identify behavioral drivers and address barriers to uptake, while ensuring that products meet users' needs and preferences. The agenda for innovators should also include building strong evidence-based cases for digital mental health, moving away from a one-size-fits-all well-being approach to embrace the development of comprehensive digital diagnostics and validated digital tools. In particular, innovators have the opportunity to make their clinical evaluations more insightful by assessing effectiveness and feasibility in the intended context of use. Finally, innovators should adhere to standardized evaluation frameworks introduced by regulators and health care providers, as this can facilitate transparency and guide health care professionals toward clinically safe and effective technologies. By laying these foundations, digital services can become integrated into clinical practice, thus facilitating deeper technology-enabled changes.

(J Med Internet Res 2021;23(10):e27507) doi:10.2196/27507

KEYWORDS

digital implementation; digital mental health; digital psychiatry; digital technology; viewpoint

Introduction

Mental health disorders represent the leading cause of disability worldwide, with over one third of the world's population being affected by a mental health condition in their lifetime [1]. Given the increasing pressures on mental health care budgets globally, as well as shortages of health care professionals who are facing an overwhelming growing burden of chronic and recurring mental health conditions, prevention strategies and improvements in early identification and treatment are essential. To this end, mobile health (mHealth) tools such as apps can offer innovative ways to help develop and facilitate mental health care provision. Indeed, the unmet need for psychiatric services has sparked significant interest in developing apps to meet patient demand. As a result, more than 10,000 apps addressing mental health symptoms such as anxiety, low mood, and insomnia are now available on the Apple App Store and Google Play Store [2]. Furthermore, despite its catastrophic effects, the COVID-19 pandemic has been a pivot point for

digital health implementation, with some describing it as years' worth of digital transformation in just a few months [3-5]. This transformation has been made possible by the recognition from governments and regulatory bodies of the need for speed, flexibility, and action in response to the pandemic [6-8]. However, many current mHealth technologies rely heavily on patients' ability to self-diagnose and use self-help rather than being fully integrated into the clinic. In building a holistic and comprehensive digital mental health ecosystem that *augments* rather than *replaces* traditional mental health care approaches, strategies for the design, validation, and implementation of mHealth technologies should be carefully considered by innovators moving forward.

mHealth: Promise of Dimensional Psychiatry and Digital Phenotyping

Psychiatric nosology, as described in gold-standard diagnostic manuals, has often been criticized for being overly narrow, failing to capture the wide range of clinical symptoms that are endorsed by individuals with mental health conditions [9,10]. In this regard, moving away from a categorical, symptom-count approach to a transdiagnostic view of psychopathology may facilitate early and accurate diagnosis, and allow for personalized treatment strategies. Given that time is a premium in clinical settings, where relying on brief symptom-count checklists such as the Patient Health Questionnaire-9 [11] and the seven-item Generalized Anxiety Disorder questionnaire [12] is common practice, digital technologies may offer an innovative way to improve diagnostic accuracy and advance mental health care provisions. Indeed, digital technologies such as mHealth apps have the capacity for the collection of a vast range of key transdiagnostic data via active reporting (ie, self-administered symptom reporting or monitoring via an app), providing valuable information on relevant social and demographic factors as well as current and past symptom profiles [13]. Digital technologies, including adaptive/nonlinear questionnaires where patients are required to answer questions based on previous answers, can further personalize and streamline the collection of active cross-disorder symptom data.

Moving beyond active symptom data, passive data, which measure aspects of daily living and can be collected continuously in the background, such as activity rhythms, text and call logs, sleep quality, keyboard reaction time [14], speech phonation, coherence, sentiment, and language patterns [15], can be applied to better inform mental health assessments in a process called digital phenotyping [16]. Data platforms that gather information from various digital sources have started emerging with the goal of collecting large-scale epidemiological and longitudinal data to identify risk factors and invest in mental health resources effectively. An example of this approach is offered by the recently launched Mental Health America (MHA) dashboard supported by MHA and Lundbeck. The dashboard collects publicly available data and anonymized data from web screening tools offered by MHA to highlight mental health hotspots across the United States. The objective is to implement advanced targeted interventions, address disparities in a timely manner, and design policies for at-risk populations [17,18]. Furthermore, advances in mHealth platforms (eg, Apple HealthKit or Google Fit) also allow for the bundling of data from different sensors such as accelerometers, microphones, GPS sensors, and gyroscopes, enabling the collection of physical and mental health information [19].

Critically, digital phenotyping for mental health is still in its infancy, with further research into how these data fit into or build upon the current diagnostic guidelines and criteria being a necessity moving forward. Both active as well as passive data collection also raise potential serious ethical concerns, which will need to be legislated and regulated. In the future, combining data from active and passive monitoring sources may feed new data into mental state examinations, allowing for more comprehensive data collection outside the time-pressured environment of clinical consultations. In turn, mHealth technologies have the potential to streamline mental health care delivery and management by reducing the number of in-person appointments, thereby freeing up clinician time, resulting in cost savings for both the provider and patient. However, these technologies usually require patients and clinicians to engage in new behaviors, which may be accompanied by reluctance to change. As a result, uptake of mHealth tools critically depends on perceived relevance, effectiveness, confidentiality, and engagement of the end users. A way to address this is by utilizing behavior theories and engaging in codesigning approaches when designing digital mental health care solutions.

Designing mHealth Tools With Patients and Clinicians in Mind

Understanding the target patient population's need is pivotal to the design of effective mHealth tools. To this end, behavior theories such as the Health Belief Model (HBM) [20], Behavioral Intervention Technology (BIT) model [21], and behavior model for Just-in-Time Adaptive Interventions (JITAI) [22] offer frameworks that help identify key drivers and barriers to the uptake of new technology, starting from app subscription to adherence to recommendations delivered by mHealth tools [23-25]. For instance, the HBM can help mHealth developers better understand their target patient population by investigating their inclination to seeking mental health support and what drives their adherence to treatment (eg, trust in clinicians, severity of illness, social stigma). The HBM is based on the theory that a person's willingness to engage in treatment is primarily due to their health perceptions. Indeed, mental health studies have shown that the perceived severity of mental illnesses impacts treatment outcomes (eg, adherence and beliefs about treatments) [26]. Thus, using the HBM can inform developers on how to best engage patients in adhering to the use of a novel digital tool.

Similarly, the BIT model combines the understanding of technological features of mHealth tools with behavioral principles to guide innovators on how to design and implement technology that is useful and usable. For instance, motivational enhancement strategies (behavior principle) can be used in the form of push notifications (technological feature) to prompt patients to track their mood regularly. The BIT model approach can be especially valuable for researchers looking to develop

multifunctional tools for mental health, including tools for mood monitoring and treatment such as electronic cognitive behavioral therapy, as well as tracking adherence to medication. In these cases, the stepwise, structured approach of the BIT model can best inform developers on how to use different behavioral principles and technologies to achieve their clinical aims.

The JITAI model can be employed by researchers developing adaptive sampling and ecological momentary assessments aimed at collecting information at sufficient frequencies and quantities to be useful without overburdening mental health patients. Critically, a delicate balance between utility and tolerability must be sought when developing mHealth monitoring tools in mental health, especially in conditions with specific triggers (eg, social anxiety triggered by meeting strangers, harmful alcohol use triggered by stress). The behavioral model for JITAI offers a framework to systematically identify when patients are most vulnerable to experiencing symptoms and pinpoints circumstances in which patients may be most receptive to the digital intervention.

Transitioning into the digital mental health care space will also involve cultivating a collaborative approach that includes academics, software developers, computer and data scientists, investors, and health care professionals. Indeed, moving forward, the concept of codesigning will need to be at the forefront of development, involving *all* potential users and stakeholders as active collaborators in the creation of digital diagnostics and interventions. This is especially critical for mental health technologies, which require clinical support from health care professionals to encourage engagement. Moving from the view of designing a digital product to designing a technology-enabled service means that the aims, role of the practitioner, and technology are designed and evaluated simultaneously in the relevant context [27]. Shifting from the practice of designing *for* to designing *with* users can help humanize digital health care, and *augment* rather than *replace* traditional health care delivery and management systems. By laying these foundations, technology-enabled services can become a part of everyday clinical practice and set the scene for deeper technology-enabled changes. Critically, as well as leveraging on behavior theories and codesigning approaches to increase engagement and relevance, mHealth developers should also pursue implementation studies to validate and assess the efficacy of their interventions.

Clinical Trials to Build a Strong Evidence-Based Case for Psychiatric mHealth Tools

Recent studies show that the most commonly downloaded mental health apps mainly fall under the umbrella of well-being apps offering relaxation, meditation, or mindfulness skills rather than validated treatments, and that the scientific language used to support claims often lacks corresponding evidence in the literature [28,29]. Although these apps can be incredibly useful and helpful for some patients, they may not be effective for more severe mental health concerns. Furthermore, many of these mHealth tools rely heavily on patients' ability to self-diagnose and use self-help rather than being fully integrated into clinical service delivery. In building a holistic and comprehensive digital mental health ecosystem, innovators are encouraged to design technology in a more sophisticated manner, moving away from a one-size-fits-all well-being approach to embrace the development of clinically validated and integrated treatments.

Clinical trials and observational studies can help mHealth developers build a strong clinical case for their technologies and differentiate them from digital wellness apps. Moving forward, it should be noted that randomized controlled clinical trials typically bear little resemblance to clinical settings. Often, recruitment favors clinicians and patients who are interested in using digital health technologies and are therefore more likely to adhere to them. However, these participants may represent only a small proportion of potential users. It should then come as no surprise if digital technologies fail in general health care settings. Thus, innovators should carefully consider the opportunity to conduct the evaluation of digital mental health technologies within the intended context, and to assess both its effectiveness and implementation requirements [27]. This so-called hybrid pragmatic trial design can assess both treatment effects and implementation strategies [30] in real-world patient populations with outcome measurements routinely used in clinical practice [31]. Thus, implementation and sustainment strategies become integral to the clinical evaluation process and part of the intervention design from the very beginning. Importantly, performing clinical studies and citing published literature in the descriptions of digital tools can only partially address the difficulties encountered by providers, payors, and patients in assessing and choosing effective digital interventions. Standardized evaluation frameworks introduced by regulators and health care providers can create transparency and guide health care professionals toward clinically meaningful technologies, with efforts being made in this direction.

Defining Appraisal Frameworks for mHealth Solutions

Several academic groups have contributed to the design of evaluation frameworks. For instance, the American Psychiatric Association (APA) has developed an app evaluation framework [32] that has been used by the New York Department of Health in the construction of an app library. Following the APA model, Lagan et al [33] designed a health app evaluation model, which harmonized 45 preexisting frameworks and where answers can be either binary or numeric to permit an objective evaluation of (1) background and access, (2) data safety and privacy, (3) app effectiveness and clinical foundation, (4) user engagement, and (5) data integration. The app evaluation model can be used by clinicians and patients to inform decisions on the suitability of candidate apps for the intended clinical use. Mental health apps often require patients to disclose sensitive personal information; thus, data safety and privacy are fundamental. To this end, the framework assesses apps on the possibility of deleting data, the format in which data are shared (nonanonymized, deidentified, anonymized), and whether data are shared with third parties. Thus, data handling procedures should be clearly stated in an app's description or privacy policy.

Ultimately, compliance with data security statutes and regulations (eg, Health Insurance Portability and Accountability Act and General Data Protection Regulation) is paramount for mental health apps collecting sensitive patient data. Clinicians will only recommend apps that fulfill data security requirements and that can easily be integrated into existing clinical workflows.

Regulators and health care providers are introducing additional ways of evaluating and certifying digital health technologies that do not fall under hardware medical device regulations. Innovators should work with regulatory bodies and be aware of the innovation plans for appraisal frameworks. The Food and Drug Administration (FDA) developed the Digital Health Innovation Action Plan to provide a more streamlined and efficient regulatory oversight of digital health technologies [34]. The Plan includes the Software Precertification Pilot Program, which precertifies organizations, not individual products [35]. An "excellence appraisal" is conducted to give the FDA the assurance that the organization is capable of producing high-quality, safe, and effective medical software products so that individual product reviews are carried out more rapidly and swiftly. An international approach to one-off regulatory appraisals also emerged with the Medical Device Single Audit Program (MDSAP), piloted in the United States, Australia, Brazil, Canada, and Japan. The objective of the MDSAP is to develop a single regulatory audit process that can satisfy the needs of multiple regulatory jurisdictions [36]. Europe stands as an observer in the MDSAP, and the new Medical Device Regulation mainly focuses on hardware devices; however, steps have been taken toward a more software-specific approach [37]. In the United Kingdom, the National Health Service (NHS) X developed the beta version of Digital Technology Assessment Criteria for health and social care, which outlines baseline assessment criteria to validate the suitability and function of digital health technologies for use by the NHS, social care staff, or directly by citizens [38]. Trade associations such as the Digital Therapeutics Alliance (DTA) are promoting industry standards and evaluation frameworks to enhance the understanding, adoption, integration, and reimbursement of clinically evaluated digital therapeutics into mainstream health care. The DTA's member companies include large pharmaceutical companies such as Novartis and Otsuka, as well as digital therapeutics developers such as Pear Therapeutics and SilverCloud, along with technology companies such as Philips, demonstrating the interdisciplinary efforts in pushing forward digital health technologies.

Compliance with data security regulations, assessment against commonly used frameworks, and certifications issued by regulatory bodies can help innovators build transparency and trust with their stakeholders. Importantly, presenting efficacy data and privacy matters in clear terms is essential to engage providers in the uptake of novel digital tools. In other areas of medicine, a clear evidence base, ease of use, and swift integration into clinical workflows have proven to facilitate the uptake of digital health technologies [39,40]. Experts have suggested that the introduction of novel technologies in health care could be facilitated by a dedicated health care profession called a "digital navigator," whose task in hospitals and clinics would be to select evidence-based apps, troubleshoot, and interpret digital data outputs in a clinically meaningful way [41]. However, mHealth developers should not heavily rely on the presence of specialized personnel; rather, they should focus on communicating evidence transparently, designing tools that fit into current care pathways and information technology system provisions, and providing actionable patient data. Consequently, the use of digital tools will not be limited to settings that have specialized staff members, and as uptake of these apps increases, it will become cost-effective to train and employ digital navigators to support further, more complex implementation of digital health tools.

Conclusion

Digital mental health technologies can enable the early identification of conditions, and facilitate disease management and treatment while empowering patients to make better informed decisions about their own health. Critically, there has been a movement toward technology that replaces the human element, rather than complementing and enhancing the invaluable involvement of health professionals in patient care. However, it is important not to throw decades or even centuries of clinical knowledge overboard. The way forward to building a digital mental health ecosystem must be more nuanced and sensitive. By developing patient-centric solutions, we can enable earlier and more accurate diagnoses, and in turn connect people with the most appropriate and effective treatment and care. As we move toward a postpandemic world, building the digital mental health ecosystem will require optimizing and adapting traditional mental health care services with considered step-wise innovation approaches, allowing health care professionals and patients to adapt to change. Innovative solutions will not come from leaps forward in technology but rather from new models of codesign, careful use of behavior theories, new pragmatic approaches to clinical evaluation, unprecedented collaboration in defining industry standards, and an open mind to trial-and-error implementation approaches.

The long-term goal for digital mental health technology innovators should be to design tools and services that can easily become so engrained in the health care ecosystem that what is now referred to as *digital health* will simply be called *health*.

Acknowledgments

This Viewpoint was funded by the Stanley Medical Research Institute (grant number 07R-1888).

Conflicts of Interest

SB is a director of Psynova Neurotech Ltd and Psyomics Ltd. SB and NAMK have financial interests in Psyomics Ltd. BS reports no conflicts of interest.

References

1. Steel Z, Marnane C, Iranpour C, Chey T, Jackson JW, Patel V, et al. The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013. Int J Epidemiol 2014 Apr;43(2):476-493 [FREE Full text] [doi: 10.1093/ije/dyu038] [Medline: 24648481]
2. Torous J, Roberts LW. Needed innovation in digital health and smartphone applications for mental health: transparency and trust. JAMA Psychiatry 2017 May 01;74(5):437-438. [doi: 10.1001/jamapsychiatry.2017.0262] [Medline: 28384700]
3. Nadella S. Empowering every developer, with Satya Nadella | Microsoft Build 2020 #MSBuild. Samir Dadoui's Technical Blog. URL: https://daoudisamir.com/empowering-every-developer-with-satya-nadella-microsoft-build-2020/ [accessed 2020-12-23]
4. Moltu C, Stefansen J, Svisdahl M, Veseth M. Negotiating the coresearcher mandate - service users' experiences of doing collaborative research on mental health. Disabil Rehabil 2012;34(19):1608-1616. [doi: 10.3109/09638288.2012.656792] [Medline: 22489612]
5. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
6. Oliver N, Lepri B, Sterly H, Lambiotte R, Deletaille S, De Nadai M, et al. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. Sci Adv 2020 Jun;6(23):eabc0764 [FREE Full text] [doi: 10.1126/sciadv.abc0764] [Medline: 32548274]
7. COVID-19 information governance advice for staff working in health and care organisations. NHSX. URL: https://perma.cc/HA25-ZRSY [accessed 2020-08-04]
8. Notification of enforcement discretion for telehealth remote communications during the COVID-19 nationwide public health emergency. HHS.gov. URL: https://www.hhs.gov/hipaa/for-professionals/special-topics/emergency-preparedness/notification-enforcement-discretion-telehealth/index.html [accessed 2021-05-17]
9. Dalgleish T, Black M, Johnston D, Bevan A. Transdiagnostic approaches to mental health problems: Current status and future directions. J Consult Clin Psychol 2020 Mar;88(3):179-195 [FREE Full text] [doi: 10.1037/ccp0000482] [Medline: 32068421]
10. Kendler KS. The nature of psychiatric disorders. World Psychiatry 2016 Feb;15(1):5-12. [doi: 10.1002/wps.20292] [Medline: 26833596]
11. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
12. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
13. Clarke DE, Kuhl EA. DSM-5 cross-cutting symptom measures: a step towards the future of psychiatric care? World Psychiatry 2014 Oct;13(3):314-316. [doi: 10.1002/wps.20154] [Medline: 25273306]
14. Mastoras R, Iakovakis D, Hadjidimitriou S, Charisis V, Kassie S, Alsaadi T, et al. Touchscreen typing pattern analysis for remote detection of the depressive tendency. Sci Rep 2019 Sep 16;9(1):13414. [doi: 10.1038/s41598-019-50002-9] [Medline: 31527640]
15. Scherer S, Lucas GM, Gratch J, Rizzo A, Morency LP. Self-reported symptoms of depression and PTSD are associated with reduced vowel space in screening interviews. IEEE Trans Affective Comput 2016 Jan 1;7(1):59-73. [doi: 10.1109/taffc.2015.2440264]
16. Torous J, Kiang MV, Lorme J, Onnela J. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR Ment Health 2016 May 05;3(2):e16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
17. Snyder BB. Lundbeck taps Mental Health America—and its big data—to create a national crisis dashboard. Fierce Pharma. URL: https://www.fiercepharma.com/marketing/lundbeck-taps-big-data-partnering-mental-health-america-to-create-a-national-mental [accessed 2021-05-14]
18. Identifying and addressing mental illness hot spots caused by the Covid-19 pandemic. Newsroom. URL: https://newsroom.lundbeckus.com/news-release/2020/identifying-and-addressing-mental-illness-hot-spots-caused-by-covid19-pandemic [accessed 2021-05-14]
19. Hermes S, Riasanow T, Clemons EK, Böhm M, Krcmar H. The digital transformation of the healthcare industry: exploring the rise of emerging platform ecosystems and their influence on the role of patients. Bus Res 2020 Sep 11;13(3):1033-1069. [doi: 10.1007/s40685-020-00125-x]

20. Baranowski T, Cullen KW, Nicklas T, Thompson D, Baranowski J. Are current health behavioral change models helpful in guiding prevention of weight gain efforts? Obes Res 2003 Oct;11(Suppl):23S-43S. [doi: 10.1038/oby.2003.222] [Medline: 14569036]
21. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 05;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
22. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. Health Psychol 2015 Dec;34S:1209-1219 [FREE Full text] [doi: 10.1037/hea0000306] [Medline: 26651462]
23. Naslund JA, Aschbrenner KA, Kim SJ, McHugo GJ, Unützer J, Bartels SJ, et al. Health behavior models for informing digital technology interventions for individuals with mental illness. Psychiatr Rehabil J 2017 Sep;40(3):325-335 [FREE Full text] [doi: 10.1037/prj0000246] [Medline: 28182469]
24. An S, Lee H. Adoption of mobile apps for mental health: socio-psychological and technological factors. In: Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering. 2018 Presented at: 7th International Conference, MobiHealth 2017; November 14-15, 2017; Vienna, Austria p. 29-37. [doi: 10.1007/978-3-319-98551-0_4]
25. Taj F, Klein MCA, van Halteren A. Digital health behavior change technology: bibliometric and scoping review of two decades of research. JMIR Mhealth Uhealth 2019 Dec 13;7(12):e13311 [FREE Full text] [doi: 10.2196/13311] [Medline: 31833836]
26. Averous P, Charbonnier E, Dany L. Relationship between illness representations, psychosocial adjustment, and treatment outcomes in mental disorders: a mini review. Front Psychol 2020;11:1167. [doi: 10.3389/fpsyg.2020.01167] [Medline: 32612557]
27. Mohr DC, Weingardt KR, Reddy M, Schueller SM. Three problems with current digital mental health research . . . and three things we can do about them. Psychiatr Serv 2017 May 01;68(5):427-429 [FREE Full text] [doi: 10.1176/appi.ps.201600541] [Medline: 28412890]
28. Larsen ME, Huckvale K, Nicholas J, Torous J, Birrell L, Li E, et al. Using science to sell apps: Evaluation of mental health app store quality claims. NPJ Digit Med 2019;2:18. [doi: 10.1038/s41746-019-0093-1] [Medline: 31304366]
29. Carlo AD, Hosseini Ghomi R, Renn BN, Areán PA. By the numbers: ratings and utilization of behavioral health mobile applications. NPJ Digit Med 2019;2:54. [doi: 10.1038/s41746-019-0129-6] [Medline: 31304400]
30. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012 Mar;50(3):217-226 [FREE Full text] [doi: 10.1097/MLR.0b013e3182408812] [Medline: 22310560]
31. Glasgow RE. What does it mean to be pragmatic? Pragmatic methods, measures, and models to facilitate research translation. Health Educ Behav 2013 Jun;40(3):257-265. [doi: 10.1177/1090198113486805] [Medline: 23709579]
32. The App Evaluation Model. American Psychiatric Association. 2018. URL: https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/the-app-evaluation-model [accessed 2020-12-23]
33. Lagan S, Aquino P, Emerson MR, Fortuna K, Walker R, Torous J. Actionable health app evaluation: translating expert frameworks into objective metrics. NPJ Digit Med 2020;3:100. [doi: 10.1038/s41746-020-00312-4] [Medline: 32821855]
34. Digital Health Innovation Action Plan. US Food and Drug Administration Center for Devices and Radiological Health (CDRH). 2017. URL: https://www.fda.gov/media/106331/download [accessed 2020-12-23]
35. Developing a Software Precertification Program: A Working Model. US Food and Drug Association. 2019. URL: https://www.fda.gov/media/119722/download [accessed 2020-12-23]
36. Medical Device Single Audit Program (MDSAP). TÜV SUD. URL: https://tinyurl.com/27uhv6d7 [accessed 2020-12-23]
37. Reflection Paper on regulatory frameworks for digital health technologies in Europe. Digital Europe. 2019 May 04. URL: https://www.digitaleurope.org/wp/wp-content/uploads/2019/04/DIGITALEUROPE-Reflection-on-regulatory-frameworks-for-digital-health-technologies-in-Europe-.pdf [accessed 2020-12-23]
38. Digital Technology Assessment Criteria – beta. NHSX. URL: https://www.nhsx.nhs.uk/key-tools-and-info/designing-and-building-products-and-services/ [accessed 2020-12-23]
39. Palacholla RS, Fischer N, Coleman A, Agboola S, Kirley K, Felsted J, et al. Provider- and patient-related barriers to and facilitators of digital health technology adoption for hypertension management: scoping review. JMIR Cardio 2019 Mar 26;3(1):e11951 [FREE Full text] [doi: 10.2196/11951] [Medline: 31758771]
40. Slevin P, Kessie T, Cullen J, Butler MW, Donnelly SC, Caulfield B. Exploring the barriers and facilitators for the use of digital health technologies for the management of COPD: a qualitative study of clinician perceptions. QJM 2020 Mar 01;113(3):163-172. [doi: 10.1093/qjmed/hcz241] [Medline: 31545374]
41. Wisniewski H, Torous J. Digital navigators to implement smartphone and digital tools in care. Acta Psychiatr Scand 2020 Apr;141(4):350-355 [FREE Full text] [doi: 10.1111/acps.13149] [Medline: 31930477]

Abbreviations

APA: American Psychiatric Association
BIT: Behavioral Intervention Technology model
DTA: Digital Therapeutics Alliance
FDA: Food and Drug Administration
HBM: Health Belief Model
JITAI: Just-in-Time Adaptive Interventions
MDSAP: Medical Device Single Audit Program
MHA: Mental Health America
mHealth: mobile health
NHS: National Health Service

Edited by R Kukafka; submitted 28.01.21; peer-reviewed by M Gottschalk, K Kalimeri; comments to author 26.03.21; revised version received 20.05.21; accepted 05.07.21; published 13.10.21.

Please cite as:
Spadaro B, Martin-Key NA, Bahn S
Building the Digital Mental Health Ecosystem: Opportunities and Challenges for Mobile Health Innovators
J Med Internet Res 2021;23(10):e27507
URL: https://www.jmir.org/2021/10/e27507
doi:10.2196/27507
PMID:34643537

Original Paper

Mobile Health Apps for Pregnant Women: Systematic Search, Evaluation, and Analysis of Features

Gabriela Frid[1], BS; Kelly Bogaert[1], MD; Katherine T Chen[1], MD, MPH

Icahn School of Medicine at Mount Sinai, New York, NY, United States

Corresponding Author:
Gabriela Frid, BS
Icahn School of Medicine at Mount Sinai
1 Gustace L. Levy Pl
New York, NY, 10029
United States
Phone: 1 3478664514
Email: gabriela.frid@icahn.mssm.edu

Abstract

Background: Many pregnant women use the internet to obtain information about pregnancy and childbirth. Over 50% of pregnant women use pregnancy apps and must search through thousands of pregnancy or women's health–related apps available on app stores. The COVID-19 pandemic is changing how women receive prenatal care. Mobile health apps may help maintain women's satisfaction with their prenatal care.

Objective: Our objective is to identify pregnancy mobile apps and to evaluate the apps using a modified APPLICATIONS (app comprehensiveness, price, privacy, literature used, in-app purchases, connectivity, advertisements, text search field, images/videos, other special features, navigation ease, subjective presentation) scoring system.

Methods: A list of pregnancy apps was identified in the first 20 Google search results using the search term "pregnancy app." After excluding irrelevant, inaccurate, malfunctioning, or no longer available apps, all unique apps were downloaded and evaluated with the modified APPLICATIONS scoring system, which includes both objective and subjective criteria and evaluation of special features.

Results: A list of 57 unique pregnancy apps was generated. After 28 apps were excluded, the remaining 29 apps were evaluated, with a mean score of 9.4 points out of a maximum of 16. The highest scoring app scored 15 points. Over 60% (18/29) of apps did not have comprehensive information for every stage of pregnancy or did not contain all four desired components of pregnancy apps: health promotion/patient education, communication, health tracking, and notifications and reminders. Only 24% (7/29) of apps included a text search field, and only 28% (8/29) of apps cited literature.

Conclusions: Our search yielded many high-scoring apps, but few contained all desired components and features. This list of identified and rated apps can lessen the burden on pregnant women and providers to find available apps on their own. Although health care providers should continue to vet apps before recommending them to patients, these findings also highlight that a Google search is a successful way for patients and providers to find useful and comprehensive pregnancy apps.

(J Med Internet Res 2021;23(10):e25667) doi:10.2196/25667

KEYWORDS

app; app store; mHealth; mobile health; prenatal; pregnancy; women's health

Introduction

The internet and smartphones are increasingly popular both as a means to access health care information and as tools for health care management [1]. During the COVID-19 pandemic, in-person health care visits dropped and patients sought care via different methods [2]. Recent data shows that 21% of prenatal visits in March 2020 were impacted (ie, cancelled, rescheduled, or conducted remotely) [3]. In April 2020, 26% of prenatal visits were impacted [3]. For those that had remote phone or video visits, less than half felt they received the same amount or more information and care as compared to that received during a traditional visit [3]. The pandemic has also caused a 25% increase in mobile health app downloads as compared to the same time last year [4].

Internet use during pregnancy is common [5]. The majority of women in the United States own a smartphone and past studies have shown that over 50% of pregnant women download apps focused on pregnancy, downloading an average of three during the prenatal period [6-8]. Women who use mobile prenatal care apps and have fewer prenatal visits have no reduction in patient satisfaction [9]. Health care–related apps account for a large market share of the apps available on the US iOS App Store, with almost 100,000 apps as of 2017 [10]. A 2013 study showed that pregnancy apps are the most used health apps [11]. Although 94% of pregnant women reported that a smartphone changed their lives for the better, there are many apps to choose from and the market for apps is constantly growing [6].

Due to the number of apps available to pregnant women, knowing what apps are available, how to select among them, and what information and features women are seeking is important. Wang et al [12] surveyed 535 women and showed that the most common reasons women used apps were to monitor fetal development (83%) and to obtain information on nutrition (26.2%) and antenatal care (23.9%). Lee et al [8] evaluated 47 apps that were identified by surveying 193 women and reported that most women decided to download an app after evaluating its content themselves (61.7%), and that the most frequently cited benefit was convenience (35.8%), while the most common weakness was lack of credibility (39%).

These studies provide information about why and how women use apps, but there have been no critical quantitative analyses done to evaluate and rate these apps in terms of their usefulness and benefit to pregnant women. The purpose of this study is to identify and evaluate pregnancy apps recommended to women on the internet. Our study additionally sought to create a comprehensive list of app features and a quantitative measure of comprehensiveness.

Methods

This study did not require review by the Institutional Review Board at the Icahn School of Medicine at Mount Sinai as it did not involve human subjects. We used the web search engine developed by Google, using the search term “pregnancy app” in March 2019. Search engines and online reading are top ways that consumers discover apps, and we selected Google’s search engine because it accounts for over 90% of all search queries [13,14]. The first page, with an average of 10 search results, accounts for 91.5% of traffic, while the second page accounts for 4.8% [15]. We confirmed that the results displayed were nearly identical when searched through different devices (eg, computer, iPad, and iPhone) and from different geographical locations. Google Incognito mode was used to limit variation of results based on the user’s browser cookies.

We downloaded all apps mentioned within the first 20 search results. We noted characteristic information—such as app name, app store, and app icon—and evaluated app content using an adapted APPLICATIONS (app comprehensiveness, price, privacy, literature used, in-app purchases, connectivity, advertisements, text search field, images/videos, other special features, navigation ease, subjective presentation) scoring system. APPLICATIONS is an acronym for components of apps that can be scored to help determine the rating of the app [16]. The usefulness and benefit of the apps was determined by the app comprehensiveness score, which judges the ability of an app to provide useful pregnancy information, and by the other components of the APPLICATIONS scoring system, which judge other aspects that have been shown to make apps successful [16]. This score was devised based on availability of information on the four distinct portions of pregnancy care, as well as the top four functions of the most commonly used pregnancy apps, as per Lee et al [8].

Other modified components of the original APPLICATIONS scoring system are described here. “Price” and “Paid subscription” were combined into one “Price” score. “Privacy” was added because while the initial APPLICATIONS scoring system was created with providers in mind, this study is evaluating patient-facing apps, and patients have expressed security and privacy concerns as an important consideration when choosing whether to use an app [17]. “Interdevice compatibility” was removed as each app denotes the proper platform to use. “Images/videos” was added, which was part of “Other components” previously. The “Other components” category was therefore renamed “Other special features” and the scoring was expanded due to the numerous special features provided on pregnancy apps, as the initial APPLICATIONS scoring system was intended for pregnancy wheel dating apps, which serve a more targeted, specific function and are less likely to have numerous other features.

One author (GF) checked the apps in the app stores to see if there were in-app purchases and also downloaded and opened the apps in airplane mode to evaluate the connectivity component and determine if functionality was dependent on internet access. All three authors evaluated the remaining features using the APPLICATIONS scoring system, as shown in Table 1. App comprehensiveness was determined as shown in Textbox 1. The “Other special features” category was tabulated in Multimedia Appendix 1. Navigation ease and subjective appearance were scored using a Likert scale with 1=poor, 2=below average, 3=average, 4=above average, and 5=excellent.

Table 1. The APPLICATIONS scoring system [16,17].

Component	Maximum score	Description
App comprehensiveness	3	0=none, 1=1-2 components, 2=3-5 components, 3=6-8 components
Price	1	0=priced, 1=free
Privacy	1	0=none, 1=privacy statement or login
Literature used	1	0=no references, 1=references used
In-app purchases	1	0=present, 1=absent
Connectivity	1	0=internet required, 1=internet not required
Advertisements	1	0=present, 1=absent
Text search field	1	0=no search field, 1=search field present
Images/videos	2	0=absent, 1=images or videos, 2=images and videos
Other special features	2	0=absent, 1=1-4 special features, 2=5-9 special features
Navigation ease	1	0=ease of navigation score <3, 1=ease of navigation score ≥3
Subjective presentation	1	0=subjective presentation score <3, 1=subjective presentation ≥3
Total	16	Sum of all scores

Textbox 1. App comprehensiveness criteria [8].

Criteria:

- Health promotion/patient education
- Patient communication
- Health tracking
- Notifications and reminders
- Preconception information
- Antepartum information
- Intrapartum information
- Postpartum information

Scoring: 0 points if no components were present, 1 point for 1-2 components, 2 points for 3-5 components, and 3 points for 6-8 components.

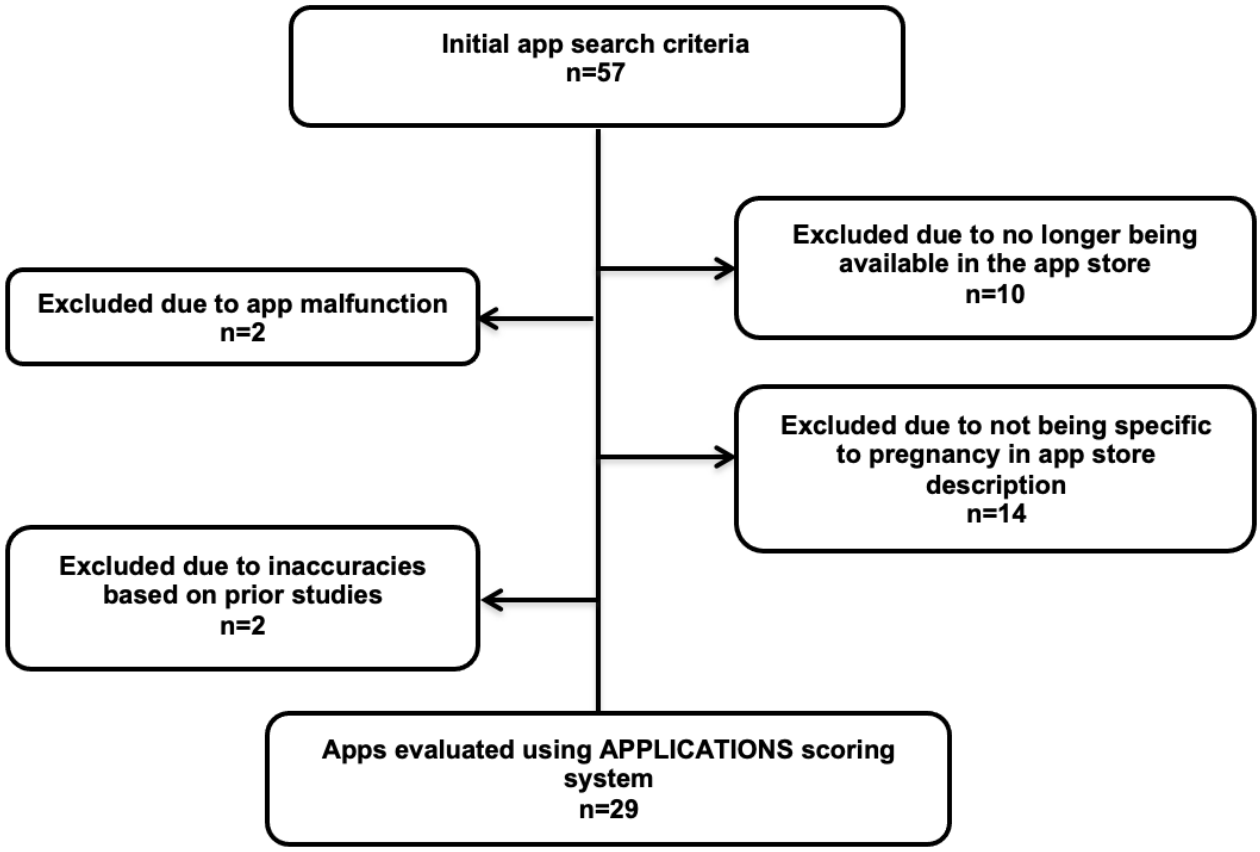
To account for any interobserver differences, we reconciled ratings and recorded objective errors. Our reconciliation process included a meeting of all authors during which we discussed each feature of each app. When there was a discrepancy, we determined whether it was a transcription or misclassification mistake and arrived at 100% consensus for the objective components of each app. For navigation ease and subjective presentation, we averaged reviewers' scores and awarded 0 points for an average rating of <3 and 1 point for an average rating of ≥3. We subsequently calculated a final total score for each app.

Results

The first 20 search results from the first two pages of a Google search on a computer for the term "pregnancy app" were recorded. The results were either specific apps (n=4) or articles about apps (n=16). All the apps listed in each result were noted and this search yielded 57 apps.

A total of 28 apps were excluded for the following reasons (Figure 1): 12 apps were no longer available or did not work, 14 apps were not related to pregnancy based on not having the word "pregnancy" in the description on the app store, and 2 were deemed inaccurate based on prior studies [18,19]. The remaining 29 apps were all downloaded on the Apple App Store or Google Play Store and evaluated between July-November 2019.

Figure 1. Google search for pregnancy apps. APPLICATIONS: app comprehensiveness, price, privacy, literature used, in-app purchases, connectivity, advertisements, text search field, images/videos, other special features, navigation ease, subjective presentation.

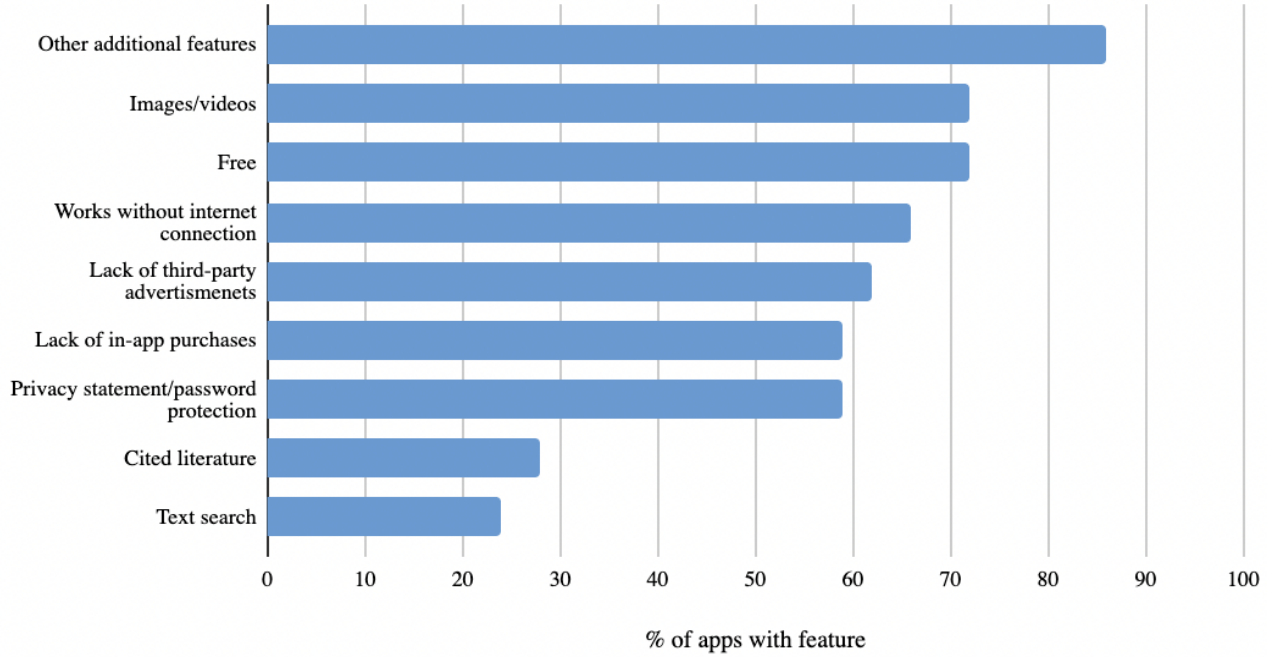


The 29 remaining apps were scored using the APPLICATIONS scoring system by all authors (Multimedia Appendix 1). The highest scoring app earned 15 out of 16 possible points. In addition, 11 apps scored above the mean score of 9.4. The lowest scoring app earned 4 points.

We determined that 41% (12/29) of apps included ≥6 comprehensiveness components out of 8, while 24% (7/29) of apps only included 1-2 comprehensiveness components out of 8, signifying missing information with regard to all stages of pregnancy or desired functionalities of pregnancy apps: health promotion/patient education, communication, health tracking, and notifications and reminders.

Evaluated features and functionality are shown in Figure 2. Common features and functionality included special features (25/29, 86%), free cost (21/29, 72%), images and/or videos (21/29, 72%), offline functionality (19/29, 66%), including a privacy statement or password protection (17/29, 59%), lack of third-party advertisements (18/29, 62%), and lack of in-app purchases (17/29, 59%). Use of cited literature (8/29, 28%) and text search (7/29, 24%) were rare features.

Figure 2. Features and functionalities of apps.



The objective component reporting error rate was 31 of 435 (7.1%), meaning that the authors independently assigned the same score to a given objective component of an app 92.9% of the time.

Discussion

Principal Results

In this study, we simulated how pregnant women would find apps by performing a Google search and then we evaluated and rated the identified apps. Although our results present several high-scoring apps, few contain all the components and features that make up an accurate and comprehensive app. More research is needed to survey patients and providers about which features are the most highly desired and needed. Many apps were missing cited sources, making it difficult to interpret accuracy. The rarest component was a text search, which can make it difficult for women to sift through information and quickly find what they are looking for. The most common app-specific features were contraction timers, journaling/photo uploads, and appointment trackers, checklists, and calendars. The least common were tools for obtaining safety information and health/fitness, despite these being common reasons pregnant women seek out apps [12].

Comparison With Prior Work

Prior studies have evaluated apps by asking pregnant women directly what apps they are using or through a systematic search of the app store. We sought to evaluate the most likely common apps that pregnant women are to encounter, which we did through a Google search [14]. Furthermore, to our knowledge, pregnancy apps have not previously been scored with the APPLICATIONS scoring system. App stores have an abundance of options, requiring consumers to sift through apps that may not be relevant (such as humorous apps for fake pregnancy tests), which can easily result in them feeling overloaded [20]. As most of the Google search results were written articles describing apps, erroneous and irrelevant apps from the app store are less likely to be included. This study is able to provide patients and providers with a curated list of rated apps and their features.

Clinical Implications

The information found in this study may be particularly useful to women during the COVID-19 pandemic. Pregnant women often seek information on the internet and do not discuss much of what they find with their physicians [21]. Due to the pandemic, half of adults say that they or someone in their household has skipped seeking medical care due to anxiety about contracting COVID-19 [22]. Prenatal visits have needed to be rescheduled or modified, and many women are not satisfied with the level of information they are receiving through phone or video visits. The combination of these factors has likely led to the increase in pregnancy app downloads, which has occurred at a time when pregnant women have fewer touchpoints with doctors where false information could be corrected, emphasizing the need for high-quality and accurate pregnancy apps [4].

Limitations

Given the dynamic nature of Google search results and the app store, several factors limit our analysis. This was evident in our search, as 12 of the apps initially identified were either no longer available or did not work by the time we tried to evaluate them. Additionally, we decided to use only the first two pages of Google search results as they account for most views. It is possible that other apps present on later pages were useful and comprehensive.

Conclusions

The current method of app selection by the majority of women is to download the app and search the content themselves [12]. Although this may yield good results, it also means women must download multiple apps, as well as pay for many of these

apps. On average, Lee showed that users download over three apps and are more likely to download a free app versus a paid app [8]. By using a Google search, we were able to identify many high-scoring apps that may be used during pregnancy. The identified rated apps can lessen the burden on pregnant women and providers to search for useful and comprehensive apps on their own. As telehealth continues to expand, more research is needed in the area of pregnancy app development.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of evaluated and rated pregnancy apps.
[DOCX File , 294 KB - jmir_v23i10e25667_app1.docx]

References

1. Mosa ASM, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012 Jul 10;12(1):67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
2. Ateev M. The Impact of the COVID-19 Pandemic on Outpatient Visits: A Rebound Emerges. To the Point (blog), Commonwealth Fund. 2020 May 19. URL: https://www.commonwealthfund.org/publications/2020/apr/impact-covid-19-outpatient-visits [accessed 2021-09-26]
3. Burgess A, Breman R, Bradley D, Dada S, Burcher P. Pregnant Women's Reports of the Impact of COVID-19 on Pregnancy, Prenatal Care, and Infant Feeding Plans. MCN Am J Matern Child Nurs 2021;46(1):21-29. [doi: 10.1097/NMC.0000000000000673] [Medline: 33009009]
4. ORCHA. COVID-19: Digital Health Trends Report. Organization for the Review of Care and Health Applications. 2020 Sep 22. URL: https://orchahealth.com/covid19-digital-health-trends-report/ [accessed 2021-09-26]
5. Romano AM. A Changing Landscape: Implications of Pregnant Women's Internet Use for Childbirth Educators. J Perinat Educ 2007;16(4):18-24 [FREE Full text] [doi: 10.1624/105812407X244903] [Medline: 18769519]
6. Patient Adoption of mHealth: Use, Evidence and Remaining Barriers to Mainstream Acceptance. IMS Institute for Healthcare Informatics. 2015 Sep. URL: https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/patient-adoption-of-mhealth.pdf [accessed 2021-09-26]
7. Jayaseelan R, Pichandy C, Rushandramani D. Usage of Smartphone Apps by Women on their Maternal Life. J Mass Communicat Journalism 2015 Oct 29;05(07):158-164. [doi: 10.4172/2165-7912.1000267]
8. Lee Y, Moon M. Utilization and Content Evaluation of Mobile Applications for Pregnancy, Birth, and Child Care. Healthc Inform Res 2016 Apr;22(2):73-80 [FREE Full text] [doi: 10.4258/hir.2016.22.2.73] [Medline: 27200216]
9. Marko KI, Ganju N, Krapf JM, Gaba ND, Brown JA, Benham JJ, et al. A Mobile Prenatal Care App to Reduce In-Person Visits: Prospective Controlled Trial. JMIR mHealth uHealth 2019 May 01;7(5):e10520 [FREE Full text] [doi: 10.2196/10520] [Medline: 31042154]
10. Jahns RG. The market for mHealth app services will reach $26 billion by 2017. Research 2 Guidance. 2013. URL: http://www.research2guidance.com/the-market-for-mhealth-app-services-will-reach-26-billion-by-2017/ [accessed 2021-09-26]
11. Dolan B. Report finds pregnancy apps more popular than fitness apps. Mobi Health News. URL: https://www.mobihealthnews.com/20333/report-finds-pregnancy-apps-more-popular-than-fitness-apps [accessed 2021-09-26]
12. Wang N, Deng Z, Wen LM, Ding Y, He G. Understanding the Use of Smartphone Apps for Health Information Among Pregnant Chinese Women: Mixed Methods Study. JMIR mHealth uHealth 2019 Jun 18;7(6):e12631 [FREE Full text] [doi: 10.2196/12631] [Medline: 31215516]
13. James Tiongsen. Mobile app marketing insights: How consumers really find and use your apps. Think with Google. 2015 May. URL: https://www.thinkwithgoogle.com/marketing-strategies/app-and-mobile/mobile-app-marketing-insights/ [accessed 2021-09-26]
14. Desjardins J. How Google retains more than 90% of market share. Business Insider. URL: https://www.businessinsider.com/how-google-retains-more-than-90-of-market-share-2018-4 [accessed 2021-09-26]
15. The value of google result positioning. Chitika Insights. URL: https://www.benchmark.no/wp-content/uploads/2020/04/chitikainsights-valueofgoogleresultspositioning.pdf [accessed 2021-09-26]
16. Chyjek K, Farag S, Chen K. Rating Pregnancy Wheel Applications Using the APPLICATIONS Scoring System. Obstet Gynecol 2015 Jun;125(6):1478-1483. [doi: 10.1097/AOG.0000000000000842] [Medline: 26000520]
17. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and Facilitators of the Use of Mobile Health Apps From a Security Perspective: Mixed-Methods Study. JMIR mHealth uHealth 2019 Apr 16;7(4):e11223 [FREE Full text] [doi: 10.2196/11223] [Medline: 30990458]
18. Moglia M, Nguyen H, Chyjek K, Chen K, Castaño PM. Evaluation of Smartphone Menstrual Cycle Tracking Applications Using an Adapted APPLICATIONS Scoring System. Obstet Gynecol 2016 Jun;127(6):1153-1160. [doi: 10.1097/AOG.0000000000001444] [Medline: 27159760]

19. Soffer MD, Chen KT. In Search of Accurate Fetal Heart Rate Monitoring Mobile Applications. Telemed J E Health 2019 Sep;25(9):870-877. [doi: 10.1089/tmj.2018.0104] [Medline: 30358492]
20. Cook J. A dangerous pregnancy scam is trending on the app store. The Huffington Post. URL: https://www.huffpost.com/entry/lunar-app-pregnancy-test_n_5dfb9cfae4b006dceaaa4148 [accessed 2021-09-26]
21. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: a systematic review. BMC Pregnancy Childbirth 2016 Mar 28;16:65 [FREE Full text] [doi: 10.1186/s12884-016-0856-5] [Medline: 27021727]
22. Hamel L, Kearney A, Kirzinger A, Lopes L, Muñana C, Brodie M. Impact of Coronavirus on personal health, economic and food security, and Medicaid. KFF Health Tracking Poll. 2020 May 27. URL: https://www.kff.org/report-section/kff-health-tracking-poll-may-2020-health-and-economic-impacts/ [accessed 2021-09-26]

Abbreviations

APPLICATIONS: app comprehensiveness, price, privacy, literature used, in-app purchases, connectivity, advertisements, text search field, images/videos, other special features, navigation ease, subjective presentation

Edited by R Kukafka; submitted 10.11.20; peer-reviewed by M Rodrigues, K Villinger; comments to author 28.01.21; revised version received 23.03.21; accepted 14.09.21; published 18.10.21.

Please cite as:
Frid G, Bogaert K, Chen KT
Mobile Health Apps for Pregnant Women: Systematic Search, Evaluation, and Analysis of Features
J Med Internet Res 2021;23(10):e25667
URL: https://www.jmir.org/2021/10/e25667
doi:10.2196/25667
PMID:34524100

Original Paper

Characteristics of Acute Childhood Illness Apps for Parents: Environmental Scan

James Benoit[1], PhD; Lisa Hartling[2], PhD; Michelle Chan[1], PhD; Shannon Scott[1], PhD, RN

[1]Faculty of Nursing, University of Alberta, Edmonton, AB, Canada

[2]Department of Pediatrics, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB, Canada

Corresponding Author:
Lisa Hartling, PhD
Department of Pediatrics
Faculty of Medicine and Dentistry
University of Alberta
4-472 Edmonton Clinic Health Academy
11405 87 Ave NW
Edmonton, AB, T6G 1C9
Canada
Phone: 1 780 492 6124
Email: hartling@ualberta.ca

Abstract

Background: Providing parents with resources that aid in the identification and management of acute childhood illnesses helps those parents feel better equipped to assess their children's health and significantly changes parental health-seeking behaviors. Some of these resources are limited by accessibility and scalability. Remote locations and staffing limitations create challenges for parents aiming to access their child's health information. Mobile health apps offer a scalable, accessible solution for improving health literacy by enabling access to health information through mobile devices.

Objective: The aim of our study is to create an inventory of acute childhood illness apps that are available to North American parents and caregivers, assess their quality, and identify the areas in which future apps can be improved.

Methods: We conducted an environmental scan to identify and summarize app information for parents and digital health researchers. The Google and Apple app marketplaces were used as search platforms. We built a list of search terms and searched the platforms for apps targeted at parents and related to acute pediatric illnesses in the United States and Canada. We assessed apps meeting the inclusion criteria using the Mobile App Rating Scale (MARS), a validated tool for assessing the quality of health apps. The MARS examines apps on 5 subscales: engagement, functionality, aesthetics, information quality, and subjective quality. Data were analyzed by MARS subscale averages and individual item scores.

Results: Overall, 650 unique apps were screened, and 53 (8.2%) were included. On a scale of 1-5, apps had an average engagement score of 2.82/5 (SD 0.86), functionality score of 3.98/5 (SD 0.72), aesthetics score of 3.09/5 (SD 0.87), information quality score of 2.73/5 (SD 1.32), and subjective quality score of 2.20/5 (SD 0.79). On the same scale of 1-5, app scores ranged from 2.2/5 to 4.5/5 (mean 3.2, SD 0.6). The top 3 MARS-scored apps were *Baby and Child First Aid* (4.5/5), *Ada* (4.5/5), and *HANDi Paediatric* (4.2/5). Taken together, the top 3 apps covered topics of emergency pediatric first aid, identification of (and appropriate response to) common childhood illnesses, a means of checking symptoms, and a means of responding to emergency situations. There was a lack of Canadian-based app content available to parents in both marketplaces; this space was filled with content originating primarily in the United Kingdom and the United States. In addition, published evidence of the impact of the included apps was poor: of 53 apps, only 5 (9%) had an evidence base showing that the app had been trialed for usability or efficacy.

Conclusions: There is a need for evidence-based acute childhood illness apps of Canadian origin. This environmental scan offers a comprehensive picture of the health app landscape by examining trends in acute childhood illness apps that are readily available to parents and by identifying gaps in app design.

(J Med Internet Res 2021;23(10):e29441) doi:10.2196/29441

KEYWORDS

internet; mHealth; mobile health; digital health; ehealth; app; mobile application; Android; Apple; marketplace; environmental scan; review; acute childhood illness; knowledge translation; child; parent; caregiver; mobile phone

Introduction

Background

The unexpected COVID-19 outbreak has affected how health information is communicated, how individuals seek health information and services, and how these services are delivered [1,2]. Families are delaying seeking health care for children, introducing consequences for their current and future health outcomes by delaying emergency care and wellness visits [3-6]. Notably, there has been a 57% reduction in pediatric emergency department (ED) use, with use inversely correlated with pandemic severity [7]. Although on the surface this may look encouraging, this decrease was primarily due to a drop in low-acuity visits [7]. Providing resources to parents that aid in the detection and identification of childhood illnesses helps them to feel better equipped to assess their child's health and significantly changes parental health-seeking behaviors [8]. For context, these illnesses could include conditions such as gastroenteritis, bronchiolitis, and asthma.

Traditionally, parents obtained health information primarily from friends, family, and physicians [9,10]. In clinical settings, conventional modes of communicating complex health information to parents include information sheets and pamphlets [11]. In Canada, another common health resource for parents is teletriage, accessed by calling 811 in all but one province and one territory (Manitoba and Nunavut, respectively), where nurses answer health- and illness-related questions. This service is highly used: for example, in the province of Alberta, 694,313 calls were made among a population of 4.3 million in 2018 [12]. However, more recent studies in health information–seeking behavior suggest that this may be changing, with an Australian pediatric children's hospital survey indicating that 96% of the parents use the internet and 63% use a smartphone to search for health information [13]. In conjunction with Google searches, this behavior is used to establish a cause of illness, access means of assessing symptom severity, and exercise prudence regarding visiting a physician [14]. Parents also search for support through web-based forums, and this means of information gathering provides additional reassurance and validation [15].

However, some of these information resources are limited by accessibility and scalability. Remote locations (ie, geography) and staffing limitations create challenges for parents aiming to access health information about their child [16,17]. These challenges are compounded by the existing variance in parents' health literacy and language skills [11] and their willingness to ask others for health information; for example, many parents feel unable to express their anxiety about their child's health because of fear that they will be perceived as worriers [18,19]. In addition, 30% of the Canadian parents who have children presenting to the ED have low health literacy [20], which accounted for 940,637 of the 3,135,457 children's visits to Canadian EDs in 2018 [21]. Notably, parents with low health literacy are 3 times as likely to bring their child in for nonurgent conditions [22]. It is therefore vital to the mission of delivering equitable health care that resources are made accessible to parents with lower health literacy.

Mobile health (mHealth) apps offer a potential solution for improving health literacy by enabling access to health information through a new medium: the app marketplaces found on mobile devices such as phones and tablets [23]. A large study of 4974 American adults demonstrated that significant associations exist between adequate health literacy and the use of health information technology such as patient portal apps [24]. More than 96% of Americans [25] and 87% of Canadians [26] own a mobile phone, and smartphones are owned by 81% and 78%, respectively. Of mobile phone owners, more than 97% use an Android- or Apple-based operating system [27]. In 2017, nearly a third of the Canadian adults used mHealth apps to monitor their health [28]. Importantly, increased use of health apps has been shown to have a significant correlation with improved health behaviors [29].

Previous studies examining parents' internet search patterns for health information have indicated that a variety of information-seeking strategies are used and that information is trusted differently based on its source (eg, information found from searching university- and hospital-based websites was considered far more accurate and safer than information found within public search engine results) [30]. However, little is known about parents' interactions with health apps. A recent review suggested that existing app assessment tools are targeted at expert users, and nonexperts such as parents and caregivers still lack these resources to make an informed decision about which apps to use [31]. An investigation of patient-facing apps showed that the participants were unable to complete 57% of the relevant tasks because of a lack of confidence with the app and frustration with its design and navigation, but paradoxically they remained interested in using apps to self-manage care [32].

There are differing opinions on the current use of apps for parents. It has been pointedly suggested that "apps don't help parents of sick kids" [33]. However, this viewpoint drew its conclusions from a review that included only 3 digital interventions found in scientific studies [34]. As another review of apps pointed out, there is little crossover between scientific studies and app marketplaces: in a review of pain-related apps, those seen in the marketplace are not seen in scientific studies and vice versa [35]. Given the large and increasing number of health-related apps available (more than 300,000 in 2017 [36,37]), there is a disparity between the size of the health app landscape and its representation in scientific studies. To create an actionable set of information for parents and caregivers, it will be useful to look beyond apps that appear in scientific studies.

Health App Landscape

To create an inventory of the apps available to parents in North America that provide acute pediatric health information, we conducted an environmental scan, a review technique rooted in

business that is designed to summarize information for decision-makers [38,39]. Environmental scans are conducted to identify trends and avoidable pitfalls in a specific area of inquiry, be it related to product, policy, or strategy. We used the results of this scan to create a picture of the health app landscape by examining trends in apps related to acute pediatric health that are readily available to parents and identifying the gaps in app design that can be addressed.

Methods

Overview

To carry out a structured environmental scan, we designed a search strategy that worked with the algorithms of Google's and Apple's app marketplaces. The finalized search strategy was applied by 2 reviewers (JB and MC) to these 2 app marketplaces in Canada and the United States. Next, 2 reviewers (JB and MC) screened the apps based on predefined inclusion criteria and extracted attributes about each app. We built evidence tables describing app characteristics (eg, number of reviews) and app quality (eg, quality of information) using items from the Mobile App Rating Scale (MARS) [40] and analyzed the results for trends. It should be noted that although this approach includes elements of previously suggested search strategies for apps (eg, the 7 strategies for assessing apps proposed in the study by Boudreaux et al [41]), some elements of such a strategy (eg, searching app clearinghouses) are no longer widely used, and others (eg, piloting the apps) are beyond the scope of this environmental scan.

Search Strategy

We designed a set of 17 search terms in collaboration with a research librarian to identify relevant apps in the Google Play Store and Apple App Store. We limited our searches to these stores, given that the scope of this scan was to investigate apps that are likely to be accessible to North American parents. As other environmental scans that were drawn from for our approach have discussed [42], there is little formal knowledge publicly available about the specifics of how either store's search function works, beyond testing for specific functionality (eg, the Google Play Store's search used Boolean terms, whereas the Apple App Store's did not). The results presented here will be *best available* knowledge based on internet searches and responses to information requests.

On the basis of responses to the queries sent to the Google Play Store and Apple App Store support teams, we learnt that the Google Play Store app search currently integrates data from title, publisher, and app descriptions. The results displayed are then limited by their search rank for the specific search term used. The Apple App Store app search has different input parameters: it uses app title, keywords, and primary category to search, whereas the app's promotional text does not affect search rank and the search does not use the app's full description. However, keywords are not displayed for apps, making it difficult to expand search terms directly based on visible information.

Apps not available in a particular country's store or on a particular device were not visible in the results. Similarly, personalization of results (demonstrated by 2 users entering the same search term and generating different lists of results based on hidden user metrics) makes it difficult for a researcher to be sure that their queries are returning all relevant apps. We addressed this issue by building software using the Google Play Store's and the Apple App Store's application programming interfaces (the intermediary that communicates among different pieces of software) for both the Canadian and US stores and conducted tests to ensure that personalized results were not being returned and that the same set of results would be returned when different users ran the same search. This approach to remove result personalization addresses the inconsistency of personalized search results among users.

We examined the first 10 apps relevant to our scan that appeared using plain-language search terms (eg, child illness) for vocabulary in the stores' description related to the app's intended purpose and compiled a list of 17 search terms using language that appeared frequently: *Child illness*; *Pediatric illness*; *Paediatric illness*; *Child symptoms*; *Pediatric symptoms*; *Paediatric symptoms*; *Child care illness*; *Pediatric care illness*; *Paediatric care illness*; *Child diagnosis*; *Pediatric diagnosis*; *Paediatric diagnosis*; *Parent care child*; *Parent child illness*; *Child carer*; *Pediatric carer*; *Paediatric carer*.

Search and Screening

These terms were searched 4 times on September 14, 2020, for 4 marketplace locations (the US Apple App Store, US Google Play Store, Canadian Apple App Store, and Canadian Google Play Store), and a maximum of 50 results per search term per store returned (for a maximum of 850 apps returned per store location across all searches). We built custom software in Python (Python Software Foundation) that used app marketplace application programming interfaces and stored search results in a comma-separated values database. We confirmed with test searches that the same set of results was returned on 2 different devices. The results were imported into an Excel spreadsheet (Microsoft Corporation). The MARS was adapted to a spreadsheet-ready format to allow for ease of charting.

Three inclusion criteria were used to screen apps:

1. The app contained content related to acute child (age 0-21 years) illness.
2. The app's intended audience included parents or informal carers (eg, non–health care professionals).
3. The app still existed in the Google Play Store or Apple App Store when being assessed.

In addition, ad hoc reasons were recorded when apps included in the screening process could not be assessed with the MARS (eg, once downloaded, the app contained no content).

The title and description of each app were independently assessed by 2 researchers (JB and MC) for inclusion. In cases of ambiguity or where disagreement occurred after the results were compared, photos of the app depicting examples of its appearance and content on the Google Play Store or Apple App Store were used to gather additional information for inclusion and exclusion. The 2 researchers discussed this information to reach consensus.

App Quality Assessment

We based the app assessments on the MARS, a validated tool designed for health app assessment [40,43]. The MARS has high internal consistency and interrater reliability, designed to classify and rate the quality of novel health apps; it is also used as a tool for guiding app design. The MARS assesses app quality based on 5 subscales: engagement, functionality, esthetics, information quality, and subjective quality. Each section includes a number of items that are directed questions. All items are scored out of 5, with a higher score indicating a better outcome. Each subscale score is determined by taking a mean of the subscale item scores. The MARS subscales can be found in its entirety in Multimedia Appendix 1.

To determine interrater agreement, 2 reviewers (JB and MC) conducted the MARS assessment for 10% of the apps, consistent with previous work [44]. We calculated the interrater agreement by identifying occurrences of major differences in the item scores (differences greater than 1) [42] and assessing the agreement between the 2 reviewers using the Cohen κ [45]. After excluding one app from comparison because of an unresolvable difference in app accessibility among the test devices, the interrater agreement among the remaining app items was found to be substantial at 0.69 [46], above the 0.60 threshold indicating that significant disagreements exist [47], and 1 reviewer conducted the remaining assessments.

Individual app assessments were conducted by installing each app on an Android or Apple device (if the same app was available on both devices, we compared the 2 versions for major differences and assessed the Android version). The apps were then launched and browsed for 10 minutes. When necessary, an account was created for apps that required a log in to access the app's content. After exploring the app, the MARS was applied, with the app accessed as necessary to gather information for specific scale items. The characteristics of apps that the MARS collects include information such as the app's rating, date of last update, and intended age group; a complete list is shown in Multimedia Appendix 1, Table S1. The 23 MARS items, grouped by subscale, are presented in Multimedia Appendix 1, Table S2. To determine whether the app appeared in the literature, Google Scholar was searched for occurrences of the app's name.

Analyses

The item scores for app quality were averaged across dimensions to create a set of 5 scores for each app. The app scores were averaged across items in the engagement, functionality, esthetics, and information subscales to create a final app score. Data were treated as interval level data, consistent with the recommendations by Norman around the increased robustness of parametric tests [48,49]. The scores were converted to a score out of 5 for consistency with the MARS scoring. The scores of the top 3 apps were compared with the mean score of all apps on each MARS item and subscale using two-tailed *t* tests.

The apps were then grouped based on app cost (free vs paid); the 6 paid apps averaged Can $3.66 (US $2.89) per app, with a range of Can $1.99-$6.99 (US $1.57-$5.52), and the group scores were compared among the MARS subscales. This analysis was repeated by grouping the apps into 2 categories reflecting the duration since their last update: apps last updated less than 6 months ago and apps updated more than 6 months ago.

Separate from the MARS scores, we conducted a quantitative examination of the types of apps available in each store. The apps were grouped into 3 functional categories, and these groups were compared among the 4 stores for significant differences using chi-square tests. The 3 app categories were informational, actionable, and consultatory. The apps were classified into 1 of these 3 groups based on their primary functionality. If an app contained elements of more than one category, we examined the app's description to determine how the developers intended the app to be used. Informational apps primarily provided information about children's illnesses, whereas actionable apps guided users to care recommendations, and consultatory apps provided a telehealth-based *virtual physician* service.

We also examined how the top 3 apps for childhood illnesses differed from the average app scores by dimension and scale item and compared them using *t* tests.

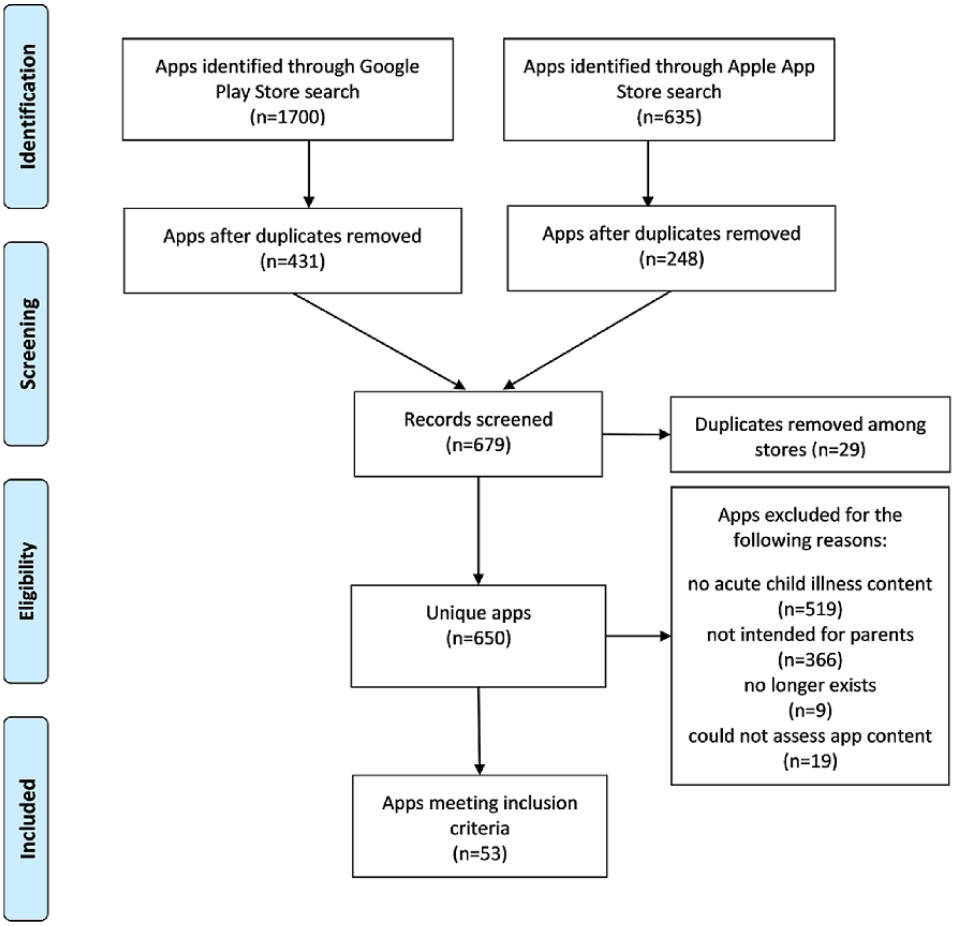
Results

Overview

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram (Figure 1) provides an overview of the search and screening process [50].

Our searches identified 2335 apps: 1700 (72.81%) from the Google Play Store and 635 (27.19%) from the Apple App Store. Of these 2335 apps, 650 (27.84%) were unique and 1685 (72.16%) were duplicates. After we applied our inclusion criteria to the 650 unique apps, 70 (10.8%) were considered potentially relevant; of these 70 apps, 17 (24%) were excluded, leaving 53 (76%) of apps for analysis. A total of 62% (33/53) of apps were available through the Google app marketplace and 55% (29/53) through the Apple app marketplace (9/53, 17%) of apps were available on both platforms).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram showing the environmental scan search and app screening process.



App Characteristics

All 53 apps focused on physical health: 19 targeted goal setting as a focus, whereas 18 targeted behavior change. A total of 50 apps had a theoretical background based in information or education delivery, 46 apps provided advice, tips, strategies, or skills training, 25 focused on assessment, and 17 dealt with information monitoring or tracking.

In all, 37 apps had a commercial affiliation, whereas 9 were government-affiliated, 3 had an affiliation with a nongovernmental organization, and one was affiliated with a university. Two apps—Asthma Action Hero and Pediatrics for All—did not fall into the MARS classification: the development of these apps was affiliated primarily with 2 individual physicians, Dr Helena K Bentley and Dr Hugo Rodrigues, respectively. Of the 53 apps assessed, 50 were targeted at young adults and adults, whereas one app was targeted at these groups as well as older adolescents, and 2 apps were targeted at general audiences, including children aged below 12 years. It should be noted that these findings are not in conflict with our inclusion criteria of an app's target audience consisting of parents: apps can target multiple user ages while including parents as an intended audience, and all age groups included in the MARS (with the exception of children) could include parents.

In terms of app requirements and functionality, 8 apps allowed password protection, 23 had sharing features (eg, options for posting to Facebook or Twitter), 19 had a visible app community, 27 required internet access to load all parts of the app (including advertisements), and 7 required users to have an account and log in. A total of 9 apps did not seem to contain any of these features.

MARS Ratings

The MARS subscales assess engagement using 5 items (mean 2.8/5, SD 0.86), functionality using 4 items (mean 4.0/5, SD 0.72), esthetics using 3 items (mean 3.1/5, SD 0.87), information quality using 7 items (mean 2.7/5, SD 1.32), and subjective quality using 4 items (mean 2.2/5, SD 0.79). The item scores for each dimension are summarized in Table 1.

Table 1. Mobile App Rating Scale item and subscale score averages for the assessed apps (N=53).

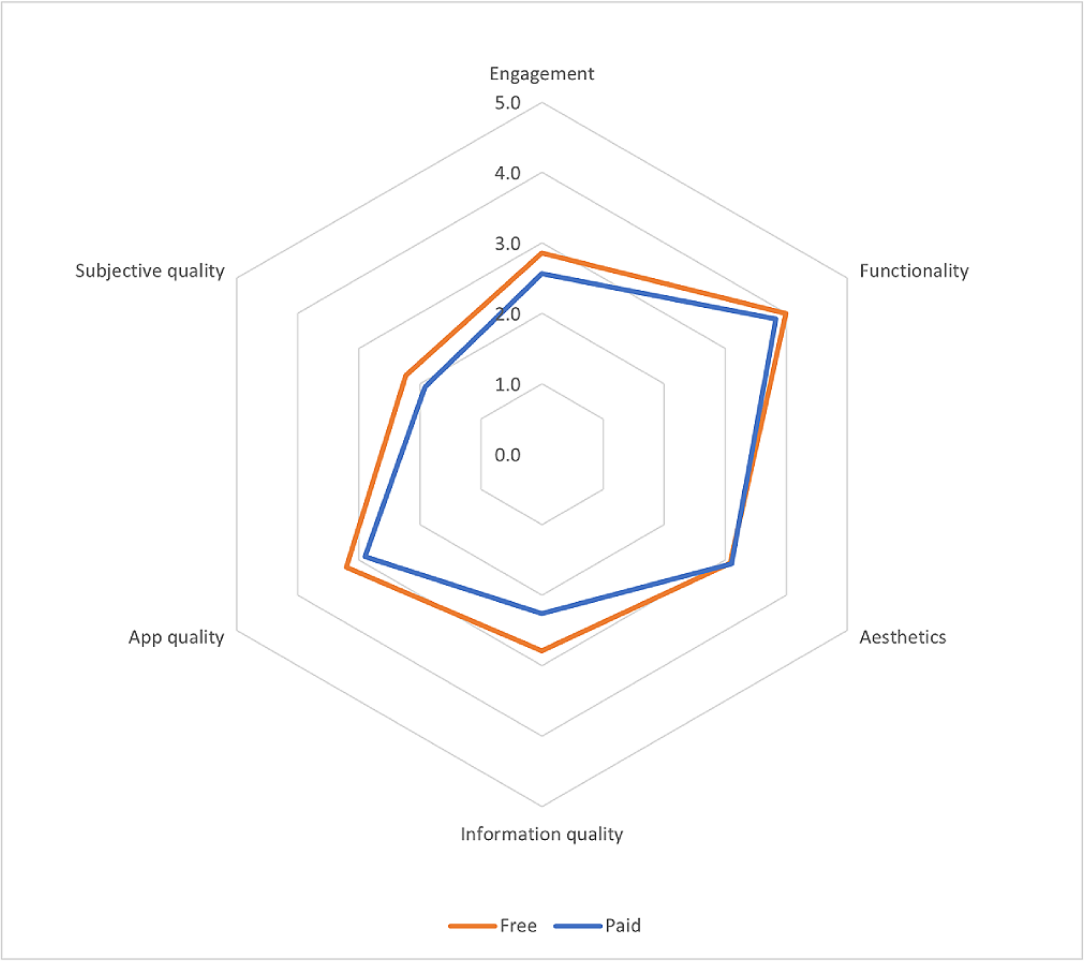
Subscale and item	Values, mean (SD)
Engagement	
Entertainment	2.2 (0.57)
Interest	3.1 (0.97)
Customization	2.0 (1.07)
Interactivity	3.3 (0.69)
Target audience	3.7 (0.92)
Engagement average	2.8 (0.86)
Functionality	
Performance	4.0 (0.81)
Ease of use	4.0 (0.68)
Navigation	3.9 (0.52)
Gestural design	4.1 (0.81)
Functionality average	4.0 (0.72)
Esthetics	
Layout	3.5 (0.75)
Graphics	3.0 (1.08)
Visual appeal	2.7 (0.72)
Aesthetics average	3.1 (0.87)
Information quality	
Description accuracy	4.3 (0.91)
Goal setting	2.5 (2.24)
Information quality	3.6 (0.95)
Information quantity	3.0 (1.09)
Visual information	3.0 (1.14)
Credibility	2.6 (1.47)
Evidence base	0.3 (0.89)
Information average	2.7 (1.32)
Subjective quality	
Recommended app	2.8 (1.31)
Frequency of app use	1.5 (0.54)
Willingness to pay for app	1.2 (0.71)
Subjective star rating	3.3 (1.17)
Subjective average	2.2 (0.79)

Of the 53 apps assessed, the average user rating across all versions of all apps was 4.4/5, and the apps had been rated a total of 386,024 times (median 6, IQR 0-68). Because there was a large SD of 39,730 ratings, we normalized the rating by n ratings for a weighted average rating of 4.7/5 across all apps, excluding the apps without ratings. Of the 53 apps, 6 had to be purchased for use, and all other apps were free to use. The MARS subscale summary scores, grouped by paid versus free apps, are presented in Table 2. A comparison between the paid and free app scores by MARS subscale is shown in Figure 2. The free apps scored slightly higher on the overall MARS score and on 4 of the 5 subscales, although the differences were not significant.

Table 2. Mobile App Rating Scale subscale summary scores for app cost and duration since last update (N=53).

Mobile App Rating Scale subscale	Cost			Duration since update		
	Free (n=47), mean (SD)	Paid (n=6), mean (SD)	*P* value	Less than 6 months, mean (SD)	More than 6 months, mean (SD)	*P* value
Engagement	2.9 (0.63)	2.6 (0.61)	.32	3.0 (0.80)	2.7 (0.48)	.16
Functionality	4.0 (0.58)	3.8 (0.58)	.53	4.0 (0.41)	4.0 (0.49)	.95
Esthetics	3.1 (0.70)	3.1 (0.81)	.94	3.3 (0.51)	3.0 (0.35)	.24
Information quality	2.8 (0.80)	2.3 (0.74)	.15	2.9 (1.26)	2.6 (1.02)	.38
Subjective quality	2.2 (0.79)	1.9 (0.86)	.42	2.4 (0.75)	2.1 (0.54)	.32

Figure 2. Comparison between paid and free app scores by Mobile App Rating Scale subscale.



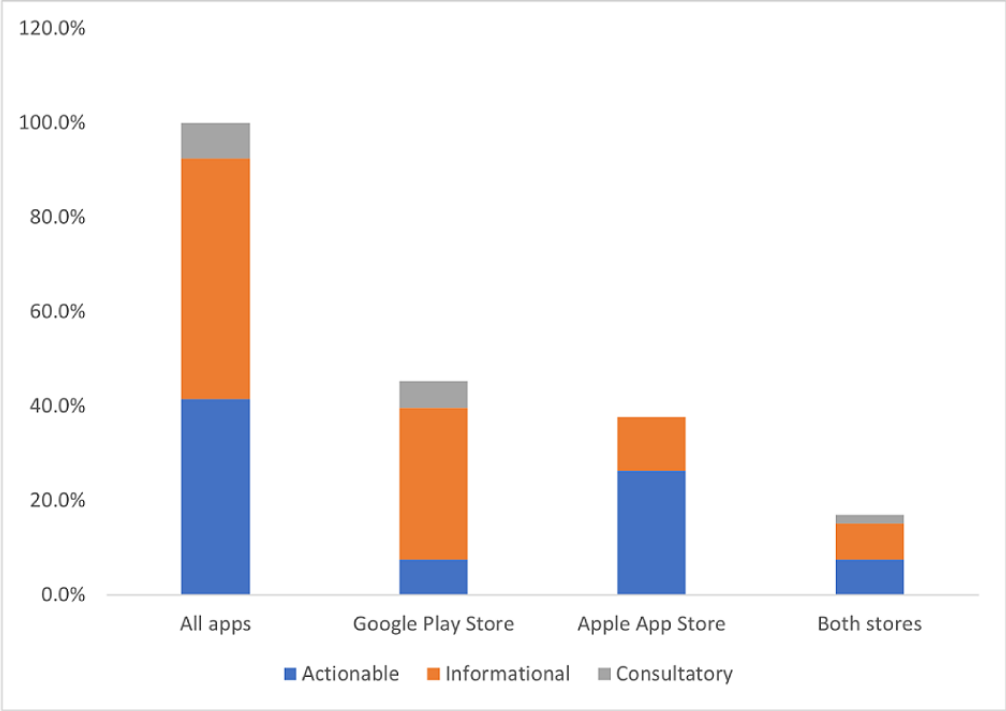
The apps were last updated, on average, on June 19, 2019, with a range of most recent updates between February 9, 2016, and August 31, 2020. Of the 53 apps assessed, 20 (38%) had been updated within 6 months at the time the marketplace searches were conducted. The differences in the app scores when grouped by duration since the last update are shown in Table 2. Those updated within 6 months scored slightly higher overall and on 4 of the 5 MARS subscales, but the scores were not significantly different.

Outside of the MARS assessments, we found that all the assessed apps could be sorted into one of three categories: primarily informational (eg, e-books), actionable (eg, tools), and consultatory (eg, telemedicine). On examining the distribution of these 3 categories between the Apple and Google app marketplaces, we found that the 2 stores had a different division of app types, as shown in Table 3 and Figure 3.

Table 3. Proportion of acute childhood illness apps available by store, grouped by primary functional category (N=53).

	Actionable	Informational	Consultatory
All apps, n (%)	22 (42)	27 (51)	4 (8)
Google Play Store	4 (8)	17 (32)	3 (6)
Apple App Store	14 (26)	6 (11)	0 (0)
Both app stores	4 (8)	4 (8)	1 (2)
Difference, *P* value	.002	.03	.08

Figure 3. Types of apps available in each marketplace.



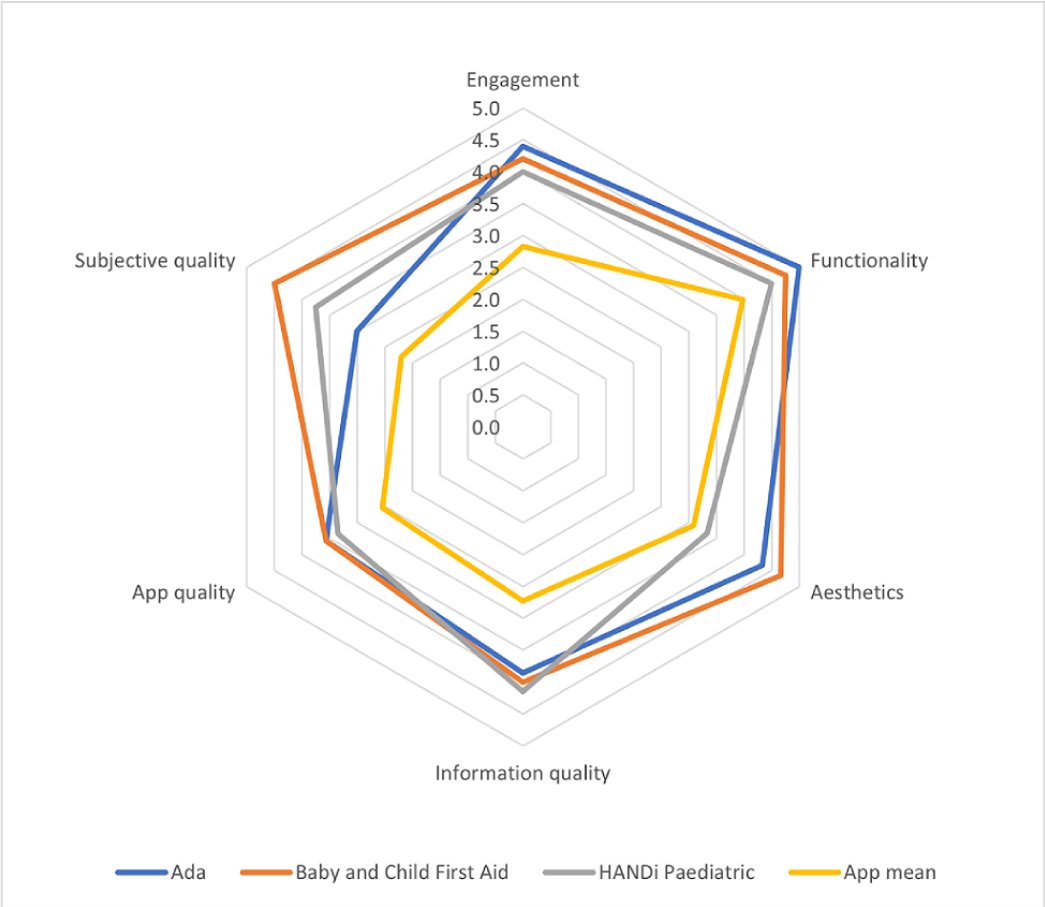
Of the 53 apps assessed, the top 3 on the MARS scale were Baby and Child First Aid by the British Red Cross (4.5/5), Ada by Ada Health (4.5/5), and HANDi Paediatric by Musgrove Park Hospital (4.2/5). Baby and Child First Aid provides common and emergency first aid scenarios and advice to parents. HANDi Paediatric has a similar focus on children's health, but it provides illness-related information rather than first aid advice. Ada is an artificial intelligence–enabled symptom checker that asks a series of diagnostic questions to users to suggest illnesses, and it can be used for third parties (eg, children). A visual representation of all MARS items is presented in Figure 4, which shows the average rating of all apps (in blue), and a comparison of the 3 apps that had the top ratings on each subdimension of the MARS. The top 3 apps were also examined by MARS subscale scores (Table 4; Figure 5).

Figure 4. Mobile App Rating Scale app radar for top three apps versus the mean score of all 53 apps.

Table 4. Top apps by Mobile App Rating Scale subscale score.

Mobile App Rating Scale subscale	App mean (SD)	Subscale score			Top 3 mean versus app mean, *P* value
		Baby and Child First Aid	Ada	HANDi Paediatric	
Engagement	2.8 (0.86)	4.2	4.4	4	<.001
Functionality	4 (0.72)	4.8	5	4.5	.01
Esthetics	3.1 (0.87)	4.7	4.3	3.3	.11
Information quality	2.7 (1.32)	4	3.9	4.1	<.001
Subjective quality	2.2 (0.79)	4.5	3	3.8	.06

Figure 5. Radar graph of top apps by Mobile App Rating Scale subscale.



Regarding the information provided within the apps, the source of information was often unlisted or unclear. However, we noted that some apps had matching information: 26% (14/53) of the apps contained content matching text from the Schmitt-Thompson Clinical Content nursing triage guidelines [51] and were similar in appearance and functionality. Predominantly, these apps were owned by individual children's hospitals in the United States (10/14, 71%). The 4 other apps used guidelines developed by the UK-based National Health Service clinical commissioning groups; each was developed for region-specific deployment (eg, Suffolk).

App Marketplaces

With regard to the differences between the app marketplaces, after pooling the results from both marketplaces, we noted that 4 apps were available in Canada but not in the United States, whereas 14 were available in the United States but not in Canada. However, with the exception of Baby and Child First Aid, the apps available in Canada but not in the United States were of low quality (mean MARS score 2.4/5, SD 0.20). Furthermore, of the apps available in the United States but not in Canada, 71% (10/14) were local hospital apps.

Discussion

Principal Findings

The purpose of conducting this environmental scan was to generate a comprehensive picture of the app landscape for parents in North America trying to make the best health decisions for their acutely ill child and for health care researchers to identify gaps in the app ecosystem. Given the observed trend of parents not seeking health care services for their children during the COVID-19 pandemic [7], it is important to address the gap we observed in Canadian context–specific evidence-based information to support parental health care decisions for their children. With the large changes in pediatric ED use during the COVID-19 pandemic [7], there is a need to create information access that identifies when interventions are necessary, rather than the current trend of parents delaying treatment of their children's less serious illnesses.

Our scan of the Apple App Store and Google Play Store in Canada and the United States identified 53 apps that we appraised using the validated MARS tool. The mean MARS score was 3.2/5 (SD 0.60), scoring slightly higher than the scale's *average* category of 3/5. The apps had a high overall functionality score, with especially strong gestural design (ie, whether interactions such as taps and other screen-based gestures in the app were consistent and intuitive when examining all app pages). Gestures worked in the apps with few errors. The esthetics score rated slightly below the app mean score, primarily because of a low visual appeal score, but it was bolstered by strong app layouts.

The app design choices were often outdated, an unsurprising finding considering the long average update time of 453 days. If only those apps that were updated within 6 months are

considered, the visual appeal score of the 20 included apps is slightly higher (although not significantly so). The low average information quality of apps was caused primarily by a low evidence base score (ie, proof that the app has been trialed, as verified by evidence in published scientific studies). This seems to diverge from mHealth app studies in other health disciplines such as pharmacy, which found that information quality had the highest scores among the subscales [52]. When we examined the literature for occurrences of app names, we found mention of only 4 apps: Ada, Kinsa (a digital thermometer companion), HANDi Paediatric, and WebMD Baby. Information quality was buoyed by a strong description accuracy (mean 4.3, SD 0.91; ie, whether the app's described content matched its actual content) and information quality (mean 3.6, SD 0.95). The average scores in subjective quality app ratings (mean 3.3, SD 1.16) were similar to the overall assessed scores, whereas willingness to pay for an app (mean 1.2, SD 0.71) and projected app use (mean 1.5, SD 0.54) were low. However, these low scores may better reflect the fact that most of the apps were designed primarily for ad hoc use during acute illnesses, rather than for regular use in chronic health situations.

We found that the apps included in each store seemed to have different foci. On closer examination, the Google marketplace apps were noted to be primarily information-focused (eg, e-books), with 4 times as many (17/53, 32%) information-focused apps as action-oriented apps (eg, symptom-tracking tools; 4/53, 8%). The Apple marketplace apps had the opposite trend: there were more than twice as many action-oriented apps (14/53, 26%) as information-focused apps (6/53, 11%). This finding suggests that a parent's choice of mobile device brand may not be incidental: beyond this choice leading to different information availability, there is a difference in the format of information presented. Specifically, unless parents have access to both Android and Apple devices, they will only have access to apps available on one app store. Depending on the device available, these differences may affect how effectively parents are able to find and use an app to help their child. For example, the Apple and Google app stores differ in how apps are assessed for quality and how their search algorithms work, as well as in quality requirements for app developers [53]. As has been previously suggested, the mode in which information is presented matters [54-56], and this difference has the potential to affect how parents learn more about their child's illness. In the same manner, the fact that there are apps that are available in one country but not in another also affects parents' ability to find health information. User characteristics also present a possible barrier to accessing health information: a Dutch study concluded that app use differed based on user demographics (eg, app use was higher in younger, more educated users) [57].

When we examined app pay practices, we found that the paid apps scored slightly lower (although not significantly so, based on two-tailed *t* tests assuming equal variances based on similar domain and overall variances) on 4 of the 5 MARS subscales—engagement (P=.29), functionality (P=.51), information quality (P=.13), and subjective quality (P=.36)—and higher on one subscale: aesthetics (P=.93). Limiting the interpretation of this finding is the small number of paid apps compared with that of free apps (6 vs 47, respectively). We also examined whether the apps that were updated more than 6 months ago differed in quality from the more recently updated apps. Although only small differences were seen among the apps, it will be interesting to investigate in future studies whether, as noted with previous work [58], some apps worsen with time in domains such as privacy. Trust is an increasingly important dimension of apps, given their status as an emerging health technology [59], and it has been shown to be an important consideration in app adoption in some groups such as the Deaf [60]. A limitation of the MARS is that it does not assess app credibility to the depth suggested by other app assessment frameworks such as the mHealth App Trustworthiness checklist [59]. Future research could assess apps using such a checklist to discover how items important to accessibility, such as customization, correlate with app trustworthiness dimensions such as user autonomy and empowerment.

On examining the top 3 apps, we noted that Baby and Child First Aid presents 17 common first aid scenarios (eg, *allergic reaction*) to parents, with a set of causes and steps to follow (eg, basic triage, ideal action to take, and action to take in an emergency). This app also allows parents to self-assess the effectiveness of the teaching videos and other materials available through quizzes. This app is highly relevant, comes from a credible organization, and is updated regularly (at the time of writing, the most recent update was within 6 months, on October 12, 2020). However, this app's credibility could be improved by demonstrating its effectiveness in peer-reviewed studies. Ada is an artificial intelligence–based symptom checker that uses a series of questions to reach a list of possible diagnoses, ranked in order of likelihood. It provides options to check another person's symptoms, track symptoms, and centralize health information (eg, medications and insurance). Ada benefits from a presence in scientific studies [61,62]. HANDi Paediatric provides parents with home assessment guidelines and information about common childhood illnesses. Notably, it includes nearby ED locations and the ability to call emergency services if needed. Taken together, these 3 apps give parents information related to emergency pediatric first aid, identification of (and appropriate response to) common childhood illnesses, 2 ways to check symptoms, and a means of responding to emergency situations. Two of the top apps (Ada and HANDi Paediatric) have a substantial, published evidence base [63-65], and there is a significant difference between their scores and the mean score of all apps for this item (Ada=3/5, HANDi Paediatric=3/5; item mean 0.28/5, SD 0.89; P<.001). A comparison of these top 3 apps with the app average shows that the top 3 apps are significantly different in terms of engagement (P<.001), information quality (P<.001), and functionality (P=.01), with no significant differences for subjective quality and esthetics.

Among all the apps reviewed, there was a lack of app presence in published studies (MARS item evidence base: Has the app been trialed or tested? Must be verified by evidence [in published scientific literature]; mean 0.28/5, SD 0.89). However, information quality and quantity were above average and average (mean scores 3.6 and 3.0, respectively). Therefore, there seems to be a gap between the app's provision of quality content

and the ability to make published claims about the effectiveness of accessing that content. This is an area that has been examined previously: a 2014 study found that there was no crossover between academic and commercial app offerings for pain-related apps [35]. The lack of rigor in assessing apps gave rise to the concern that a lack of peer-reviewed apps has resulted in an unregulated and uncertified set of apps available to consumers and that this could be addressed by scientists leveraging decreasing app development times and costs to offer more evidence-supported apps [66]. As one recent investigation into symptom checkers concluded, apps must be assessed on the accuracy of their results in the context of coverage of available conditions and patient populations if these apps do not give results for some subgroups (eg, children) [67].

In assessing these apps, a key gap for parents to be aware of, and for developers to note, was the lack of Canadian-developed content available to parents. Although apps for children's illnesses were developed using knowledge from US and UK sources, there were no apps that seemed to use Canadian-based information. Among the top 3 apps, Ada develops medical content through an internally peer-reviewed process, using a team of physicians in the United States [68]. HANDi Paediatric bases its information on consultants and general practitioner and hospital clinical guidelines in the United Kingdom [69]. Baby and Child First Aid does not list a source of information. This strongly suggests that although high-quality information for childcare is available in apps, Canadian parents must rely on guidelines developed elsewhere. It also means that Canadian interests, values, and approaches to pediatric illnesses are not well represented in the app marketplaces. This lack of representation is concerning and suggests a gap in availability of an app for parents based on Canadian evidence-based health guidelines.

Limitations

When we scrutinized the app stores, we found that the Google Play Store does not recommend similar apps if the app being examined is not compatible with the user's devices, and the Apple App Store does not show apps that are unavailable on the user's current device. Spellings and idioms were important: searching *pediatric* returns apps from North America first, whereas *paediatric* returns UK- and Australia-based apps. The mechanisms behind each store's search parameters are unique and undisclosed, and the differences in search mechanisms that are described do not support the use of a single systematic process used to search both stores. Therefore, there is no way to guarantee that all relevant and currently available apps have been captured during the Play Store and App Store searches. We attempted to mitigate this limitation through our use of multiple search terms co-designed with a research librarian and having 2 reviewers assess all apps for inclusion. In addition, the testing of each app was conducted by 1 evaluator (with a second evaluator assessing 5/53, 10% of apps); having more evaluators comparing ratings would decrease the possibility that the rating agreements occurred by chance. Our examination of peer-reviewed study availability relating to single apps was limited to a single search engine (Google Scholar), and rigor could have been added to this method by triangulating results with PubMed and directly checking developer websites.

Conclusions

This study examined 53 apps for parents related to acute childhood illness using the MARS assessment tool. Overall app quality was rated slightly above average, driven by high scores in functionality, whereas there was scope for the apps to improve their information content through increased presence in peer-reviewed studies. There was a strong need identified for evidence-based apps of Canadian origin. These apps should be developed in a way that ensures accessibility of content, transparency of information sources, and effectiveness as demonstrated through peer-reviewed studies. In the case that developers are unable to test for effectiveness, demonstrating that the app was created using evidence-based, peer-reviewed information would give parents a basis for trusting the app's content. Considering these findings in relation to the current practice of most parents of using internet searches as their clinician-adjunct source of health information, there is a clear opportunity for clinicians and scientists to collaborate to create apps that increase parents' confidence in their health knowledge.

Acknowledgments

The authors thank Ms Liz Dennett at the University of Alberta's JW Scott Health Sciences Library for her help in designing a search strategy and examining the characteristics of app marketplaces. The authors also thank the Stollery Children's Hospital Foundation, which supported this study through the Stollery Science Lab. LH is supported by a Canada Research Chair in Knowledge Synthesis and Translation. SS is supported by a Canada Research Chair in Knowledge Translation in Child Health. LH and SS are distinguished researchers with the Stollery Science Lab supported by the Stollery Children's Hospital Foundation. JB is supported by a Women and Children's Health Research Institute Postdoctoral Fellowship.

Conflicts of Interest

None declared.

Multimedia Appendix 1
App quality ratings by Mobile App Rating Scale domain.
[DOC File , 185 KB - jmir_v23i10e29441_app1.doc]

References

1. Nguyen HC, Nguyen MH, Do BN, Tran CQ, Nguyen TT, Pham KM, et al. People with suspected Covid-19 symptoms were more likely depressed and had lower health-related quality of life: The potential benefit of health literacy. J Clin Med 2020 Mar 31;9(4):965 [FREE Full text] [doi: 10.3390/jcm9040965] [Medline: 32244415]
2. Zhang Y, Ma ZF. Impact of the Covid-19 pandemic on mental health and quality of life among local residents in Liaoning Province, China: A cross-sectional study. Int J Environ Res Public Health 2020 Mar 31;17(7):2381 [FREE Full text] [doi: 10.3390/ijerph17072381] [Medline: 32244498]
3. Chanchlani N, Buchanan F, Gill PJ. Addressing the indirect effects of COVID-19 on the health of children and young people. Can Med Asso J 2020 Aug 10;192(32):921-927 [FREE Full text] [doi: 10.1503/cmaj.201008] [Medline: 32586838]
4. Dann L, Fitzsimons J, Gorman KM, Hourihane J, Okafor I. Disappearing act: COVID-19 and paediatric emergency department attendances. Arch Dis Child 2020 Aug;105(8):810-811 [FREE Full text] [doi: 10.1136/archdischild-2020-319654] [Medline: 32518141]
5. Isba R, Edge R, Jenner R, Broughton E, Francis N, Butler J. Where have all the children gone? Decreases in paediatric emergency department attendances at the start of the COVID-19 pandemic of 2020. Arch Dis Child 2020 Jul;105(7):704-770. [doi: 10.1136/archdischild-2020-319385] [Medline: 32376695]
6. Lazzerini M, Barbi E, Apicella A, Marchetti F, Cardinale F, Trobia G. Delayed access or provision of care in Italy resulting from fear of COVID-19. Lancet Child Adolesc Health 2020 May;4(5):10-11 [FREE Full text] [doi: 10.1016/S2352-4642(20)30108-5] [Medline: 32278365]
7. Goldman RD, Grafstein E, Barclay N, Irvine MA, Portales-Casamar E. Paediatric patients seen in 18 emergency departments during the COVID-19 pandemic. Emerg Med J 2020 Dec;37(12):773-777 [FREE Full text] [doi: 10.1136/emermed-2020-210273] [Medline: 33127743]
8. Lim E, Mistry RD, Battersby A, Dockerty K, Koshy A, Chopra MN, et al. "How to Recognize if Your Child Is Seriously Ill" During COVID-19 Lockdown: An evaluation of parents' confidence and health-seeking behaviors. Front Pediatr 2020 Nov 17;8:580323 [FREE Full text] [doi: 10.3389/fped.2020.580323] [Medline: 33313025]
9. Davis DW, Jones VF, Logsdon MC, Ryan L, Wilkerson-McMahon M. Health promotion in pediatric primary care: importance of health literacy and communication practices. Clin Pediatr (Phila) 2013 Dec;52(12):1127-1134. [doi: 10.1177/0009922813506607] [Medline: 24146229]
10. Scott SD, Albrecht L, Given LM, Hartling L, Johnson DW, Jabbour M, et al. Pediatric information seeking behaviour, information needs, and information preferences of health care professionals in general emergency departments: Results from the Translating Emergency Knowledge for Kids (TREKK) Needs Assessment. Can J Emerg Med 2018 Jan;20(1):89-99. [doi: 10.1017/cem.2016.406] [Medline: 28067181]
11. DeWalt DA, Hink A. Health literacy and child health outcomes: a systematic review of the literature. Pediatrics 2009 Nov;124 Suppl 3:265-274. [doi: 10.1542/peds.2009-1162B] [Medline: 19861480]
12. Annual Report 2018-2019. Alberta Health Services. URL: https://www.albertahealthservices.ca/assets/about/publications/2018-19-annual-report-web-version.pdf [accessed 2021-09-09]
13. Yardi S, Caldwell PH, Barnes EH, Scott KM. How do parents search for information on their child's medical condition? J Paediatr Child Health 2017 May 08;53(Suppl 3):26-27. [doi: 10.1111/jpc.13597_13]
14. Benedicta B, Caldwell PH, Scott KM. How parents use, search for and appraise online health information on their child's medical condition: A pilot study. J Paediatr Child Health 2020 Feb 31;56(2):252-258. [doi: 10.1111/jpc.14575] [Medline: 31365171]
15. Treadgold BM, Teasdale E, Muller I, Roberts A, Coulson N, Santer M. Parents and carers' experiences of seeking health information and support online for long-term physical childhood conditions: a systematic review and thematic synthesis of qualitative research. BMJ Open 2020 Dec 28;10(12):e042139 [FREE Full text] [doi: 10.1136/bmjopen-2020-042139] [Medline: 33372077]
16. Amram O, Schuurman N, Pike I, Friger M, Yanchar NL. Assessing access to paediatric trauma centres in Canada, and the impact of the golden hour on length of stay at the hospital: an observational study. BMJ Open 2016 Jan 08;6(1):e010274 [FREE Full text] [doi: 10.1136/bmjopen-2015-010274] [Medline: 26747041]
17. Franklin M. Alberta Health Link switchboard swamped with COVID-19 calls. CTV News. URL: https://calgary.ctvnews.ca/alberta-health-link-switchboard-swamped-with-covid-19-calls-1.4847260 [accessed 2021-09-09]
18. De S, Tong A, Isaacs D, Craig JC. Parental perspectives on evaluation and management of fever in young infants: an interview study. Arch Dis Child 2014 Aug;99(8):717-723. [doi: 10.1136/archdischild-2013-305736] [Medline: 24849214]
19. Thompson AP, Nesari M, Hartling L, Scott SD. Parents' experiences and information needs related to childhood fever: A systematic review. Patient Educ Couns 2020 Apr;103(4):750-763. [doi: 10.1016/j.pec.2019.10.004] [Medline: 31668490]
20. Morrison AK, Myrvik MP, Brousseau DC, Hoffmann RG, Stanley RM. The relationship between parent health literacy and pediatric emergency department utilization: a systematic review. Acad Pediatr 2013;13(5):421-429 [FREE Full text] [doi: 10.1016/j.acap.2013.03.001] [Medline: 23680294]
21. NACRS emergency department visits and length of stay, 2018-2019. Canadian Institute for Health Information. URL: https://www.cihi.ca/en/nacrs-emergency-department-visits-and-length-of-stay-2018-2019-xlxs [accessed 2021-09-09]

22. Morrison A, Schapira M, Gorelick M, Hoffmann R, Brousseau D. Low caregiver health literacy is associated with higher pediatric emergency department use and nonurgent visits. Acad Pediatr 2014;14(3):309-314 [FREE Full text] [doi: 10.1016/j.acap.2014.01.004] [Medline: 24767784]
23. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014;21(5):427-436 [FREE Full text] [doi: 10.1002/cpp.1855] [Medline: 23918764]
24. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: The potential for a new digital divide. J Med Internet Res 2016 Oct 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
25. Demographics of mobile device ownership and adoption in the United States. Pew Research Center. URL: https://www.qure4u.com/demographics-of-mobile-device-ownership-and-adoption-in-the-united-states/ [accessed 2021-09-09]
26. Communications Monitoring Report 2018. Canadian Radio-Television and Telecommunications Commission. URL: https://crtc.gc.ca/pubs/cmr2018-en.pdf [accessed 2021-09-09]
27. Sambaraju A, Pantula M, Kuppusamy K. Analyzing user awareness on security in android smartphone devices. In: Proceedings of the Second International Conference on Computer Networks and Communication Technologies. Second International Conference on Computer Networks and Communication Technologies: Springer; 2020 Presented at: Second International Conference on Computer Networks and Communication Technologies; May 23–24, 2019; Coimbatore, India p. 213-221. [doi: 10.1007/978-3-030-37051-0_24]
28. Diffusion of smart devices for health in Canada – Final Report. Canada Health Infoway. 2017. URL: https://www.infoway-inforoute.ca/en/component/edocman/resources/reports/benefits-evaluation/3366-the-diffusion-of-smart-devices-for-health-in-canada-study-final-report [accessed 2021-09-09]
29. Ernsting C, Dombrowski SU, Oedekoven M, Kanzler M, Kuhlmey A, Gellert P. Using smartphones and health apps to change and manage health behaviors: A population-based survey. J Med Internet Res 2017 Apr 05;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]
30. Pehora C, Gajaria N, Stoute M, Fracassa S, Serebale-O'Sullivan R, Matava CT. Are Parents Getting it Right? A survey of parents' internet use for children's health care information. Interact J Med Res 2015 Jun 22;4(2):e12 [FREE Full text] [doi: 10.2196/ijmr.3790] [Medline: 26099207]
31. Azad-Khaneghah P, Neubauer N, Cruz A, Liu L. Mobile health app usability and quality rating scales: a systematic review. Disabil Rehabil Assist Technol 2021 Oct 08;16(7):712-721. [doi: 10.1080/17483107.2019.1701103] [Medline: 31910687]
32. Sarkar U, Gourley GI, Lyles CR, Tieu L, Clarity C, Newmark L, et al. Usability of commercially available mobile applications for diverse patients. J Gen Intern Med 2016 Dec;31(12):1417-1426. [doi: 10.1007/s11606-016-3771-6] [Medline: 27418347]
33. Miller S. The Age of Hubris. Br J Gen Pract 2020 Apr;70(693):194 [FREE Full text] [doi: 10.3399/bjgp20X709217] [Medline: 32217597]
34. Donovan E, Wilcox CR, Patel S, Hay AD, Little P, Willcox ML. Digital interventions for parents of acutely ill children and their treatment-seeking behaviour: a systematic review. Br J Gen Pract 2020 Mar;70(692):172-178 [FREE Full text] [doi: 10.3399/bjgp20X708209] [Medline: 32041764]
35. de la Vega R, Miró J. mHealth: a strategic field without a solid scientific soul. a systematic review of pain-related apps. PLoS One 2014;9(7):e101312 [FREE Full text] [doi: 10.1371/journal.pone.0101312] [Medline: 24999983]
36. Fougerouse P, Yasini M, Marchand G, Aalami OO. A cross-sectional study of prominent us mobile health applications: Evaluating the current landscape. AMIA Annu Symp Proc 2018;2017:715-723 [FREE Full text] [Medline: 29854137]
37. The growing value of digital health: Evidence and impact on human health and the healthcare system. IQVIA Institute. 2017. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/the-growing-value-of-digital-health [accessed 2021-09-09]
38. Graham PT, Evitts T, Thomas-MacLean R. Environmental scans: How useful are they for primary care research? Can Fam Physician 2008 Jul;54(7):1022-1023 [FREE Full text] [Medline: 18625830]
39. Rowel R, Moore ND, Nowrojee S, Memiah P, Bronner Y. The utility of the environmental scan for public health practice: Lessons from an urban program to increase cancer screening. J Natl Med Assoc 2005 Apr;97(4):527-534. [Medline: 15868772]
40. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: A new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
41. Boudreaux ED, Waring ME, Hayes RB, Sadasivam RS, Mullen S, Pagoto S. Evaluating and selecting mobile health apps: Strategies for healthcare providers and healthcare organizations. Transl Behav Med 2014 Dec 24;4(4):363-371 [FREE Full text] [doi: 10.1007/s13142-014-0293-9] [Medline: 25584085]
42. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the quality of mobile apps for the management of pain: Systematic search and evaluation using the mobile app rating scale. JMIR Mhealth Uhealth 2018 Oct 25;6(10):e10718 [FREE Full text] [doi: 10.2196/10718] [Medline: 30361196]

43. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
44. Ramsey RR, Caromody JK, Voorhees SE, Warning A, Cushing CC, Guilbert TW, et al. A systematic evaluation of asthma management apps examining behavior change techniques. J Allergy Clin Immunol Pract 2019 Nov;7(8):2583-2591 [FREE Full text] [doi: 10.1016/j.jaip.2019.03.041] [Medline: 30954644]
45. Hallgren KA. Computing inter-rater reliability for observational data: An overview and tutorial. Tutor Quant Methods Psychol 2012;8(1):23-34 [FREE Full text] [doi: 10.20982/tqmp.08.1.p023] [Medline: 22833776]
46. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159. [doi: 10.2307/2529310]
47. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: Measures of agreement. Perspect Clin Res 2017;8(4):187-191 [FREE Full text] [doi: 10.4103/picr.PICR_123_17] [Medline: 29109937]
48. Sullivan GM, Artino AR. Analyzing and interpreting data from likert-type scales. J Grad Med Educ 2013 Dec;5(4):541-542 [FREE Full text] [doi: 10.4300/JGME-5-4-18] [Medline: 24454995]
49. Norman G. Likert scales, levels of measurement and the "laws" of statistics. Adv Health Sci Educ Theory Pract 2010 Dec 10;15(5):625-632. [doi: 10.1007/s10459-010-9222-y] [Medline: 20146096]
50. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
51. Nurse triage guidelines. Schmitt-Thompson Clinical Content. URL: https://www.stcc-triage.com/ [accessed 2021-09-09]
52. Kim BY, Sharafoddini A, Tran N, Wen EY, Lee J. Consumer mobile apps for potential drug-drug interaction check: Systematic review and content analysis using the Mobile App Rating Scale (MARS). JMIR Mhealth Uhealth 2018 Mar 28;6(3):e74 [FREE Full text] [doi: 10.2196/mhealth.8613] [Medline: 29592848]
53. Grundy QH, Wang Z, Bero LA. Challenges in assessing mobile health app quality: A systematic review of prevalent and innovative methods. Am J Prev Med 2016 Dec;51(6):1051-1059. [doi: 10.1016/j.amepre.2016.07.009] [Medline: 27659122]
54. Roter DL. Oral literacy demand of health care communication: challenges and solutions. Nurs Outlook 2011;59(2):79-84. [doi: 10.1016/j.outlook.2010.11.005] [Medline: 21402203]
55. Bernhardt JM, Felter EM. Online pediatric information seeking among mothers of young children: Results from a qualitative study using focus groups. J Med Internet Res 2004 Mar 1;6(1):e7 [FREE Full text] [doi: 10.2196/jmir.6.1.e7] [Medline: 15111273]
56. Baumann I, Jaks R, Robin D, Juvalta S, Dratva J. Parents' health information seeking behaviour - does the child's health status play a role? BMC Fam Pract 2020 Dec 10;21(1):266 [FREE Full text] [doi: 10.1186/s12875-020-01342-3] [Medline: 33302881]
57. Bol N, Helberger N, Weert JC. Differences in mobile health app use: A source of new digital inequalities? The Information Society 2018 Apr 26;34(3):183-193. [doi: 10.1080/01972243.2018.1438550]
58. Taylor VF, Martinovic I. To update or not to update: Insights from a two-year study of android app evolution. In: Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security. 2017 Presented at: ACM on Asia Conference on Computer and Communications Security; April 2 - 6, 2017; Abu Dhabi p. 45-57 URL: https://doi.org/10.1145/3052973.3052990 [doi: 10.1145/3052973.3052990]
59. van Haasteren A, Gille F, Fadda M, Vayena E. Development of the mHealth app trustworthiness checklist. Digit Health 2019 Nov 21;5:2055207619886463 [FREE Full text] [doi: 10.1177/2055207619886463] [Medline: 31803490]
60. Jacob SA, Chong EY, Goh SL, Palanisamy UD. Design suggestions for an mHealth app to facilitate communication between pharmacists and the Deaf: perspective of the Deaf community (HEARD Project). Mhealth 2021;7:29 [FREE Full text] [doi: 10.21037/mhealth.2020.01.04] [Medline: 33898598]
61. Jungmann SM, Klan T, Kuhn S, Jungmann F. Accuracy of a Chatbot (Ada) in the diagnosis of mental disorders: Comparative case study with lay and expert users. JMIR Form Res 2019 Oct 29;3(4):e13863. [doi: 10.2196/13863] [Medline: 31663858]
62. Holdener M, Gut A, Angerer A. Applicability of the user engagement scale to mobile health: A survey-based quantitative study. JMIR Mhealth Uhealth 2020 Jan 03;8(1):e13244 [FREE Full text] [doi: 10.2196/13244] [Medline: 31899454]
63. Ronicke S, Hirsch MC, Türk E, Larionov K, Tientcheu D, Wagner AD. Can a decision support system accelerate rare disease diagnosis? Evaluating the potential impact of Ada DX in a retrospective study. Orphanet J Rare Dis 2019 Dec 21;14(1):69 [FREE Full text] [doi: 10.1186/s13023-019-1040-6] [Medline: 30898118]
64. Miller S, Gilbert S, Virani V, Wicks P. Patients' utilization and perception of an artificial intelligence-based symptom assessment and advice technology in a British Primary Care waiting room: Exploratory pilot study. JMIR Hum Factors 2020 Jul 10;7(3):e19713 [FREE Full text] [doi: 10.2196/19713] [Medline: 32540836]
65. Whiting A, Bridges S. G105 Handi paediatric: smartphone application providing integrated-care pathways for common childhood conditions. Arch Dis Child 2019;104:A42. [doi: 10.1136/archdischild-2019-rcpch.101]
66. Shing S, Yuan B. Apps developed by academics. J Edu Pract 2016;7(33):1-9 [FREE Full text]

67. Gilbert S, Mehl A, Baluch A, Cawley C, Challiner J, Fraser H, et al. How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs. BMJ Open 2020 Dec 16;10(12):e040269 [FREE Full text] [doi: 10.1136/bmjopen-2020-040269] [Medline: 33328258]
68. Our medical content process. Ada. URL: https://ada.com/medical-content-creation-process/ [accessed 2021-02-17]
69. HANDi Paediatric app: advice for common childhood conditions. NHS Devon. URL: https://devonccg.nhs.uk/health-services/children-and-young-people/handi-app [accessed 2021-09-08]

Abbreviations

ED: emergency department
MARS: Mobile App Rating Scale
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by R Kukafka; submitted 07.04.21; peer-reviewed by S Gilbert, J Avis, C Lokker, I Pike; comments to author 31.05.21; revised version received 19.07.21; accepted 12.08.21; published 19.10.21.

Please cite as:
Benoit J, Hartling L, Chan M, Scott S
Characteristics of Acute Childhood Illness Apps for Parents: Environmental Scan
J Med Internet Res 2021;23(10):e29441
URL: https://www.jmir.org/2021/10/e29441
doi:10.2196/29441
PMID:34665144

Original Paper

Effectiveness of Smartphone-Based Community Case Management on the Urgent Referral, Reconsultation, and Hospitalization of Children Aged Under 5 Years in Malawi: Cluster-Randomized, Stepped-Wedge Trial

Griphin Baxter Chirambo[1,2], BSc, MSc, PhD; Matthew Thompson[3], BSc, MSc, PhD; Victoria Hardy[3], BSc, MSc; Nicole Ide[3], BA, MA; Phillip H Hwang[3], BSc, MSc, PhD; Kanika Dharmayat[4], BSc, MPH; Nikolaos Mastellos[4], BSc, MSc, PhD; Ciara Heavin[5], BSc, MSc, PhD; Yvonne O'Connor[5], BSc, MSc, PhD; Adamson S Muula[6], BSc, MSc, MPhil, PhD; Bo Andersson[7], BSc, MSc, DPhil, PhD; Sven Carlsson[7], BSc, MSc, PhD; Tammy Tran[4], BSc, MSc; Jenny Chen-Ling Hsieh[8], BSc, MSc; Hsin-Yi Lee[8], BA, MA; Annette Fitzpatrick[3], BSc, MSc, PhD; Tsung-Shu Joseph Wu[8], BSc, MSc, PhD; John O'Donoghue[2,4], BSc, MSc, PhD

[1]Faculty of Health Sciences, Mzuzu University, Mzuzu, Malawi

[2]Malawi eHealth Research Center, University College Cork, Cork, Ireland

[3]University of Washington, Seatle, WA, United States

[4]Imperial College London, London, United Kingdom

[5]University College Cork, Cork, Ireland

[6]College of Medicine, University of Malawi, Blantyre, Malawi

[7]Lund University, Lund, Sweden

[8]Luke International Norway, Mzuzu, Malawi

Corresponding Author:
Griphin Baxter Chirambo, BSc, MSc, PhD
Faculty of Health Sciences
Mzuzu University
Private Bag 201, Luwinga
Mzuzu 2
Mzuzu
Malawi
Phone: 265 0999236092
Email: gbchirambo@yahoo.co.uk

Abstract

Background: Integrated community case management (CCM) has led to reductions in child mortality in Malawi resulting from illnesses such as malaria, pneumonia, and diarrhea. However, adherence to CCM guidelines is often poor, potentially leading to inappropriate clinical decisions and poor outcomes. We determined the impact of an e-CCM app on the referral, reconsultation, and hospitalization rates of children presenting to village clinics in Malawi.

Objective: We determined the impact of an electronic version of a smartphone-based CCM (e-CCM) app on the referral, reconsultation, and hospitalization rates of children presenting to village clinics in Malawi.

Methods: We used a stepped-wedge, cluster-randomized trial to compare paper-based CCM (control) with and without the use of an e-CCM app on smartphones from November 2016 to February 2017. A total of 102 village clinics from 2 districts in northern Malawi were assigned to 1 of 6 clusters, which were randomized on the sequencing of the crossover from the control phase to the intervention phase as well as the duration of exposure in each phase. Children aged ≥2 months to <5 years who presented with acute illness were enrolled consecutively by health surveillance assistants. The primary outcome of urgent referrals to higher-level facilities was evaluated by using multilevel mixed effects models. A logistic regression model with the random effects of the cluster and the fixed effects for each step was fitted. The adjustment for potential confounders included baseline factors, such as patient age, sex, and the geographical location of the village clinics. Calendar time was adjusted for in the analysis.

Results: A total of 6965 children were recruited—49.11% (3421/6965) in the control phase and 50.88% (3544/6965) in the intervention phase. After adjusting for calendar time, children in the intervention phase were more likely to be urgently referred to a higher-level health facility than children in the control phase (odds ratio [OR] 2.02, 95% CI 1.27-3.23; *P*=.003). Overall, children in the intervention arm had lower odds of attending a repeat health surveillance assistant consultation (OR 0.45, 95% CI 0.34-0.59; *P*<.001) or being admitted to a hospital (OR 0.75, 95% CI 0.62-0.90; *P*=.002), but after adjusting for time, these differences were not significant (*P*=.07 for consultation; *P*=.30 for hospital admission).

Conclusions: The addition of e-CCM decision support by using smartphones led to a greater proportion of children being referred to higher-level facilities, with no apparent increase in hospital admissions or repeat consultations in village clinics. Our findings provide support for the implementation of e-CCM tools in Malawi and other low- and middle-income countries with a need for ongoing assessments of effectiveness and integration with national digital health strategies.

Trial Registration: ClinicalTrials.gov NCT02763345; https://clinicaltrials.gov/ct2/show/NCT02763345

(J Med Internet Res 2021;23(10):e25777) doi:10.2196/25777

KEYWORDS

community case management; mobile health; pediatrics; childhood infection; mobile phone

Introduction

Malawi has one of the world's lowest doctor-to-patient ratios [1], with ≤1 doctor per 50,000 people [2-5], which is considerably lower than the World Health Organization's recommendation of 1 doctor per 5000 people. In order to improve access to essential health services for many of Malawi's marginalized communities, the Malawi Ministry of Health started an Emergency Human Resources Program in 2005 when it recruited and trained selected community members to become health surveillance assistants (HSAs) [6]. HSAs are allocated to hard-to-reach areas, each serving a population of approximately 1000 [7]. They have several duties they are expected to carry out (eg, family planning, environmental health, and HIV counseling), of which community case management (CCM) is integral [8]. HSAs are encumbered by an overwhelming workload, insufficient day-to-day professional support (including training opportunities), and inadequate remuneration [7]. These competing pressures may lead to further progress in disease control and childhood survival. CCM is a paper-based clinical decision aid that enables frontline community health workers known as HSAs in Malawi to recognize and manage the major causes of mortality among children aged ≤5 years in low- and middle-income countries (LMICs), specifically malaria, pneumonia, diarrhea, and malnutrition [9]. CCM uses simple signs and symptoms to help HSAs identify self-limiting conditions that can be managed in the community and more serious illnesses that require medical attention from clinically trained health personnel (eg, nurses) at higher-level health care facilities or hospitals. The introduction of this strategy has provided access to basic health care that many children would otherwise have been denied [10].

While mortality rates among the under-fives have steadily declined since the introduction of CCM in 2009, concerns have begun to surface about inconsistent standards of delivery of CCM; it has been observed that HSAs do not always follow all the steps in the protocol when assessing children, potentially resulting in inappropriate clinical management decisions [11]. In Malawi, HSAs deliver CCM from village clinics; in a country where 84% of the population resides in rural locations [12] more than 8 km from facility-based health providers, CCM has helped provide equitable access to health care for many communities that were previously underserved [13]. CCM is widely implemented in LMICs and has demonstrated some success in improving the management of children with acute illness and reducing mortality. For example, the CCM of pneumonia reduced pneumonia-associated mortality by 70% in some developing countries [14,15]. As the adoption of CCM in Malawi in 2008, this strategy has contributed to significant reductions in under-five mortality from 97.8 per 1000 live births to 52.7 in 2017 [16].

The implementation of CCM is based on the diagnostic and prognostic value of certain signs and symptoms; children with uncomplicated illness are managed in the community with symptomatic treatment, whereas those with *danger* signs are urgently referred to a health facility that can provide a higher level of care. Accurate and timely urgent referral is an integral component of CCM and is a prerequisite for improving health outcomes among acutely unwell children. Aside from preventable mortality, failure to refer at the appropriate time can overburden constrained resources at higher-level facilities needed for treatment of severely unwell children, lead to unnecessary hospital admissions, or involve caregivers repeating inconvenient journeys to village clinics because their child's condition has not resolved or is deteriorating.

However, there is considerable evidence of suboptimal management decisions from field studies, including poor HSA adherence to the required steps of CCM [17]. HSAs do not always follow all the steps in the protocol when assessing children, potentially resulting in inappropriate clinical management decisions [11]. Indeed, because CCM is delivered using a paper-based tool, it is easy for relevant fields to be overlooked or for clinically invalid data to be recorded, contributing to inappropriate management decisions. Poor adherence to CCM is further compounded by factors such as transportation costs or distance to higher-level facilities or medication supply chain issues that further impede the potential impact of CCM.

The proliferation of mobile phones and cellular networks in LMICs has prompted the development of several mobile health (mHealth) technologies as alternative platforms for delivering

childhood intervention strategies [18]. Decision support systems within these technologies can facilitate clinical decision-making and offer potential advantages over paper-based tools [19,20]. The potential advantages of e-CCM over paper-based CCM include shorter consultation time and a greater likelihood of correct diagnostic and treatment decisions, as noted in previous studies [21,22]. However, adopting mHealth tools as an adjunct (or replacement) for current paper-based assessment and reporting tools requires considerable initial and sustained investment. Therefore, demonstrating the impact of mHealth on health care outcomes is crucial. This study aims to determine the effectiveness of smartphone-based CCM on urgent referral, reconsultation, and hospitalization of children aged ≤5 years in Malawi.

Methods

Study Design

The full study protocol has been published [23]. In summary, we used a stepped-wedge, cluster-randomized design trial to compare paper-based CCM with a mobile phone app version of the CCM that was developed by the Supporting LIFE research team (SL e-CCM app) [24]. Village clinics were grouped into 6 clusters based on geographic proximity, and clusters were randomized to determine the sequence of crossover from the control (using paper CCM alone) to the intervention (using paper CCM as well as SL e-CCM) and the duration of exposure (2-7 weeks) in each phase. Children aged ≥2 months to <5 years were triaged using paper CCM or paper CCM+SL e-CCM app depending on when they presented to village clinics.

This study was conducted in 102 village clinics across Nkhata Bay and Rumphi districts in northern Malawi, where Chichewa, Chitumbuka, and Tonga are the principal languages spoken. Cassava farming is the main occupation of the 215,429 inhabitants of the Nkhata Bay District, situated on Lake Malawi. Rumphi district, which extends west from Lake Malawi to the Zambian border, is principally a tobacco farming community with 166,460 inhabitants. According to the National Statistic Office of Malawi, 23.4% and 17.4% of the adult population (defined as those aged ≥15 years) are illiterate in Nkhata Bay and Rumphi, respectively. Village clinics are typically basic community structures equipped only with rudimentary diagnostic aids (eg, stopwatch) and medical supplies (eg, oral rehydration therapy and artemisin-based combination therapy). Each clinic is operated by a single HSA, who is a government-employed community health worker responsible for assessing and managing acutely unwell children using CCM. HSAs are people who possess a Malawi School Certificate of Education or Junior Certificate of Education. They are frontline health workers who work in the village clinic [7]. A village clinic serves a catchment area of approximately 1000 people [10,25]. Both districts are predominantly rural, with 30%-35% of the population residing more than 8 km from a health facility with qualified clinical staff [26].

Participants

A total of 102 HSAs were recruited to participate (men: 77/102, 75.4%; women: 24/102, 23.5%; not disclosed: 1/102, 0.9%). The ages of these HSAs ranged from 27 to 59 years, and they had been working as HSAs for between 1 and 27 years (mean 10.1 years). Before the trial, 101 HSAs already had phones, of which 36.6% (37/101) were smartphones.

HSAs consecutively enrolled eligible children aged ≥2 months to <5 years from November 2016 to February 2017, who presented with acute illness but were not unconscious, convulsing, or previously enrolled in the study. Because of high rates of illiteracy in Nkhata Bay and Rumphi districts, caregivers were required to verbally consent in their preferred language for their child to be enrolled to avoid limiting participation by literacy levels.

Randomization and Blinding

Village clinics were grouped into 6 clusters based on geographic proximity. Clusters were then randomized using a web-based random number generator to determine the sequencing of crossover of clusters from the control phase (paper CCM) to the intervention phase (paper CCM+SL e-CCM), as well as the duration of exposure (2-7 weeks) in each phase. Neither the participants nor the researchers were blinded to allocation. Baseline data were collected during the index visit and recorded in the village clinic register (VCR) in the control arm and in the SL e-CCM app during the intervention phase.

HSA Training

Before commencing recruitment, we engaged with national and district-level health authorities to explain to them the objectives of the clinical trial and the procedures to be done. HSAs recruited by the study team attended a 1-day training workshop to learn study procedures for the control and intervention phases of the trial, as well as how to operate the SL e-CCM app and the smartphone (details of the hardware and development and functionality of the app are described in full elsewhere [23]). Before crossover from the control to the intervention phase, HSAs attended a further 2-day training workshop to familiarize them with study procedures and the technology specific to the intervention phase. Training workshops were conducted in English with a member of the research team fluent in the regional dialects of Chichewa, Tonga, and Chitumbuka to facilitate communication. HSAs were permitted to use their clinical discretion if they disagreed with the care recommendations provided by the SL e-CCM app.

Procedures

During the control phase, HSAs used paper-based CCM to assess and manage children for 2-7 weeks (depending on the assigned order of clusters after randomization). Details of clinical presentations were recorded by hand in the VCR, as per standard practice. During the intervention phase, HSAs recorded clinical presentations synchronously in both the VCR and SL e-CCM apps. Form validation ensured that HSAs were required to complete every field on the e-CCM app [11]. A complete inventory of the data collected by HSAs during each phase of the trial has been reported elsewhere [23]. Outcome data were collected retrospectively using local nursing students or graduates who traveled to village clinics and higher-level health facilities within an assigned catchment area on a weekly basis to manually abstract data from patient records. Enrolled children were reidentified at clinical sites by cross-referencing each

child's full name, date of enrollment, date of birth, and sex along with caregiver name and contact details prepopulated on clinical research forms. Given that caregivers can represent to any village clinic and higher-level health facility, child identifiers and subsequent attendances at health facilities were verified by caregivers via cell phone beforehand, to improve the efficiency of data collection. Clinical data were recorded in the VCR during both the control and intervention phases of the study in order to fulfill local reporting requirements for village clinics and district health authorities.

Outcomes

The primary outcomes were urgent referrals to higher-level health facilities, reconsultations to village clinics, and hospitalization within 7 days of study enrollment.

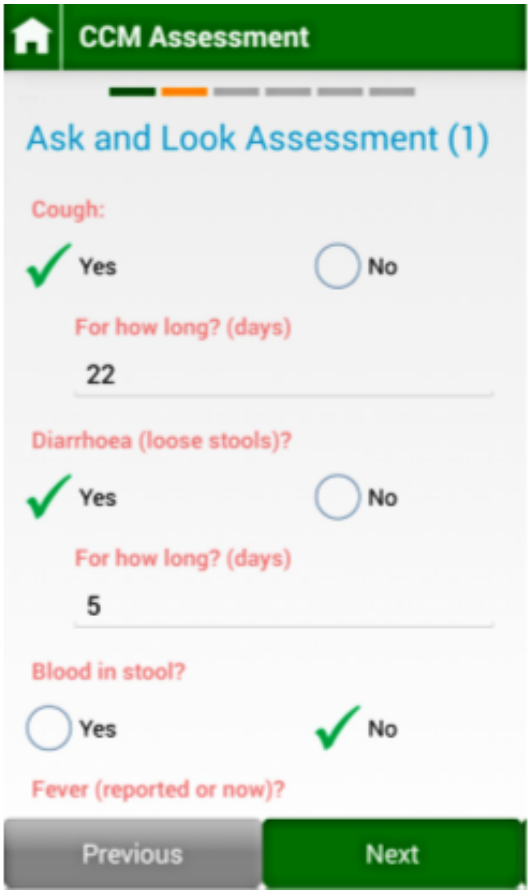
Outcomes were determined by case note review of VCR and health facility records undertaken 2-weeks after study enrollment. Reconsultations were defined as reattendances to any village clinic for a health concern related to the reason for presentation recorded in the VCR at baseline, and hospital admissions related to an inpatient stay in a secondary or tertiary care facility for ≥1 day because of deterioration of illness recorded in the VCR.

Description of SL e-CCM App

The SL e-CCM App was developed and implemented by the Supporting LIFE research team and is an android-based smartphone app that replicates the paper-based CCM decision support tool routinely used by HSAs in Malawi and across sub-Saharan Africa [27]. The tool enables HSAs to enter the information they would usually gather using the paper CCM form, answer a series of clinical questions, and obtain selected clinical measurements (eg, breathing rate). Data are entered directly into the app via touch screen technology, either by selecting the appropriate option or by entering free text [23,28]. The app provides treatment recommendations, including medications, referral to health facilities, and health education for parents. The app requires that users complete all steps (ie, questions and prompts) before the user can proceed to the next step. Figures 1 and 2 show the dashboard of the Supporting LIFE App.

Figure 1. The Supporting LIFE App. CCM: community case management; IMCI: Integrated Management of Childhood Illnesses.

Figure 2. CCM assessment using the Supporting LIFE App. CCM: community case management.



Statistical Analysis

Descriptive analyses were used to examine the individual-level baseline characteristics of children between the control and intervention periods across all clusters. Participant characteristics included the child's age, sex, the district in which the child was recruited, and the treatment decision by the HSA at the index visit. For continuous measures, we reported the mean and SD. For categorical measures, we reported counts and proportions. The 2-tailed *t* tests and chi-square tests were used to determine bivariate differences for continuous and categorical measures, respectively.

Following an intention-to-treat principle, we evaluated overall differences in the proportion of children who were urgently referred to a higher-level health facility by HSAs between the control and intervention periods. For the primary analysis, urgent referrals to higher-level facilities were evaluated at the patient level using multilevel mixed effects models [29]. As all outcomes were binary, a logistic regression model with random effect of cluster and fixed effect for each step was fitted [30,31]. Potential confounders included baseline factors of age, sex, and geographical location of village clinics (ie, urban and rural). Calendar time was also adjusted for in the analysis, and the effect of time was modeled as a categorical variable [30]. Odds ratios (ORs) with 95% CI were estimated and reported. Potential effect modification by age and sex was assessed by stratification and inclusion of interaction terms with the intervention period in the model. Other main outcomes of reconsultations at village clinics and hospitalizations were also analyzed using mixed models with the same potential confounders and effect modifiers. For secondary analyses, we explored the heterogeneity in treatment effects between clusters using within-cluster comparisons between the intervention and control periods. Statistical significance was set at $P<.05$, and all tests were two-sided. All analyses were conducted using Stata version 14 (StataCorp). Patient flow, including the number and reasons for exclusion and withdrawal, was reported according to CONSORT recommendations.

Dissemination of Results

The results of this study will be disseminated through presentations in research conferences, and copies of this study will be put in the Libraries of Mzuzu University, College of Medicine, University of Washington, University College Cork, Lund University, and Imperial College London.

Results

Participants Recruited and Baseline Characteristics

A total of 6965 children were recruited, of whom 49.11% (3421/6965) were included in the paper CCM (control) phase and 50.88% (3544/6965) were included in the e-CCM (intervention) phase (Table 1). A larger number of children were recruited from the Nkhata Bay district than from Rumphi in both trial phases, reflecting the higher population in the former, as noted above. Children recruited in the control phase were significantly younger (mean age 24.6, SD 15.4 months) than those in the intervention phase (26.9, SD 16.1 months). Most children in both the control (3217/3421, 94.03%) and in the

intervention phases (3335/3544, 94.10%) were treated at home after their assessment by the HSA (Table 1).

Table 1. Baseline characteristics of participants (N=6965).

Characteristic	Control arm (paper-CCM[a]; n=3421)	Intervention arm (e-CCM app+paper CCM; n=3544)	*P* value
Age (months), mean (SD)	24.6 (15.4)	26.9 (16.1)	<.001
Sex, n (%)			
Male	1741 (50.89)	1783 (50.31)	.63
Female	1680 (49.10)	1759 (49.63)	.66
Not recorded	0 (0)	2 (0.06)	.16
District[b], n (%)			
Rumphi	1590 (46.47)	1535 (43.31)	.01
Nkhata Bay	1777 (51.94)	1962 (55.36)	<.001
HSA[c] treatment decision at index visit, n (%)			
Urgent referral	179 (5.23)	198 (5.58)	.51
Treated at home or advice given to caregiver	3217 (94.03)	3335 (94.10)	.91
Nothing recorded in village clinic register	25 (0.73)	11 (0.31)	.01

[a]CCM: community case management.

[b]Information on districts was missing for 54 participants in the control group (1.6%) and 47 participants in the intervention group (1.3%).

[c]HSA: health surveillance assistant.

Effect of Intervention on Urgent Referral

A total of 5.23% (179/3421) of children in the control phase were urgently referred to a higher-level health care facility, compared with 5.58% (198/3544) in the intervention phase (OR 1.04, 95% CI 0.83-1.29; *P*=.75; Table 2). After adjusting for the effects of time (model 1), the OR was 2.02 (95% CI 1.27-3.23; *P*=.003), in favor of the intervention. Further adjustment for age, sex, and district (models 3 and 4) provided similar strengths of association. We found no significant effect modification in the tested models. Examination of cluster-specific treatment effects, adjusted for calendar time, across the 6 clusters showed that the direction of effect was identical to the overall effect (ie, favoring intervention) in 4 of the 6 clusters but was statistically significant in only one of these (cluster 2; *P*<.001) most likely affected by sample size. In 2 of the 6 clusters (clusters 3 and 4), the direction of effect favored the control arm but was statistically significant in only one of these clusters (cluster 4; *P*=.009).

Table 2. Effects of intervention on urgent referral (N=6965).

Model	Control arm (n=3421), n (%)	Intervention arm (n=3544), n (%)	OR[a] (95% CI)	*P* value	ICC[b] (95% CI)
Urgent referral	179 (5.23)	198 (5.58)			
Model 1: unadjusted	N/A[c]	N/A	1.04 (0.83-1.29)	.75	0.128 (0.042-0.326)
Model 2: adjusted for time	N/A	N/A	2.02 (1.27-3.23)	.003	0.127 (0.042-0.325)
Model 3: adjusted for time, age, and sex	N/A	N/A	2.01 (1.26-3.21)	.003	0.125 (0.041-0.321)
Model 4: adjusted for time, age, sex, and district	N/A	N/A	2.15 (1.44-3.22)	<.001	0.024 (0.000-0.054)

[a]OR: odds ratio.

[b]ICC: intracluster correlation coefficient.

[c]N/A: not applicable.

Effect of Intervention on Repeat Consultation

Repeat consultations with an HSA within 7 days occurred for 5.26% (180/3421) of children in the control phase, compared with 2.42% (86/3544) in the intervention phase (OR 0.45, 95% CI 0.34-0.59; *P*<.001; Table 3). After adjusting for the effects of time (model 1), the OR was 0.57 (95% CI 0.32-1.04), in favor of the intervention arm, but this was not statistically significant (*P*=.07). Further adjustment for age, sex, and district (models 3 and 4) provided similar ORs. We found no significant effect modification in these models. Examination of cluster-specific treatment effects showed that the direction of effect for the outcome of repeat consultation favored the control arm in 4 clusters (clusters 1, 3, 4, and 5) and was statistically significant

in only one of these (cluster 4; *P*=.04), whereas in clusters 2 and 6 it favored the intervention, but was not statistically significant (cluster 2, *P*=.79; cluster 6, *P*=.09).

A total of 219 clinical features were recorded on the children attending repeat consultations, although not all children had clinical symptoms recorded, and more than one clinical feature could be recorded in one child. The most common symptoms reported were fever for <7 days, cough for 21 or more days, diarrhea for <14 days without blood, and fast breathing (Table 3).

Table 3. Effects of intervention on repeat consultation (N=6965).

Model	Control arm (n=3421), n (%)	Intervention arm (n=3544), n (%)	OR[a] (95% CI)	*P* value	ICC[b] (95% CI)
Repeat consultation	180 (5.3)	86 (2.4)			
Model 3: adjusted for time, age, and sex	N/A[c]	N/A	0.57 (0.32-1.04)	.07	0.186 (0.063-0.435)
Model 4: adjusted for time, age, sex, and district	N/A	N/A	0.58 (0.32-1.05)	.07	0.084 (0.024-0.250)

[a]OR: odds ratio.

[b]ICC: intracluster correlation coefficient.

[c]N/A: not applicable.

Effect of Intervention on Hospital Admission

A total of 565 children were admitted to the hospital, 9.35% (320/3421) in the control arm and 6.91% (245/3544) in the intervention arm (OR 0.75, 95% CI 0.62-0.90; *P*=.002; Table 4). Of these, 51.3% (290/565) were urgently referred by their HSA, and 47.8% (270/565) were taken to the hospital by their parents (information on origination was not available for 5 children). After adjusting for the effects of time (model 2), the OR was 1.23 (95% CI 0.83-1.81), in favor of the control arm, but was not statistically significant (*P*=.30). Further adjustment for age, sex, and district (models 3 and 4) provided similar ORs. We found no significant effect modification in these models. Examination of cluster-specific treatment effects showed that the direction of effect on hospital admission favored the control arm in 4 clusters (clusters 1, 2, 5, and 6) and was statistically significant in only one of these (cluster 2; *P*=.002), and the intervention arm in 2 clusters (3 and 4), but was not statistically significant (cluster 3, *P*=.49; cluster 4, *P*=.10).

Table 4. Effect of intervention on hospital admissions (N=6965).

Model	Control arm (n=3421), n (%)	Intervention arm (n=3544), n (%)	OR[a] (95% CI)	*P* value	ICC[b] (95% CI)
Hospitalization	320 (9.4)	245 (6.9)			
Model 3: adjusted for time, age, and sex	N/A[c]	N/A	1.23 (0.83-1.82)	.29	0.034 (0.010-0.113)
Model 4: adjusted for time, age, sex, and district	N/A	N/A	1.14 (0.78-1.67)	.50	0.004 (0.000-0.036)

[a]OR: odds ratio.

[b]ICC: intracluster correlation coefficient.

[c]N/A: not applicable.

Discussion

Principal Findings

We found that the addition of e-CCM to the usual practice of assessing and treating children with HSAs using the paper-based assessment tool led to a significant increase in the proportion of children referred urgently to higher-level health care facilities. This direction of effect of the intervention was found in most but not all of the 6 clusters of clinics. The intervention was also associated with smaller proportions of children who attended a repeat consultation at the village clinic or who needed to be admitted to the hospital, but these were not statistically significant after full adjustment in the models.

We speculate that the modest effect of e-CCM on increasing the proportion of children who were urgently referred may have been because of greater adherence to the CCM decision support algorithm [32,33], as the smartphone app encouraged adherence to CCM [33], or that the smartphone app reduced errors by providing more accurate assessment than paper CCM alone. This is because when using e-CCM, there is less opportunity for the HSA to miss assessment items, as they are forced to complete all items before progressing in the assessment tool. As such, this reduces errors that may have occurred when HSAs use paper-based CCM alone, where they can choose to fill in some questions and leave out others. The effects could, however, merely represent the impact of the requirement in our trial for HSAs to double enter data (ie, on paper format and then using the app) with replication of assessment providing more opportunity to conduct the required assessment. Children in the intervention group were less likely to return for repeat consultation, which could be because of their assessment being more thorough or correct at their initial visit to the village clinic, thus reducing the need to return to the clinic for the same illness.

Our trial also demonstrated that 94.1% (6552/6965) of children aged ≤5 years presenting with acute illness are managed in the community by HSAs, with only approximately 1 in 20 being urgently referred to a higher-level health care facility. Some caregivers appear to bypass the village clinic and attend the hospital directly without being assessed at the village clinic. This implies dissatisfaction with the services provided by lower-level health care facilities, such as village clinics [25]. One reason that caregivers may bypass village clinics and go directly to higher-level facilities may be because caregivers perceive that their child needs urgent attention and may feel that going to the village clinic wastes time [34]. Indeed, previous studies have found that some community members do not have confidence in the care they receive from first-level health care providers, such as community health workers [35,36].

A major strength of our study is the use of a rigorous stepped-wedge design, which aims to minimize potential biases. Previous research has established considerable support by HSAs and the local community for planned interventions [7,10,37]. We could not conduct an individual randomized study at the child or HSA level, following guidance from local investigators. A stepped-wedge design has significant advantages over a simple before and after design as it attempts to minimize potential temporal biases and provides a more powerful study design using a control group. We selected referrals as the main trial outcome as this is a major driver of health care use, and there is currently evidence that underreferral contributes to child morbidity [38,39]. We believe that this study provides a high degree of generalizability to similar settings in sub-Saharan Africa, where CCM is used.

This study had some limitations. One involved the requirement of our study design for HSAs to continue to use the paper tool during the intervention phase, which meant that we could not assess the independent impact of e-CCM but rather its benefit. This occurred because the SL e-CCM App was not yet endorsed by the Malawi Ministry of Health, so it could not replace the paper tool. In addition, assessing outcomes was challenging because of the lack of unique patient identifiers and incomplete records at health care facilities. The study may have been underpowered to detect effects on repeat consultations and hospital admission, and we noted heterogeneity between clusters in direction and significance of effects.

Our results add to previous qualitative studies that have shown that HSAs feel empowered using apps such as the one evaluated here [21] and that these tools might support greater adherence to CCM [32]. These results support wider implementation and potential addition of further functionalities to support HSAs using smartphones to facilitate other tasks, such as reporting requirements or drug stocking. For policy makers, our study provides robust evidence on the effectiveness of this e-CCM tool that few other studies have provided to date and could therefore add evidence to support national digital health strategies in developing an integrated community health information system. However, we acknowledge that further data are needed to determine the costs (and cost-effectiveness) of the initial setup (training, smartphones, and servers), and ongoing support (technical support, hardware replacement, and software updates). For researchers in this area, evidence for the independent effects of e-CCM on outcomes is needed. These results call for additional studies on e-CCM apps and similar mHealth tools to determine the impact on other clinical outcomes, such as duration of illness and resource use. In addition, there is a need to ensure data and workflow integration of e-CCM as part of community health service delivery, management, and existing digital tools used at other levels of the health system, such as the District Health Information System, electronic medical record systems, and real-time surveillance at the community level.

Conclusions

The vast majority of children assessed by HSAs in these 2 large rural areas of Malawi are managed by HSAs in their communities. Adding an e-CCM that involved an app on smartphones to HSAs' usual paper-based CCM tool led to a significant increase in the proportion of children referred from village clinics to higher-level health care facilities. Although the effects on decreased hospital admission and decreased repeat consultation were suggestive but inconclusive, they support the hypothesis that the e-CCM tool improved decision-making at the HSA level. Combined with existing qualitative literature showing high levels of acceptability of mHealth versions of CCM by community health workers, our findings support further efforts to deploy smartphone-based tools as part of integrated digital health strategies in Malawi and similar countries in sub-Saharan Africa, with ongoing evaluation of effectiveness, cost-effectiveness, and acceptability by health care workers and community members.

Acknowledgments

The authors would like to thank the University of Malawi (College of Medicine), Mzuzu University, University of Washington, and Imperial College London for giving permission to conduct this study. This work was supported by the European Union's Seventh Framework Programme for Research, Technological Development, and Demonstration (grant 305292). The funder of this study had no role in the design of the study, collection of data, analysis of the results, interpretation of results, preparation of the manuscript, or decision to publish.

Authors' Contributions

GBC helped in the development of the questionnaire and interview guide, data collection, and writing of the manuscript. NM and TT contributed to the trial design and management and commented on this manuscript. KD contributed to the conduct of the study and commented on this manuscript. JOD conceived the study, developed the research question, participated in trial design

and management, and commented on this manuscript. PH developed the statistical analysis plan and analyzed and interpreted the data. NI managed the field data collection and contributed to the writing of the manuscript. JTSW assisted in the study implementation and communication with national and district health authorities and commented on this manuscript. All authors contributed to the interpretation of the findings, drafting of the manuscript, and approval of the final version.

Conflicts of Interest

None declared.

References

1. Mandeville KL, Ulaya G, Lagarde M, Muula AS, Dzowela T, Hanson K. The use of specialty training to retain doctors in Malawi: a discrete choice experiment. Soc Sci Med 2016 Nov;169:109-118 [FREE Full text] [doi: 10.1016/j.socscimed.2016.09.034] [Medline: 27716548]
2. McAuliffe E, Bowie C, Manafa O, Maseko F, MacLachlan M, Hevey D, et al. Measuring and managing the work environment of the mid-level provider--the neglected human resource. Hum Resour Health 2009 Feb 19;7(1):13 [FREE Full text] [doi: 10.1186/1478-4491-7-13] [Medline: 19228391]
3. Schmiedeknecht K, Perera M, Schell E, Jere J, Geoffroy E, Rankin S. Predictors of workforce retention among Malawian nurse graduates of a scholarship program: a mixed-methods study. Glob Health Sci Pract 2015 Mar 5;3(1):85-96 [FREE Full text] [doi: 10.9745/GHSP-D-14-00170] [Medline: 25745122]
4. Samb B, Celletti F, Holloway J, Van Damme W, De Cock KM, Dybul M. Rapid expansion of the health workforce in response to the HIV epidemic. N Engl J Med 2007 Dec 13;357(24):2510-2514. [doi: 10.1056/NEJMsb071889] [Medline: 18077816]
5. Medhanyie AA, Little A, Yebyo H, Spigt M, Tadesse K, Blanco R, et al. Health workers' experiences, barriers, preferences and motivating factors in using mHealth forms in Ethiopia. Hum Resour Health 2015 Jan 15;13(1):2 [FREE Full text] [doi: 10.1186/1478-4491-13-2] [Medline: 25588973]
6. Kok MC, Muula AS. Motivation and job satisfaction of health surveillance assistants in Mwanza, Malawi: an explorative study. Malawi Med J 2013 Mar;25(1):5-11 [FREE Full text] [Medline: 23717748]
7. Chikaphupha KR, Kok MC, Nyirenda L, Namakhoma I, Theobald S. Motivation of health surveillance assistants in Malawi: a qualitative study. Malawi Med J 2016 Jun 02;28(2):37-42 [FREE Full text] [doi: 10.4314/mmj.v28i2.2] [Medline: 27895826]
8. Kadzandira JM, Chilowa WR. The role of heath surveillance assistants (HSAs) in the delivery of health services and immunization in Malawi. Eldis. 2001. URL: https://www.eldis.org/document/A56114 [accessed 2021-09-03]
9. Young M, Wolfheim C, Marsh DR, Hammamy D. World Health Organization/United Nations Children's Fund joint statement on integrated community case management: an equity-focused strategy to improve access to essential treatment services for children. Am J Trop Med Hyg 2012 Nov;87(5 Suppl):6-10 [FREE Full text] [doi: 10.4269/ajtmh.2012.12-0221] [Medline: 23136272]
10. Kok MC, Namakhoma I, Nyirenda L, Chikaphupha K, Broerse JE, Dieleman M, et al. Health surveillance assistants as intermediates between the community and health sector in Malawi: exploring how relationships influence performance. BMC Health Serv Res 2016 May 03;16(1):164 [FREE Full text] [doi: 10.1186/s12913-016-1402-x] [Medline: 27142944]
11. O' Connor Y, Hardy V, Thompson M, Mastellos N, Tran T, O' Donoghue J, et al. The importance of form field validation: lessons learnt from a feasibility study of an mHealth application in Malawi, Africa. 2016 Presented at: Proceeding of the 20th Pacific Asia Conference on Information Systems (PACIS 2016); Jun 27 - Jul 1, 2016; Chiayi, Taiwan.
12. Malawi population at 17.5 million: 2018 PHC preliminary results. Malawi Government. 2019. URL: http://populationmalawi.org/wp1/malawi-population-at-17-5-million-2018-phc-preliminary-results/ [accessed 2021-01-02]
13. Dharmayat KI, Tran T, Hardy V, Chirambo BG, Thompson MJ, Ide N, et al. Sustainability of 'mHealth' interventions in sub- Saharan Africa: a stakeholder analysis of an electronic community case management project in Malawi. Malawi Med J 2019 Sep;31(3):177-183 [FREE Full text] [doi: 10.4314/mmj.v31i3.3] [Medline: 31839886]
14. Young M, Wolfheim C, Marsh DR, Hammamy D. World Health Organization/United Nations Children's Fund joint statement on integrated community case management: an equity-focused strategy to improve access to essential treatment services for children. Am J Trop Med Hyg 2012 Nov;87(5 Suppl):6-10 [FREE Full text] [doi: 10.4269/ajtmh.2012.12-0221] [Medline: 23136272]
15. Theodoratou E, Al-Jilaihawi S, Woodward F, Ferguson J, Jhass A, Balliet M, et al. The effect of case management on childhood pneumonia mortality in developing countries. Int J Epidemiol 2010 Apr;39 Suppl 1:i155-i171 [FREE Full text] [doi: 10.1093/ije/dyq032] [Medline: 20348118]
16. Kruk ME, Chukwuma A, Mbaruku G, Leslie HH. Variation in quality of primary-care services in Kenya, Malawi, Namibia, Rwanda, Senegal, Uganda and the United Republic of Tanzania. Bull World Health Organ 2017 Jun 01;95(6):408-418 [FREE Full text] [doi: 10.2471/BLT.16.175869] [Medline: 28603307]
17. O'Connor Y, Heavin C, O' Connor S, Gallagher J, Wu J, O'Donoghue J. Service blueprint for improving clinical guideline adherence via mobile health technology. World Academy of Science, Engineering and Technology. 2015. URL: https:/

/publications.waset.org/10002844/service-blueprint-for-improving-clinical-guideline-adherence-via-mobile-health-technology [accessed 2021-09-03]
18. Hardy V, Thompson M, Muula AS. Evaluating mobile solutions of integrated Community Case Management (iCCM): making the final connection. Malawi Med J 2017 Dec;29(4):332-334 [FREE Full text] [doi: 10.4314/mmj.v29i4.11] [Medline: 29963291]
19. Chirwa Z, Kumwenda B, Msiska B. Impact of mHealth solutions on health care services provided by community health workers: a case of Nkhotakota district. In: Proceedings of the IST-Africa 2016 Conference. 2016 Presented at: IST-Africa 2016 Conference; May 11-13, 2016; Durban, South Africa.
20. Kabakyenga J, Barigye C, Brenner J, Maling S, Buchner D, Nettle-Aquirre A, et al. A demonstration of mobile phone deployment to support the treatment of acutely ill children under five in Bushenyi district, Uganda. Afr Health Sci 2016 Mar;16(1):89-96 [FREE Full text] [doi: 10.4314/ahs.v16i1.12] [Medline: 27358618]
21. Ide N, Hardy V, Chirambo G, Heavin C, O'Connor Y, O'Donoghue J, et al. People welcomed this innovation with two hands: a qualitative report of an mhealth intervention for community case management in Malawi. Ann Glob Health 2019 Apr 25;85(1):61 [FREE Full text] [doi: 10.5334/aogh.919] [Medline: 31025838]
22. Chirambo GB, Muula AS, Thompson M, Hardy VE, Heavin C, Connor YO, et al. End-user perspectives of two mHealth decision support tools: electronic community case management in Northern Malawi. Int J Med Inform 2021 Jan;145:104323. [doi: 10.1016/j.ijmedinf.2020.104323] [Medline: 33232917]
23. Hardy V, O'Connor Y, Heavin C, Mastellos N, Tran T, O'Donoghue J, et al. The added value of a mobile application of community case management on referral, re-consultation and hospitalization rates of children aged under 5 years in two districts in Northern Malawi: study protocol for a pragmatic, stepped-wedge cluster-randomized controlled trial. Trials 2017 Oct 11;18(1):475 [FREE Full text] [doi: 10.1186/s13063-017-2213-z] [Medline: 29020976]
24. O' Connor Y, Hardy V, Heavin C, Gallagher J, O' Donoghue J. Supporting LIFE: mobile health application for classifying, treating and monitoring disease outbreaks of sick children in developing countries. In: New Horizons in Design Science: Broadening the Research Agenda. Cham: Springer; 2015:366-370.
25. Zembe-Mkabile WZ, Jackson D, Sanders D, Besada D, Daniels K, Zamasiya T, et al. The 'community' in community case management of childhood illnesses in Malawi. Glob Health Action 2016 Jan 27;9:29177 [FREE Full text] [doi: 10.3402/gha.v9.29177] [Medline: 26823049]
26. Nsona H, Mtimuni A, Daelmans B, Callaghan-Koru JA, Gilroy K, Mgalula L, et al. Scaling up integrated community case management of childhood illness: update from Malawi. Am J Trop Med Hyg 2012 Nov;87(5 Suppl):54-60 [FREE Full text] [doi: 10.4269/ajtmh.2012.11-0759] [Medline: 23136278]
27. D-Tree : examining the efficiency of a community case management mobile medical diagnostic tool. University of Cape Town. 2015. URL: https://open.uct.ac.za/handle/11427/20565 [accessed 2021-09-03]
28. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ 2015 Feb 06;350:h391 [FREE Full text] [doi: 10.1136/bmj.h391] [Medline: 25662947]
29. Hemming K, Taljaard M, Forbes A. Analysis of cluster randomised stepped wedge trials with repeated cross-sectional samples. Trials 2017 Mar 04;18(1):101 [FREE Full text] [doi: 10.1186/s13063-017-1833-7] [Medline: 28259174]
30. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. Contemp Clin Trials 2007 Feb;28(2):182-191. [doi: 10.1016/j.cct.2006.05.007] [Medline: 16829207]
31. Bakibinga P, Kamande E, Omuya M, Ziraba AK, Kyobutungi C. The role of a decision-support smartphone application in enhancing community health volunteers' effectiveness to improve maternal and newborn outcomes in Nairobi, Kenya: quasi-experimental research protocol. BMJ Open 2017 Jul 20;7(7):e014896 [FREE Full text] [doi: 10.1136/bmjopen-2016-014896] [Medline: 28729309]
32. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ 2005 Apr 02;330(7494):765 [FREE Full text] [doi: 10.1136/bmj.38398.500764.8F] [Medline: 15767266]
33. Kahabuka C, Moland KM, Kvåle G, Hinderaker SG. Unfulfilled expectations to services offered at primary health care facilities: experiences of caretakers of underfive children in rural Tanzania. BMC Health Serv Res 2012 Jun 14;12(1):158 [FREE Full text] [doi: 10.1186/1472-6963-12-158] [Medline: 22697458]
34. Kruk ME, Mbaruku G, McCord CW, Moran M, Rockers PC, Galea S. Bypassing primary care facilities for childbirth: a population-based study in rural Tanzania. Health Policy Plan 2009 Jul 20;24(4):279-288. [doi: 10.1093/heapol/czp011] [Medline: 19304785]
35. Visconti RM, Larocca A, Marconi M. Accessibility to First-Mile health services: a time-cost model for rural Uganda. Soc Sci Med 2020 Nov;265:113410. [doi: 10.1016/j.socscimed.2020.113410] [Medline: 33045653]
36. Bastawrous A, Armstrong MJ. Mobile health use in low- and high-income countries: an overview of the peer-reviewed literature. J R Soc Med 2013 Apr;106(4):130-142 [FREE Full text] [doi: 10.1177/0141076812472620] [Medline: 23564897]
37. Macintyre KC, Littrell M, Hotchkiss DR, Mndzebele S, Nkambule R, Malima B, et al. Barriers to referral in Swaziland: perceptions from providers and clients of a system under stress. World Med Health Policy 2011;3(4):1-29 [FREE Full text] [doi: 10.2202/1948-4682.1183]

38. Disease Control Priorities in Developing Countries, 2nd edition. New York: Oxford University Press; 2006:1245-1261-1293-1307.
39. Chirambo GB, Muula AS, Thompson M. Factors affecting sustainability of mHealth decision support tools and mHealth technologies in Malawi. Inform Med Unlocked 2019;17:100261. [doi: 10.1016/j.imu.2019.100261]

Abbreviations

CCM: community case management
HSA: health surveillance assistant
LMIC: low- and middle-income country
mHealth: mobile health
OR: odds ratio
SL e-CCM: Supporting LIFE e-community case management
VCR: village clinic register

Edited by R Kukafka; submitted 15.11.20; peer-reviewed by S Taylor-Robinson, B Tilahun, P Delir Haghighi, H Mehdizadeh; comments to author 29.12.20; revised version received 11.02.21; accepted 10.08.21; published 20.10.21.

Please cite as:
Chirambo GB, Thompson M, Hardy V, Ide N, Hwang PH, Dharmayat K, Mastellos N, Heavin C, O'Connor Y, Muula AS, Andersson B, Carlsson S, Tran T, Hsieh JCL, Lee HY, Fitzpatrick A, Joseph Wu TS, O'Donoghue J
Effectiveness of Smartphone-Based Community Case Management on the Urgent Referral, Reconsultation, and Hospitalization of Children Aged Under 5 Years in Malawi: Cluster-Randomized, Stepped-Wedge Trial
J Med Internet Res 2021;23(10):e25777
URL: https://www.jmir.org/2021/10/e25777
doi:10.2196/25777
PMID:34668872

Original Paper

The Use of Cannabinoids for Insomnia in Daily Life: Naturalistic Study

Nirushi Kuhathasan[1,2], HBsc; Luciano Minuzzi[1,3], MD, PhD; James MacKillop[2,3,4], PhD; Benicio N Frey[1,2,3], MSc, MD, PhD

[1]Mood Disorders Program and Women's Health Concerns Clinic, St. Joseph's Healthcare Hamilton, Hamilton, ON, Canada

[2]Michael G DeGroote Centre for Medicinal Cannabis Research, McMaster University, Hamilton, ON, Canada

[3]Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, ON, Canada

[4]Peter Boris Centre for Addictions Research, McMaster University/St. Joseph's Healthcare Hamilton, Hamilton, ON, Canada

Corresponding Author:
Benicio N Frey, MSc, MD, PhD
Mood Disorders Program and Women's Health Concerns Clinic
St. Joseph's Healthcare Hamilton
100 West 5th Street
Hamilton, ON, L8N 3K7
Canada
Phone: 1 905 522 1155
Email: freybn@mcmaster.ca

Abstract

Background: Insomnia is a prevalent condition that presents itself at both the symptom and diagnostic levels. Although insomnia is one of the main reasons individuals seek medicinal cannabis, little is known about the profile of cannabinoid use or the perceived benefit of the use of cannabinoids in daily life.

Objective: We conducted a retrospective study of medicinal cannabis users to investigate the use profile and perceived efficacy of cannabinoids for the management of insomnia.

Methods: Data were collected using the Strainprint app, which allows medicinal cannabis users to log conditions and symptoms, track cannabis use, and monitor symptom severity pre- and postcannabis use. Our analyses examined 991 medicinal cannabis users with insomnia across 24,189 tracked cannabis use sessions. Sessions were analyzed, and both descriptive statistics and linear mixed-effects modeling were completed to examine use patterns and perceived efficacy.

Results: Overall, cannabinoids were perceived to be efficacious across all genders and ages, and no significant differences were found among product forms, ingestion methods, or gender groups. Although all strain categories were perceived as efficacious, predominant *indica* strains were found to reduce insomnia symptomology more than cannabidiol (CBD) strains (estimated mean difference 0.59, SE 0.11; 95% CI 0.36-0.81; adjusted P<.001) and predominant *sativa* strains (estimated mean difference 0.74, SE 0.16; 95% CI 0.43-1.06; adjusted P<.001). Indica hybrid strains also presented a greater reduction in insomnia symptomology than CBD strains (mean difference 0.52, SE 0.12; 95% CI 0.29-0.74; adjusted P<.001) and predominant sativa strains (mean difference 0.67, SE 0.16; 95% CI 0.34-1.00; adjusted P=.002).

Conclusions: Medicinal cannabis users perceive a significant improvement in insomnia with cannabinoid use, and this study suggests a possible advantage with the use of predominant indica strains compared with predominant sativa strains and exclusively CBD in this population. This study emphasizes the need for randomized placebo-controlled trials assessing the efficacy and safety profile of cannabinoids for the treatment of insomnia.

(J Med Internet Res 2021;23(10):e25730) doi:10.2196/25730

KEYWORDS

medicinal cannabis; insomnia; symptom management; linear mixed-effects

Introduction

Background

With the growing interest in the therapeutic and medicinal uses of cannabis, there is an increased need to better understand the harms and benefits of acute and long-term therapeutic use of cannabinoids. Among individuals who use medicinal cannabis in Canada, 42% report using cannabis 2-3 times a day, with 40% of users reporting their consumption to be >14 grams per week [1]. In fact, rates of medicinal cannabis authorization in Canada rose from 8000 in 2014 to 340,000 in 2018 [2]. Similarly, with nationwide cannabis legalization in October 2018, general cannabis use rates in Canada increased from 14% to 18% between 2018 and 2019 [3].

Despite the paucity of randomized placebo-controlled trials, both recreational and medicinal cannabis users report perceptions of a broad spectrum of benefits from cannabis. Among these benefits of the use of cannabis is aiding sleep [4]. In fact, in addition to pain and anxiety, insomnia has commonly been reported to be among the top reasons individuals seek medicinal cannabis [2]. This association is very relevant considering the high rates of insomnia in the general population. It is estimated that approximately 10% of adults experience chronic insomnia [5], and nearly one-third of all adults suffer from occasional or intermittent insomnia symptoms annually [5]. Longitudinal studies have found that nearly 70% of individuals reporting insomnia symptoms at baseline continue to report symptoms a year later [6], and 50% continue to report having symptoms 3 years later. Insomnia is also one of the most common complaints in primary care, often presenting itself at both symptom and diagnostic levels [6]. Characterized by difficulty in falling asleep, staying asleep, or having a nonrestorative sleep, insomnia negatively affects functioning, quality of life, and mental health [7]. In addition, insomnia often co-occurs with common medical and psychiatric conditions [6]. Individuals experiencing these comorbidities report greater impairments in psychosocial and cognitive functioning compared with individuals without sleep disturbances [6-10].

Cannabinoids for Insomnia

Recent reviews have concluded that the current evidence of the benefits of using cannabinoids for insomnia symptoms are largely driven by clinical trials that used cannabinoids for the treatment of other conditions, such as pain or multiple sclerosis [11-13]. Similarly, although some previous studies have examined recreational and medicinal cannabis use in naturalistic samples, very few have focused on insomnia as a primary outcome [14,15]. In one study, 95 medicinal cannabis users were surveyed on the effects of cannabis products used for various conditions and symptoms [14]. The results indicated a statistically significant preference toward *Cannabis indica* products to help with sedation and sleep [14]. In addition, the same study reported that users also preferred these products for insomnia, encouraging further research focusing on the condition [14]. In another study, a mobile app collecting data on medicinal cannabis in naturalistic conditions was used to measure the self-reported effectiveness and side effects of cannabis [15]. The study examined 2332 users across 10,535 tracked cannabis sessions [15]. The results indicated significant reductions in symptom severity across all reported symptoms, with significantly more relief in anxiety- and depression-related symptoms than pain symptoms [15]. Notably, in this particular study, insomnia was examined as a symptom of anxiety and presented the largest symptom relief score across all examined symptoms following cannabis consumption [15].

Most relevantly, a recent naturalistic study that examined cannabis use for insomnia in a sample of 409 participants across 1056 sessions reported significant reductions in symptom severity; however, these findings were limited to raw, natural medical cannabis flowers and lacked information on the perceived efficacy of various cannabis product forms. Furthermore, this study was limited by a lack of information on patient demographics, as the information collected from users did not include key demographic data, such as age and gender [16]. Since the legalization of cannabis in Canada in 2018, research regulations for the drug remain quite stringent [1]. Similarly, because of its status as a schedule 1 drug in the United States, it is underinvestigated for therapeutic purposes [17-20]. Therefore, not only is there a major gap in studies assessing insomnia as the primary outcome, but also a lack of scientific literature on the use of cannabis products that are currently being consumed by the general public [11-13]. To help address these gaps, we conducted a retrospective study to investigate the perceived effectiveness of the use of cannabinoids in treating insomnia symptoms in a large, naturalistic sample of Canadians. We also describe the key demographic characteristics of these individuals, such as age and gender distribution, types of cannabinoid use, and methods of ingestion.

Methods

Overview of Strainprint App

We conducted a retrospective study examining cannabis use for the management of insomnia symptoms using anonymous archival data obtained from the medicinal cannabis–tracking app Strainprint (Strainprint Technologies Ltd). Strainprint is a Canadian app with a large database of medical and recreational cannabis users with >90 million data points and 2 million reported patient outcomes. The app allows users to track and monitor changes in their symptoms as a function of different doses, strains, and forms of cannabis. It engages users through a loyalty rewards system where users earn points for tracking sessions of cannabis use. Through Strainprint, users are able to record medical conditions, symptoms being treated, methods of ingestion, doses, emotive effects, pre- and postmedication ratings, and cannabis product constituents by batch for each tracked session. Tracked information can also be shared with health care providers. On initial use of the app, individuals are prompted to enter basic demographic information, such as year and month of birth, gender, and the conditions and symptoms that they wish to treat. When individuals are ready to track their medication session, they open the app before using cannabis and select the relevant symptoms they wish to treat from a dropdown list of their previously chosen symptoms. Users are then taken through a set of steps where they are first prompted

to rate the severity of their symptoms on a 11-point numeric rating scale (0=least severe and 10=very severe) before medication. Next, individuals select the cannabis they are using by product name and batch. Strainprint prepopulates the app with lab-verified cannabis constituents by batch for all medical cannabis products sold by licensed producers in Canada. Data on cannabis content are pulled directly from cannabis distributors. Users then select the product form (flower, oil, capsule, edible, vape pen, or concentrate), route of administration (vape, oil, smoke, edible, pill, tincture, spray, concentrate, dab bubble, dab portable, oral, topical, or transdermal), and dose (drops, mg, ml, or puffs) for that specific session. After an onset period defined by the chosen route of administration (eg, 20 minutes for smoke and 60 minutes for pill or edible), users are prompted with a push notification (8 hours later for sleep) to complete their session by rating their symptom severity postmedication on the same 11-point numeric scale.

Strainprint also provides individuals with a complete history of their use, along with product recommendations based on other users' experiences with the same symptoms. As part of Strainprint's terms of service, individuals agree to share their anonymous information for research and other purposes. In this study, we examined the data of individuals who used medicinal cannabis to manage the severity of insomnia symptoms for the condition of insomnia. Specific variables for this study were determined before data extraction, and the information was subsequently provided by Strainprint stripped of identifiers.

Study Sample

Our study included all tracked sessions between February 27, 2017, and February 28, 2020. The final sample consisted of 991 Canadian medicinal cannabis users with insomnia who used the app to monitor changes in insomnia symptoms across 24,189 recorded sessions. The sample comprised 42.6% (422/991) self-identified male participants and 56.1% (556/991) self-identified female participants (13/991, 1.3% of users did not report gender), ranging in age from 18 to 74 years (mean 36.32, SD 11.65). Additional descriptive statistics on the sample are presented in Figure S1 of Multimedia Appendix 1.

Statistical Analysis

First, we completed a descriptive analysis of the data set by generating information on the specific cannabis use profile for the management of insomnia symptoms. In particular, we examined the frequencies of categorical cannabis use variables such as use time of day, strain categories, product forms, and ingestion methods. These data were further stratified to investigate cannabis use trends by both age and gender. For inferential analyses, our primary analysis focused on the perceived efficacy of cannabis for the management of insomnia symptoms. Efficacy was calculated as the change in insomnia symptomology between pre- and postmedication rates, as reported by users.

Generally, this type of statistical modeling would be completed as a standard regression analysis; however, a standard regression analysis assumes that observations are independent. In this particular data set, users reported multiple observations, and a standard regression analysis would not account for between-person variability in tracked sessions across users. Therefore, we used linear mixed-effects modeling, a type of regression model that estimates random effects (accounting for between-subject variability) in addition to standard fixed effects (accounting for within-subject variability) regardless of differences in the number of reported observations per user. In essence, this mixed-modeling method estimates random intercepts and slopes, which are then used to make more accurate inferences at the fixed-effects level without violating the independence assumption.

Assumptions for each model were checked to ensure the validity of the models used. Residual plots were examined and were determined to not deviate from the assumptions of linearity, normality, and homoscedasticity. The assumption of independence was met by accounting for tracked session nesting within participants using mixed-effects models. For this analysis, linear mixed-effects modeling was used to predict changes in the perceived efficacy of cannabis use with regard to demographic information (ie, age and gender) and cannabis use information (ie, use time of day, product form, and strain category) across tracked sessions. In addition, although several studies have challenged the labeling of strain categories in commercial products, in this naturalistic study, we analyzed this variable as a commercialized label influencing purchasing choices. In all analyses, P values were corrected for multiple comparisons using the stringent Bonferroni correction ($P<.05$, Bonferroni corrected).

Ethical Approval

Ethical approval for this research was granted by the Hamilton Integrated Research Ethics Board (project #7162). The study was designed to be compliant with the Health and Information Protection Act, 2016.

Results

Strain Categories for Insomnia

Descriptive statistics examining the percentage of each strain category (ie, predominant sativa, sativa hybrid, predominant indica, indica hybrid, balanced hybrid, or cannabidiol [CBD]) used for the management of insomnia symptoms across 24,189 tracked sessions are presented in Table 1. Overall, predominant indica and indica hybrid strains were the most commonly used strains for insomnia, whereas predominant sativa and sativa hybrid strains were used least for the management of insomnia symptoms. Notably, although CBD is not traditionally considered a strain category, Strainprint recognizes the variations in tetrahydrocannabinol (THC) or CBD content across different strains and presents a CBD-predominant product category as a strain on the app. Further descriptive statistics of strain categories stratified by age and gender are presented in Figures S2 and S3 in Multimedia Appendix 1.

Table 1. Descriptive information on frequency of strain categories used across 24,189 tracked sessions (N=24,189).

Stain category	Sessions, n (%)
Indica	9263 (38.29)
Indica hybrid	6468 (26.74)
CBD[a]	3327 (13.75)
Balanced hybrid	3068 (12.68)
Sativa	1098 (4.54)
Sativa hybrid	605 (2.5)

[a]CBD: cannabidiol.

Cannabis Product Forms and Ingestion Methods for Insomnia

Descriptive statistics examining the frequencies of cannabis product forms (ie, flower, oil, capsule, edible, vape pen, or concentrate) used for the management of insomnia symptoms across 24,189 tracked sessions are presented in Table 2. Because of the relatively small number of data points, products in the form of vape pens and concentrates were combined to form an *other* group. Across all age groups and genders, cannabis was most often used in the form of flowers, followed by oil products, for the management of insomnia symptoms. Table 3 presents descriptive statistics examining the frequencies of cannabis ingestion methods (ie, vape, oil, smoke, edible, pill, tincture, spray, concentrate dab bubbler, dab portable, oral, topical, or transdermal) across all tracked sessions. Again, because of the relatively small number of data points, the categories of concentrate, dab bubbler, dab portable, oral, topical, and transdermal were combined to form a single category. Vaping was the most popular ingestion method across all age groups and genders. All reported results were stratified by age and gender.

Table 2. Frequency and percentage of cannabis product forms used across 24,189 sessions between genders and age groups (N=24,189).

Product form	Sessions, n (%)								
	By gender			By age (years)					Overall
	Female	Male	Unknown	18-24	25-34	35-44	45-54	>55	
Flower	6767 (43.82)	8517 (55.2)	160 (1.04)	1712 (11.09)	3954 (25.6)	5622 (36.4)	1934 (12.5)	1997 (12.9)	15,444 (100)
Oil	5407 (70.2)	2222 (28.9)	70 (0.9)	336 (4.4)	1332 (17.3)	2401 (31.2)	1609 (20.9)	1887 (24.5)	7699 (100)
Capsule	793 (94.5)	44 (5.2)	2 (0.2)	1 (0.1)	341 (40.6)	179 (21.3)	47 (5.6)	269 (32.1)	839 (100)
Edible	121 (96)	5 (4)	0 (0)	15 (11.9)	21 (16.7)	6 (4.8)	0 (0)	73 (57.9)	126 (100)
Other	36 (44.4)	45 (55.6)	0 (0)	2 (2.5)	6 (7.4)	38 (46.9)	18 (22.2)	17 (21)	81 (100)

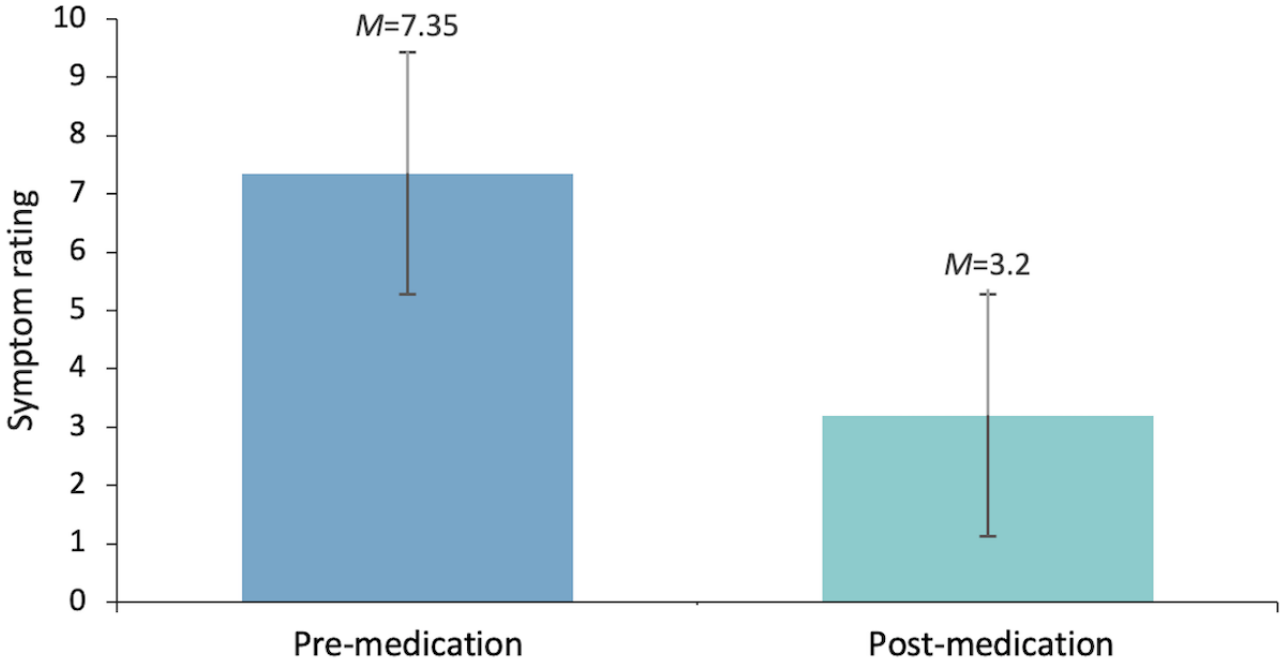
Table 3. Frequency and percentage of cannabis ingestion methods across 24,189 sessions between genders and age groups (N=24,189).

Ingestion method	Sessions, n (%)								
	By gender			By age (years)					Overall
	Female	Male	Unknown	18-24	25-34	35-44	45-54	>55	
Vape	3985 (46.9)	4419 (52)	91 (1.1)	719 (8.5)	1773 (20.9)	3180 (37.4)	1315 (15.5)	1389 (16.4)	8495 (100)
Oil	5277 (71.4)	2044 (27.7)	68 (0.9)	321 (4.3)	1282 (17.4)	2208 (29.9)	1570 (21.2)	1840 (24.9)	7389 (100)
Smoke	2379 (38.3)	3774 (60.7)	62 (1)	947 (15.2)	2062 (33.2)	2186 (35.2)	438 (7)	521 (8.4)	6215 (100)
Edible	546 (58.6)	383 (41.1)	3 (0.3)	37 (4)	165 (17.7)	425 (45.6)	135 (14.5)	152 (16.3)	932 (100)
Pill	635 (93.5)	44 (6.5)	0 (0)	2 (0.3)	255 (37.6)	96 (14.1)	46 (6.8)	278 (40.9)	679 (100)
Tincture	121 (77.1)	32 (20.4)	4 (2.5)	12 (7.6)	37 (23.6)	41 (26.1)	54 (34.4)	12 (7.6)	157 (100)
Spray	108 (85.7)	15 (11.9)	3 (2.3)	0 (0)	16 (12.7)	51 (40.5)	25 (19.8)	32 (25.4)	126 (100)
Other	73 (37.2)	122 (62.2)	1 (0.5)	28 (14.3)	64 (32.7)	59 (30.1)	25 (12.8)	19 (9.7)	196 (100)

Symptom Severity Ratings

Mean symptom severity ratings were examined before and after cannabis use across tracked sessions (N=991 users across 24,819 sessions; Figure 1). Before cannabis use, the mean symptom severity rating across sessions was 7.35 (SD 1.88), whereas the mean symptom severity rating after use was 3.20 (SD 2.37).

Figure 1. Mean symptom severity ratings pre- (mean 7.35, SD 1.88) and post-(mean 3.20, SD 2.37) cannabis use.



Linear Mixed-Effects Model Predictions of Perceived Efficacy

We first examined the perceived efficacy of cannabinoid use for insomnia as a function of gender and found it to be significant across both genders (Table 4). The perceived efficacy of cannabinoid use for insomnia was also significant across all age groups (Table 5). Comparisons between gender and age can be found in Multimedia Appendix 1 (Tables S1 and S2).

Table 4. Efficacy by gender. The efficacy was tested using linear mixed modeling (β coefficient was not standardized).

Gender	Estimate (SE)	*t* test[a] (*df*)	*P* value
Female	3.4289 (0.1696)	26.814 (852.4)	<.001
Male	3.5282 (0.1288)	27.399 (868.3)	<.001

[a]Tests were two-tailed.

Table 5. Efficacy by age groups. The efficacy was tested using linear mixed modeling (β coefficient was not standardized).

Age (years)	Estimate (SE)	*t* test[a] (*df*)	*P* value
18-24	3.43291 (0.16118)	21.298 (2073.5)	<.001
25-34	3.35234 (0.11098)	30.206 (1656.2)	<.001
35-44	3.72108 (0.10826)	34.370 (1604.7)	<.001
45-54	3.12995 (0.13658)	22.916 (1894.7)	<.001
>55	3.76299 (0.15390)	24.251 (2242.4)	<.001

[a]Tests were two-tailed.

Next, we examined whether the time of cannabis use predicted perceived efficacy and found that efficacy was significant regardless of the time of day (Table 6). Because of the nature of the data, information on shift work was not available; therefore, we did not compare efficacy across different use times during the day. More detailed frequency and percentage information of sessions for each use time of the day can be found in Table S3 of Multimedia Appendix 1.

Table 6. Efficacy by use time of day. The efficacy was tested using linear mixed modeling (β coefficient was not standardized).

Time of day	Estimate (SE)	*t* test[a] (*df*)	*P* value
Morning	3.668345 (0.102629)	35.744 (488.1)	<.001
Afternoon	2.754512 (0.225454)	12.218 (120.1)	<.001
Evening	3.369924 (0.119686)	28.156 (391.6)	<.001
Overnight	3.449696 (0.089454)	38.564 (739.5)	<.001

[a]Tests were two-tailed.

We also examined perceived efficacy as a function of cannabis product forms and found that all product forms were perceived as efficacious (Table 7). Notably, for some product forms (ie, vape pen and concentrate), there were too few observations to warrant inclusion in primary analyses, even when combined to form a single category. Therefore, of all available product forms, only those making up at least 0.005% of the data set were included in the analyses. There were no significant differences in efficacy among product forms (all *P*>.05; Table S4, Multimedia Appendix 1).

Table 7. Efficacy by product form. The efficacy was tested using linear mixed modeling (β coefficient was not standardized).

Product form	Estimate (SE)	*t* test[a] (*df*)	*P* value
Capsules	3.79417 (0.34792)	10.9054 (25.7)	<.001
Edible	4.15951 (0.56358)	7.3806 (7.9)	<.001
Flower	3.43969 (0.08728)	39.4100 (737.0)	<.001
Oil	3.47823 (0.11693)	29.7470 (374.8)	<.001

[a]Tests were two-tailed.

Finally, we examined perceived efficacy as a function of strain category and found that cannabis was efficacious regardless of the specific strain being used (Table 8). Interestingly, predominant indica strains were found to be more efficacious than CBD (estimated mean difference 0.59, SE 0.11; 95% CI 0.36-0.81; adjusted *P*<.001) and predominant sativa strains (estimated mean difference 0.74, SE 0.16; 95% CI 0.43-1.06; adjusted *P*<.001). Indica hybrid strains were also found to be more efficacious than CBD (estimated mean difference 0.52, SE 0.12; 95% CI 0.29-0.74; adjusted *P*<.001) and predominant sativa strains (estimated mean difference 0.67, SE 0.16; 95% CI 0.34-1.00; adjusted *P*=.002). Balanced hybrid strains were also found to be more efficacious than CBD (estimated mean difference 0.39, SE 0.13; 95% CI 0.14-0.64; adjusted *P*=.03) and sativa strains (estimated mean difference 0.54, SE 0.17; 95% CI 0.20-0.88; adjusted *P*=.03; Table 9).

Table 8. Efficacy by strain categories. The efficacy was tested using linear mixed modeling (β coefficient was not standardized).

Strain categories	Estimate (SE)	*t* test[b] (*df*)	*P* value
Balanced hybrid	3.461359 (0.112701)	30.713 (300.8)	<.001
CBD[a]	3.074027 (0.114947)	26.743 (367.7)	<.001
Indica	3.661426 (0.094507)	38.742 (673.9)	<.001
Indica hybrid	3.589259 (0.097939)	36.648 (469.1)	<.001
Sativa	2.916945 (0.163117)	17.883 (86.4)	<.001
Sativa hybrid	3.470149 (0.171074)	20.285 (92.0)	<.001

[a]CBD: cannabidiol.

[b]Tests were two-tailed.

Table 9. Efficacy comparisons between strain categories. The efficacy was tested using linear mixed modeling (β coefficient was not standardized).

Strain categories	Estimate (SE)	*t* test[b] (*df*)	*P* value
Balanced hybrid versus CBD[a]	0.387332 (0.125752)	3.080 (167.3)	.03
Indica versus balanced hybrid	0.200067 (0.107863)	1.855 (208.6)	.98
Indica hybrid versus balanced hybrid	0.127900 (0.104839)	1.220 (128.9)	.99
Balanced hybrid versus sativa	0.544414 (0.170767)	3.188 (72.0)	.03
Sativa hybrid versus balanced hybrid	0.008790 (0.181109)	0.049 (75.1)	.99
Indica versus CBD	0.587400 (0.114173)	5.145 (216.2)	<.001
Indica hybrid versus CBD	0.515232 (0.115805)	4.450 (197.1)	<.001
CBD versus sativa	0.157082 (0.163729)	0.959 (58.6)	.99
Sativa hybrid versus CBD	0.396122 (0.181141)	2.187 (82.9)	.48
Indica versus indica hybrid	0.072168 (0.083424)	0.865 (183.5)	.99
Indica versus sativa	0.744481 (0.159664)	4.663 (69.6)	<.001
Indica versus sativa hybrid	0.191277 (0.162713)	1.176 (68.4)	.99
Indica hybrid versus sativa	0.672314 (0.164905)	4.077 (73.6)	<.001
Indica hybrid versus sativa hybrid	0.119110 (0.166813)	0.714 (82.5)	.99
Sativa hybrid versus sativa	0.553204 (0.201776)	2.742 (65.5)	.12

[a]CBD: cannabidiol.

[b]Tests were two-tailed.

Discussion

Principal Findings

Results from this large naturalistic sample of medicinal cannabis users who tracked their insomnia symptoms before and after cannabis use suggest significant improvements in insomnia symptoms, with no gender differences in perceived efficacy. Notably, this study uses a naturalistic design by analyzing crowdsourced data from a medicinal cannabis–tracking mobile app. With increasing advances in technology, this study presents a unique perspective on a health management self-monitoring tool that examines data on a population scale.

Analyses of product forms and ingestion methods found that cannabis was most often used in the form of flowers or oils and most often ingested via vapes, oils, or smoking. In addition, although all strains were reported to be beneficial for the management of insomnia, predominant indica and indica hybrid strains were found to be more efficacious than CBD and predominant sativa strains. This finding is in contrast with those of a previous study reporting that strains with significantly higher concentrations of CBD were generally preferred by individuals using cannabis to treat symptoms of insomnia [21]. Despite this, our findings are in line with results from previous studies that have reported indica and hybrid strains to be among the most frequently used strains for insomnia [16]. This same study reported that the most used strains were fairly high in THC content and were combined with high to moderate CBD content.

Another study investigating multiple doses of cannabinoids for sleep reported that administration of both 5 mg/5 mg and 15 mg/15 mg of THC/CBD demonstrated a decrease in stage 3 sleep when compared with placebo, with the higher dose also showing increased states of wakefulness [22]. THC administration on its own demonstrated no significant changes to sleep architecture from placebo; however, the same study found that high doses of THC alone or in combination with CBD resulted in increased subjective sleepiness [22]. From this, the researchers concluded that CBD may have dose-dependent effects on alertness and that the activating and sedating properties of CBD and THC, respectively, could work together to induce sleep and counteract daytime sleepiness [22]. Although few clinical trials have objectively analyzed cannabinoids for sleep with sleep outcomes as primary measures, some preliminary trials have shown that administration of THC and THC-derivatives, alone or in combination with CBD, were associated with subjective improvement in sleep outcomes [11-13]. In addition, previous studies examining strain preferences have also reported increased preferences toward indica strains for sleep [14,23,24]. In one study, indica was preferred for sedation and sleep, whereas sativa was preferred to increase energy [14]. Another study investigating qualitative responses reported that patients using medicinal cannabis preferred using indica at night to improve sleep [24]. In essence, to better understand the efficacy of cannabinoids for insomnia, randomized placebo-controlled studies are needed.

The human endocannabinoid (eCB) system has been increasingly implicated in body and brain homeostasis, including sleep. For instance, the eCB system is thought to play an active role in regulatory processes, such as pain perception, memory, and sleep modulation [25,26]. Although the neurobiological basis of cannabis for sleep is still being understood, overlaps between the neuronal circuitry of sleep and wake states and the eCB system suggest that cannabinoids can contribute to

sleep-related mechanisms and physiology [27-29]. Therefore, the eCB system has become a growing target in sleep research [25-30]. Despite the perceived benefits of cannabinoids, there remains a lack of placebo-controlled trials that have examined the effects of the drug using validated sleep measures or objective sleep outcomes [11-13]. In addition, the current literature on the existence of potential risks, harms, and side effects associated with cannabinoid treatments remain extremely sparse for sleep disorders; however, there is growing evidence that suggests an increased risk of both acute and chronic cognitive impairments [31,32]. Although these risks are poorly understood, research suggests that the prevalence of these effects is increasing [31]. Future clinical trials should focus on the benefits and potential harms through the use of validated objective and subjective measures. Because of the highly comorbid nature of insomnia and other sleep disorders, additional variables such as medication interactions, potential side effects, and comorbid diagnoses are also worth investigating.

Limitations

Some limitations should be considered when interpreting the results of this study. First, individual conditions and symptoms were subjectively reported by users on the Strainprint app. Therefore, it is unknown whether subjects will meet the full criteria for insomnia or any other sleep-related disorder. Moreover, individual user data are restricted to the information collected by Strainprint; therefore, additional information that may affect cannabinoid efficacy (eg, medical history, body size, other concurrent medications, or tolerance) could not be assessed. Another limitation of this study is the lack of a placebo control group. Because data were collected from a sample of medicinal cannabis users, it is possible that individual expectations of cannabinoid efficacy may have attributed to positive postmedication ratings. In other words, the large magnitude effect-size observed reflects pharmacological effects and response expectancy (placebo) effects, and the proportionate contribution of each, fundamentally, cannot be ascertained. It is also possible that this study examined the data of individuals who were more likely to find cannabis to be effective, as the Strainprint app is geared toward individuals who wish to improve therapeutic outcomes by tracking their cannabis use. As a result, the sample may disproportionately represent users who benefit from using cannabis. In addition, the Strainprint app primarily collects data on cannabis use and has very limited data on its potential side effects. Therefore, beyond perceived efficacy, it was not possible to ascertain from the available data whether users experienced any negative side effects from cannabis use.

This study also examined various strain categories; however, distinctions between these strains remain the subject of much debate [33-35]. Cannabis has historically been classified into two separate species (*C.sativa* and *C. indica*) with distinct biological effects. However, years of breeding and hybridization have rendered potential distinctions often meaningless [34-37]. As recreational cannabis use has become increasingly popular, commercialization of the plant has led to the emergence of products marketed as derivatives or hybrids of these species [23,35,36]. Among consumers, the terms *sativa*, *indica*, and *hybrid* are used colloquially and are associated with perceived effects [37]. Sativa has been associated with stimulating effects, indica with sedating effects, and hybrids are perceived to be bred from the former two to fit the more personalized needs of consumers [23,24,35,37]. Interestingly, a recent study collected data characterizing various commercial products classified as *sativa*, *indica*, or *hybrid* and used supervised and unsupervised machine learning algorithms to subjective effect tags [36]. The models indicated a clear division among *sativas* and *indicas*, with *hybrids* in between, suggesting distinct subjective effects among the categories [36].

Despite these perceived effects, strain categories are largely baseless [34-37]. Instead, many researchers hold that the differences in perceived effects between strain categories may be owing to other components of cannabis (ie, terpenes), which are rarely accurately reported to consumers [34,36,38,39]. Interestingly, some studies have found that products labeled as *indica* and *sativa* have similar concentrations of major cannabinoids but distinctly different concentrations of terpenes [23,40,41]. To date, hundreds of different cannabinoids and terpenes have been identified, all with varying pharmacological properties and outcomes [38,42]. These cannabinoids and terpenes are also known to interact synergistically with one another to exert *entourage effects*, which can have enhanced therapeutic benefits for consumers [38,39,42-44]. Given this, it remains unclear whether the perceived therapeutic effects are a result of the individual components of cannabis products or the combined effects of interacting cannabinoids and terpenes. Although strain categories are largely arbitrary, many researchers continue to examine their perceived effects to better understand consumer choices [16,24,44-47]. The nature of our data allowed us to do the same, providing insight into the naturalistic setting of cannabis use. In essence, although the analysis of strain categories in this study provides valuable research on how efficacious various strains are perceived to be, it is worth noting that the perceived efficacy and differences between strains may be driven, at least in part, by self-selection and placebo effects.

Furthermore, previous studies have found inconsistencies between product labels and content, as well as differences in cannabinoid content reporting among labs [48,49]. With recreational cannabis products, the accuracy of product labels relies heavily on growers, suppliers, and dispensaries; however, there are currently no standardized procedures or reliable methods for verifying strains or cannabis content in commercialized products [35]. Despite this, consumers greatly rely on product labels for information on the cannabis content of a product, often using these labels to communicate preferences for desired effects [24,35,36]. One study even reported that demand for *indica* and *sativa* products was similar, with hypothetical purchasing tasks suggesting that consumer decisions were determined by the perceived effects of each strain in the context or setting of the typical activity-based purpose [23]. Unfortunately, because the Strainprint app prepopulates product data from multiple sources, variability across products is an issue, and we were unable to measure the accuracy of cannabinoid content for each product.

For addressing the limitations discussed above, similar future studies should investigate the effects of various terpenes and cannabinoids on perceived efficacy. Previous research has also suggested addressing strain variability by classifying cannabis products according to chemical phenotypes and pharmacological characteristics [35,43,50]. A necessary next step toward accurately classifying cannabis subgroups and creating more precise product labels for consumers is a better understanding of the association between the chemical composition of individual products and the perceived effects experienced by cannabis users. As the colloquial use of strain categories is likely to persist in the commercial marketplace, it is also necessary that future studies attempt to genetically profile samples of commercialized cannabis products, such that genotypes of the same strain are at least comparable. In addition, randomized placebo-controlled trials are necessary to ultimately test the efficacy and safety of cannabis-based treatments for insomnia.

Despite these limitations, this study is strengthened by its ecological validity, as data were obtained from a large naturalistic registry of medicinal cannabis users who prospectively tracked changes in their insomnia symptoms before and after cannabis use. The results of this study can help in designing future clinical trials to ultimately test the efficacy and safety profile of different cannabinoids in the management of insomnia.

Conclusions

The results of this study suggest that individuals using medicinal cannabis to manage insomnia symptoms report significant symptom reduction after use. This general perceived improvement in insomnia symptoms highlights the potential for cannabis to be used as a treatment option for sleep disorders. Future research should investigate the benefits and harms of cannabinoids for insomnia through rigorous randomized placebo-controlled trials.

Conflicts of Interest

JM is a principal and senior scientist in BEAM Diagnostics, Inc, and a consultant to Clairvoyant Therapeutics, Inc. This study was also funded by a Michael G DeGroote Center for Medicinal Cannabis Research Graduate Scholarship (NK). The other authors have no conflicts of interest to declare.

Multimedia Appendix 1
Supplementary material.
[DOCX File , 29 KB - jmir_v23i10e25730_app1.docx]

References

1. Ko GD, Bober SL, Mindra S, Moreau JM. Medical cannabis - the Canadian perspective. J Pain Res 2016;9:735-744. [doi: 10.2147/JPR.S98182] [Medline: 27757048]
2. Turna J, Balodis I, Munn C, Van Ameringen M, Busse J, MacKillop J. Overlapping patterns of recreational and medical cannabis use in a large community sample of cannabis users. Compr Psychiatry 2020 Oct;102:152188 [FREE Full text] [doi: 10.1016/j.comppsych.2020.152188] [Medline: 32653594]
3. Rotermann M. Analysis of trends in the prevalence of cannabis use and related metrics in Canada. Health Rep 2019 Jun 19;30(6):3-13 [FREE Full text] [doi: 10.25318/82-003-x201900600001-eng] [Medline: 31216047]
4. Walsh Z, Callaway R, Belle-Isle L, Capler R, Kay R, Lucas P, et al. Cannabis for therapeutic purposes: patient characteristics, access, and reasons for use. Int J Drug Policy 2013 Nov;24(6):511-516. [doi: 10.1016/j.drugpo.2013.08.010] [Medline: 24095000]
5. Ferrie JE, Kumari M, Salo P, Singh-Manoux A, Kivimäki M. Sleep epidemiology— a rapidly growing field. Int J Epidemiol 2011 Dec;40(6):1431-1437 [FREE Full text] [doi: 10.1093/ije/dyr203] [Medline: 22158659]
6. Morin CM, Benca R. Chronic insomnia. Lancet 2012 Mar 24;379(9821):1129-1141. [doi: 10.1016/S0140-6736(11)60750-2] [Medline: 22265700]
7. Levenson JC, Kay DB, Buysse DJ. The pathophysiology of insomnia. Chest 2015 Apr;147(4):1179-1192 [FREE Full text] [doi: 10.1378/chest.14-1617] [Medline: 25846534]
8. Ohayon MM. Epidemiology of insomnia: what we know and what we still need to learn. Sleep Med Rev 2002 Apr;6(2):97-111. [doi: 10.1053/smrv.2002.0186] [Medline: 12531146]
9. Buysse DJ, Angst J, Gamma A, Ajdacic V, Eich D, Rössler W. Prevalence, course, and comorbidity of insomnia and depression in young adults. Sleep 2008 Apr;31(4):473-480 [FREE Full text] [doi: 10.1093/sleep/31.4.473] [Medline: 18457234]
10. Chen P, Huang CL, Weng S, Wu M, Ho C, Wang J, et al. Relapse insomnia increases greater risk of anxiety and depression: evidence from a population-based 4-year cohort study. Sleep Med 2017 Oct;38:122-129. [doi: 10.1016/j.sleep.2017.07.016] [Medline: 29031746]
11. Babson KA, Sottile J, Morabito D. Cannabis, cannabinoids, and sleep: a review of the literature. Curr Psychiatry Rep 2017 Apr;19(4):23. [doi: 10.1007/s11920-017-0775-9] [Medline: 28349316]
12. Gates PJ, Albertella L, Copeland J. The effects of cannabinoid administration on sleep: a systematic review of human studies. Sleep Med Rev 2014 Dec;18(6):477-487. [doi: 10.1016/j.smrv.2014.02.005] [Medline: 24726015]

13. Kuhathasan N, Dufort A, MacKillop J, Gottschalk R, Minuzzi L, Frey BN. The use of cannabinoids for sleep: a critical review on clinical trials. Exp Clin Psychopharmacol 2019 Aug;27(4):383-401. [doi: 10.1037/pha0000285] [Medline: 31120284]
14. Pearce DD, Mitsouras K, Irizarry KJ. Discriminating the effects of Cannabis sativa and Cannabis indica: a web survey of medical cannabis users. J Altern Complement Med 2014 Oct;20(10):787-791. [doi: 10.1089/acm.2013.0190] [Medline: 25191852]
15. Stith SS, Vigil JM, Brockelman F, Keeling K, Hall B. Patient-reported symptom relief following medical cannabis consumption. Front Pharmacol 2018;9:916 [FREE Full text] [doi: 10.3389/fphar.2018.00916] [Medline: 30210337]
16. Vigil JM, Stith SS, Diviant JP, Brockelman F, Keeling K, Hall B. Effectiveness of raw, natural medical flower for treating insomnia under naturalistic conditions. Medicines (Basel) 2018 Jul 11;5(3):75 [FREE Full text] [doi: 10.3390/medicines5030075] [Medline: 29997343]
17. Rhodes E, Andreae M, Bourgiose T, Indyk D, Rhodes R, Sacks H. Stakeholders' views on barriers to research on controversial controlled substances. J Clin Ethics 2016;27(4):308-321 [FREE Full text] [doi: 10.3109/15360288.2010.529988] [Medline: 28001138]
18. Mead A. Legal and regulatory issues governing cannabis and cannabis-derived products in the United States. Front Plant Sci 2019;10:697 [FREE Full text] [doi: 10.3389/fpls.2019.00697] [Medline: 31263468]
19. Shen H. Federal red tape ties up marijuana research. Nature 2014 Mar 27;507(7493):407-408. [doi: 10.1038/507407a] [Medline: 24670737]
20. Stith SS, Vigil JM. Federal barriers to Cannabis research. Science 2016 Jun 03;352(6290):1182. [doi: 10.1126/science.aaf7450] [Medline: 27257247]
21. Belendiuk KA, Babson KA, Vandrey R, Bonn-Miller MO. Cannabis species and cannabinoid concentration preference among sleep-disturbed medicinal cannabis users. Addict Behav 2015 Nov;50:178-181. [doi: 10.1016/j.addbeh.2015.06.032] [Medline: 26151582]
22. Nicholson AN, Turner C, Stone BM, Robson PJ. Effect of Delta-9-tetrahydrocannabinol and cannabidiol on nocturnal sleep and early-morning behavior in young adults. J Clin Psychopharmacol 2004 Jun;24(3):305-313. [doi: 10.1097/01.jcp.0000125688.05091.8f] [Medline: 15118485]
23. Sholler DJ, Moran MB, Dolan SB, Borodovsky JT, Alonso F, Vandrey R, et al. Use patterns, beliefs, experiences, and behavioral economic demand of indica and sativa cannabis: A cross-sectional survey of cannabis users. Exp Clin Psychopharmacol 2021 Apr 15:A. [doi: 10.1037/pha0000462] [Medline: 33856822]
24. Piper BJ. Mother of Berries, ACDC, or Chocolope: Examination of the strains used by medical cannabis patients in New England. J Psychoactive Drugs 2018;50(2):95-104 [FREE Full text] [doi: 10.1080/02791072.2017.1390179] [Medline: 29064777]
25. Stasiulewicz A, Znajdek K, Grudzień M, Pawiński T, Sulkowska AJ. A guide to targeting the endocannabinoid system in drug design. Int J Mol Sci 2020 Apr 16;21(8):2778 [FREE Full text] [doi: 10.3390/ijms21082778] [Medline: 32316328]
26. Pacher P, Bátkai S, Kunos G. The endocannabinoid system as an emerging target of pharmacotherapy. Pharmacol Rev 2006 Sep;58(3):389-462 [FREE Full text] [doi: 10.1124/pr.58.3.2] [Medline: 16968947]
27. Murillo-Rodríguez E. The role of the CB1 receptor in the regulation of sleep. Prog Neuropsychopharmacol Biol Psychiatry 2008 Aug 01;32(6):1420-1427. [doi: 10.1016/j.pnpbp.2008.04.008] [Medline: 18514375]
28. Murillo-Rodriguez E, Poot-Ake A, Arias-Carrion O, Pacheco-Pantoja E, Fuente-Ortegon AD, Arankowsky-Sandoval G. The emerging role of the endocannabinoid system in the sleep-wake cycle modulation. Cent Nerv Syst Agents Med Chem 2011 Sep 01;11(3):189-196. [doi: 10.2174/187152411798047780] [Medline: 21919868]
29. Kesner AJ, Lovinger DM. Cannabinoids, endocannabinoids and sleep. Front Mol Neurosci 2020;13:125 [FREE Full text] [doi: 10.3389/fnmol.2020.00125] [Medline: 32774241]
30. Di Marzo V. New approaches and challenges to targeting the endocannabinoid system. Nat Rev Drug Discov 2018 Sep;17(9):623-639. [doi: 10.1038/nrd.2018.115] [Medline: 30116049]
31. MacKillop J. Priorities for a comprehensive post-legalization cannabis research agenda in Canada. Can J Addict 2019;10(3):67-71. [doi: 10.1097/CXA.0000000000000057]
32. Crean RD, Crane NA, Mason BJ. An evidence based review of acute and long-term effects of cannabis use on executive cognitive functions. J Addict Med 2011 Mar;5(1):1-8 [FREE Full text] [doi: 10.1097/ADM.0b013e31820c23fa] [Medline: 21321675]
33. Sawler J, Stout JM, Gardner KM, Hudson D, Vidmar J, Butler L, et al. The genetic structure of marijuana and hemp. PLoS One 2015;10(8):e0133292 [FREE Full text] [doi: 10.1371/journal.pone.0133292] [Medline: 26308334]
34. Piomelli D, Russo EB. The Cannabis sativa Versus Cannabis indica Debate: An interview with Ethan Russo, MD. Cannabis Cannabinoid Res 2016;1(1):44-46 [FREE Full text] [doi: 10.1089/can.2015.29003.ebr] [Medline: 28861479]
35. Schwabe AL, McGlaughlin ME. Genetic tools weed out misconceptions of strain reliability in Cannabis sativa: implications for a budding industry. J Cannabis Res 2019 Jun 07;1(1):3 [FREE Full text] [doi: 10.1186/s42238-019-0001-1] [Medline: 33526091]

36. de la Fuente A, Zamberlan F, Ferrán A, Carrillo F, Tagliazucchi E, Pallavicini C. Relationship among subjective responses, flavor, and chemical composition across more than 800 commercial cannabis varieties. J Cannabis Res 2020 Jul 17;2(1):21 [FREE Full text] [doi: 10.1186/s42238-020-00028-y] [Medline: 33526118]
37. McPartland JM. Cannabis systematics at the levels of family, genus, and species. Cannabis Cannabinoid Res 2018;3(1):203-212 [FREE Full text] [doi: 10.1089/can.2018.0039] [Medline: 30426073]
38. Stasiłowicz A, Tomala A, Podolak I, Cielecka-Piontek J. Cannabis sativa L. as a natural drug meeting the criteria of a multitarget approach to treatment. Int J Mol Sci 2021 Jan 14;22(2):778 [FREE Full text] [doi: 10.3390/ijms22020778] [Medline: 33466734]
39. Booth JK, Bohlmann J. Terpenes in Cannabis sativa - From plant genome to humans. Plant Sci 2019 Jul;284:67-72 [FREE Full text] [doi: 10.1016/j.plantsci.2019.03.022] [Medline: 31084880]
40. Russo EB. Taming THC: potential cannabis synergy and phytocannabinoid-terpenoid entourage effects. Br J Pharmacol 2011 Aug;163(7):1344-1364 [FREE Full text] [doi: 10.1111/j.1476-5381.2011.01238.x] [Medline: 21749363]
41. Hazekamp A, Tejkalová K, Papadimitriou S. Cannabis: From Cultivar to Chemovar II—A metabolomics approach to cannabis classification. Cannabis Cannabinoid Res 2016 Dec;1(1):202-215. [doi: 10.1089/can.2016.0017]
42. Baron EP, Lucas P, Eades J, Hogue O. Patterns of medicinal cannabis use, strain analysis, and substitution effect among patients with migraine, headache, arthritis, and chronic pain in a medicinal cannabis cohort. J Headache Pain 2018 May 24;19(1):37 [FREE Full text] [doi: 10.1186/s10194-018-0862-2] [Medline: 29797104]
43. Russo EB. The case for the entourage effect and conventional breeding of clinical cannabis: No "Strain," no gain. Front Plant Sci 2018;9:1969 [FREE Full text] [doi: 10.3389/fpls.2018.01969] [Medline: 30687364]
44. Pellati F, Borgonetti V, Brighenti V, Biagi M, Benvenuti S, Corsi L. Cannabis sativa L. and nonpsychoactive cannabinoids: their chemistry and role against oxidative stress, inflammation, and cancer. Biomed Res Int 2018;2018:1691428 [FREE Full text] [doi: 10.1155/2018/1691428] [Medline: 30627539]
45. Stith SS, Diviant JP, Brockelman F, Keeling K, Hall B, Lucern S, et al. Alleviative effects of Cannabis flower on migraine and headache. J Integr Med 2020 Sep;18(5):416-424. [doi: 10.1016/j.joim.2020.07.004] [Medline: 32758396]
46. Li X, Diviant JP, Stith SS, Brockelman F, Keeling K, Hall B, et al. The effectiveness of cannabis flower for immediate relief from symptoms of depression. Yale J Biol Med 2020 Jun;93(2):251-264 [FREE Full text] [Medline: 32607086]
47. Stith SS, Vigil JM, Brockelman F, Keeling K, Hall B. The association between cannabis product characteristics and symptom relief. Sci Rep 2019 Feb 25;9(1):2712 [FREE Full text] [doi: 10.1038/s41598-019-39462-1] [Medline: 30804402]
48. Vandrey R, Raber JC, Raber ME, Douglass B, Miller C, Bonn-Miller MO. Cannabinoid dose and label accuracy in edible medical cannabis products. J Am Med Assoc 2015;313(24):2491-2493. [doi: 10.1001/jama.2015.6613] [Medline: 26103034]
49. Jikomes N, Zoorob M. The cannabinoid content of legal cannabis in washington state varies systematically across testing facilities and popular consumer products. Sci Rep 2018 Mar 14;8(1):4519 [FREE Full text] [doi: 10.1038/s41598-018-22755-2] [Medline: 29540728]
50. Schilling S, Melzer R, McCabe PF. Cannabis sativa. Curr Biol 2020 Jan 06;30(1):8-9 [FREE Full text] [doi: 10.1016/j.cub.2019.10.039] [Medline: 31910378]

Abbreviations

CBD: cannabidiol
eCB: endocannabinoid
THC: tetrahydrocannabinol

Edited by R Kukafka; submitted 12.11.20; peer-reviewed by J Thrul, C Cuttler; comments to author 15.03.21; revised version received 10.05.21; accepted 27.07.21; published 27.10.21.

Please cite as:
Kuhathasan N, Minuzzi L, MacKillop J, Frey BN
The Use of Cannabinoids for Insomnia in Daily Life: Naturalistic Study
J Med Internet Res 2021;23(10):e25730
URL: https://www.jmir.org/2021/10/e25730
doi:10.2196/25730
PMID:34704957

Original Paper

Young People's Experiences of Engaging With Fitspiration on Instagram: Gendered Perspective

Joanne Mayoh[1], BSc, PhD; Ian Jones[1], BSc, MSc, PhD

Department of Sport and Event Management, Bournemouth University, Poole, United Kingdom

Corresponding Author:
Joanne Mayoh, BSc, PhD
Department of Sport and Event Management
Bournemouth University
Fern Barrow
Poole, BH12 5BB
United Kingdom
Phone: 44 1202968385
Email: jmayoh@bournemouth.ac.uk

Abstract

Background: Fitness inspiration or *fitspiration* is a term used to describe web-based images of fit people, people in the gym, health foods, or inspirational quotes relating to diet and fitness being shared and consumed via visual social media. The popularity of this content is most notable via the *Instagram* platform. Currently, the majority of fitspiration research has focused on women's experiences; however, increasingly, studies have pointed to the need to explore the gendered ways by which people engage with this content.

Objective: The aim of this study is to explore how young men and women engage in fitspiration content on Instagram and provide a gendered analysis of how and why they consume this content.

Methods: This study used a cross-sectional web-based survey (N=1213) of UK-based fitspiration users aged 18-24 years consisting of closed-ended questions to capture quantitative data.

Results: The majority actively using Instagram for fitspiration (therefore eligible participants) were women (826/1175, 70.30%). Men were more likely to view content posted by athletes ($\chi^2_{1, N=1153}$=71.8; *P*=.001) and bodybuilders ($\chi^2_{1, N=1153}$=32.8; *P*<.001), whereas women were more likely to view content related to weight loss ($\chi^2_{1, N=1153}$=36.8; *P*<.001), diet plans ($\chi^2_{1, N=1153}$=11.9; *P*<.001), and celebrities' content ($\chi^2_{1, N=1153}$=33.5; *P*<.001). Men were more likely to use fitspiration as a source of inspiration to exercise to gain muscle or get stronger ($\chi^2_{1, N=1147}$=17.9; *P*<.001), whereas women were more likely to use fitspiration as inspiration for healthy eating ($\chi^2_{1, N=1147}$=37.7; *P*<.001), or to exercise to diet or lose weight ($\chi^2_{1, N=1147}$=13.5; *P*<.001). Women were more likely to engage in passive behaviors such as viewing content on their feed ($\chi^2_{1, N=1139}$=7.9; *P*=.005) or scrolling through accounts ($\chi^2_{1, N=1139}$=15.2; *P*<.001), whereas men were more likely to engage in active consumption by tagging fitspiration accounts in posts ($\chi^2_{1, N=1139}$=7.2; *P*=.007), commenting on posts ($\chi^2_{1, N=1139}$=8.1; *P*=.004), and posting fitspiration content ($\chi^2_{1, N=1139}$=6.4; *P*=.01).

Conclusions: Female fitspiration consumers engaged with content that reinforced the feminine *thin but shapely* ideal, whereas male users sought out content that reinforced the masculine *muscular* ideal. Male users were more likely to engage actively with content (eg, posting fitspiration content), while female users were more likely to engage passively (eg, scrolling through accounts, posts, or images). Future research should consider how fitspiration consumption reflects and reproduces oppressive gender ideology.

(J Med Internet Res 2021;23(10):e17811) doi:10.2196/17811

KEYWORDS

social media; gender; physical fitness; women's health; men's health; body ideals

Introduction

Background

Social media is becoming embedded within our everyday lives, with 68% of people and 98% of those aged 16-24 years within the United Kingdom using web-based social networking platforms [1]. Among their many purposes, web-based social media platforms are used for communication, community building, information gathering, and as a source of inspiration within multiple contexts. One such context is that of *fitspiration* or *fitspo*, commonly used abbreviations for the term *fitness inspiration* [2], which describes web-based content such as images of fit people, people in the gym, health foods, or inspirational quotes relating to diet and fitness shared and consumed habitually via social media. Increasingly, academic research has demonstrated numerous perceived positive outcomes associated with consuming the fitspiration content on social media [2-5]. These studies explore how web-based fitspiration can provide wider access to reliable health information [2-5] and motivate individuals to engage in healthy behaviors [3], including by providing motivation to exercise [4] and in helping to create a sense of community [2].

While fitspiration is widespread on multitudinous social media platforms, it is particularly common on Instagram [6]. Instagram is a predominantly mobile app, with over 400 million global users [7] including 29% of women and 26% of men in the United Kingdom [8]. With regard to fitspiration, a recent Instagram search (July 11, 2019) of #fitspo yielded over 66 million posts [9] demonstrating the popularity of this specific content. A key reason for complementarity between this visual platform and fitspiration is that Instagram has a unique focus on sharing and consuming *snapshot photography* or self-selected images in which individuals are portrayed in an attractive light [10] making it particularly harmonious with imagery that showcases fit and healthy *ideal* bodies and lives, reinforcing social norms regarding health and fitness. Furthermore, Instagram offers users the opportunity to edit and apply filters to images to distort reality into an ideal to be shared, making the potential impact on well-being and appearance concerns more pronounced when compared with other platforms [11]. It is therefore imperative that research focuses on this unique platform to adequately inform policy regarding the health and well-being of young adults (<24 years).

Users of Instagram are exposed to multiple *idealized* images of peers and other people or organizations (with public accounts) actively shaping visual ideas about beauty regarding *acceptable* or *desirable* bodies [12]. Those who engage with fitspiration on Instagram specifically (eg, by following or scrolling through specific accounts or searching for hashtags) are exposed to idealized *fit* bodies and unsubstantiated, often unrealistic (and occasionally incorrect) advice regarding diet and fitness, shaping the body's ideal narrative and reinforcing social norms regarding the body. Research exploring the nature of fitspiration on social media and Instagram has demonstrated that the content displays multiple potentially harmful themes and includes objectifying images that depict idealized bodies [2-4], can depict extreme and often restrictive healthy-eating content linked to disordered eating [13], and can promote increased symptoms of orthorexia nervosa [14], a term coined to describe an obsession with proper or *healthful* eating [15]. Fitspiration content also promotes exercising for appearance-related reasons [16], which has been associated with negative body image [17], higher measures of disordered eating [18], and positive correlations with higher depressive and erectile dysfunction symptoms [19]. The aforementioned research often draws upon theoretical knowledge such as objectification [6,11] and social comparison theories [4,20] to explore the potentially negative implications of consuming fitspiration content on both mental and physical health.

The paradoxical findings on the relationship between fitspiration consumption and well-being reflect a small yet emerging body of research exploring the impact of fitspiration content on the lives and experiences of consumers. Despite this, research conducted on fitspiration to date focuses on the nature of the fitspiration content as opposed to the individuals that consume it [6,16,21]. While exploration into fitspiration content has valuable implications for policy, researchers have called for further inquiry exploring how individuals use fitspiration and the role it plays in their lives [2] to better support those who value it as meaningful content.

Currently, the majority of research exploring fitspiration focuses specifically on women's experiences [11,22,23] and the nature of woman-centered content [4,24]. This is because women are more likely to use visual social media platforms [25,26] and are the biggest consumers of fitspiration images [21], and because fitspiration images are more likely to be women, with (36.4%) posts depicting female participants compared with 27.5% of posts depicting a male participant [21]. Associated findings demonstrate that exposure to fitspiration on Instagram may negatively influence women's appearance-related concerns [11] and promote an often unrealistic and unachievable fit or thin body ideal [4,23,24]. These findings reflect broader cultural and media trends surrounding the ideal female body of one that is simultaneously very thin, and exceptionally fit and toned without displaying excess muscularity [16,27]. While fitspiration may be considered a healthy alternative to *thinspiration*, which communicates an unhealthy thin body ideal [24], the fitspiration ideal communicated to women within these contexts could be seen as oppressive, narrow, and contradictory and can promote women to engage in extreme exercise behavior [13]. There is also an emergent body of academic research that focuses on men's experiences with fitspiration and the content that they are exposed to [21,28,29]. This research is valuable, as data suggest that a significant minority (30%) of fitspiration images on social media targets men [4,21]. Reviews of fitspiration content aimed at men reflect a heavily gendered muscular ideal [21,29] that provides a stark contrast to the message communicated to women through cultural discourse. Palmer [28] recognizes the pervasive nature of this form of social media content, demonstrating that it has the ability to influence men's ideas regarding masculinity, and can encourage users to compare physiques to dominant *ideals*. This dominant discourse regarding male body ideology depicts participants who are highly muscular and nearly always have visible abdominal, bicep, and pectoral muscles, and compared with women these ideals demonstrate

less emphasis on thinness or weight loss [21]. Therefore, the culturally pervasive ideological body identified in fitspiration targeting men reflects an ongoing cultural trend in the media, which has demonstrated a rapid increase in male muscle mass over time [30], aggregating the pressure on men to conform to increasingly unrealistic and unachievable body ideals.

Gendered analyses have also shown significant differences in the nature of the fitspiration images depicting men and women's bodies, with women being more likely to be objectified and sexualized, and images with a greater focus on the participant's buttocks [21], whereas posts of men were more likely to show the participant's face [4]. Evidence also suggests that male fitspiration content on Instagram is more likely to focus on fitness as opposed to diet and alternative weight loss methods, with Tiggemann and Zaccardo [4] noting that 68.2% of male fitspiration focused on fitness-related activities, compared with only 58% for women. There are also clear gendered differences in how genders interact with and experience fitspiration content. In their study exploring male fitspiration use on Instagram, Fatt et al [29] identified that the frequency of viewing fitspiration posts was not significantly correlated with body satisfaction, appearance-based exercise motivation, and health-based exercise motivation in men. These results contrast with previous findings that demonstrated positive correlations between viewing frequency, body dissatisfaction, and disordered eating for women, signifying that fitspiration may psychologically impact men and women differently [11,13]. However, researchers have identified that fitspiration content consumption is linked to less body satisfaction when men internalize the muscular ideal and make more appearance comparisons [29]; therefore, certain men may experience the negative impacts of fitspiration in a similar way to women if they internalized male body ideals and made body comparisons.

Objectives

The aim of this study is to explore how young (18-24 years) people engage with fitspiration content on Instagram and provide a gendered analysis of how and why they consume this content. This is in response to the aforementioned research that demonstrates the gendered nature of their consumption experiences in terms of the ideologies communicated through fitspiration, the composition of images, and the ways in which men and women interact with the content. While previous studies provide some light in terms of exploring how fitspiration consumption and the impact of fitspiration on individuals are gendered, because of the popularity and cultural relevance of this form of consumption, there is a great need for further research exploring the motivations for engagement, patterns of consumption, and perceived impact of fitspiration from a gendered perspective. This research will allow for greater exploration into how and why this experience is gendered, which can be used to help support digital literacy training and support for consumers of this content.

Methods

Study Design

The research used a cross-sectional web-based survey consisting of closed-ended questions to capture quantitative data to meet the study's aim of exploring how those aged 18-24 years in the United Kingdom engage with fitspiration on Instagram using a gendered perspective. The survey was informed by the existing literature (outlined above), specifically by a previous survey-based fitspiration research study [2], and took approximately 10 minutes to complete. Ethical approval for the study was granted by an institutional ethical review (Bournemouth University).

Setting

Data were collected during a 1-week period in May 2019 using a web-based insight exchange platform (Cint database) to request a representative sample from appropriate users registered to a range of web-based survey panels. A minimum sample size of 1000 was requested, which was achieved and exceeded, a benefit of using the insight exchange platform. However, it is important to note that the number of participants contacted is unknown owing to the nature of the participant recruitment platform.

Participants

The eligibility criteria were that participants (n=1213) had to be aged 18-24 years (in line with previous research [21,31,32]) and broader demographic data [7,33], lived in the United Kingdom, and had reported viewing fitspiration on Instagram. In line with previous research, for the purposes of this survey, fitspiration was defined as photographs of fit people, people in the gym, health foods, or inspirational quotes relating to diet and fitness. Self-reported eligible participants also completed screening questions to ensure that they met the inclusion criteria (reporting a frequency other than *not at all* for the question "How often do you check Instagram?" or a frequency other than *never* to the question "How often do you view *fitspiration* on Instagram?").

Variables

Tested Variables

The following variables were measured (a copy of the questionnaire is available on request from the authors): demographic characteristics (age, BMI, gender, sexuality, and educational level), use (frequency of use and duration of use), and content and engagement (content engaged with while online, means of engaging with content, and reasons for engagement).

Demographic Characteristics

To allow for adequate description of the sample, a number of demographic characteristics were collected, including gender, age, sexual identity, and highest level of education completed. BMI was calculated using participants' self-reported height and weight.

Instagram Use

To determine how the sample used Instagram, participants indicated how often they checked the platform using a 9-point scale adapted from previous research [11,17,29] (0: *not at all*; 9: every 5 minutes). In addition, participants were asked to specify how long they spent on Instagram on a typical day (*1-5 minutes or less* to *8 hours or more*) on an 8-point scale.

Fitspiration Use

To ascertain how the sample used fitspiration, participants indicated how often they viewed this content (as defined above) using scales adapted from previous literature [11] based on pilot study feedback on Instagram. The final scale had 8 points (1: never to 8: every time).

Fitspiration Content

Participants indicated the types of fitspiration they engaged with on the platform by selecting from a checkbox list informed by previous research [21] listing the following types of content: weight loss or fitness or body transformation journeys, personal trainers, athletes, celebrities or models, bodybuilders or strength content, clean eating, cleanses or detox, diet plans, fitness challenges, influencers, and fitspiration quote pages. Prevalent examples of fitspiration content were included, and participants were able to tick as many options that applied to them as it was anticipated (based on pilot study feedback) that they may have engaged with multiple forms of fitspiration.

Fitspiration Engagement

In line with previous fitspiration research [2] and based on feedback from the pilot study, participants identified how they engaged with fitspiration content by selecting from a list of nine engagement behaviors that ranged from active (post content; comment on posts; tag fitspiration accounts in posts; tag friends in posts; share content with friends; and like posts) to passive (follow accounts or view on Instagram feed, scroll through individual accounts posts or images, and search hashtags). Collecting these data helped address a research gap identified by Raggatt et al [2], who drew on the work of Jenkins-Guarnieri et al [34] to explain that previous studies investigating social media have not adequately captured the different ways in which people interact with social media content.

Reasons for Engaging With Fitspiration Content

Participants' reasons for engaging with fitspiration content were collected using a predefined checkbox list of 8 reasons developed using existing literature [2] and based on feedback from piloting that included inspiration for health and well-being; healthy eating; to exercise to improve body shape, tone, or size; to gain muscle or get stronger; to diet or lose weight; to improve health and well-being; facilitate improvements in appearance; to help learn about health and well-being or because friends or peers view or like it.

Bias

A small incentive was offered in terms of points that participants could accumulate to gain future cash rewards, rather than simply offering cash incentives to try to minimize any subsequent bias. This is designed to discourage *professional* respondents whose primary motivation is to gain payment for completion.

Analysis

Data were analyzed using IBM SPSS for Windows (version 25), with charts produced using Microsoft Excel. Given the nonparametric nature of the data, basic descriptive statistics were used to describe the nature of the sample, and chi-square analyses were undertaken to explore the relationships between gender and Instagram use in terms of the variables outlined earlier.

Results

Description of Sample

Following the screening questions, the survey was completed by 1213 eligible participants with a median age of 21 years. Response rates to individual questions varied and are reported as such. Where there were missing data, responses were deleted. A full breakdown is provided in Table 1; however, to summarize, most (68.10%) participants identified as female, compared with 28.19% identifying as male and 0.57% identifying as nonbinary or gender queer. In terms of sexuality, most of the sample identified as straight (83.47%). BMI calculations showed that over half of the respondents had a BMI within the normal range, 11.10% were categorized as underweight, and almost one-third (32.85%) were overweight or obese. Finally, most were educated to either Advanced or Advanced Subsidiary level (UK qualifications offered across a range of participants to school-leavers; 29.6%) or degree level or higher (33.70%).

Table 1. Description of sample.

Characteristics[a]	Values, n (%)
Gender (n=1213)	
Female	826 (68.1)
Male	342 (28.2)
Nonbinary or genderqueer	7 (0.6)
Sexuality (n=1090)	
Female straight	647 (59.4)
Male straight	263 (24.1)
Female bisexual or pansexual	106 (9.7)
Male bisexual or pansexual	23 (2.1)
Female gay	21 (1.9)
Male gay	15 (1.4)
Other	6 (0.6)
Nonbinary and genderqueer, bisexual, and pansexual	4 (0.4)
Nonbinary and genderqueer and gay	2 (0.2)
Nonbinary and genderqueer and other	1 (0.1)
BMI category (n=1108)	
Underweight	621 (56.0)
Normal	232 (20.9)
Overweight	132 (12.0)
Obese	113 (11.1)
Highest educational qualification (n=1184)	
Undergraduate degree or higher	399 (33.7)
A or AS level	350 (25.3)
GCSE[b] or O level	207 (17.5)
Higher qualification below degree level	181 (15.3)
Postgraduate degree	113 (9.5)
Other	47 (3.97)

[a]Responses are ordered from most to least common.

[b]GCSE: General Certificate of Secondary Education.

Fitspiration Content

With regard to the fitspiration content on Instagram, 49.38% (599/1213) viewed content at least once per day, 32.32% (392/1213) at least once per week, and 18.30% (222/1213) less often. A total of 56.86% (714/1213) reported checking the platform at least once every hour, while 88.13% (1069/1213) checked it at least once every few hours. On average, 67.39% (812/1205) of participants spent an hour or more per day on Instagram with 18.50% (223/1205) spending over 4 hours and 2.99% (36/1205) spending over 8 hours on the platform. A categorical gender comparison between men and women (there were insufficient data to test differences with those who identified as nonbinary or genderqueer for any variables) using a chi-square test of independence demonstrated that women were more likely to check Instagram more regularly than their male counterparts ($\chi^2_{1,\ N=1168}=19.1$; $P=.008$); however, there was no significant difference between the amount of time spent on Instagram per day for men and women ($\chi^2_{1,\ N=1160}=13.8$; $P=.06$).

The most common types of fitspiration viewed by all participants were content from celebrities or models (717/1196, 59.95%), weight loss or fitness journeys or body transformation stories (641/1196, 53.60%), and influencers (473/1196, 39.55%). Gender differences in terms of engaging with specific forms of content are outlined in Table 2. Categorical gender comparisons between men and women using chi-square tests of independence showed male participants were more likely to view content posted by athletes ($\chi^2_{1,\ N=1153}=71.8$; $P<.001$) and bodybuilders ($\chi^2_{1,\ N=1153}=32.82$; $P<.001$) than their female counterparts and women were more likely to view content identified as weight

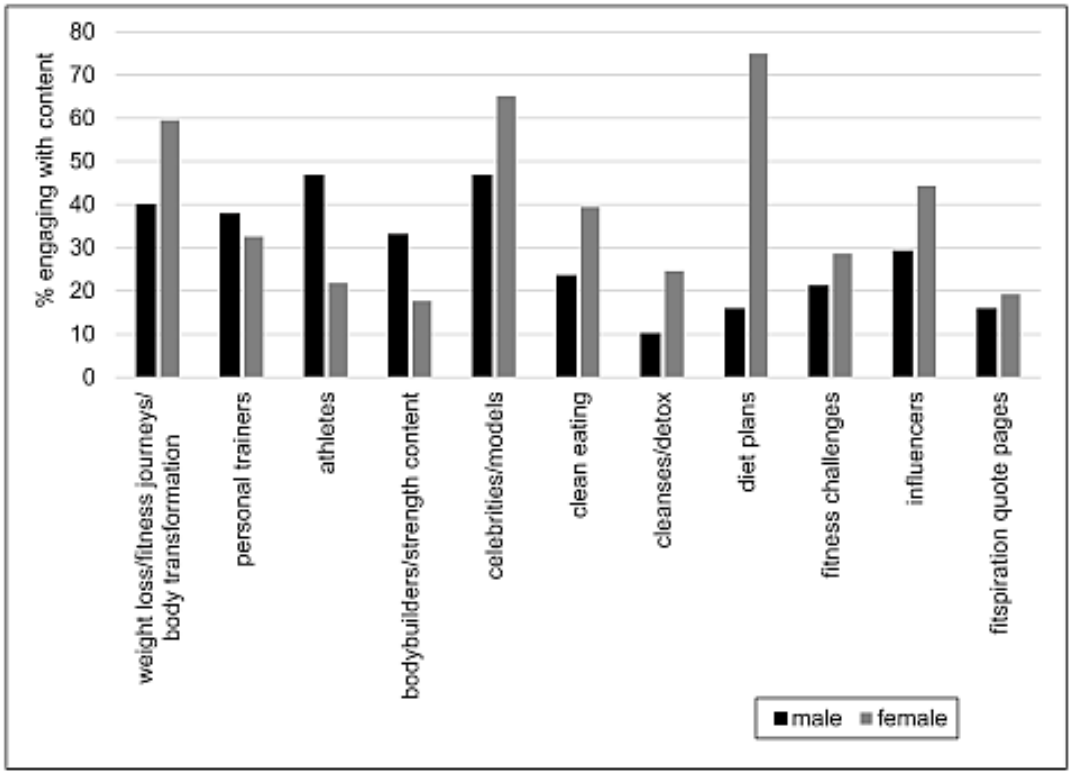
loss or fitness journeys or body transformation stories ($\chi^2_{1, N=1153}$=36.8; *P*<.001), celebrities or models ($\chi^2_{1, N=1153}$=33.5; *P*<.001), clean eating ($\chi^2_{1, N=1153}$=26.5; *P*<.001), cleanses or detox ($\chi^2_{1, N=1153}$=30.8; *P*<.001), diet plans ($\chi^2_{1, N=1153}$=11.9; *P*=.001), fitness challenges ($\chi^2_{1, N=1153}$=6.9; *P*=.009) and from influencers ($\chi^2_{1, N=1153}$=22.8; *P*<.001). The results are detailed in Table 2.

Figure 1 shows the gender and fitspiration content.

Table 2. Gender and fitspiration content (n=1153).

Type of engagement	Total (n=1153), n (%)	Gender, n (%)		*P* value	Φ coefficient
		Male	Female		
Engagement with weight loss or fitness journeys or body transformation content				<.001	−0.179
Do engage	620 (53.8)	134 (39.9)	486 (59.5)		
Do not engage	533 (46.2)	202 (60.1)	331 (40.5)		
Engagement with personal trainers				.09	−0.050
Do engage	393 (34.1)	127 (37.8)	266 (32.6)		
Do not engage	760 (65.9)	209 (62.2)	551 (67.4)		
Engagement with athletes				<.001	−0.250
Do engage	335 (29.1)	157 (46.7)	178 (21.8)		
Do not engage	818 (70.9)	179 (53.3)	639 (78.2)		
Engagement with bodybuilders or strength content				<.001	0.169
Do engage	255 (22.1)	111 (33)	144 (17.6)		
Do not engage	898 (77.9)	225 (67)	673 (82.4)		
Engagement with celebrities or models				<.001	−0.170
Do engage	689 (59.8)	157 (46.7)	532 (65.1)		
Do not engage	464 (40.2)	179 (53.3)	285 (34.9)		
Engagement with clean eating				<.001	−0.152
Do engage	401 (34.8)	79 (23.5)	322 (39.4)		
Do not engage	752 (65.2)	257 (76.5)	495 (60.6)		
Engagement with cleanse or detox				<.001	−0.163
Do engage	235 (20.4)	34 (10.1)	201 (24.6)		
Do not engage	918 (79.6)	302 (89.9)	616 (75.4)		
Engagement with diet plans				.001	−0.102
Do engage	258 (22.4)	53 (15.8)	612 (74.9)		
Do not engage	895 (77.6)	283 (84.2)	205 (25.1)		
Engagement with fitness challenges				.009	−0.077
Do engage	305 (26.5)	71 (21.1)	234 (28.6)		
Do not engage	848 (73.5)	265 (78.9)	583 (71.4)		
Engagement with influencers				<.001	−0.141
Do engage	460 (39.9)	98 (29.2)	362 (44.3)		
Do not engage	693 (60.1)	238 (70.8)	455 (55.7)		
Engagement with fitspiration quote pages				.17	−0.041
Do engage	210 (18.2)	53 (15.8)	157 (19.2)		
Do not engage	943 (81.8)	283 (84.2)	660 (80.8)		

Figure 1. Gender and fitspiration content.



How Participants Engaged With Fitspiration on Instagram

In terms of gender and engagement with fitspiration, a number of differences emerged, as highlighted in Table 3.

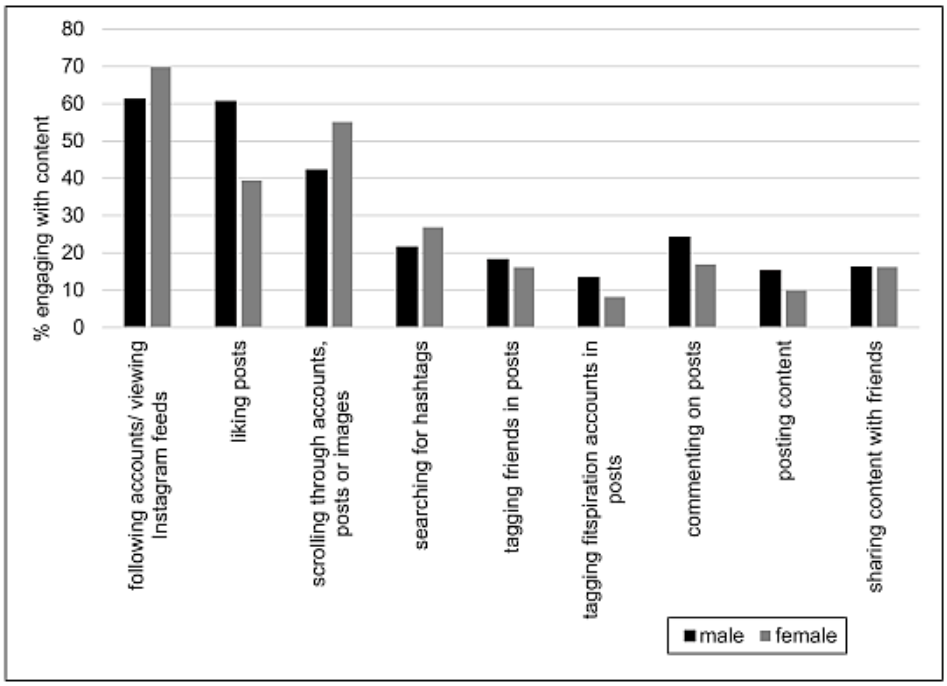
Figure 2 shows gender and fitspiration engagement.

Women were significantly more likely than men to engage in passive consumption behaviors such as follow accounts or views on Instagram feed ($\chi^2_{1,\ N=1139}$=7.9; *P*=.005) or scroll through individual accounts, posts, or images ($\chi^2_{1,\ N=1139}$=15.2; *P*<.001). Conversely, men were more likely than women to engage in active behaviors such as tag fitspiration accounts in posts ($\chi^2_{1,\ N=1139}$=7.2; *P*=.007), comment on posts ($\chi^2_{1,\ N=1139}$=8.1; *P*=.004), and post fitspiration content ($\chi^2_{1,\ N=1139}$=6.4; *P*=.01).

Table 3. Gender and engagement with fitspiration (n=1139).

	Total (n=1139), n (%)	Gender, n (%)		*P* value	Φ coefficient
		Male	Female		
Engagement through following accounts or view on Instagram feed				.005	−0.083
Yes	767 (67.30)	202 (61.20)	565 (69.80)		
No	372 (32.70)	128 (38.80)	244 (30.20)		
Engagement through liking posts				.11	−0.048
Yes	731 (64.20)	200 (60.60)	531 (65.60)		
No	408 (35.80)	130 (39.40)	278 (34.40)		
Engagement by scrolling through individual accounts, posts, or images				<.001	−0.115
Yes	586 (51.40)	140 (42.20)	446 (55.10)		
No	553 (48.60)	190 (57.60)	363 (44.90)		
Engagement through searching for hashtags				.06	−0.057
Yes	289 (25.40)	71 (21.50)	218 (26.90)		
No	850 (74.60)	259 (78.50)	591 (73.10)		
Engagement through tagging friends in posts				.39	0.026
Yes	190 (16.70)	60 (18.20)	130 (16.10)		
No	949 (83.30)	270 (81.80)	679 (83.90)		
Engagement through tagging fitspiration accounts in posts				.007	0.079
Yes	110 (9.70)	44 (13.30)	66 (8.20)		
No	1029 (90.30)	286 (86.70)	743 (91.80)		
Engagement through commenting on posts				.004	0.084
Yes	217 (19.10)	80 (24.20)	137 (16.90)		
No	922 (80.90)	250 (75.80)	672 (83.10)		
Engagement through posting content				.01	0.075
Yes	130 (11.40)	50 (15.20)	80 (9.90)		
No	1009 (88.60)	280 (84.80)	729 (90.10)		
Engagement through sharing content with friends				.96	0.002
Yes	184 (16.20)	53 (16.10)	131 (16.20)		
No	955 (83.80)	277 (83.90)	678 (83.80)		

Figure 2. Gender and fitspiration engagement.



Why Participants Engaged With Fitspiration on Instagram

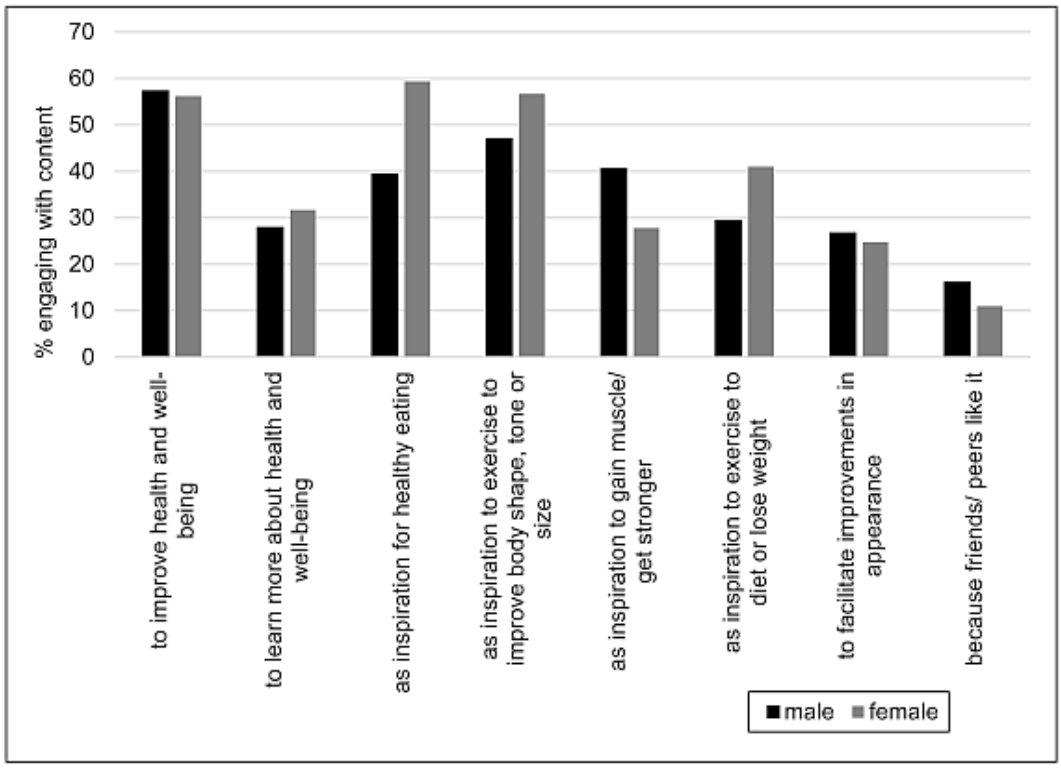
The main reasons for viewing fitspiration content selected by participants were to improve their health and well-being (669/1189, 56.27%); inspiration to exercise to improve body shape, tone, or size (636/1189, 53.49%); and inspiration for healthy eating (630/1189, 52.99%). There were clear gender differences between men and women in terms of the reasons for viewing fitspiration. Men were more likely to use fitspiration as an inspiration to exercise to gain muscle or get stronger ($\chi^2_{1, N=1147}$=17.9; *P*<.001) or because friends or peers view or like it ($\chi^2_{1, N=1147}$=5.5; *P*=.02) than women, and women were more likely to use fitspiration as inspiration for healthy eating ($\chi^2_{1, N=1147}$=37.7; *P*<.001); inspiration to exercise to improve body shape, tone, or size ($\chi^2_{1, N=1147}$=8.9; *P*<.001); or inspiration to exercise to diet or lose weight ($\chi^2_{1, N=1147}$=13.5; *P*<.001) than their male counterparts. The reasons for engagement with fitspiration by gender are listed in Table 4.

Figure 3 shows gender and reasons for fitspiration engagement.

Table 4. Gender and reasons for engagement (n=1147).

Reason for engagement	Total (n=1147), n (%)	Gender, n (%)		*P* value	Φ coefficient
		Male	Female		
To improve health and well-being				.74	0.010
Yes	648 (56.50)	189 (57.30)	459 (56.20)		
No	499 (43.50)	141 (42.70)	358 (43.80)		
To learn more about health and well-being				.20	−0.038
Yes	351 (30.60)	92 (27.90)	259 (31.70)		
No	796 (69.40)	238 (72.10)	558 (68.30)		
As inspiration for healthy eating				<.001	−0.181
Yes	615 (53.60)	130 (39.40)	485 (59.40)		
No	532 (46.40)	200 (60.60)	332 (40.60)		
As inspiration to exercise to improve body shape, tone or size				.003	−0.088
Yes	618 (53.90)	155 (47)	463 (56.70)		
No	529 (46.10)	175 (53)	354 (43.30)		
As inspiration to exercise to gain muscle or get stronger				<.001	0.125
Yes	361 (31.50)	134 (40.60)	227 (27.80)		
No	786 (68.50)	196 (59.40)	590 (72.20)		
As inspiration to exercise to diet or lose weight				<.001	−0.108
Yes	432 (37.70)	97 (29.40)	335 (41)		
No	715 (62.30)	233 (70.60)	482 (59)		
To facilitate improvements in appearance				.52	−0.019
Yes	291 (25.40)	88 (26.70)	203 (24.80)		
No	856 (74.60)	242 (73.30)	614 (75.20)		
Because friends or peers like it				.02	−0.069
Yes	143 (12.50)	53 (16.10)	90 (11)		
No	1004 (87.50)	277 (83.90)	727 (89)		

Figure 3. Gender and reasons for fitspiration engagement.



Discussion

Principal Findings

This quantitative survey of young Instagram fitspiration users reveals important insights regarding how and why consumers in the United Kingdom engage in this form of digital content. Most participants actively using Instagram for fitspiration (therefore eligible to take part in the study) were women (826/1175, 70.30%), consistent with previous academic research [21] and large-scale surveys [26], which suggests that women are more likely to engage with Instagram as a platform than men. However, this study makes a unique contribution to the existing body of knowledge in demonstrating that men and women have significantly different experiences in terms of what, how, and why they engage with the fitspiration content on Instagram. The findings relating to content preferences and motivations for viewing are gendered in a way that reflects clear dualistic gender ideals that are communicated through and reinforced by the digital content men and women consume, and the associated consumption behaviors.

The Masculine Muscular Ideal

Historically, the dominant body ideal for men has increasingly shifted toward muscularity, with a specific emphasis on muscular definition and tone [35-37]. This trend has emerged as the masculine *muscular ideal*, which has been suggested to affect the ways in which many men view and critique their bodies [35]. In line with this, the study demonstrates that men are explicitly attracted to fitspiration content from athletes and bodybuilders, as inspiration to exercise to gain muscle and become stronger. These findings align with previous research that suggests web-based fitspiration images of men are far more likely to show bodies with visibly high levels of muscularity than those of women [21]. The study demonstrates that male fitspiration users are more likely to engage with content from professional sportsmen as opposed to models, celebrities, and influencers. This suggests that men may privilege fitspiration content from sports professionals with active performing bodies, as opposed to celebrities, models, and influencers who have perhaps constructed visually esthetic bodies to be admired and idealized.

Some scholarly work exploring the masculine body ideal has suggested that in addition to promoting muscularity, men are increasingly interested in reducing body fat to promote *leanness* [38]. Specifically, research has identified two separate dimensions of fat loss as a drive for thinness and a drive for leanness with men, perhaps having a tendency to strive toward a lean toned body as opposed to the thin ideal, which may be more attractive to women [39]. Despite research suggesting that men are increasingly interested in fat loss, the findings indicate that men are less likely to engage with content promoting weight loss and diet or use fitspiration to support healthy eating, diet, or weight loss than women. It could therefore be hypothesized that either fitspiration content promotes the *thin* as opposed to *lean* ideal or that diet or weight loss fitspiration content was predominantly focused on women and their bodies or lifestyles.

Moreover, it could be suggested that fitspiration users are more likely to privilege a large muscular (as opposed to lean body) ideal compared with the general population. This aligns with previous research suggesting that male fitspiration users are significantly more likely to internalize a muscular body ideal when compared with nonusers [29]. While muscularity is still privileged in terms of male fitspiration, this could be down to this content focusing uniquely on achieving peak physical fitness, thus identifying a cultural prejudice toward traditional representations of masculinity [28].

The Feminine Thin but Shapely Ideal

Critical research has suggested that the dominant body ideal for women is a contradictory body [4,23,24,27], simultaneously very thin, and exceptionally fit and toned without displaying excess muscularity [16]. This demonstrates a contrast to the masculine ideal body presented above, which honors muscular size. The notion that women privilege thinness is reinforced in this study, with female users being significantly more likely to consume fitspiration content promoting thinness through diet and weight loss than male users. This is also clear through women's consumption motivation behavior, as they are significantly more likely to use fitspiration as inspiration to lose weight through diet and exercise than men. These preference and behavioral findings align with existing literature that demonstrates that the content of fitspiration aimed at women is more likely to show thin bodies and reflect a thin ideal than those aimed at men [4,21].

The findings also demonstrate that women are significantly more likely to consume fitspiration content from celebrities or models and influencers than their male counterparts, who, as noted above, are more likely to engage with athletes and bodybuilders. While numerous athletes and sportspeople monetize their use of Instagram [40,41], it could be argued that the promotion of a consumptive ethic and level of brand endorsement is more common within the accounts of influencers for whom this is often their sole intention or form of employment, and celebrities or models who increasingly yield salaries exclusively or in the main using social media. It could therefore be suggested that women who use fitspiration are likely to be marketed to or actively purchase goods to conform to an ideal body constructed in line with consumptive ethic, thus reinforcing the economic oppression that they have historically faced within Western society [42]. It could also be seen as reflective of the oppressive nature of the consumptive ethic [42] that emphasizes (active) male production and (passive) female consumption of goods and services, locating the power firmly in the hands of men.

Gender and Active or Passive Engagement

While women are significantly more likely to engage in passive fitspiration consumption behaviors such as following, viewing, and scrolling through individual accounts, posts, or images, men are more likely to engage in active engagement, such as tagging fitspiration accounts in posts, commenting, and posting fitspiration content. This is in contrast to previous research [43], which suggests that women are consistently more likely to engage in social media activities by commenting, sharing, or liking. Therefore, based on these findings, it could be posited that fitspiration provides a unique social media context in which men engage more actively than women. This may be because of the emphasis on fitness and muscularity within fitspiration images reinforcing traditional masculine ideals, as opposed to content such as thinspiration, which promotes a thin ideal typically associated with femininity.

Ultimately, the finding that men are significantly more likely to report posting fitspiration content than women provides an interesting texture to the existing research narrative, as previous content research demonstrates that the majority (67%) of fitspiration images depict women as opposed to men [4]. There are a number of potential explanations for this. First, it may be that while men post more fitspiration content, it could be of images other than their own bodies, especially given that the fitspiration definition adopted is relatively broad (photographs of fit people, people in the gym, health foods, or inspirational quotes relating to diet and fitness). This could also be explained by drawing on research that explores the gendered nature of (self) objectification [11,39], which suggests that women are more likely to view their bodies as objects to be evaluated and may therefore be more inclined to offer images of their bodies for assessment. It may be that fewer women posted a higher volume of fitspiration images. This could be explained by the fact that women who meet gendered body ideals may be more likely to post a higher volume of images. Women not meeting gendered body ideals may be less inclined to label images of their bodies as fitspiration or failing to accurately report their active engagement or that men were posting images of women and coding them as fitspiration content.

Study Strengths and Limitations

This study has numerous strengths and limitations that require consideration to adequately interpret the study findings. First, these findings reflect a large sample (n=1213) of fitspiration users in the United Kingdom and shed valuable light on their motivations for engagement, patterns of consumption, and the perceived impact of fitspiration from a gendered perspective. These user-centered data are especially enlightening, as they can be used to support the largely content-driven literature in this area. However, while this provides an alternative perspective, this study does not explain the relationship between content and consumption in terms of how they intersect, and to what extent consumption is driven by content. Furthermore, the research method provides a breadth of existing knowledge but sacrifices some depth that could potentially be achieved through qualitative interviews and questionnaires, or naturalistic observation and walk-through methods.

While the consumer-driven focus of the study is rationalized based on the absence of this focus from the existing research narrative, there are aspects of the consumer experience that have been neglected. More in-depth exploration could have involved exploring the gender differences in terms of body-part focus to build on existing literature [21] and by exploring the extent to which consumers engage with fitspiration images of bodies representing a gender other than their own; for example, women looking at images of men's bodies. While there is a rationale for research in these areas, it was not considered to be the focus

of the current research and therefore could be considered a limitation.

Generalizability

As with any such survey, the representativeness of any sample can always be questioned. Although a random sample was drawn from a larger panel, it is difficult to judge the extent to which such samples are truly representative of the actual population; for example, it may be that women may be more likely to engage with web-based surveys; hence, the results must be treated with some caution. The use of nonvalidated measures could also be identified as a potential issue, even given the descriptive nature of the items, and this could be addressed in future research.

Third, this study explores the gendered nature of fitspiration consumption and yields some important findings regarding the differences in what fitspiration young UK-based men and women were consuming, and how and why they chose to consume it. However, because of the statistical nature of the analysis, this study did not go into any depth regarding individuals identified as nonbinary or genderqueer (7/1175, 0.6%) as opposed to males or females. This 0.6% is comparable with estimates for the general UK population that suggest 0.4% of adults in the United Kingdom identify as nonbinary or genderqueer when faced with it as an option alongside males and females [44]. Nonbinary or genderqueer individuals are likely to have unique experiences of consuming content largely related to the gendered bodies on the internet, and this was neglected in this study.

Finally, as with any study involving multiple analyses, the risk of false positives is always present. No adjustments were made for this within the analysis for two reasons: first, the research provides findings that are aligned with previous empirical work and theory, and second, that there was a general degree of consistency within the findings. While the risk remains, it would seem to be small, and given the exploratory nature of the study, further work and analysis may not be necessary to strengthen confidence in the analyses presented here.

Future Directions

This study and the evaluation above highlight several valuable directions for future fitspiration research. First, there is a need for research to explore the relationship between content and consumption in terms of how they drive each other and the uniquely gendered nature of each. This could take the form of qualitative research exploring the fitspiration consumption process in more depth, or scholarly work focusing on fitspiration producers, who are likely to also be actively engaged consumers.

Furthermore, this survey could be used as a template to explore generational and cross-cultural differences and compare the unique ways in which these consumption patterns are gendered. It would be especially interesting to explore if and how this content is consumed in non-Western, majority world countries, with alternative ideas regarding body ideals and gender roles to make comparisons. In addition, to provide a comprehensive understanding of how representative this and other research on fitspiration use and UK consumers, there is a need for large-scale quantitative work that explores the prevalence of this web-based behavior both in the United Kingdom and globally.

While this research provides an exploration into the fitspiration consumption experiences of young people aged 18-24 years who represent around 24.1% of all UK-based Instagram users, additional research should focus on younger participants and adolescents who, despite using Instagram less (representing 8.1% of UK-based users [33]) may be more vulnerable to body ideals communicated via social media, and the negative aspects of fitspiration as they are at a crucial stage in the development of positive or negative body image [45].

There is also a need to explore how people who identify as nonbinary or genderqueer as opposed to male or female consume fitspiration, and the challenges that they face in being confronted by large volumes of this content that largely seems to be coded as either male or female. There is also a need for content-driven research exploring fitspiration content created by and for genderqueer and nonbinary individuals. It is also important that future research looks to further unpack the finding that women were more likely to consume fitspiration on Instagram, while men were more likely to create content. There is a need to explore possible explanations to ascertain whether men are less likely to post images of their bodies as fitspiration than women, such as posting images of women and coding them as fitspiration content that women fail to accurately report their active engagement, or suggest an alternative explanation. Finally, future research needs to explore whether male fitspiration users are indeed a unique subculture, and whether the male muscular ideal is broadly still dominant, as theorists have suggested that there is movement away from hypermasculine male stereotypes in the general media, with the strong hard man ideal becoming diminished in favor of more diverse forms of masculinity [46].

Conclusions

The purpose of this study was to explore how young men and women engage with fitspiration content on Instagram and to provide a gendered analysis of how and why they consume this content. The key findings of this study achieved this aim in demonstrating the gendered way in which consumers engage with the fitspiration content on Instagram. Specifically, while all fitspiration consumers engaged with content that reinforced traditional body ideals, ideals were extremely gendered. Furthermore, male users were more likely to engage actively with content than female users who consumed content more passively, indicating that the experiences of men and women consuming fitspiration are vastly different. Therefore, based on these findings, it is suggested that any interventions designed to address the potential harm of this web-based content for young people should be gender-specific to adequately address the heavily gendered body ideals and experiential differences for young men and women.

Acknowledgments

This study was funded by Bournemouth University's internal research funding.

Conflicts of Interest

None declared.

References

1. Internet access households and individuals. Office for National Statistics. 2019. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/datasets/internetaccesshouseholdsandindividualsreferencetables [accessed 2021-09-07]
2. Raggatt M, Wright CJ, Carrotte E, Jenkinson R, Mulgrew K, Prichard I, et al. "I aspire to look and feel healthy like the posts convey": engagement with fitness inspiration on social media and perceptions of its influence on health and wellbeing. BMC Public Health 2018 Aug 10;18(1):1002 [FREE Full text] [doi: 10.1186/s12889-018-5930-7] [Medline: 30097034]
3. Vaterlaus JM, Patten EV, Roche C, Young JA. #Gettinghealthy: the perceived influence of social media on young adult health behaviors. Comput Human Behav 2015 Apr;45:151-157. [doi: 10.1016/j.chb.2014.12.013]
4. Tiggemann M, Zaccardo M. "Exercise to be fit, not skinny": the effect of fitspiration imagery on women's body image. Body Image 2015 Sep;15:61-67. [doi: 10.1016/j.bodyim.2015.06.003] [Medline: 26176993]
5. Easton S, Morton K, Tappy Z, Francis D, Dennison L. Young people's experiences of viewing the fitspiration social media trend: qualitative study. J Med Internet Res 2018 Jun 18;20(6):e219 [FREE Full text] [doi: 10.2196/jmir.9156] [Medline: 29914859]
6. Tiggemann M, Zaccardo M. 'Strong is the new skinny': a content analysis of #fitspiration images on Instagram. J Health Psychol 2018 Jul;23(8):1003-1011. [doi: 10.1177/1359105316639436] [Medline: 27611630]
7. Number of monthly active Instagram users from January 2013 to June 2018. Statista. 2013. URL: https://www.statista.com/statistics/253577/number-of-monthly-active-instagram-users/ [accessed 2021-09-07]
8. Technology tracker Q1 2018. Ipsos. 2018. URL: https://www.ipsos.com/en/technology-tracker-q1-2018 [accessed 2021-09-07]
9. #Fitspo. Instagram Explore. URL: https://www.instagram.com/explore/tagsfitspo [accessed 2021-09-07]
10. Manago AM, Graham MB, Greenfield PM, Salimkhan G. Self-presentation and gender on MySpace. J Appl Dev Psychol 2008;29(6):446-458. [doi: 10.1016/j.appdev.2008.07.001]
11. Fardouly J, Willburger BK, Vartanian LR. Instagram use and young women's body image concerns and self-objectification: testing mediational pathways. New Media Soc 2017 Feb 01;20(4):1380-1395. [doi: 10.1177/1461444817694499]
12. Mayoh J. Perfect pregnancy? Pregnant bodies, digital leisure and the presentation of self. Leisure Stud 2019 Jan 16;38(2):204-217. [doi: 10.1080/02614367.2018.1562492]
13. Holland G, Tiggemann M. "Strong beats skinny every time": disordered eating and compulsive exercise in women who post fitspiration on Instagram. Int J Eat Disord 2017 Jan;50(1):76-79. [doi: 10.1002/eat.22559] [Medline: 27302867]
14. Turner PG, Lefevre CE. Instagram use is linked to increased symptoms of orthorexia nervosa. Eat Weight Disord 2017 Jun;22(2):277-284 [FREE Full text] [doi: 10.1007/s40519-017-0364-2] [Medline: 28251592]
15. Orthorexia. National Eating Disorders Association. URL: https://www.nationaleatingdisorders.org/learn/by-eating-disorder/other/orthorexia [accessed 2021-09-07]
16. Boepple L, Ata RN, Rum R, Thompson JK. Strong is the new skinny: a content analysis of fitspiration websites. Body Image 2016 Jun;17:132-135. [doi: 10.1016/j.bodyim.2016.03.001] [Medline: 27045871]
17. Vartanian LR, Wharton CM, Green EB. Appearance vs. health motives for exercise and for weight loss. Psychol Sport Exerc 2012 May;13(3):251-256. [doi: 10.1016/j.psychsport.2011.12.005]
18. Adkins EC, Keel PK. Does "excessive" or "compulsive" best describe exercise as a symptom of bulimia nervosa? Int J Eat Disord 2005 Jul;38(1):24-29. [doi: 10.1002/eat.20140] [Medline: 15991218]
19. DiBartolo P, Lin L, Montoya S, Neal H, Shaffer C. Are there healthy and unhealthy reasons for exercise? Examining individual differences in exercise motivations using the function of exercise scale. J Clin Sport Psychol 2007;1(2):93-120. [doi: 10.1123/jcsp.1.2.93]
20. Peng C, Wu T, Chen Y, Atkin DJ. Comparing and modeling via social media: the social influences of fitspiration on male Instagram users' work out intention. Comput Human Behav 2019 Oct;99:156-167. [doi: 10.1016/j.chb.2019.05.011]
21. Carrotte ER, Prichard I, Lim MS. "Fitspiration" on social media: a content analysis of gendered images. J Med Internet Res 2017 Mar 29;19(3):e95 [FREE Full text] [doi: 10.2196/jmir.6368] [Medline: 28356239]
22. Robinson L, Prichard I, Nikolaidis A, Drummond C, Drummond M, Tiggemann M. Idealised media images: the effect of fitspiration imagery on body satisfaction and exercise behaviour. Body Image 2017 Sep;22:65-71. [doi: 10.1016/j.bodyim.2017.06.001] [Medline: 28654826]
23. Sumter SR, Cingel DP, Antonis D. "To be able to change, you have to take risks #fitspo": exploring correlates of fitspirational social media use among young women. Telemat Inform 2018 Aug;35(5):1166-1175. [doi: 10.1016/j.tele.2018.01.013]
24. Alberga AS, Withnell SJ, von Ranson KM. Fitspiration and thinspiration: a comparison across three social networking sites. J Eat Disord 2018 Nov 26;6:39 [FREE Full text] [doi: 10.1186/s40337-018-0227-x] [Medline: 30534376]

25. Duggan M, Ellison N, Lampe C, Lenhart A, Madden M. Social media update 2014. Pew Research Center. 2015. URL: https://www.pewresearch.org/internet/2015/01/09/social-media-update-2014/ [accessed 2021-09-07]
26. Social media fact sheet. Pew Research Center. 2021. URL: https://www.pewinternet.org/fact-sheet/social-media/ [accessed 2021-09-07]
27. Markula P. Firm but shapely, fit but sexy, strong but thin: the postmodern aerobicizing female bodies. Sociol Sport J 1995;12(4):424-453. [doi: 10.1123/ssj.12.4.424]
28. Palmer L. Poppin bottles, getting wheysted:exploring young men's engagement with fitspiration content and its consequential influences on attitudes and behaviour. J Promotional Commun 2015;3(3):426-445.
29. Fatt SJ, Fardouly J, Rapee RM. #malefitspo: links between viewing fitspiration posts, muscular-ideal internalisation, appearance comparisons, body satisfaction, and exercise motivation in men. New Media Soc 2019 Jan 01;21(6):1311-1325. [doi: 10.1177/1461444818821064]
30. Law C, Labre MP. Cultural standards of attractiveness: a thirty-year look at changes in male images in magazines. Journal Mass Commun Q 2002;79(3):697-711. [doi: 10.1177/107769900207900310]
31. Bell BT, Cassarly JA, Dunbar L. Selfie-objectification: self-objectification and positive feedback ("Likes") are associated with frequency of posting sexually objectifying self-images on social media. Body Image 2018 Sep;26:83-89. [doi: 10.1016/j.bodyim.2018.06.005] [Medline: 30005293]
32. Jong ST, Drummond MJ. Exploring online fitness culture and young females. Leisure Stud 2016 May 08;35(6):758-770. [doi: 10.1080/02614367.2016.1182202]
33. UK: Instagram users by age. Statista. URL: https://www.statista.com/statistics/1018012/instagram-users-united-kingdom/#:%7E:text=This%20statistic%20shows%20the%20breakdown,and%2034%20years%20of%20age [accessed 2021-09-07]
34. Jenkins-Guarnieri MA, Wright SL, Johnson B. Development and validation of a social media use integration scale. Psychol Pop Media Cult 2013 Jan;2(1):38-50. [doi: 10.1037/a0030277]
35. McCreary D. Gender and age differences in the relationship between body mass index and perceived weight: exploring the paradox. Int J Men's Health 2002 Jan 1;1(1):31-42. [doi: 10.3149/jmh.0101.31]
36. Darcy AM, Doyle AC, Lock J, Peebles R, Doyle P, Le Grange D. The eating disorders examination in adolescent males with anorexia nervosa: how does it compare to adolescent females? Int J Eat Disord 2012 Jan;45(1):110-114 [FREE Full text] [doi: 10.1002/eat.20896] [Medline: 22170022]
37. Mellor D, Waterhouse M, Mamat NH, Xu X, Cochrane J, McCabe M, et al. Which body features are associated with female adolescents' body dissatisfaction? A cross-cultural study in Australia, China and Malaysia. Body Image 2013 Jan;10(1):54-61. [doi: 10.1016/j.bodyim.2012.10.002] [Medline: 23228484]
38. CD M, GJ G. Body image concerns of male rugby players, with specific focus on muscularity and body fat. J Obes Overweig 2017 May;3(1):103-110. [doi: 10.15744/2455-7633.3.103]
39. Smolak L, Murnen SK. Drive for leanness: assessment and relationship to gender, gender role and objectification. Body Image 2008 Sep;5(3):251-260. [doi: 10.1016/j.bodyim.2008.03.004] [Medline: 18585105]
40. Smith LR, Sanderson J. I'm Going to Instagram It! An analysis of athlete self-presentation on Instagram. J Broadcast Electron Media 2015 May 28;59(2):342-358. [doi: 10.1080/08838151.2015.1029125]
41. Toffoletti K, Thorpe H. The athletic labour of femininity: the branding and consumption of global celebrity sportswomen on Instagram. J Consum Cult 2018 Apr 27;18(2):298-316. [doi: 10.1177/1469540517747068]
42. Wolf N. The Beauty Myth. New York: Random House; 2013.
43. Blank G, Lutz C. The social structuration of six major social media platforms in the United Kingdom: Facebook, LinkedIn, Twitter, Instagram, Google+ and Pinterest. In: Proceedings of the 7th 2016 International Conference on Social Media. 2016 Presented at: 7th 2016 International Conference on Social Media & Society; Jul 11-13,2016; London. [doi: 10.1145/2930971.2930979]
44. Glen F, Hurrell K. Technical Note: Measuring Gender Identity. In: Manchester: Equality and Human Rights Commission 2012.
45. Voelker DK, Reel JJ, Greenleaf C. Weight status and body image perceptions in adolescents: current perspectives. Adolesc Health Med Ther 2015 Aug 25;6:149-158 [FREE Full text] [doi: 10.2147/AHMT.S68344] [Medline: 26347007]
46. Hanke R. The "mock‐macho" situation comedy: hegemonic masculinity and its reiteration. West J Commun 1998 Mar;62(1):74-93. [doi: 10.1080/10570319809374598]

Edited by R Kukafka, G Eysenbach; submitted 14.01.20; peer-reviewed by E Carrotte, M Lim, S Musy; comments to author 18.04.20; revised version received 30.09.20; accepted 13.07.21; published 04.10.21.

Please cite as:
Mayoh J, Jones I
Young People's Experiences of Engaging With Fitspiration on Instagram: Gendered Perspective
J Med Internet Res 2021;23(10):e17811
URL: https://www.jmir.org/2021/10/e17811
doi:10.2196/17811
PMID:34605768

Original Paper

e-Cigarette Cessation: Content Analysis of a Quit Vaping Community on Reddit

Laura Struik[1], BSCN, MSN, PhD; Youjin Yang[1], BSN

School of Nursing, Department of Health and Social Development, University of British Columbia Okanagan, Kelowna, BC, Canada

Corresponding Author:
Laura Struik, BSCN, MSN, PhD
School of Nursing
Department of Health and Social Development
University of British Columbia Okanagan
1147 Research Road
Kelowna, BC, V1V 1V7
Canada
Phone: 1 2508079972
Email: laura.struik@ubc.ca

Abstract

Background: e-Cigarette use, also known as vaping, has increased dramatically over the past few years, especially among younger demographics. However, researchers have found that a large number of e-cigarette users want to quit. Little is known about the unique aspects of vaping cessation, which is critical to informing the development of relevant resources and interventions for e-cigarette users who want to quit. Social media forums such as Reddit provide opportunities to understand the experiences of behavior change such as quitting vaping from the perspective of end users.

Objective: This study aims to examine a quit vaping subreddit to understand how e-cigarette users are experiencing and approaching vaping cessation. Specifically, we examine methods used to approach quitting, reasons for quitting, and barriers and facilitators to quitting.

Methods: A total of 1228 posts were collected. The posts were inductively coded to generate categories and subcategories using conventional content analysis. Data were analyzed using the NVivo 12 qualitative data analysis software.

Results: Most users reported a preference for approaching quitting through gradual reduction, particularly through the use of their own devices by tapering the nicotine content. Their reasons for quitting were primarily related to experiencing negative physical consequences associated with vaping, especially in relation to their lungs (eg, tight chest), and tired of feeling stuck to the vape because of nicotine addiction. Top barriers to quitting were related to withdrawal symptoms and intensity of addiction. The top facilitators to quitting were related to using distraction techniques (eg, hobby, gaming, and mindfulness exercises), as well as having a positive mindset.

Conclusions: The findings of this study reveal unique aspects that encompass the process of quitting vaping. These findings have significant implications for both policy and intervention development.

(J Med Internet Res 2021;23(10):e28303) doi:10.2196/28303

KEYWORDS

qualitative research; electronic nicotine delivery systems; vaping; cessation; social media

Introduction

Vaping with nicotine, or using e-cigarettes, has become a popular global phenomenon in recent years. Compared with 2011, whereby 7 million people worldwide used e-cigarettes, the number of e-cigarette users had increased to 41 million in 2018 [1], with youth and young adults representing the largest users of e-cigarettes [2]. Despite how vaping has grown in popularity, recent evidence indicates that a large number of e-cigarette users actually want to quit [3-6]. A focus group study in Canada found that >40% of youth and young adults who vaped indicated a strong desire to quit, and >40% had tried to quit in the past year [5]. Another study in the United States found that 45% of youth were interested in quitting, and 25% had tried to quit in the past year [4].

Although it is encouraging that many e-cigarette users want to quit, there remains a lack of cessation services to support their quit efforts. Researchers acknowledge that most cessation advice mimics smoking cessation [7] and that there is a dearth of resources tailored specifically to vaping [4,5]. The 2020 Surgeon General's report on smoking cessation called for research to develop and understand safe and effective e-cigarette cessation guidelines and interventions [8]. A critical step toward developing effective cessation resources is to first understand the experiences and preferences of e-cigarette users as it relates to quitting vaping. However, no studies to date have examined how vapers quit and what might help them with quitting.

Web-based forums serve as particularly fruitful platforms for gathering information on end user experiences and preferences. Reddit, the most comprehensive forum on the internet [9], has several e-cigarette–related discussion topics. The use of social media such as Reddit to understand vaping cessation experiences is particularly important given that engagement on social media platforms for sharing and gathering health information is highest among youth and young adults [10,11]. Researchers have studied Reddit discussions to generate important information about e-cigarette flavors [12], the health effects of e-cigarette use [13,14], e-cigarette use patterns [9,15,16], reasons and attitudes surrounding underage e-cigarette use [17], and underage buying and selling [18]. However, no researchers have examined the quit vaping communities that are now on Reddit. This user-generated content around vaping holds significant potential to inform best practices for cessation guidelines and services. Therefore, we examined a quit vaping subreddit to understand how e-cigarette users experience and approach vaping cessation. Specifically, we examined methods for approaching quitting, reasons for quitting, and barriers and facilitators to quitting.

Methods

Overview

Similar to other studies involving Reddit [17-19], the examined subreddit contains publicly available threads and comments. No personal information (eg, account details) was included, and the usernames of the Reddit posters were not presented in this study. As such, this study was classified by the University of British Columbia Okanagan's Behavioural Research Ethics Board as research not involving human subjects and was, therefore, not subject to institutional review board jurisdiction.

Data Collection

We examined the subreddit r/QuitVaping community for this study. This subreddit was created in 2015 and designed to help vapers motivate each other to quit. This subreddit contained 4700 subscribers. Posts (including original posts and all responses) were collected retrospectively during a 4-week period (August 1, 2020, to August 31, 2020) by using the *top posts of this month* feature. The *top posts of this month* means that the posts with the most upvotes over this period were selected for analysis. This approach to data collection is supported by other content analyses of subreddit communities (eg, see work by Sowles et al [19]). The posts were first copied and pasted onto Microsoft Word, which were then uploaded onto the NVivo 12 qualitative analytic software (QSR International Pty Ltd) for coding. A total of 1228 posts were collected and analyzed.

Data Analysis

The posts were inductively coded to generate categories and subcategories using conventional content analysis [20]. Both authors engaged in collaborative coding via UBC Zoom to code three original posts and 10 responses (13 total), which resulted in the assignment of 44 codes. The authors developed a coding legend and assigned preliminary codes directly onto the Microsoft Word document via the *comments* feature. The authors then independently reviewed an additional set of three original posts and eight responses (11 total). Each author assigned >35 codes. Only six discrepancies in coding were noted, which were then discussed, resolved, and worked into the coding framework. The second author then independently coded the remaining posts, consulting with the first author regularly as needed. Once all the posts were coded, authors met again and refined and collapsed the framework with subcategories under the four major categories of quitting methods, reasons for quitting, barriers to quitting, and facilitators to quitting.

Results

Users

In total, 318 unique users were represented in this sample of posts. A little less than 10% (29/318, 9.1%) of the represented sample indicated their age, with most (24/29, 83%) of them in the young adult age range (aged between 18 and 35 years), and the remaining outliers were aged <18 years (2/29, 7%) and >35 years (3/29, 10%). This may indicate a relatively young sample overall.

Posts

Of the 1228 posts, 189 (15.39%) were original posts, and 1039 (84.61%) were responses to these posts. The average number of responses per original post was 5.5 (SD 5.55), and the range in the number of comments per post was 0-50, with most posts within the range of 5-10 and only one post with 50 comments. This one post is noteworthy, as the engagement was garnered through a call to action for mutual support and accountability. The post is as follows:

> *Everyone who is actively quitting/trying to quit/already did quit and wants to keep it that way, comment on this post. I will send each of you a message on reddit every single day checking up on you, and sending you motivational things to aid in your journey. Only asking that you try and do the same for me :) let me know!!!*

The data within these 1228 posts are broken down into the following four categories: quitting methods, reasons for quitting, barriers to quitting, and facilitators to quitting (Table 1). The most represented categories of post content in this data set were related to barriers and facilitators to quitting.

Table 1. Representation of post categories (N=1228).

Category	Values, n (%)
Quitting methods	175 (14.25)
Reasons for quitting	211 (17.18)
Barriers to quitting	441 (35.91)
Facilitators to quitting	570 (46.42)

Quitting Methods

e-Cigarette users' method of quitting was reflected in 14.25% (175/1228) of posts. Most of these posts (117/175, 66.9%) reflected a preference for a gradual reduction approach. This was accomplished primarily by tapering the nicotine concentration (55/175, 31.4%), which was followed by the use of different sources of nicotine replacement therapy (NRT; 33/175, 18.9%; eg, patches, gum, lozenges, toothpicks, and Nicorette inhalers). The most popular NRT used were patches and gum, which are often used together. Overall, 13 users reported using nicotine pouches to help them quit, and 6 stated using the vape less frequently. The *cold turkey* approach was represented in 33.1% (58/175) of the posts.

Reasons for Quitting

The reasons for quitting vaping were provided in 17.18% (211/1228) of posts (Table 2). There were eight reasons for quitting, as listed by the posters. The number one reason for quitting vaping was experiencing negative physical health effects of vaping, which included shortness of breath, chest pain, wheezing, coughing, sore throat, poor oral health, poor skin health, and poor sleep quality. The second most commonly cited reason to quit was that individuals were tired of feeling stuck to their e-cigarettes, indicating that they wanted to be free to live their lives without a dependency on a vaping device. The other most common reasons for quitting are related to the mental health and financial cost of vaping. These top reasons were then followed by a desire to improve physical health (eg, to improve running time), accommodate life changes (eg, having a baby), to gain time they would be otherwise using to vape or think about vaping, and to improve their intimate relationships (eg, spending time with partners who do not vape or smoke).

Table 2. Reasons for quitting (N=211).

Reason	Values, n (%)	Representative quotes
Experiencing negative physical health consequences	59 (28.0)	"I vaped for 3 years too, very consistently, and I really didn't have a problem with it until I felt the health issues—constant dry throat, never ending phlegm, tonsil stones, fatigue, swollen lips (from vape juice spitting), irritability, and horrible acne. I also felt like I hadn't been myself for awhile."
Tired of feeling stuck to the vape or nicotine	47 (22.3)	"Last night, I had a moment of clarity. My whole life basically revolves around my vape. It's my first thought when I wake up every morning, I have to constantly have it with me, and everything ends with a vape. Anytime I'm without it, I'm just looking forward to the time I can use it. It's honestly hard to imagine life without it, but I know I don't want to be a 70-year-old vaper."
Experiencing negative mental health consequences	28 (13.3)	"I'm fed up with feeling that anxiety and irritability when I'm not vaping."
Financial loss	28 (13.3)	"I had to get to a point where I hated what it was doing to me and my finances."
Wanting to improve physical health	17 (8.1)	"I was hoping quitting would help me with my run times before being winded (which currently doesn't take too long)."
Need to adapt to life changes	17 (8.1)	"But with Covid and everything happening, I want to kick this addiction to the curb."
Improve use of time	6 (2.8)	"Vaping sucks so much it makes you procrastinate, it makes you feel lazy and it makes you run away from things that need to get done."
Prioritizing intimate relationships	6 (2.8)	"Now I'm married and have a 1 year old child and my wife hates me vaping especially since she found out how poisonous the vape juice is for children and she's always concerned about me accidently leaving my vape around our child."

Barriers to Quitting

Barriers to quitting were reflected in 35.91% (441/1228) of the posts. Nine barriers were found, which were listed in the order of most saturated to least saturated (Table 3). The most challenging barrier was related to nicotine withdrawal, with reports of intense withdrawal symptoms associated with trying to quit. The second most common barrier was related to nicotine dependence, with many users citing heavy addiction because of the nicotine salt devices that many of the posters were using (eg, JUUL). The third barrier was related to proximity to others who vape as a major barrier to quitting, which included both web-based (eg, seeing friends vaping on social media) and offline interactions. The fourth barrier to quitting was related to existing mental health challenges, whereby users described

how their stress, anxiety, or depression made it harder for them to quit. The fifth barrier was related to the use of other substances, including smoking traditional cigarettes and other drugs. The sixth barrier was related to vaping becoming habitual and associated with routine things such as commuting to school or work. The seventh cited barrier was related to the accessibility of vaping from both a purchase standpoint as well as from a use standpoint, as users could vape in areas where smoking is prohibited. Other barriers to quitting included the enjoyment of vaping and experiencing weight-related issues after stopping.

Table 3. Barriers to quitting (N=441).

Barrier	Values, n (%)	Representative quote
Intense withdrawal symptoms		
Insomnia; fatigue; sweats; headache; brain fog; chest pain; cough; sore throat; dizziness; anxiety; depression; irritability	137 (31.1)	"I'm 2 days in from daily use but my pulse is around 90-100 bpm when I'm just sitting is this normal? People already told me being anxious 24/7 is common but I feel like my heart won't stop beating fast all day and night. I feel like I can't sleep because I'm so anxious about nothing and my heart won't stop beating quick."
Dependency on high nicotine		
JUUL; using >20mg/ml	116 (26.3)	"I'm telling you, the nic salts will get you every time. Juuls are just icing on the cake."
Proximity to those who vape		
Face-to-face interactions; web-based interactions	41 (9.3)	"I used to see vapes on peoples social media and I would get insane cravings out of nowhere."
Mental health challenges		
High stress; suffer from anxiety or depression	39 (8.8)	"Convinced myself it would help with the stress and anxiety of everything going on."
Other substance use		
Smoking cigarettes; other drugs	34 (7.7)	"Add in that I took Adderall daily, which if anyone doesn't know, makes your desire to smoke increase x10000000. So breaking into pod #3 in a day wasn't unheard of."
Habit or routine associations		
Commute; other day-to-day habits	25 (5.7)	"I'm a delivery driver and I think I connect vaping with driving also. But of course now, I vape all the time. In bed, at least 20 minutes after waking up, want it after a meal and an absolute must have when I'm drinking alcohol."
Accessibility of e-cigarettes		
Easy to purchase; can use anywhere	19 (4.3)	"Vaping I did 10x more than smoking! So easy to do anywhere."
Other		
Enjoyment of vaping	13 (2.9)	"Also a non-drinker and I feel it's my only guilty pleasure [...] I haven't tried to stop yet."
Increased appetite or weight gain after stopping	13 (2.9)	"When does the binge eating stop? I've been 2 weeks nicotine free and still wanna eat constantly and gained so much weight."

Facilitators to Quitting

Approximately 46.42% (570/1228) of posts provided a description of facilitators to quit vaping (Table 4). Nine facilitators to quitting were identified. The most cited facilitator to quitting was distractions, with the number one distraction strategy relating to keeping their minds and hands busy through hobbies, movies, games, and mindfulness exercises. Another common distraction strategy was by replacing puffing on an e-cigarette with eating, chewing gum, drinking water or tea, or chewing on flavored toothpicks. Finally, users spoke about the key role that exercise plays in their quitting efforts from running to weight lifting to yoga.

The second most cited facilitator to quitting was related to having a positive mindset or a positive view of self. Individuals described the need to believe in themselves, encourage themselves, care for their mind and body, grace with themselves, and reward themselves. In addition to self-love was the value of social support for quitting vaping. Although the primary source of support drawn upon was others in the Reddit community, some described the positive support received from family, friends, and even quit buddies to help them quit.

Making vaping less available or desirable was another facilitator to quitting vaping, which was achieved through a variety of strategies, including tossing the device, having someone hide the device, and purposely using undesirable flavors to dissuade them from vaping. Another facilitator to quitting was through the use of behavioral support interventions, which was primarily sought through the *Quit Vaping* app, followed by Alan Carr's

book "Quit smoking the easy way," and then websites that provide information on vaping cessation.

Experiencing the negative effects of vaping, being reminded of their reasons to quit vaping, experiencing the benefits of quitting, and changing their environment (eg, staying away from people who vape, switching up routine, and taking a vacation) were also listed as facilitators to remain vape free.

Table 4. Facilitators to quitting (N=570).

Facilitator	Values, n (%)	Representative quote
Distractions		
Keeping mind or hands busy (hobby, entertainment like movies and gaming, and mindfulness exercises); eating, chewing gum, drinking water or tea, and toothpicks; exercise	279 (48.9)	"I am trying to just live in the moment instead of having to pacify every single discomfort with nicotine. A workout program I follow always says 'get used to the discomfort' and that thought has been on repeat in my mind. Deep breaths and meditation have been helping a lot too."
Positive mindset or positive self-concept		
Resilient mindset; progress mindset (winning streak); self-care, grace, or rewarding	63 (11.1)	"This time feels a lot different though. I am being nice to myself. Letting myself be irritated, eat as many snacks as I want, and really feel my feelings."
Social support		
Support from others (Reddit community, friends, and family); quitting with someone	54 (9.5)	"This little community has made a huge difference on my mindset on quitting!! We all want the best for each other."
Making vaping less accessible or less desirable		
Tossing device; having someone hide device; using undesirable flavors	54 (9.5)	"When I used zero [nicotine], I just kept my device somewhere completely across the house so it was never instantly accessible. It's much easier to say no when you think about how little you actually get for the effort of going to grab it."
Behavioral support		
App (*Quit Vaping*); Allen Carr *Quit smoking the easy way;* quit help websites	34 (6.0)	"And use some app like quit vaping [app name]. Being able to easily see milestones and achievements is very motivating at the start."
Other		
Awareness of negative effects of vaping	23 (4.0)	"I always thought it helped with my anxiety but it honestly only ever made it worse."
Reminders of reasons to quit	17 (3.0)	"I was actually in one of those zones thinking, 'It wasn't so bad, maybe I just "like" vaping' when I started writing this post. By the end of writing it, I remembered how bad my experience really was and haven't had a craving since."
Experiencing benefits of not vaping	11 (1.9)	"I quit two months ago and have only recently started to really feel the positive benefits, but from what I've experienced, it's totally worth it."
Changing environment (social or routine)	11 (1.9)	"I am moving into a house with some very supportive people and I've found that when in a new setting, cravings are a lot lower."

Discussion

Principal Findings

This study is the first of its kind to examine a quit vaping community on Reddit to understand the process of quitting vaping from the perspective of e-cigarette users. The findings of this study shed light on what quitting vaping is like and how it aligns with quitting smoking. The findings reveal that the process of quitting vaping with nicotine, although similar to quitting combustible cigarette smoking in many respects, is also unique in several noteworthy ways.

Quitting Methods

Both cold turkey and gradual reduction are relevant to vaping cessation. However, significantly more e-cigarette users opted for gradual reduction compared with cold turkey. This is unique compared with smoking, whereby most smokers opt for the cold turkey approach [21]. Interestingly, a large portion of those who chose gradual reduction chose to do so by tapering their vaping device. The fact that e-cigarette users are trying to figure out how to quit on their own in this way reflects the dearth of evidence-based support available to those who want to quit and is indicative of the urgent need to develop resources and guidelines to help e-cigarette users with quitting. There are concerns about e-cigarette users turning to their devices to quit. One particular concern is that of ongoing use. In studies that examined the efficacy of using e-cigarettes to quit smoking, although some users were able to successfully quit smoking, almost all were still vaping a year later [22,23], indicating that the use of e-cigarettes promoted ongoing addiction to nicotine compared with the use of approved pharmacological cessation support (eg, NRT).

Similar to this study, an analysis of >3000 Twitter tweets revealed that a gradual approach to quitting by JUUL users was popular, especially by tapering the nicotine concentration of their pods or changing their devices [24]. In contrast, adolescents in the study by Kong et al [25] used the cold turkey approach more frequently. This difference may be because of the length of nicotine product use, where adolescents may more easily quit cold turkeys because of a shorter timeline of use. The differences found indicate that guidelines and interventions should be tailored to different age groups, with attention to the time using nicotine products as well as the type of products used. The use of high-nicotine salt devices such as JUUL appears to lead to more intense addiction, which may make a cold turkey approach more difficult. The variability of nicotine concentration in vaping devices leads to complexity in relation to cessation, which is different from combustible smoking.

Reasons for Quitting

In this study, the most commonly cited reasons for quitting vaping were related to experiences of negative health consequences and addiction. Several studies reveal that experiencing negative health consequences and dependence are consistently among the top few reasons for quitting e-cigarettes [3,5,24,25]. Of particular note is the finding around current health effects. In an analysis of 2000 text responses to a question about reasons to quit vaping from the evidence-based e-cigarette cessation program for youth and young adults, "This is Quitting," researchers found similar findings [3]. They found that the top-rated reason for quitting was health, especially current and general health, compared with future health [3]. The study by Ungar et al [24] also found that experiencing negative health consequences was the top-rated reason to quit. This reason to quit vaping, which is experiencing negative health consequences in the present, is different from the top-rated reason to quit combustible smoking, which is future health [26]. The fact that e-cigarette users are experiencing adverse health effects early in their vaping trajectory brings forward the need to pay more attention to the immediate health effects of vaping. There needs to be particular attention on the effect of vaping on the lungs given that users in this study primarily reported adverse pulmonary outcomes from vaping, including tight chest, sleep apnea, cough, wheezing, and breathlessness.

Barriers and Facilitators to Quitting

In this study, users reported intense withdrawal symptoms and dependency on high nicotine levels as the top barriers to quitting. This is not surprising given that the most popular e-cigarette devices enable the delivery of very high concentrations of nicotine to the brain (eg, JUUL is typically approximately 50 mg/ml) [27], the amount of nicotine that you would expect to see a heavy smoker consuming. These high-nicotine devices appear to be particularly problematic when trying to quit. This is not surprising given that the largest population of e-cigarette users are youth and young adults [28,29], those at an age when the brain is still developing [30,31], and that a developing brain is most vulnerable to long-term and intense nicotine addiction, which makes quitting much harder [27,31-33]. Countries such as Canada, which are taking a strong stance in limiting the nicotine concentration of e-cigarettes [34], are playing a critical role in protecting developing brains from early and long-term addiction. The findings of this study confirm that taking action to limit the nicotine concentration of these devices is critical. However, cessation services and advice must be reflective of the current state of use, which is when high-nicotine concentrations are allowed, as well as be ready to adapt to future use. Researchers should explore and evaluate how cessation advice and support may best accommodate shifts in use behaviors because of policy changes.

It is also interesting that not only offline interactions but also web-based interactions with other e-cigarette users served as a key barrier to quitting. Vaping has progressed in popularity at a time when social media use and social media influencers are rampant. e-Cigarette companies have successfully tapped into this trend and are not only lending to the uptake of e-cigarettes [35] but are also serving as a barrier to quitting, as revealed in this study. Vaping cessation efforts also need to capitalize on these web-based networks, and there are some innovative efforts being made in this regard (eg, This is Quitting by the Truth campaign) [7,36].

The findings also reveal that mental health challenges are a major barrier to quitting. Tobacco use is known to coincide with mental health issues, including among young adults [37]. Mental health disorders are increasing and are an urgent concern [38], especially in the context of the COVID-19 pandemic [39]. In light of this, attention to and screening for mental health among e-cigarette users is warranted. In addition, cessation interventions that promote positive mental health appear to align with stated facilitators to quitting, which include resiliency, self-grace, and self-love.

Another barrier to quitting is related to the accessibility of e-cigarettes. Vaping with nicotine is easily accessible via web-based sales [40], as well as the recent establishment of a large number of vape retailers that sell these products [41]. For example, in Canada, it was found that 76% of retail outlets sold vaping products in 2014 [42]. In the province of British Columbia alone, there are an estimated 90,000 vape retailers compared with only 6000 retailers that sell combustible cigarettes [43]. In addition, the policies around vaping are different from smoking; for example, you cannot be smoking indoors, but you can still vape in many indoor spaces despite several jurisdictions implementing vape-free indoor policies [44]. This is combined with the discreet and appealing designs of e-cigarettes and that e-cigarette users are left to their own devices to regulate their use. e-Cigarette advertising capitalizes on these freedoms associated with vaping and promotes users to use their vape at all times [45]. For example, a Blu commercial spends the entire 60 second advertising time to tell potential e-cigarette users that they can "smoke blue virtually anywhere" [46]. According to the users who posted on the Reddit community, being able to vape anywhere at any point in time has left them doing exactly that. This constant use has left them feeling "stuck to the vape" and is the primary reason for quitting. These findings have important implications for banning advertisements that promote vaping in this way, as well as policies for limiting where vaping can occur. The findings also hold implications in limiting the accessibility of both web-based and offline sales of e-cigarettes.

The top facilitator to quitting was the use of distractions, which ranged from things that kept the mind and hands busy to replacing the e-cigarette with other things, such as toothpicks and exercise. Similar to smoking cessation interventions such as Crush the Crave [47], the importance of including tips and opportunities to distract those who are trying to quit vaping is key. There is an opportunity for cessation interventions to capitalize on new digital technologies available to develop distractions for young vapers. Given that youth and young adults, the highest demographic of vapers, are ubiquitous on the web, focusing on innovative efforts in this context to help distract them through their cravings would likely prove quite fruitful, especially given their receptivity to receiving interventions for cessation through these media [5].

Future Research

These findings bring forward several areas for future research. First, the unique aspects of vaping cessation call for urgent research focused on developing cessation guidelines tailored to e-cigarette use. The development of these guidelines should establish both pharmacotherapy and behavioral support recommendations. For pharmacotherapy, it is interesting that a large portion of vapers used NRT to support their quit attempts, and they primarily used the nicotine patch and gum in tandem. There is a need to explore the right *recipe* for vapers who are trying to quit. This is especially critical, as the common use of high nicotine concentrations in e-cigarettes may mean that vapers need a different approach to NRT compared with smokers. For behavioral support interventions, the reported reasons for quitting, barriers and facilitators to quitting, and benefits of quitting could inform the development of interventions that resonate with vapers. Indeed, the findings offer a beginning framework and structure for exploring and developing recommendations.

Longitudinal research that tracks how cessation is approached is needed. It would be interesting to explore whether approaches to quitting shift over time with the introduction of new resources and policies. For example, whether fewer vapers resort to tapering their own devices to quit if pharmacotherapy options are tailored to vaping. It would also be interesting to conduct research on which facilitators are linked to successful long-term abstinence rates and for whom so that various populations have access to more tailored options to help them quit. Finally, there is a need to explore predictors of vaping so that we can identify high-risk populations before engaging in e-cigarette use.

Limitations

The findings of the study are based on data during a time when there was a pandemic, which may have influenced the experience of quitting. In addition, this study was conducted when few resources were available to support the cessation efforts of e-cigarette users. In addition, the findings are limited to data provided on one platform during a 4-week period. Another limitation of this study is the inability to definitively determine the demographic details of the Reddit community users, limiting the ability to note nuances in experiences and preferences based on demographic variables. However, Reddit users are known to skew toward being young and male [48], indicating that the results may represent the experiences of young males more than any other group. Finally, we could not determine how many users were still vaping and how many were vape free nor we could link particular facilitators to quitting success.

Conclusions

The findings of this study reveal the unique aspects that encompass the process of quitting vaping from the perspective of e-cigarette users. This work not only validates the need for more stringent policies around e-cigarette consumption and use but also brings forward some important gaps in relation to efforts to meet the needs of e-cigarette users who want to quit. The findings offer some recommendations and a beginning framework for steering e-cigarette policy and intervention efforts so that they resonate with e-cigarette users.

Conflicts of Interest

None declared.

References

1. Jones L. Vaping: how popular are e-cigarettes? BBC News. 2019. URL: https://www.bbc.com/news/business-44295336 [accessed 2021-02-27]
2. Tobacco use by youth is rising. Vital Signs CDC. URL: https://www.cdc.gov/vitalsigns/youth-tobacco-use/ [accessed 2021-02-27]
3. Amato MS, Bottcher MM, Cha S, Jacobs MA, Pearson JL, Graham AL. "It's really addictive and I'm trapped:" a qualitative analysis of the reasons for quitting vaping among treatment-seeking young people. Addict Behav 2021 Jan;112:106599 [FREE Full text] [doi: 10.1016/j.addbeh.2020.106599] [Medline: 32950927]
4. Smith TT, Nahhas GJ, Carpenter MJ, Squeglia LM, Diaz VA, Leventhal AM, et al. Intention to quit vaping among United States adolescents. JAMA Pediatr 2021 Jan 01;175(1):97-99. [doi: 10.1001/jamapediatrics.2020.2348] [Medline: 32804194]
5. Sanchez S, Kaufman P, Pelletier H, Baskerville B, Feng P, O'Connor S, et al. Is vaping cessation like smoking cessation? A qualitative study exploring the responses of youth and young adults who vape e-cigarettes. Addict Behav 2021 Feb;113:106687. [doi: 10.1016/j.addbeh.2020.106687] [Medline: 33045643]
6. National Academies of Sciences, Engineering, and Medicine. Public Health Consequences of E-cigarettes. Washington, DC: National Academies Press; Jan 23, 2018.

7. Graham AL, Jacobs MA, Amato MS. Engagement and 3-month outcomes from a digital e-cigarette cessation program in a cohort of 27 000 teens and young adults. Nicotine Tob Res 2020 Apr 21;22(5):859-860 [FREE Full text] [doi: 10.1093/ntr/ntz097] [Medline: 31197320]
8. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. Smoking Cessation: A Report of the Surgeon General. Washington (DC): US Department of Health and Human Services; 2020.
9. Zhan Y, Zhang Z, Okamoto JM, Zeng DD, Leischow SJ. Underage juul use patterns: content analysis of reddit messages. J Med Internet Res 2019 Sep 09;21(9):e13038 [FREE Full text] [doi: 10.2196/13038] [Medline: 31502542]
10. Hausmann JS, Touloumtzis C, White MT, Colbert JA, Gooding HC. Adolescent and young adult use of social media for health and its implications. J Adolesc Health 2017 Jun;60(6):714-719 [FREE Full text] [doi: 10.1016/j.jadohealth.2016.12.025] [Medline: 28259620]
11. Goodyear VA, Armour KM, Wood H. Young people and their engagement with health-related social media: new perspectives. Sport Educ Soc 2018 Jan 25;24(7):673-688 [FREE Full text] [doi: 10.1080/13573322.2017.1423464] [Medline: 31814804]
12. Wang L, Zhan Y, Li Q, Zeng DD, Leischow SJ, Okamoto J. An examination of electronic cigarette content on social media: analysis of e-cigarette flavor content on reddit. Int J Environ Res Public Health 2015 Nov 20;12(11):14916-14935 [FREE Full text] [doi: 10.3390/ijerph121114916] [Medline: 26610541]
13. Hua M, Sadah S, Hristidis V, Talbot P. Health effects associated with electronic cigarette use: automated mining of online forums. J Med Internet Res 2020 Jan 03;22(1):e15684 [FREE Full text] [doi: 10.2196/15684] [Medline: 31899452]
14. Li Q, Zhan Y, Wang L, Leischow SJ, Zeng DD. Analysis of symptoms and their potential associations with e-liquids' components: a social media study. BMC Public Health 2016 Jul 30;16:674 [FREE Full text] [doi: 10.1186/s12889-016-3326-0] [Medline: 27475060]
15. Chen AT, Zhu S, Conway M. What online communities can tell us about electronic cigarettes and hookah use: a study using text mining and visualization techniques. J Med Internet Res 2015 Sep 29;17(9):e220 [FREE Full text] [doi: 10.2196/jmir.4517] [Medline: 26420469]
16. Kavuluru R, Han S, Hahn EJ. On the popularity of the USB flash drive-shaped electronic cigarette juul. Tob Control 2019 Jan;28(1):110-112 [FREE Full text] [doi: 10.1136/tobaccocontrol-2018-054259] [Medline: 29654121]
17. Brett EI, Stevens EM, Wagener TL, Leavens EL, Morgan TL, Cotton WD, et al. A content analysis of juul discussions on social media: using reddit to understand patterns and perceptions of juul use. Drug Alcohol Depend 2019 Jan 01;194:358-362. [doi: 10.1016/j.drugalcdep.2018.10.014] [Medline: 30472576]
18. Liu H, Li Q, Zhan Y, Zhang Z, Zeng DD, Leischow SJ. Characterizing social media messages related to underage juul e-cigarette buying and selling: cross-sectional analysis of reddit subreddits. J Med Internet Res 2020 Jul 20;22(7):e16962 [FREE Full text] [doi: 10.2196/16962] [Medline: 32706661]
19. Sowles SJ, McLeary M, Optican A, Cahn E, Krauss MJ, Fitzsimmons-Craft EE, et al. A content analysis of an online pro-eating disorder community on reddit. Body Image 2018 Mar;24:137-144 [FREE Full text] [doi: 10.1016/j.bodyim.2018.01.001] [Medline: 29414146]
20. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
21. What you need to know to quit smoking. Truth Initiatve. 2018. URL: https://truthinitiative.org/research-resources/quitting-smoking-vaping/what-you-need-know-quit-smoking [accessed 2021-02-28]
22. Chen R, Pierce JP, Leas EC, White MM, Kealey S, Strong DR, et al. Use of electronic cigarettes to aid long-term smoking cessation in the united states: prospective evidence from the path cohort study. Am J Epidemiol 2020 Dec 01;189(12):1529-1537 [FREE Full text] [doi: 10.1093/aje/kwaa161] [Medline: 32715314]
23. Pierce JP, Benmarhnia T, Chen R, White M, Abrams DB, Ambrose BK, et al. Role of e-cigarettes and pharmacotherapy during attempts to quit cigarette smoking: the PATH study 2013-16. PLoS One 2020 Sep 02;15(9):e0237938 [FREE Full text] [doi: 10.1371/journal.pone.0237938] [Medline: 32877429]
24. Unger JB, Rogers C, Barrington-Trimis J, Majmundar A, Sussman S, Allem J, et al. "I'm using cigarettes to quit juul": an analysis of twitter posts about juul cessation. Addict Behav Rep 2020 May 12;12:100286 [FREE Full text] [doi: 10.1016/j.abrep.2020.100286] [Medline: 32637562]
25. Kong G, Bold KW, Cavallo DA, Davis DR, Jackson A, Krishnan-Sarin S. Informing the development of adolescent e-cigarette cessation interventions: a qualitative study. Addict Behav 2021 Mar;114:106720. [doi: 10.1016/j.addbeh.2020.106720] [Medline: 33162230]
26. Wellman RJ, O'Loughlin EK, Dugas EN, Montreuil A, Dutczak H, O'Loughlin J. Reasons for quitting smoking in young adult cigarette smokers. Addict Behav 2018 Feb;77:28-33. [doi: 10.1016/j.addbeh.2017.09.010] [Medline: 28946012]
27. Goniewicz ML, Boykan R, Messina CR, Eliscu A, Tolentino J. High exposure to nicotine among adolescents who use juul and other vape pod systems ('pods'). Tob Control 2019 Nov;28(6):676-677 [FREE Full text] [doi: 10.1136/tobaccocontrol-2018-054565] [Medline: 30194085]
28. Hammond D, Rynard VL, Reid JL. Changes in prevalence of vaping among youths in the United States, Canada, and England from 2017 to 2019. JAMA Pediatr 2020 Aug 01;174(8):797-800 [FREE Full text] [doi: 10.1001/jamapediatrics.2020.0901] [Medline: 32364581]

29. Hammond D, Reid JL, Cole AG, Leatherdale ST. Electronic cigarette use and smoking initiation among youth: a longitudinal cohort study. CMAJ 2017 Oct 30;189(43):E1328-E1336 [FREE Full text] [doi: 10.1503/cmaj.161002] [Medline: 29084759]
30. Scott JC, Slomiak ST, Jones JD, Rosen AF, Moore TM, Gur RC. Association of cannabis with cognitive functioning in adolescents and young adults: a systematic review and meta-analysis. JAMA Psychiatry 2018 Jun 01;75(6):585-595 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.0335] [Medline: 29710074]
31. Counotte DS, Goriounova NA, Moretti M, Smoluch MT, Irth H, Clementi F, et al. Adolescent nicotine exposure transiently increases high-affinity nicotinic receptors and modulates inhibitory synaptic transmission in rat medial prefrontal cortex. FASEB J 2012 May;26(5):1810-1820 [FREE Full text] [doi: 10.1096/fj.11-198994] [Medline: 22308197]
32. Benowitz NL. Nicotine addiction. N Engl J Med 2010 Jun 17;362(24):2295-2303 [FREE Full text] [doi: 10.1056/NEJMra0809890] [Medline: 20554984]
33. Goriounova NA, Mansvelder HD. Short- and long-term consequences of nicotine exposure during adolescence for prefrontal cortex neuronal network function. Cold Spring Harb Perspect Med 2012 Dec 01;2(12):a012120 [FREE Full text] [doi: 10.1101/cshperspect.a012120] [Medline: 22983224]
34. Health canada proposes reducing nicotine concentration limit in vaping products. CTV News. 2020. URL: https://www.ctvnews.ca/health/health-canada-proposes-reducing-nicotine-concentration-limit-in-vaping-products-1.5237744 [accessed 2021-02-28]
35. Chu K, Colditz JB, Primack BA, Shensa A, Allem J, Miller E, et al. Juul: spreading online and offline. J Adolesc Health 2018 Nov;63(5):582-586 [FREE Full text] [doi: 10.1016/j.jadohealth.2018.08.002] [Medline: 30348280]
36. Graham AL, Jacobs MA, Amato MS, Cha S, Bottcher MM, Papandonatos GD. Effectiveness of a quit vaping text message program in promoting abstinence among young adult e-cigarette users: protocol for a randomized controlled trial. JMIR Res Protoc 2020 May 01;9(5):e18327 [FREE Full text] [doi: 10.2196/18327] [Medline: 32356774]
37. King JL, Reboussin BA, Spangler J, Cornacchione Ross J, Sutfin EL. Tobacco product use and mental health status among young adults. Addict Behav 2018 Feb;77:67-72 [FREE Full text] [doi: 10.1016/j.addbeh.2017.09.012] [Medline: 28965069]
38. Mental health disorder statistics. Johns Hopkins Medicine. URL: https://www.hopkinsmedicine.org/health/wellness-and-prevention/mental-health-disorder-statistics [accessed 2021-02-28]
39. Panchal N, Rabah K, Cox C, Garfield R. The Implications of COVID-19 for Mental Health and Substance Use. Kaiser Family Foundation. URL: https://www.kff.org/coronavirus-covid-19/issue-brief/the-implications-of-covid-19-for-mental-health-and-substance-use/ [accessed 2021-02-28]
40. Dai H, Hao J. Online popularity of juul and puff bars in the USA: 2019-2020. Tob Control 2020 Oct 13:tobaccocontrol-2020-055727 (forthcoming). [doi: 10.1136/tobaccocontrol-2020-055727] [Medline: 33051277]
41. D'Angelo H, Rose SW, Golden SD, Queen T, Ribisl KM. E-cigarette availability, price promotions and marketing at the point-of sale in the contiguous United States (2014-2015): national estimates and multilevel correlates. Prev Med Rep 2020 Jun 26;19:101152 [FREE Full text] [doi: 10.1016/j.pmedr.2020.101152] [Medline: 32670780]
42. Hammond D, White CM, Czoli CD, Martin CL, Magennis P, Shiplo S. Retail availability and marketing of electronic cigarettes in Canada. Can J Public Health 2015 Oct 09;106(6):e408-e412 [FREE Full text] [doi: 10.17269/cjph.106.5105] [Medline: 26680433]
43. Health canada confirms ban of advertising for vaping products wherever they can be seen or heard by youth. Health Canada. URL: https://www.canada.ca/en/health-canada/news/2020/07/health-canada-confirms-ban-of-advertising-for-vaping-products-wherever-they-can-be-seen-or-heard-by-youth.html [accessed 2020-11-09]
44. Wilson N, Hoek J, Thomson G, Edwards R. Should e-cigarette use be included in indoor smoking bans? Bull World Health Organ 2017 Jul 01;95(7):540-541 [FREE Full text] [doi: 10.2471/BLT.16.186536] [Medline: 28670020]
45. Struik LL, Dow-Fleisner S, Belliveau M, Thompson D, Janke R. Tactics for drawing youth to vaping: content analysis of electronic cigarette advertisements. J Med Internet Res 2020 Aug 14;22(8):e18943 [FREE Full text] [doi: 10.2196/18943] [Medline: 32663163]
46. Blu E Cigarette. YouTube. URL: https://www.youtube.com/watch?v=oQelKE-aBNg [accessed 2021-02-28]
47. Baskerville NB, Struik LL, Dash D. Crush the crave: development and formative evaluation of a smartphone app for smoking cessation. JMIR Mhealth Uhealth 2018 Mar 02;6(3):e52 [FREE Full text] [doi: 10.2196/mhealth.9011] [Medline: 29500157]
48. Sattelberg W. The demographics of reddit: who uses the site? ALPHR. 2021. URL: https://www.alphr.com/demographics-reddit/ [accessed 2021-04-15]

Abbreviations

NRT: nicotine replacement therapy

Edited by R Kukafka; submitted 01.03.21; peer-reviewed by S Sanchez, N Jiang; comments to author 23.03.21; revised version received 20.04.21; accepted 19.07.21; published 25.10.21.

Please cite as:
Struik L, Yang Y
e-Cigarette Cessation: Content Analysis of a Quit Vaping Community on Reddit
J Med Internet Res 2021;23(10):e28303
URL: https://www.jmir.org/2021/10/e28303
doi:10.2196/28303
PMID:34694229

Original Paper

Impulsivity, Self-control, Interpersonal Influences, and Maladaptive Cognitions as Factors of Internet Gaming Disorder Among Adolescents in China: Cross-sectional Mediation Study

Yanqiu Yu[1], PhD; Phoenix Kit-Han Mo[1], PhD; Jianxin Zhang[2], MD; Jibin Li[3], PhD; Joseph Tak-Fai Lau[1], PhD

[1]Center for Health Behaviours Research, Jockey Club School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong, Hong Kong

[2]West China School of Public Health, Sichuan University, Chengdu, China

[3]Department of Clinical Research, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou, China

Corresponding Author:
Joseph Tak-Fai Lau, PhD
Center for Health Behaviours Research
Jockey Club School of Public Health and Primary Care
The Chinese University of Hong Kong
Ngan Shing St
Sha Tin
Hong Kong
Hong Kong
Phone: 852 2252 8713
Email: jlau@cuhk.edu.hk

Abstract

Background: Gaming disorder, including internet gaming disorder (IGD), was recently defined by the World Health Organization as a mental disease in the 11th Revision of the International Classification of Diseases (ICD-11). Thus, reducing IGD is warranted. Maladaptive cognitions related to internet gaming (MCIG) have been associated with IGD, while impulsivity, self-control, parental influences, and peer influences are key risk factors of IGD. Previous literature suggests that MCIG is associated with the aforementioned 4 risk factors and IGD, and may thus mediate between these risk factors and IGD. These potential mediations, if significant, imply that modification of MCIG may possibly alleviate these risk factors' harmful impacts on increasing IGD. These mediation hypotheses were tested in this study for the first time.

Objective: This study tested the mediation effects of MCIG between intrapersonal factors (impulsivity and self-control) and IGD, and between interpersonal factors (parental influences and peer influences) and IGD among adolescents in China.

Methods: An anonymous, cross-sectional, and self-administered survey was conducted among secondary school students in classroom settings in Guangzhou and Chengdu, China. All grade 7 to 9 students (7 to 9 years of formal education) of 7 secondary schools were invited to join the study, and 3087 completed the survey. The *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-5*) checklist was used to assess IGD. MCIG was assessed by using the Chinese version of the Revised Internet Gaming Cognition Scale. Impulsivity, self-control, and parental or peer influences were measured by using the motor subscale of the Barratt Impulsiveness Scale, the Brief Self-Control Scale, and the modified interpersonal influence scale, respectively. Structural equation modeling was conducted to examine the mediation effects of MCIG between these risk factors and IGD.

Results: The prevalence of IGD was 13.57% (418/3081) and 17.67% (366/2071) among all participants and adolescent internet gamers, respectively. The 3 types of MCIG (perceived rewards of internet gaming, perceived urges for playing internet games, and perceived unwillingness to stop playing without completion of gaming tasks) were positively associated with IGD. Impulsivity, self-control, parental influences, and peer influences were all significantly associated with the 3 types of MCIG and IGD. The 3 types of MCIG partially mediated the associations between the studied factors and IGD (effect size of 30.0% to 37.8%).

Conclusions: Impulsivity, self-control, and interpersonal influences had both direct and indirect effects via MCIG on IGD. Modifications of the 3 types of MCIG can potentially reduce the harmful impacts of impulsivity and interpersonal influences on IGD and enhance the protective effect of self-control against IGD. Future longitudinal studies are warranted.

(J Med Internet Res 2021;23(10):e26810) doi:10.2196/26810

KEYWORDS

cognition; adolescent health; health risk behaviors; internet; self-control; China

Introduction

After the inclusion of internet gaming disorder (IGD) in the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (*DSM-5*) in 2013 [1], World Health Organization (WHO) recognized gaming disorder (online and offline) as a disease in the 11th Revision of *International Classification of Diseases* (*ICD-11*) in 2019 [2]. The reported range of prevalence of IGD among Chinese adolescents is wide (2.4% to 21.5%), possibly due to methodological differences [3]. Adolescents are vulnerable to IGD, which has many adverse consequences (eg, loneliness and depression [4-6]).

Maladaptive cognition is an important intrapersonal factor of addictive behaviors like pathological gambling [7] and internet addiction [8]; maladaptive cognitions related to internet gaming (MCIG) have also been associated with IGD [8-10]. A systematic review of 36 studies of MCIG proposed a 4-factor cognitive framework (ie, overvaluation of gaming rewards, maladaptive rules, gaming for self-esteem, and gaming for social acceptance) [11]; the 4 types of MCIG were all positively associated with IGD among adolescents [9]. A recent validation study modified this 4-factor model and revealed a new 3-factor model (ie, perceived rewards of internet gaming, perceived urges for playing internet games, and perceived unwillingness to stop playing without completion of gaming tasks) that demonstrated satisfactory psychometric properties [12]. The 3 domains of MCIG were all positively correlated with IGD [12]. The revised scale was used in this study.

Impulsivity and self-control are 2 important intrapersonal factors of addictive behaviors, including substance use [13,14], smoking [15], alcohol drinking [16], internet addiction [17], and IGD [4,18-21]. Impulsivity and self-control fit into the dual-process theories of decision-making for risk behaviors [22-25]. Impulsivity represents reactive, intuitive, and affective processes with high responsiveness to temptations and prompt actions without deliberation [24]. In contrast, self-control reflects the reasoned process and ability in regulating impulses deliberatively [26,27]. The 2 processes jointly affect decisions in the performance of risk behaviors [22-25]. Adolescents with high impulsivity may exhibit heightened spontaneous responses to behavioral cues to internet gaming, while those with low self-control may find it difficult to resist the temptation of playing internet games and stop playing.

Interpersonal influences are important factors of IGD. Empirical studies have reported positive associations between interpersonal influences (eg, parents' invitations of playing internet games and intensity of peers' gaming behaviors) and IGD or internet addiction among high schools students [28,29]. Interpersonal influences may affect IGD in different ways. Significant others' direct invitations to play internet games may trigger prompt engagement in internet gaming. According to the health belief model, such invitations to play internet games represent cues to action, which is an important determinant of health-related behavior [30]. Frequent invitations may increase adolescents' gaming intensity, which is associated with risk of IGD [28]. In addition, the reciprocal determinism construct of social cognitive theory (SCT) postulates that one's environment, personal factors, and health-related behaviors interact with each other [31]. Having significant others playing internet games frequently forms a social environment that may increase adolescents' gaming frequency. More importantly, SCT postulates that observational learning is an important factor of health-related behavior [31]; adolescents may play internet games frequently through observational learning from their significant others' frequent internet gaming.

Importantly, MCIG are potential mediators of the association between impulsivity or self-control and IGD. Brand's model [32] proposes that dysfunctional personality traits increase addictiveness to specific internet applications (including internet gaming), through mediation of changes in related cognitions (eg, expectations). Cognition of outcome expectancy regarding addictive behaviors has been shown to partially mediate the associations between impulsivity (a type of dysfunctional personality trait) and addictive behavior (eg, substance use and alcohol drinking) [33,34]. These studies, however, did not look at IGD. Following Brand's model, this study hence tested whether MCIG would mediate the association between impulsivity and IGD. This contention is indirectly supported by a study that reported a partial mediation of the association between maladaptive personality traits (ie, negative affectivity, detachment, and psychoticism) and IGD via the cognition of expectancies related to using internet gaming to escape from reality [35]. That study, however, did not include other types of MCIG. Furthermore, we did not locate studies that looked at MCIG as a mediator between self-control and IGD. This study thus sought to fill some of deficiencies in the relevant literature.

The mediation between interpersonal influences and IGD via MCIG was also tested in this study. The hypothesis is supported by the theory of triadic influence, which is a framework that incorporates influential factors of addictive behaviors from a number of psychological and sociological theories [36]. It distinguishes between distal or ultimate factors that influence proximal factors of changes in health-related behaviors [37]. According to the theory of triadic influence, interpersonal factors are distal or ultimate variables while cognitive factors related to a certain behavior are proximal variables [36]; the theory suggests that cognitive factors (eg, outcome expectancy) mediate between interpersonal influences and health-related behaviors [38,39]. Such mediations were found for some addictive behaviors (eg, substance use [40], alcohol drinking [41], and internet addiction [29]). One study also reported full and partial mediation of the cognition of positive outcome expectancy of internet gaming between interpersonal influences (peers' positive attitude toward gaming, intensity of peers' internet

game use, and the frequency of peers' invitation to play internet games) and IGD among high school students [28]. To our knowledge, however, no study has looked at the mediation effects of other types of MCIG.

We thus investigated the factors of IGD, including impulsivity, self-control, interpersonal influences exerted by parents and peers, and the 3 types of MCIG (perceived rewards, perceived urges, and perceived unwillingness to stop playing) among junior middle school students (grade 7 to 9 of formal education) in 2 populous cities in China. We then tested the mediation effects of MCIG on the associations between impulsivity or self-control and IGD, and between interpersonal influences and IGD, respectively. We hypothesized that the direct and indirect effects would be statistically significant.

Methods

Participants and Data Collection

The cross-sectional survey was conducted among secondary school students in Guangzhou and Chengdu in China from October 2018 to December of 2018. The 2 cities are located in southern and southwestern China, which had populations of 14.9 and 16.3 million people in 2018, respectively, and 0.36 and 0.40 million secondary school students in 2018, respectively [42]. Seven junior middle schools (4 out of 409 from Guangzhou and 3 out of 460 from Chengdu) were conveniently selected and participated in the study. All the Grade 7 students (7 years of formal education) in Guangzhou and all the Grade 7 to 9 students in Chengdu were invited to join the survey. The inclusion criteria were those who were full-time Grade 7 students and Grade 7-9 students of the participating schools in Guangzhou and Chengdu, respectively; and those willing to participate in the study. The procedure of data collection was described in a published study that used a subsample of the survey [12], and it is briefly introduced here. Under the supervision of well-trained field workers, students self-administered an anonymous structured questionnaire in the absence of teachers in classroom settings. They were briefed about the objectives of the survey, the return of completed questionnaires implied informed consent, and the students had the right to quit at any time without any negative consequences. No incentives were given to the students. The study was approved by the Survey and Behavioral Research Ethics Committee of the Chinese University of Hong Kong (#SBRE-18-430).

Of the 4350 students invited to participate in the present study, 3147 (72.34%) returned the questionnaire. Among all the returned questionnaires, 66 (2.10%) were removed from data analyses as there were more than 20% of missing data in their response items. Data obtained from the remaining 3081 students (97.90%) were used for data analysis (1126 from Guangzhou [36.55%] and 1955 from Chengdu [63.45%]).

Measures

Background Variables

Background information was collected, including sex, grades, being born in the studied city, whether living with both parents, both father's and mother's educational levels (junior middle school or below, senior middle school or equal, or college or above), household income level compared with their classmates (5 points: much higher to much lower), and self-reported academic performance (3 points: above average, average, and below average).

IGD Assessment

The 9-item *DSM-5* checklist was used to assess IGD [43]; it recorded the presence of addictive symptoms, including preoccupation, withdrawal, tolerance, inability to control internet gaming, loss of interest in other activities, psychological or social problems, deception, avoidance, and significant loss due to internet gaming. IGD is defined by endorsement of ≥ 5 items (yes-no response options). The Chinese version of *DSM-5* has been validated as having good psychometric properties and diagnostic validity [44,45]. The Cronbach α of the checklist was .79 in the present study.

Maladaptive Cognitions Related to Internet Gaming

MCIG was measured by using the Chinese version of the Revised Internet Gaming Cognition Scale (C-RIGCS). It consists of 3 subscales: perceived rewards of internet gaming, perceived urges for playing internet games, and perceived unwillingness to stop playing without completion of gaming tasks. Sample items are "I feel more in control when I play internet games," "I would feel bad if I was not able to play internet games," and "I feel uncomfortable thinking about my unfinished goals or objectives in internet games." The C-RIGCS has been validated in Chinese adolescents and has shown acceptable psychometric properties [12]. The items were rated with 5-point Likert scales (0=never to 4=always), with higher scores indicating higher levels of MCIG. The Cronbach α of the overall scale and its 3 subscales in the present study were .91, .86, .81, and .74, respectively.

Impulsivity

Impulsivity was measured by using the 10-item motor impulsiveness subscale of the Barratt Impulsiveness Scale, which indicates the tendency to act on the spur of the moment and with fast reactions [46]. The Chinese version made some cultural adaptations and showed good reliability and construct validity in Chinese adolescents [47]. A sample item is "I do things without thinking." The items were rated with 5-point Likert scales (1=completely disagree to 5=completely agree), with higher scores indicating higher levels of impulsivity. The Cronbach α of the scale was .91 in this study.

Self-Control

Self-control was measured by using the 13-item Brief Self-Control Scale [26], which demonstrated good psychometric properties in Chinese adolescents [48]. A sample item is "I am good at resisting temptation." The items were rated with 5-point Likert scales (1=never to 5=always), with higher scores indicating higher levels of self-control. The Cronbach α of the scale was .74 in this study.

Interpersonal Influences

Interpersonal influences were measured by revising the 6 items that assessed similar tendencies in a previous study [28]. The

items included frequency of being invited to play internet games from parents and peers, perceived parents' and peers' gaming intensity, and perceived parental influences and peer influences on current internet gaming behavior. Sample items were "How often do your parents invite you to play internet games?", "How often do your parents play internet games?", and "To what extent do you think your parents affect your internet gaming behavior?" Confirmatory factor analysis was conducted to examine the 2-factor structure (parental influences and peer influences) of the 6 items, which showed an acceptable goodness of fit (comparative fit index [CFI]=0.97, Tucker-Lewis index [TLI]=0.97, and root mean square error of approximation [RMSEA]=0.09). The items were rated with 4-point Likert scales (1=never/nil to 4=always/severe), with higher scores indicating higher levels of interpersonal influences. The Cronbach α of the overall scale and its 2 subscales in this study were .73, .63 (a Cronbach α >.60 was considered acceptable in previous literature [49,50]), and .71, respectively.

Statistical Analysis

IGD was used as the binary dependent variable. Univariate logistic regression analysis was conducted to establish the associations between the studied background variables and IGD; crude odds ratios (ORcs) and their respective 95% CIs were derived. Pearson correlation coefficients (r_p) and Spearman correlation coefficients (r_s) were derived for continuous and ordinal variables, respectively. The mediation effects were tested by using structural equation modeling (SEM) with weighted least square mean and variance-adjusted estimation. Three latent variables were created: (1) impulsivity or low self-control was derived from the scale scores of impulsivity and self-control (reversed scores), (2) interpersonal influences was derived from the subscale scores of parental influences and peer influences, and (3) maladaptive cognitions was derived from the subscale scores of the 3 types of MCIG. The paths between the 3 latent variables and IGD were fit to test the mediation hypotheses. Recommended goodness-of-fit indicators included CFI ≥0.90, TLI ≥0.90, and RMSEA ≤0.08. The SEM was conducted by using Mplus 7.0; other statistical analyses were performed with SPSS version 21.0 (IBM Corp). Statistical significance was defined as a 2-tailed *P* value <.05.

Results

Descriptive Statistics

More than half of the participants were males (1550/3081, 50.31%) and first-year students (1979/3081, 64.23%). More than one-fifth were not born in the city where the study was conducted (691/3081, 22.43%) and did not live with both parents (639/3081, 20.74%). Around one-fifth of the participants' fathers (594/3081, 19.28%) and mothers (566/3081, 18.37%) had received tertiary education or above; 12.56% (387/3081) self-perceived a lower or much lower household income level than did their classmates, and 19.47% (600/3081) self-reported a below-average academic performance (see Table 1). The mean of impulsivity, self-control, parental influences, and peer influences were 22.9 (SD 7.6, range 10-50), 44.4 (SD 7.6, range 13-65), 4.8 (SD 1.8, range 3-12), and 6.5 (SD 2.1, range 3-12), respectively. Similarly, the mean of the overall C-RIGCS and its 3 subscales were 17.0 (SD 11.6, range 0-60), 7.2 (SD 5.9, range 0-28), 3.7 (SD 3.5, range 0-16), and 6.1 (SD 3.7, range 0-16), respectively.

Table 1. Background characteristics of participants (N=3081).

Characteristic	Value, n (%)
Sex	
Female	1525 (49.49)
Male	1550 (50.31)
Missing data	6 (0.19)
Grade	
Seven	1979 (64.23)
Eight	579 (18.79)
Nine	523 (16.98)
Study site	
Guangzhou	1126 (36.55)
Chengdu	1955 (63.45)
Born in the city where the study was conducted	
Yes	2367 (76.83)
No	691 (22.43)
Missing data	23 (0.75)
Living with both parents	
Yes	2382 (77.31)
No	639 (20.74)
Missing data	60 (1.95)
Father's educational level	
Junior middle school or below	1489 (48.33)
Senior middle school or equal	831 (26.97)
College or above	594 (19.28)
Missing data	167 (5.42)
Mother's educational level	
Junior middle school or below	1532 (49.72)
Senior middle school or equal	803 (26.06)
College or above	566 (18.37)
Missing data	180 (5.84)
Household income level when compared with classmates	
Much higher/higher	672 (21.81)
Moderate	2002 (64.98)
Lower/much lower	387 (12.56)
Missing data	20 (0.65)
Self-reported academic performance	
Above average	1020 (33.11)
Average	1348 (43.75)
Below average	600 (19.47)
Missing data	113 (3.67)

Prevalence of IGD

The prevalence of IGD was 13.57% (418/3081; 95% CI 12.4%-14.5%) among all participants. Among those who had played internet games in the past 12 months (2071/3081, 67.22% of all participants), the prevalence of IGD was 17.67% (366/2071; 95% CI 16.0%-19.3%).

Associations Between Background Variables and IGD

The univariate logistic regression analyses showed that the background variables were all significantly associated with IGD, except for place of birth (whether born in the city where the study was conducted; Table 2). Significant factors included sex (males vs females: ORc=2.80, 95% CI 2.23-3.51), student grade (Grade 8 vs 7: ORc=1.90, 95% CI 1.49-2.44; Grade 9 vs 7: ORc=1.36, 95% CI 1.03-1.80), study site (Chengdu vs Guangzhou: ORc=2.12, 95% CI 1.66-2.69), living arrangement (not living vs living with both parents: ORc=1.54, 95% CI 1.22-1.95), household income level (self-perceived lower or much lower vs higher or much higher than other classmates: ORc=1.64, 95% CI 1.17-2.30), parental education (father's tertiary vs primary education or below: ORc=0.53, 95% CI 0.39-0.73; mothers' tertiary vs primary education or below: ORc=0.72, 95% CI 0.53-0.97), and self-reported academic performance (average vs above average: ORc=1.35, 95% CI 1.04-1.76; below average vs above average: ORc=2.53, 95% CI 1.90-3.36).

Table 2. Univariate logistic regression analysis on the associations between the studied background variables and internet gaming disorder (N=3081).

Background variables	IGD[a], n (%)	Association, ORc[b] (95% CI)
Sex		
Female[c]	119 (7.8)	N/A[d]
Male	297 (19.16)	2.80 (2.23-3.51)***
Grade		
Seven[c]	226 (11.42)	N/A
Eight	114 (19.69)	1.90 (1.49-2.44)***
Nine	78 (14.91)	1.36 (1.03-1.80)*
Study site		
Guangzhou[c]	96 (8.53)	N/A
Chengdu	322 (16.47)	2.12 (1.66-2.69)***
Born in the city where the study was conducted[e]		
Yes[c]	308 (13.01)	N/A
No	105 (15.2)	1.20 (0.94-1.52)
Living with both parents[e]		
Yes[c]	294 (12.34)	
No	114 (17.84)	1.54 (1.22-1.95)***
Father's educational level[e]		
Junior middle school or below[c]	227 (15.25)	N/A
Senior middle school or equal	109 (13.12)	0.84 (0.66-1.07)
College or above	52 (8.75)	0.53 (0.39-0.73)***
Mother's educational level[e]		
Junior middle school or below[c]	218 (14.23)	N/A
Senior middle school or equal	102 (12.7)	0.88 (0.68-1.13)
College or above	60 (10.6)	0.72 (0.53-0.97)*
Household income level when compared with classmates[e]		
Much higher/higher[c]	87 (12.95)	N/A
Moderate	246 (12.29)	0.94 (0.73-1.22)
Lower/much lower	76 (19.64)	1.64 (1.17-2.30)**
Self-reported academic performance[e]		
Above average[c]	98 (9.61)	N/A
Average	169 (12.54)	1.35 (1.04-1.76)*
Below average	127 (21.17)	2.53 (1.90-3.36)***

[a]IGD: internet gaming disorder.

[b]ORc: crude odds ratio.

[c]Reference=1.0

[d]N/A: not applicable.

[e]Missing data were excluded from the analysis.

*P<.05.

**P<.01.

***P<.001.

Correlations Among the Studied Variables

The 3 studied risk factors (impulsivity, parental influences, and peer influences) were all positively correlated with the overall C-RIGCS and its 3 subscales representing MCIG (r_p ranged from 0.19 to 0.39; *P*<.001) and IGD (r_s ranged from 0.24 to 0.26; *P*<.001), respectively. Self-control was negatively correlated with the overall C-RIGCS and its 3 subscales (r_p ranged from –0.45 to –0.27; *P*<.001) and IGD (r_s=–0.32; *P*<.001), respectively. Besides the above correlations, all the studied factors of IGD were significantly correlated with each other (see Table 3).

Table 3. Correlations among impulsivity, self-control, interpersonal influences, maladaptive cognitions, and internet gaming disorder (N=3081)[a].

Major variable	1	2	3	4	5	6	7	8	9
1. Impulsivity[b]	—[c]								
2. Self-control	0.58*	—							
3. Parental influences[b]	0.26*	0.24*	—						
4. Peer influences[b]	0.24*	0.21*	0.40*	—					
5. Overall maladaptive cognitions[b]	0.37*	0.38*	0.30*	0.39*	—				
6. Perceived rewards of internet gaming[b]	0.31*	0.32*	0.29*	0.37*	0.93*	—			
7. Perceived urges for playing internet games[b]	0.39*	0.45*	0.28*	0.35*	0.85*	0.69*	—		
8. Perceived unwillingness to stop playing without completion of gaming tasks[b]	0.28*	0.27*	0.19*	0.30*	0.85*	0.68*	0.61*	—	
9. IGD[d,e]	0.30*	0.41*	0.37*	0.42*	0.24*	0.26*	0.26*	0.32*	—

[a]Missing data were excluded from the analyses.

[b]Pearson correlation analyses.

[c]Not applicable.

[d]Spearman correlation analyses.

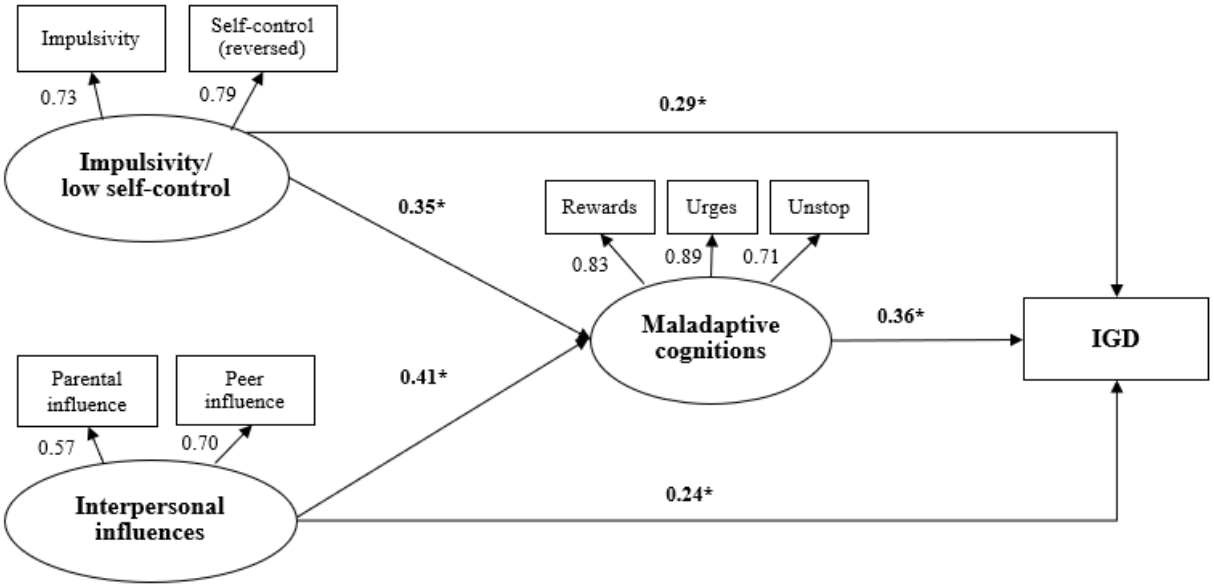
[e]IGD: internet gaming disorder.

*P<.001.

The SEM Model Testing the Mediation Hypotheses

Figure 1 presents the SEM model that demonstrated a satisfactory model fit (CFI=0.95, TLI=0.90, and RMSEA=0.08); the factor loadings of the 3 latent variables ranged from 0.57 to 0.89 (all *P*<.001). The findings revealed that maladaptive cognitions partially mediated the association between impulsivity or low self-control and IGD (mediation effect size=30.0%; *P* of Sobel test <.001), and between interpersonal influences and IGD (mediation effect size=37.8%; *P* of Sobel test <.001), respectively. Impulsivity or low self-control (standardized β=.29; *P*<.001) and interpersonal influences (standardized β=.24; *P*<.001) had significant direct effects on IGD. The standardized β values of the other paths are presented in Figure 1.

Figure 1. The structural equation model testing the mediation hypotheses. IGD: internet gaming disorder. *P<.001.



Discussion

This study revealed an alarmingly high prevalence of IGD (around 14% among all adolescent participants and 18% among adolescent internet gamers). The prevalence was comparable to the 13% among adolescents of 2 other Chinese cities [51], but much higher than the 2% among Chinese Macau adults [52] and the 3.1% among Australian adolescents [9]. These studies all used *DSM-5* criteria to assess IGD. The high prevalence of IGD signifies the need to regulate internet gaming behaviors and conduct interventions to reduce IGD among adolescents in China, as adolescent IGD has been positively associated with various behavioral problems (eg, aggression and violence [53,54]) and mental health problems (eg, loneliness and depression [4-6]).

In our study, a number of background factors were associated with a higher risk of IGD, including male sex, higher grades, not living with both parents, and self-reported below-average academic performance. First, corroborating previous studies [18,55,56], males were associated with a higher risk of IGD. This finding may be explained by the sex differences in brain responses [57], motives of internet gaming [58], and a less female-friendly gaming environment [59], among other plausible factors. Second, higher grades were positively associated with IGD in this study. Similar findings have been reported [9], but the direction of the association between age and IGD was mixed [18,60]. Future longitudinal studies are warranted to examine the associations between sex, age, grade, and IGD to develop sex- and age-specific interventions to reduce IGD. Furthermore, disadvantaged adolescents, including those of lower socioeconomic status (eg, self-perceived lower household income) and those not living with both parents were more likely than others to have IGD. Similar results were reported in previous longitudinal and cross-sectional studies [61,62]. Disadvantaged adolescents might have fewer resources for leisure activities and less parental monitoring over internet use, which may lead to more intensive internet gaming and higher risk of developing IGD [61]. A negative association between academic performance and IGD was also reported in our study, which was consistent with previous literature [18,63,64]. It is worth noting that the relationship between academic performance and IGD may be bidirectional. Frustration over poor academic performance may lead to problematic gaming as a means of escape and maladaptive coping [65], which has been associated with IGD [66]. Conversely, IGD may lead to poor academic performance [5,67]. The causal direction needs to be confirmed by longitudinal studies.

To facilitate the design of effective interventions for reducing IGD, it is important to understand the mechanisms (mediation) underlying the associations between risk or protective factors of IGD and IGD. The findings reveal that some intrapersonal (impulsivity) and interpersonal (interpersonal influences) risk factors may elevate the levels of the 3 types of MCIG, which may in turn increase the risk of developing IGD. Furthermore, self-control may reduce MCIG, which may increase IGD (a partial mediation effect). Nonetheless, the presence of significant direct effects between impulsivity, self-control, and interpersonal influences and IGD imply the existence of other unstudied mediators. For instance, coping is a potential mediator, as Brand's model postulates dysfunctional copying strategies mediate between personality traits and problematic use of the internet (including IGD) [32], while impulsivity can be considered a personality trait. Interpersonal influences may also strengthen the subjective norms of internet gaming (ie, significant others' support for internet gaming), which is a construct of the theory of planned behaviors [68]; these subjective norms are expected to be associated with MCIG. Future studies should look at other mediators.

The observed mediation effects suggest that modifications of the 3 types of MCIG can potentially reduce the harmful impacts of impulsivity or interpersonal influences on IGD and increase the protective effect of self-control against IGD. Targeted interventions to improve MCIG may include training to increase awareness of the 3 types of MCIG and skills to perform related

cognitive reconstructions [69], provision of alternative sources of rewards (eg, outdoors activities) to reduce perceived rewards of internet gaming, removal of sources of temptations and stimuli (eg, gaming devices), and introduction of distraction skills to reduce perceived urges or unwillingness to stop playing internet games [70,71].

There are also plausible explanations for why impulsivity was positively associated with the 3 types of MCIG. First, impulsivity may increase reward sensitivity that enhances adolescents' drives to seeking more rewards from addictive behaviors [72,73]; impulsive adolescents may thus possess a higher reward drive and perceive more rewards from internet gaming. Second, impulsive adolescents may be more responsive to cues of internet gaming and thus hold stronger urges for playing. Third, impulsivity in general may reduce impulse inhibitions against addictive behaviors, even in the presence of negative consequences [74-76]; adolescents with weakened inhibitions of gaming impulses may thus be less able to resist stimuli inductive to playing internet games and may perceive stronger unwillingness to stop playing. Self-control was negatively associated with the 3 types of MCIG, possibly because of the negative association between impulsivity and self-control [22-25], but other reasons may also apply. Interventions for modifying impulsivity and self-control may remove temptations and strengthen self-efficacy in regulating impulses [77]). In particular, the if-then planning intervention that specifies when, where, and how to regulate impulses is potentially useful [78]; a review reported that this type of intervention showed efficacy in reducing addictive behaviors (eg, in binge drinking and cigarette smoking) [79].

This study also found positive associations between interpersonal influences and MCIG. Having significant others who are frequent internet gamers was positively associated with MCIG. The SCT suggests that, through reciprocal determinism and observational learning, social interactions may influence both adolescents' attitudes and behavior [31]. It is likely that adolescents' parents or peers who play internet games frequently may also perceive higher levels of MCIG (eg, the 3 types of MCIG). Adolescents' maladaptive cognitions may be influenced by those of their parents via multiple means, such as social learning, subjective norms, and reinforcement. Furthermore, peers influence each other to form common perceptions regarding particular behaviors [80]. Future confirmation is needed. Family-based interventions for reducing IGD are potentially useful. A psychoeducation on maladaptive and adaptive use related to internet games showed preliminary effectiveness [81]. Another intervention that involved parental monitoring on adolescents' self-regulation practice and gaming behaviors showed short-term (3 month) efficacy in improving attitude, knowledge, self-regulation, and IGD among adolescents [82]. Although a meta-analysis showed that peer-led interventions were efficacious in reducing tobacco, alcohol, and substance use among adolescents [83], our literature search did not locate similar interventions for IGD. Future evidence-based interventions are warranted.

This study has several limitations. First, reporting bias, such as recall bias and social desirability bias, might have been introduced. Second, as the participating schools were selected based on convenience, there might have been selection bias, and generalization of the study results should be done cautiously. Third, the prevalence of IGD assessed by using the *DSM-5* checklist might have been overestimated compared with that based on the *ICD-11* criteria [84]. Fourth, we were unable to make causal inferences due to the cross-sectional nature of this study. Fifth, the revised assessment tool of interpersonal influences has not been validated although we conducted confirmatory factor analysis to test its 2-factor structure, which showed an acceptable goodness of fit. Finally, this study only investigated internet gaming but not offline video games, while both online and offline video games are included in the *ICD-11* under the category of gaming disorder, as the 2 types of gaming disorder differ in etiology, epidemiology, and treatment.

In conclusion, MCIG partially mediated the associations between impulsivity, self-control, or interpersonal influences and IGD. Modifications of the 3 types of MCIG may effectively reduce the harmful impacts of impulsivity or interpersonal influences on IGD and increase the protective effect of self-control against IGD. Future longitudinal studies are warranted to verify these findings and explore other potential mediators.

Acknowledgments

This research was funded by National Natural Science Foundation of China (grant #81373021). We would like to thank all the participants for their contribution.

Authors' Contributions

JTFL and YY conceived the study. YY, PKHM, and JTFL were responsible for the methodology. JZ and JL conducted the investigation. YY was responsible for the software and data curation, and conducted the formal analysis. PKHM and JTFL were responsible for validation. JZ, JL, and JTFL acquired the study resources. YY and JTFL wrote the original draft. YY, PKHM, and JTFL reviewed and edited the draft. JTFL conducted supervision. JL and JTFL acquired funding for the study.

Conflicts of Interest

None declared.

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM–5). Arlington: American Psychiatric Association; 2013.
2. World Health Organization. The Eleventh Revision of the International Classification of Diseases. Geneva: World Health Organization; 2018.
3. Paulus FW, Ohmann S, von Gontard A, Popow C. Internet gaming disorder in children and adolescents: a systematic review. Dev Med Child Neurol 2018 Apr 06;60(7):645-659. [doi: 10.1111/dmcn.13754] [Medline: 29633243]
4. Gentile DA, Choo H, Liau A, Sim T, Li D, Fung D, et al. Pathological video game use among youths: a two-year longitudinal study. PEDIATRICS 2011 Jan 17;127(2):e319-e329. [doi: 10.1542/peds.2010-1353] [Medline: 21242221]
5. Brunborg GS, Mentzoni RA, Frøyland LR. Is video gaming, or video game addiction, associated with depression, academic achievement, heavy episodic drinking, or conduct problems? Journal of Behavioral Addictions 2014 Mar;3(1):27-32. [doi: 10.1556/jba.3.2014.002] [Medline: 25215212]
6. Liau AK, Choo H, Li D, Gentile DA, Sim T, Khoo A. Pathological video-gaming among youth: A prospective study examining dynamic protective factors. Addiction Research & Theory 2014 Dec 19;23(4):301-308. [doi: 10.3109/16066359.2014.987759]
7. Goodie AS, Fortune EE. Measuring cognitive distortions in pathological gambling: Review and meta-analyses. Psychology of Addictive Behaviors 2013 Sep;27(3):730-743. [doi: 10.1037/a0031892] [Medline: 23438249]
8. Davis R. A cognitive-behavioral model of pathological Internet use. Computers in Human Behavior 2001 Mar;17(2):187-195. [doi: 10.1016/s0747-5632(00)00041-8]
9. King DL, Delfabbro PH. The cognitive psychopathology of internet gaming disorder in adolescence. J Abnorm Child Psychol 2016 Feb 15;44(8):1635-1645. [doi: 10.1007/s10802-016-0135-y] [Medline: 26875565]
10. Walker MB. Irrational thinking among slot machine players. J Gambling Stud 1992;8(3):245-261. [doi: 10.1007/bf01014652]
11. King DL, Delfabbro PH. The cognitive psychology of Internet gaming disorder. Clinical Psychology Review 2014 Jun;34(4):298-308. [doi: 10.1016/j.cpr.2014.03.006] [Medline: 24786896]
12. Yu Y, Mo PK, Zhang J, Li J, Lau JT. Validation of the Chinese Version of the Revised Internet Gaming Cognition Scale among Adolescents in China: Maladaptive Cognitions as Potential Determinants of Internet Gaming Disorder. IJERPH 2019 Dec 31;17(1):290. [doi: 10.3390/ijerph17010290] [Medline: 31906174]
13. Bolles JR, Earleywine M, Gordis E. Acquired preparedness model and impulsivity moderated expectancies: Examining predictors of marijuana use. Addiction Research & Theory 2014 Mar 05;22(6):490-497. [doi: 10.3109/16066359.2014.892929]
14. Gibbons FX, O'Hara RE, Stock ML, Gerrard M, Weng C, Wills TA. The erosive effects of racism: Reduced self-control mediates the relation between perceived racial discrimination and substance use in African American adolescents. Journal of Personality and Social Psychology 2012;102(5):1089-1104. [doi: 10.1037/a0027404] [Medline: 22390225]
15. Kvaavik E, Rise J. How do impulsivity and education relate to smoking initiation and cessation among young adults? J. Stud. Alcohol Drugs 2012 Sep;73(5):804-810. [doi: 10.15288/jsad.2012.73.804] [Medline: 22846244]
16. LaBrie JW, Kenney SR, Napper LE, Miller K. Impulsivity and alcohol-related risk among college students: Examining urgency, sensation seeking and the moderating influence of beliefs about alcohol's role in the college experience. Addictive Behaviors 2014 Jan;39(1):159-164. [doi: 10.1016/j.addbeh.2013.09.018] [Medline: 24120644]
17. Cao F, Su L, Liu T, Gao X. The relationship between impulsivity and Internet addiction in a sample of Chinese adolescents. Eur. psychiatr 2007 Aug 31;22(7):466-471. [doi: 10.1016/j.eurpsy.2007.05.004] [Medline: 17765486]
18. Wittek CT, Finserås TR, Pallesen S, Mentzoni RA, Hanss D, Griffiths MD, et al. Prevalence and predictors of video game addiction: a study based on a national representative sample of gamers. Int J Ment Health Addiction 2015 Sep 23;14(5):672-686. [doi: 10.1007/s11469-015-9592-8] [Medline: 27688739]
19. Kim EJ, Namkoong K, Ku T, Kim SJ. The relationship between online game addiction and aggression, self-control and narcissistic personality traits. Eur. psychiatr 2007 Dec 31;23(3):212-218. [doi: 10.1016/j.eurpsy.2007.10.010] [Medline: 18166402]
20. Choo H, Gentile DA, Sim T, Li D, Khoo A, Liau AK. Pathological video-gaming among Singaporean youth. Ann Acad Med Singap 2010 Nov;39(11):822-829 [FREE Full text] [Medline: 21165520]
21. Rehbein F, Psych G, Kleimann M, Mediasci G, Mößle T. Prevalence and risk factors of video game dependency in adolescence: results of a German nationwide survey. Cyberpsychology, Behavior, and Social Networking 2010 Jun;13(3):269-277. [doi: 10.1089/cyber.2009.0227] [Medline: 20557246]
22. Halpern-Felsher B. Adolescent decision-making. Encyclopedia of Adolescence 2011;1:30-37. [doi: 10.1016/b978-0-12-373951-3.00010-7]
23. Reyna VF, Farley F. Risk and rationality in adolescent decision making: implications for theory, practice, and public policy. Psychol Sci Public Interest 2016 Jun 23;7(1):1-44. [doi: 10.1111/j.1529-1006.2006.00026.x] [Medline: 26158695]
24. Lieberman MD. Social cognitive neuroscience: a review of core processes. Annu. Rev. Psychol 2007 Jan;58(1):259-289. [doi: 10.1146/annurev.psych.58.110405.085654] [Medline: 17002553]
25. Cabeza R, Nyberg L. Neural bases of learning and memory: functional neuroimaging evidence. Current Opinion in Neurology 2000;13(4):415-421. [doi: 10.1097/00019052-200008000-00008] [Medline: 10970058]
26. Tangney JP, Baumeister RF, Boone AL. High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. J Personality 2004 Apr;72(2):271-324. [doi: 10.1111/j.0022-3506.2004.00263.x] [Medline: 15016066]

27. Finkenauer C, Engels RCME, Baumeister RF. Parenting behaviour and adolescent behavioural and emotional problems: The role of self-control. International Journal of Behavioral Development 2016 Jun 30;29(1):58-69. [doi: 10.1080/01650250444000333]
28. Wu JYW, Ko H, Wong T, Wu L, Oei TP. Positive outcome expectancy mediates the relationship between peer influence and Internet gaming addiction among adolescents in Taiwan. Cyberpsychology, Behavior, and Social Networking 2016 Jan;19(1):49-55. [doi: 10.1089/cyber.2015.0345] [Medline: 26716791]
29. Lin M, Wu JY, Chen C, You J. Positive outcome expectancy mediates the relationship between social influence and Internet addiction among senior high-school students. J Behav Addict 2018 Jun 27;7(2):292-300. [doi: 10.1556/2006.7.2018.56] [Medline: 29950103]
30. Janz NK, Becker MH. The health belief model: a decade later. Health Education Quarterly 2016 Sep 04;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
31. Akers R. Deviant Behavior: A Social Learning Approach. 2nd ed. Belmont: Wadsworth; 1977.
32. Brand M, Young KS, Laier C. Prefrontal control and Internet addiction: a theoretical model and review of neuropsychological and neuroimaging findings. Front. Hum. Neurosci 2014 May 27;8:375. [doi: 10.3389/fnhum.2014.00375] [Medline: 24904393]
33. Treloar HR, Morris DH, Pedersen SL, Mccarthy DM. Direct and indirect effects of impulsivity traits on drinking and driving in young adults. J. Stud. Alcohol Drugs 2012 Sep;73(5):794-803. [doi: 10.15288/jsad.2012.73.794] [Medline: 22846243]
34. Vangsness L, Bry BH, LaBouvie EW. Impulsivity, negative expectancies, and marijuana use: a test of the acquired preparedness model. Addictive Behaviors 2005 Jun;30(5):1071-1076. [doi: 10.1016/j.addbeh.2004.11.003] [Medline: 15893107]
35. Laier C, Wegmann E, Brand M. Personality and cognition in gamers: Avoidance expectancies mediate the relationship between maladaptive personality traits and symptoms of Internet-gaming disorder. Front. Psychiatry 2018 Jul 10;9:304. [doi: 10.3389/fpsyt.2018.00304] [Medline: 30042702]
36. Flay B, Petraitis J. A new theory of health behavior with implications for preventive interventions. Advances in medical sociology 1994;4:19-44.
37. Grigsby TJ, Forster M, Unger JB, Sussman S. Predictors of alcohol-related negative consequences in adolescents: A systematic review of the literature and implications for future research. Journal of Adolescence 2016 Apr;48:18-35. [doi: 10.1016/j.adolescence.2016.01.006] [Medline: 26871952]
38. Chun J. Determinants of tobacco use among Korean female adolescents: longitudinal test of the theory of triadic influence. Children and Youth Services Review 2015 Mar;50:83-87. [doi: 10.1016/j.childyouth.2015.01.009]
39. Klein VM. Cultural, social and intrapersonal factors associated with clusters of co-occurring health-related behaviours among adolescents. The European Journal of Public Health 2014;25(1):31-37. [doi: 10.1093/eurpub/cku051] [Medline: 25617353]
40. Hine DW, McKenzie-Richer A, Lewko J, Tilleczek K, Perreault L. A comparison of the mediational properties of four adolescent smoking expectancy measures. Psychology of Addictive Behaviors 2002;16(3):187-195. [doi: 10.1037/0893-164x.16.3.187] [Medline: 12236453]
41. Ouellette JA, Gerrard M, Gibbons FX, Reis-Bergan M. Parents, peers, and prototypes: antecedents of adolescent alcohol expectancies, alcohol consumption, and alcohol-related life problems in rural youth. Psychology of Addictive Behaviors 1999;13(3):183-197. [doi: 10.1037/0893-164x.13.3.183]
42. National Bureau of Statistics of China. China Population Statistics Year Book of Year 2018. Beijing: National Bureau of Statistics of China; 2019.
43. Petry NM, O'Brien CP. Internet gaming disorder and the DSM-5. Addiction 2013 Jul;108(7):1186-1187. [doi: 10.1111/add.12162] [Medline: 23668389]
44. Ko C, Yen J, Chen S, Wang P, Chen C, Yen C. Evaluation of the diagnostic criteria of Internet gaming disorder in the DSM-5 among young adults in Taiwan. J Psychiatr Res 2014 Jun;53:103-110. [doi: 10.1016/j.jpsychires.2014.02.008] [Medline: 24581573]
45. Sigerson L, Li AY, Cheung MW, Luk JW, Cheng C. Psychometric properties of the Chinese Internet Gaming Disorder Scale. Addict Behav 2017 Nov;74:20-26 [FREE Full text] [doi: 10.1016/j.addbeh.2017.05.031] [Medline: 28558336]
46. Patton JH, Stanford MS, Barratt ES. Factor structure of the Barratt impulsiveness scale. J Clin Psychol 1995 Nov;51(6):768-774. [doi: 10.1002/1097-4679(199511)51:6<768::aid-jclp2270510607>3.0.co;2-1] [Medline: 8778124]
47. Yan W. Psychometric properties of the Chinese version of the BIS-11 in three adolescent samples. Chinese Journal of Clinical Psychology 2016;24(5):869-873.
48. Hu F, Chen G, Cai T. Preliminary Study on Self-Control Scale on Chinese Middle School Students. China Journal of Health Psychology 2012;20(8):1183-1184.
49. Moss S, Prosser H, Costello H, Simpson N, Patel P, Rowe S, et al. Reliability and validity of the PAS-ADD Checklist for detecting psychiatric disorders in adults with intellectual disability. J Intellect Disabil Res 1998 Apr;42 (Pt 2)(2):173-183. [doi: 10.1046/j.1365-2788.1998.00116.x] [Medline: 9617701]
50. Hair JF. Multivariate Data Analysis. Hoboken, NJ: Prentice Hall; 2006.

51. Yang X, Jiang X, Mo PK, Cai Y, Ma L, Lau JT. Prevalence and Interpersonal Correlates of Internet Gaming Disorders among Chinese Adolescents. Int J Environ Res Public Health 2020 Jan 16;17(2):579 [FREE Full text] [doi: 10.3390/ijerph17020579] [Medline: 31963197]
52. Wu AMS, Chen JH, Tong K, Yu S, Lau JTF. Prevalence and associated factors of Internet gaming disorder among community dwelling adults in Macao, China. J Behav Addict 2018 Mar 01;7(1):62-69 [FREE Full text] [doi: 10.1556/2006.7.2018.12] [Medline: 29463097]
53. Lemmens JS, Valkenburg PM, Peter J. The effects of pathological gaming on aggressive behavior. J Youth Adolesc 2011 Jan 13;40(1):38-47 [FREE Full text] [doi: 10.1007/s10964-010-9558-x] [Medline: 20549320]
54. Anderson CA, Sakamoto A, Gentile DA, Ihori N, Shibuya A, Yukawa S, et al. Longitudinal effects of violent video games on aggression in Japan and the United States. Pediatrics 2008 Nov 31;122(5):e1067-e1072. [doi: 10.1542/peds.2008-1425] [Medline: 18977956]
55. Király O, Griffiths MD, Urbán R, Farkas J, Kökönyei G, Elekes Z, et al. Problematic internet use and problematic online gaming are not the same: findings from a large nationally representative adolescent sample. Cyberpsychol Behav Soc Netw 2014 Dec;17(12):749-754 [FREE Full text] [doi: 10.1089/cyber.2014.0475] [Medline: 25415659]
56. Wartberg L, Kriston L, Kammerl R. Associations of Social Support, Friends Only Known Through the Internet, and Health-Related Quality of Life with Internet Gaming Disorder in Adolescence. Cyberpsychol Behav Soc Netw 2017 Jul;20(7):436-441. [doi: 10.1089/cyber.2016.0535] [Medline: 28715266]
57. Dong G, Zheng H, Liu X, Wang Y, Du X, Potenza MN. Gender-related differences in cue-elicited cravings in Internet gaming disorder: The effects of deprivation. J Behav Addict 2018 Dec 01;7(4):953-964 [FREE Full text] [doi: 10.1556/2006.7.2018.118] [Medline: 30556781]
58. Ko C, Yen JY, Chen CC, Chen SH, Yen CF. Gender differences and related factors affecting online gaming addiction among Taiwanese adolescents. J Nerv Ment Dis 2005 Apr;193(4):273-277. [doi: 10.1097/01.nmd.0000158373.85150.57] [Medline: 15805824]
59. Lopez-Fernandez O, Williams AJ, Griffiths MD, Kuss DJ. Female gaming, gaming addiction, and the role of women within gaming culture: a narrative literature review. Front Psychiatry 2019 Jul 10;10:454 [FREE Full text] [doi: 10.3389/fpsyt.2019.00454] [Medline: 31354536]
60. Lemmens JS, Valkenburg PM, Gentile DA. The Internet Gaming Disorder Scale. Psychol Assess 2015 Jun;27(2):567-582. [doi: 10.1037/pas0000062] [Medline: 25558970]
61. Rehbein F, Baier D. Family-, Media-, and School-Related Risk Factors of Video Game Addiction. Journal of Media Psychology 2013 Jan 01;25(3):118-128. [doi: 10.1027/1864-1105/a000093]
62. Przybylski AK, Weinstein N. Investigating the Motivational and Psychosocial Dynamics of Dysregulated Gaming: Evidence From a Preregistered Cohort Study. Clinical Psychological Science 2019 Aug 22;7(6):1257-1265. [doi: 10.1177/2167702619859341]
63. Wang C, Chan CLW, Mak K, Ho S, Wong PWC, Ho RTH. Prevalence and correlates of video and internet gaming addiction among Hong Kong adolescents: a pilot study. ScientificWorldJournal 2014;2014:874648-874649 [FREE Full text] [doi: 10.1155/2014/874648] [Medline: 25032242]
64. Rehbein F, Kliem S, Baier D, Mößle T, Petry NM. Prevalence of Internet gaming disorder in German adolescents: diagnostic contribution of the nine DSM-5 criteria in a state-wide representative sample. Addiction 2015 May 10;110(5):842-851. [doi: 10.1111/add.12849] [Medline: 25598040]
65. Kardefelt-Winther D. The moderating role of psychosocial well-being on the relationship between escapism and excessive online gaming. Computers in Human Behavior 2014 Sep;38:68-74. [doi: 10.1016/j.chb.2014.05.020]
66. Schneider LA, King DL, Delfabbro PH. Maladaptive coping styles in adolescents with Internet gaming disorder symptoms. Int J Ment Health Addiction 2017 Apr 20;16(4):905-916. [doi: 10.1007/s11469-017-9756-9]
67. Schmitt ZL, Livingston MG. Video game addiction and college performance among males: results from a 1 year longitudinal study. Cyberpsychol Behav Soc Netw 2015 Jan;18(1):25-29. [doi: 10.1089/cyber.2014.0403] [Medline: 25584728]
68. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
69. King DL, Herd MCE, Delfabbro PH. Tolerance in Internet gaming disorder: a need for increasing gaming time or something else? J Behav Addict 2017 Dec 01;6(4):525-533 [FREE Full text] [doi: 10.1556/2006.6.2017.072] [Medline: 29137493]
70. Forman EM, Hoffman KL, McGrath KB, Herbert JD, Brandsma LL, Lowe MR. A comparison of acceptance- and control-based strategies for coping with food cravings: an analog study. Behav Res Ther 2007 Oct;45(10):2372-2386. [doi: 10.1016/j.brat.2007.04.004] [Medline: 17544361]
71. Shiffman S. Coping with temptations to smoke. Journal of Consulting and Clinical Psychology 1984 Apr;52(2):261-267. [doi: 10.1037/0022-006x.52.2.261]
72. Gullo MJ, Dawe S. Impulsivity and adolescent substance use: rashly dismissed as "all-bad"? Neurosci Biobehav Rev 2008 Oct;32(8):1507-1518. [doi: 10.1016/j.neubiorev.2008.06.003] [Medline: 18588911]
73. Gullo M, Dawe S, Kambouropoulos N, Staiger PK, Jackson CJ. Alcohol expectancies and drinking refusal self-efficacy mediate the association of impulsivity with alcohol misuse. Alcohol Clin Exp Res 2010 Aug;34(8):1386-1399. [doi: 10.1111/j.1530-0277.2010.01222.x] [Medline: 20528818]

74. Gullo MJ, Loxton NJ, Dawe S. Impulsivity: four ways five factors are not basic to addiction. Addict Behav 2014 Nov;39(11):1547-1556. [doi: 10.1016/j.addbeh.2014.01.002] [Medline: 24576666]
75. Steinberg L, Chein JM. Multiple accounts of adolescent impulsivity. Proc Natl Acad Sci U S A 2015 Jul 21;112(29):8807-8808 [FREE Full text] [doi: 10.1073/pnas.1509732112] [Medline: 26150513]
76. Steinberg L. A social neuroscience perspective on adolescent risk-taking. Dev Rev 2008 Mar;28(1):78-106 [FREE Full text] [doi: 10.1016/j.dr.2007.08.002] [Medline: 18509515]
77. Litt MD, Kadden RM. Willpower versus "skillpower": Examining how self-efficacy works in treatment for marijuana dependence. Psychol Addict Behav 2015 Sep;29(3):532-540 [FREE Full text] [doi: 10.1037/adb0000085] [Medline: 25938628]
78. Keller L, Bieleke M, Gollwitzer P. Mindset theory of action phases and if-then planning. In: Social Psychology in Action: Evidence-based Interventions from Theory to Practice. Cham: Springer; 2019:23-37.
79. Prestwich A, Conner M, Lawton RJ. Implementation intentions: can they be used to prevent and treat addiction? In: Handbook of Implicit Cognition and Addiction. CA: Sage Publications; 2006:455-469.
80. Zhou N, Fang X. Beyond peer contagion: Unique and interactive effects of multiple peer influences on Internet addiction among Chinese adolescents. Computers in Human Behavior 2015 Sep;50:231-238. [doi: 10.1016/j.chb.2015.03.083]
81. Torres-Rodríguez A, Griffiths MD, Carbonell X. The treatment of Internet gaming disorder: a brief overview of the PIPATIC program. Int J Ment Health Addict 2018 Nov 13;16(4):1000-1015 [FREE Full text] [doi: 10.1007/s11469-017-9825-0] [Medline: 30147635]
82. Apisitwasana N, Perngparn U, Cottler L. PRBM 2018 Apr;Volume 11:103-115. [doi: 10.2147/prbm.s145868]
83. Georgie J, Sean H, Deborah M, Matthew H, Rona C. Peer-led interventions to prevent tobacco, alcohol and/or drug use among young people aged 11-21 years: a systematic review and meta-analysis. Addiction 2016 Mar 09;111(3):391-407 [FREE Full text] [doi: 10.1111/add.13224] [Medline: 26518976]
84. Jo YS, Bhang SY, Choi JS, Lee HK, Lee SY, Kweon Y. Clinical characteristics of diagnosis for Internet Gaming Disorder: comparison of DSM-5 IGD and ICD-11 GD diagnosis. J Clin Med 2019 Jun 28;8(7):945 [FREE Full text] [doi: 10.3390/jcm8070945] [Medline: 31261841]

Abbreviations

CFI: comparative fit index
C-RIGCS: The Chinese version of the Revised internet Gaming Cognition Scale
***DSM-5*:** *Diagnostic and Statistical Manual of Mental Disorders, fifth edition*
***ICD-11*:** *International Classification of Disease, eleventh revision*
IGD: internet gaming disorder
MCIG: maladaptive cognitions related to internet gaming
ORc: crude odds ratio
RMSEA: root mean square error of approximation
SCT: social cognitive theory
SEM: structural equation modeling
TLI: Tucker-Lewis index
WHO: World Health Organization

Edited by R Kukafka; submitted 28.12.20; peer-reviewed by Y Miao, W Zhang; comments to author 26.03.21; revised version received 05.08.21; accepted 22.09.21; published 27.10.21.

Please cite as:
Yu Y, Mo PKH, Zhang J, Li J, Lau JTF
Impulsivity, Self-control, Interpersonal Influences, and Maladaptive Cognitions as Factors of Internet Gaming Disorder Among Adolescents in China: Cross-sectional Mediation Study
J Med Internet Res 2021;23(10):e26810
URL: https://www.jmir.org/2021/10/e26810
doi:10.2196/26810
PMID:34704960

Original Paper

Examining the Trends in Online Health Information–Seeking Behavior About Chronic Obstructive Pulmonary Disease in Singapore: Analysis of Data From Google Trends and the Global Burden of Disease Study

Yang Fang[1], PhD; Thomas A Shepherd[2], PhD; Helen E Smith[1], DM, FFPHM

[1]Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore

[2]School of Medicine, Keele University, Staffordshire, United Kingdom

Corresponding Author:
Yang Fang, PhD
Lee Kong Chian School of Medicine
Nanyang Technological University Singapore
11 Mandalay Road, Clinical Sciences Building
Singapore, 308232
Singapore
Phone: 65 6513 8572
Email: fang.yang@ntu.edu.sg

Abstract

Background: Chronic obstructive pulmonary disease (COPD) is the third leading cause of death globally, and timely health care seeking is imperative for its prevention, early detection, and management. While online health information–seeking behavior (OHISB) is increasingly popular due to widespread internet connectivity, little is known about how OHISB for COPD has changed in comparison with the COPD disease burden, particularly at a country-specific level.

Objective: This study aimed to examine the trends in OHISB for COPD and how that compared with the estimates of COPD disease burden in Singapore, a highly wired country with a steadily increasing COPD disease burden.

Methods: To examine the trends in OHISB for COPD, we performed Prais-Winsten regression analyses on monthly search volume data for COPD from January 2004 to June 2020 downloaded from Google Trends. We then conducted cross-correlational analyses to examine the relationship between annualized search volume on COPD topics and estimates of COPD morbidity and mortality reported in the Global Burden of Disease study from 2004 to 2017.

Results: From 2004 to 2020, the trend in COPD search volume was curvilinear (β=1.69, t_{194}=6.64, P<.001), with a slope change around the end of 2006. There was a negative linear trend (β=–0.53, t_{33}=–3.57, P=.001) from 2004 to 2006 and a positive linear trend (β=0.51, t_{159}=7.43, P<.001) from 2007 to 2020. Cross-correlation analyses revealed positive associations between COPD search volume and COPD disease burden indicators: positive correlations between search volume and prevalence, incidence, years living with disability (YLD) at lag 0, and positive correlations between search volume and prevalence, YLD at lag 1.

Conclusions: Google search volume on COPD increased from 2007 to 2020; this trend correlated with the upward trajectory of several COPD morbidity estimates, suggesting increasing engagement in OHISB for COPD in Singapore. These findings underscore the importance of making high-quality, web-based information accessible to the public, particularly COPD patients and their carers.

(J Med Internet Res 2021;23(10):e19307) doi:10.2196/19307

KEYWORDS

online health information seeking; infodemiology; Google Trends; Global Burden of Disease study; chronic obstructive pulmonary disease; respiratory health

Introduction

Chronic obstructive pulmonary disease (COPD) is the third leading cause of death globally [1] and affected approximately 251 million of the world's population in 2016 [2]. In Singapore, it is the tenth leading cause of death [3] and affects approximately 5.9% of the general population [4] and 26% of the population aged 55 years and above [5]. The number of deaths caused by COPD in the South East Asia region is projected to increase from 1.04 million in 2016 to 1.43 million in 2030 [6]. Despite its significance, COPD still has poor awareness and understanding by both the public and health practitioners [7]. Therefore, understanding whether and how health information on COPD is accessed in proportion to the size of the disease burden is critical.

Barriers to health care seeking in COPD often lead to fewer prompt diagnoses and poorer disease management among those already diagnosed. For instance, misdiagnoses in primary care settings have been attributed to underuse of and lack of expertise with spirometry and COPD diagnostic guidelines [8]. Furthermore, patients' knowledge of COPD may be suboptimal because of the complexity of its name, not only in English but also in other languages [9]. Arguably, these barriers can be at least partially attributed to inadequate knowledge of COPD, which can be improved by making high-quality health information more accessible. According to the Health Information National Trends Survey in 2012, the internet was the first source of health information for 70% of adult US internet users [10], suggesting its increasing influence as a source of general and potentially, disease-specific health information.

Disease-specific online health information–seeking behavior (OHISB) among patients with COPD has been studied in several US-based surveys. A postal survey of 1077 patients with COPD in 2007-2008 found that 65% had internet access and 25% of this group used the internet to seek information on COPD at least once weekly. Their frequency of seeking COPD information online was associated with experiencing exacerbations or dissatisfaction with health service providers and treatments [11]. A web-based survey of 445 patients with COPD in 2016 found that physicians were the primary source of COPD information followed by the internet. The patients' online health information needs were primarily related to symptom control and COPD treatments; also, over 60% of the patients had discussed COPD information on the internet with their health care providers [12]. Another web-based survey of 176 COPD patients found that eHealth literacy, defined as the capacity to seek, locate, understand, evaluate, and apply health information from the internet, was higher in patients with more severe COPD [13]. These findings suggest that, at least in the United States, OHISB is common among COPD patients with internet access, and more frequent OHISB appears to be associated with more unmet needs in disease management.

A frequent method for examining OHISB is Google search data [14]. It is freely downloadable from Google Trends, a publicly accessible portal hosting data on aggregate search activity on the Google search engine delineated by time periods and regions. Since Google is the most popular search engine worldwide, its aggregate data are used as a surrogate indicator for OHISB at the population level.

In 2019, Boehm and colleagues [15] published a study using worldwide Google Trends data that found no change in search volume for COPD in the 15 years from 2004 to 2018. Yet, the number of deaths due to COPD was estimated to have increased by 11.7% from 1990 to 2015, despite a decrease of 41.9% in the age-standardized death rate in the general population [16]. This contrasts with other leading causes of death, such as diabetes and stroke, which have evidenced increased search volumes over the same period. The clear divergence between the rising prevalence of COPD and stagnant OHISB pattern for COPD is a cause for concern, but there are limitations in the study's methods. The analyses by Boehm et al [15] were performed on worldwide COPD search data without accounting for variation in COPD disease burden between countries and other country-level characteristics such as internet coverage; both factors are essential considerations for developing and implementing country-specific policy changes.

Our study improves upon the work by Boehm et al [15] by situating the enquiry within a single country, Singapore, and comparing Google Trends data with COPD disease burden estimates. Singapore is an opportune context for this research because of high internet penetration, popularity of Google search, and high COPD prevalence [4,5]. Furthermore, Singapore has a rapidly aging population with an increasing burden of chronic diseases. With this in mind, this study was designed to address 2 research questions. First, we aimed to examine the trend in OHISB for COPD (indicated by monthly Google search data) in Singapore from 2004 to 2020. Second, we aimed to compare online search volume with disease burden estimates to assess the extent to which OHISB for COPD reflects the disease burden reported in the Global Burden of Disease study (GBD 2017) over time. To our knowledge, this is the first study that examined trends in OHISB in conjunction with disease burden indicators in COPD.

Methods

Search Volume Data From Google Trends

Search query data relating to COPD were obtained from Google Trends, an online portal that displays search queries made worldwide on the Google search engine since 2004. The tool aggregates monthly volumes of search queries, delineated by time period and region, into a metric known as relative search interest (RSI). RSI is computed as a function of a search query's own highest query share. Its values range from 0 to 100, where 100 represents peak popularity. Google Trends excludes duplicate searches made by the same person over a short period of time. Search queries on Google Trends can be defined either as a search term (the exact search query, accounting for plural and singular forms and spelling mistakes) or a topic (groups of terms that share the same concept in any language).

Following the approach in Boehm et al [15], all search queries were defined as a topic and downloaded as monthly search data. In total, there were 198 monthly data points. When annual data

were required for comparison with GBD data, the annual averages were computed from monthly averages.

Disease Burden Data From GBD 2017

The following data relevant to COPD were obtained from GBD 2017: prevalence, incidence, disability-adjusted life years (DALY; the sum of years lived with disability [YLD] and the years of life lost [YLL]), YLD, YLL, number of deaths (mortality). GBD studies are conducted annually by The Institute for Health Metrics and Evaluation, and reports are typically issued every 2 years. These data are accessible online [17]. Currently, the data come from 195 countries and include 354 diseases and injuries. The data sources are diverse, including published literature, hospital and clinical data, surveillance and survey data, and inpatient and outpatient medical records. A detailed description of the methodology of the GBD 2017 can be found in the article by the GBD 2017 Disease and Injury Incidence and Prevalence Collaborators [2]. The most recent and publicly available GBD data were for 2017, and so the time period used for our study spanned from 2004, the first year for which Google Trends data were available, to 2017.

Statistical Analyses

As noted in previous works [18], performing Google Trend queries with similar parameters at different times can produce somewhat different data. To mitigate this problem, we performed the same query over 7 consecutive days, from June 22, 2020 to June 28, 2020. To assess the reliability of the data, two-way random model intraclass correlation coefficients (ICCs) were computed. The 7 time series data were averaged to produce a single time series for further analyses. Due to the nature of autocorrelated residuals in time series data, the Prais-Winsten estimation method was used to examine the trend in search volume. To compare search volume with disease burden estimates, a cross-correlation function was used to examine the correlation between the 2 annual time series data. The annual time series for search volume was computed by averaging the monthly search volume over the 12 months of each year.

Results

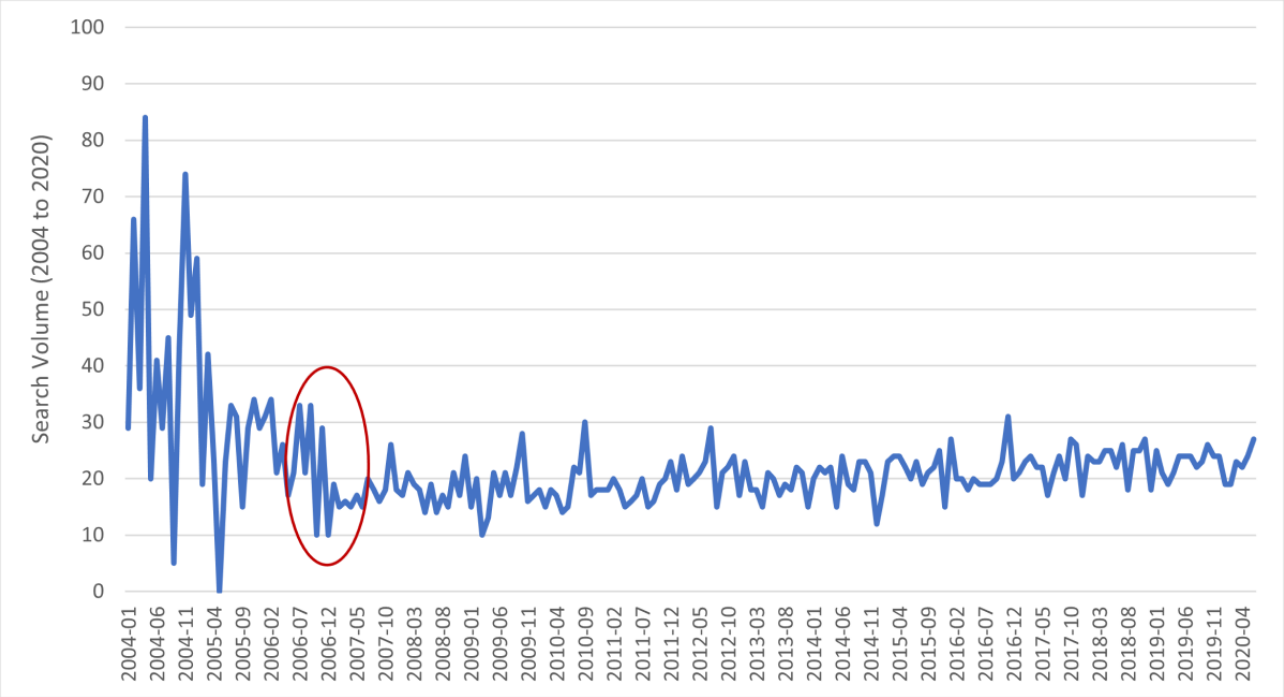
Reliability

The reliability of search volume data was moderate for the single time series data (ICC=0.55) and strong for the averaged time series data (ICC=0.90). The subsequent analyses were performed on the averaged time series data.

Trends in Monthly Search Volume for COPD From January 2004 to June 2020

The trend in monthly search volume for the COPD topic, shown in Figure 1, was examined using Prais-Winsten regression.

Figure 1. Relative search volume for chronic obstructive pulmonary disease (COPD) from 2004 to 2020. The red circle indicates an apparent slope change in search volume.



The Durbin Watson statistic [19] was 2.05. The autocorrelation coefficient was 0.09 (SE 0.72). For the overall model, R^2 was 0.22 (SE 8.16). Time demonstrated a quadratic effect on search volume (β=1.69, t_{194}=6.64, P<.001), as shown in Table 1. A quadratic effect suggests that there was 1 slope change over the entire period.

Table 1. Prais-Winsten regression examining the effect of time on monthly search volume for chronic obstructive pulmonary disease (COPD; 2004-2020).

Predictors	β[a]	*t* (*df*=194)	*P* value
time	–1.84	–7.22	<.001
time^2	1.69	6.64	<.001

[a]Standardized coefficients.

In Figure 1, the slope change appears to have happened at the end of 2006 and beginning of 2007, circled in red. To examine the trend before and after the apparent slope change, Prais-Winsten regressions were conducted for the 2 periods: 2004-2006 (36 months) and 2007-2020 (162 months). In addition, due to improvements in Google's algorithms for search volume from 2011, the trend from 2011 to 2020 was also examined. The analyses are presented in Table 2.

Table 2. Prais-Winsten regression examining the effect of time on monthly search volume during 3 different periods of analysis.

Predictor: time	β[a]	*t*	*P* value
2004 to 2006	–0.53	–3.57 (*df*=33)	.001
2007 to 2020	0.51	7.43 (*df*=159)	<.001
2011 to 2020	0.45	5.27 (*df*=111)	<.001

[a]Standardized coefficients.

From 2004 to 2006, the Durbin Watson statistic was 1.91. The autocorrelation coefficient was –0.17 (SE 0.17). For the overall model, R^2 was 0.28 (SE 16.08). Time demonstrated a negative linear effect on search volume (β=–0.53, t_{33}=–3.57, P=.001), suggesting a reduction in monthly search volume during this period.

From 2007 to 2020, the Durbin Watson statistic was 1.91. The autocorrelation coefficient was –0.03 (SE 0.08). For the overall model, R^2 was 0.26 (SE 3.25). Time demonstrated a positive linear effect on search volume (β=0.51, t_{159}=7.43, P<.001), suggesting that monthly search volume exhibited a positive linear trend from 2007 to 2020.

From 2011 to 2020, the Durbin Watson statistic was 1.99. The autocorrelation coefficient was –0.07 (SE 0.10). For the overall model, R^2 was 0.20 (SE 3.08). Time demonstrated a positive linear effect on search volume (β=0.45, t_{111}=5.27, P<.001), suggesting a positive linear trend consistent with that from 2007 to 2020, despite improvements in Google Trend's algorithms.

Cross-Correlation Between Annual Search Volume and Disease Burden Indicators (2004-2017)

Table 3 presents cross-correlations between annual COPD search volume and disease burden indicators at lags 0, –1, and 1. There were positive correlations between search volume and prevalence, incidence, and YLD at lag 0, suggesting that these pairs of variables were contemporaneously correlated. There were also positive correlations between search volume and prevalence, as well as between search volume and YLD at lag 1, suggesting that higher prevalence and YLD coincided with higher search volume 1 year later. Correlations at all other lags were nonsignificant.

Table 3. Cross-correlation analysis of annual chronic obstructive pulmonary disease (COPD) search volume and COPD disease burden indicators (2004-2017).

Disease burden indicator	Annual search volume		
	Lag –1	Lag 0	Lag 1
Prevalence	0.31	0.84[a]	0.60[a]
Incidence	0.36	0.92[a]	0.50
Disability-adjusted life years (DALY)	0.04	0.06	0.49
Years living with disability (YLD)	0.29	0.84[a]	0.59[a]
Years of life lost (YLL)	–0.00	0.08	0.41
Mortality	–0.01	0.00	0.40

[a]Exceeds the 95% CI threshold.

Discussion

This study examined the trends in OHISB for COPD in Singapore using Google search volume data between 2004 and 2020 and compared them with the trends in disease burden indicators for the same time period.

Principal Findings

The first objective was to examine the trend in search volume in COPD from 2004 to 2020. During this period, the trend in COPD search volume was curvilinear, with a slope change at the end of 2006. Further analyses examining the trends before and after this slope change revealed a downward trend from 2004 to 2006 and an upward trend from 2007 to 2020. The downward trend from 2004 to 2006 was surprising. We speculate the reason to be noisy data due to low and inconsistent search volume from 2004 to 2006, a period during which personal computing and high-speed internet connectivity were growing rapidly but not yet widespread in Singapore. From 2007 to 2020, the search volume exhibited a positive linear trend. A search of news and events from 2004 to 2020 revealed no significant events (eg, COPD health campaigns) to explain this trend. Our finding contrasts with an analysis of global data that showed no change in the trend in COPD OHIBS from 2004 to 2018 [15] and highlights the value of adopting a country-specific approach for revealing patterns that might be diluted when countries are aggregated.

The second research objective was to compare the trend in Google search volume with the disease burden indicators of COPD. From 2004 to 2017, there were positive correlations between search volume and several COPD disease burden indicators (prevalence, incidence, YLD) in the GBD study, suggesting that the trend in COPD search volume reflected the increase in some COPD morbidity indicators in the country. Search volume was not correlated with COPD mortality and YLL due to the disease, suggesting that Google search volume might be driven by the need to manage the illness for those living with COPD. This is consistent with the observation that improved health care has prolonged the living years of patients living with chronic illnesses, such as COPD.

The positive correlations between OHISB trends and COPD disease burden indicators (particularly prevalence and YLD) suggest that increased prevalence of COPD morbidity may have manifested in increased OHISB. COPD patients might be searching for health information on the internet to cope with their illness. Some of these increased searches might also have been performed by carers and family members of patients with COPD [20] or health care workers, especially those still in training. Patients with COPD appear to rely on their physicians as their primary source of information [12], but OHISB may still have a significant role when access to formal health care is limited. In Singapore, it is common for consultations with general practitioners to last for ≤5 minutes, so patients need to be well informed to optimize the brief consultation. Several patient-related barriers to timely COPD diagnoses have been reported in the literature. Patients tend to adapt to and underreport their respiratory symptoms, leading to potential underdiagnosis of COPD [21]. Older patients may also mistake symptoms such as shortness of breath as normal signs of aging [22]. Increased awareness of the significance of these symptoms among high-risk patients (eg, smokers) can increase the likelihood of a more timely diagnosis and prompt management of exacerbations.

Recommendations should be given to providers of online health information to make their materials more credible and user-friendly, reducing the barrier to timely health care seeking. Health knowledge can empower COPD patients to actively manage their own illnesses and make informed decisions about their conditions. For patients who seek health information online regularly, health care professionals can recommend trustworthy websites to complement their illness management [12].

In adopting a country-specific approach, this study seeks to present a nuanced picture of OHISB in Singapore's rapidly aging population with high internet penetration and increasing COPD burden. Heterogeneous patterns may be concealed when countries are aggregated for analysis. Furthermore, findings from single-country studies are more useful for formulating policies, which need to be tailored to the specific conditions of the target country.

Limitations

The findings of this study should be considered in the light of its limitations. First, while Google is the most popular search engine by a large margin, it nevertheless does not encompass all OHISB for COPD. Future research needs to study COPD OHISB on other platforms, such as social media, to understand how those OHISB change in accordance with COPD morbidity. Second, although Google indicated that Google Trends data should be understood as a metric of interest relative to searches on other topics, the specific way in which Google Trends data are derived is still unknown. Third, we note that the positive linear trend in COPD search volume is only moderate, and the interest level in this condition is low relative to some other health conditions. Finally, we acknowledge that this study is observational in nature, and the observed relationships were associative rather than causal. An additional analysis of search volume data on tuberculosis, a condition with a mostly static disease burden in Singapore, also revealed an upward trend, but to a lesser degree than COPD, suggesting a possible general increase in OHISB across medical conditions. Hence, increasing disease burden may be only one of the many factors driving OHISB in COPD.

Conclusion

Using Google search data, this study found an increasing trend in OHISB for COPD from 2007 to 2020 in Singapore, consistent with the increases in COPD morbidity estimates over the same time period. This suggests increasing engagement in OHISB for COPD in the population, many of whom may be COPD patients and their carers. The COPD disease burden is increasing, and timely seeking of health care is imperative for its prevention, early detection, and management. Greater public awareness is essential for minimizing the disease burden. Therefore, improving access to high-quality, web-based information on COPD is recommended for fulfilling COPD patients' information needs and improving their health outcome.

Acknowledgments

The authors acknowledge The Academic Respiratory Initiative for Pulmonary Health (TARIPH) for financial and collaboration support. Yang Fang is supported by TARIPH at Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore.

Conflicts of Interest

None declared.

References

1. The top 10 causes of death. World Health Organization. 2020 Dec 9. URL: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death [accessed 2019-11-08]
2. James S, Abate D, Abate K, Abay S, Abbafati C, Abbasi N, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Global Health Metrics 2018 Nov 10;392(10159):1789-1858. [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
3. Principal causes of death. Ministry of Health Singapore. 2018. URL: https://www.moh.gov.sg/resources-statistics/singapore-health-facts/principal-causes-of-death [accessed 2019-11-08]
4. Lim S, Lam DC, Muttalif AR, Yunus F, Wongtim S, Lan LTT, et al. Impact of chronic obstructive pulmonary disease (COPD) in the Asia-Pacific region: the EPIC Asia population-based survey. Asia Pac Fam Med 2015;14(1):4 [FREE Full text] [doi: 10.1186/s12930-015-0020-9] [Medline: 25937817]
5. Ng TP. Epidemiology of chronic obstructive pulmonary disease. The Singapore Family Physician 2013;39(2):8-10. [doi: 10.5005/jp/books/12063_1]
6. Global health estimates: Leading causes of death. World Health Organization. 2018 Oct. URL: https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death [accessed 2020-03-15]
7. Bhatia R, Fromer L. Diagnosing and treating COPD: understanding the challenges and finding solutions. IJGM 2011 Oct:729. [doi: 10.2147/ijgm.s21387] [Medline: 22114517]
8. Diab N, Gershon AS, Sin DD, Tan WC, Bourbeau J, Boulet L, et al. Underdiagnosis and Overdiagnosis of Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med 2018 Nov 01;198(9):1130-1139. [doi: 10.1164/rccm.201804-0621CI] [Medline: 29979608]
9. Wong SSL, Abdullah N, Abdullah A, Liew S, Ching S, Khoo E, et al. Unmet needs of patients with chronic obstructive pulmonary disease (COPD): a qualitative study on patients and doctors. BMC Fam Pract 2014 Apr 16;15:67 [FREE Full text] [doi: 10.1186/1471-2296-15-67] [Medline: 24739595]
10. Prestin A, Vieux SN, Chou WS. Is Online Health Activity Alive and Well or Flatlining? Findings From 10 Years of the Health Information National Trends Survey. J Health Commun 2015;20(7):790-798. [doi: 10.1080/10810730.2015.1018590] [Medline: 26042588]
11. Martinez CH, St Jean BL, Plauschinat CA, Rogers B, Beresford J, Martinez FJ, et al. Internet access and use by COPD patients in the National Emphysema/COPD Association Survey. BMC Pulm Med 2014 Apr 22;14:66 [FREE Full text] [doi: 10.1186/1471-2466-14-66] [Medline: 24755090]
12. Mahler DA, Cerasoli F, Della L, Rudzinski M. Internet Health Behaviors of Patients with Chronic Obstructive Pulmonary Disease and Assessment of Two Disease Websites. Chronic Obstr Pulm Dis 2018 May 07;5(3):158-166 [FREE Full text] [doi: 10.15326/jcopdf.5.3.2017.0173] [Medline: 30584579]
13. Stellefson ML, Shuster JJ, Chaney BH, Paige SR, Alber JM, Chaney JD, et al. Web-based Health Information Seeking and eHealth Literacy among Patients Living with Chronic Obstructive Pulmonary Disease (COPD). Health Commun 2018 Dec;33(12):1410-1424 [FREE Full text] [doi: 10.1080/10410236.2017.1353868] [Medline: 28872905]
14. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of google trends in health care research: a systematic review. PLoS One 2014;9(10):e109583 [FREE Full text] [doi: 10.1371/journal.pone.0109583] [Medline: 25337815]
15. Boehm A, Pizzini A, Sonnweber T, Loeffler-Ragg J, Lamina C, Weiss G, et al. Assessing global COPD awareness with Google Trends. Eur Respir J 2019 Jun;53(6):1-7 [FREE Full text] [doi: 10.1183/13993003.00351-2019] [Medline: 31097517]
16. GBD 2015 Chronic Respiratory Disease Collaborators. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet Respir Med 2017 Sep;5(9):691-706 [FREE Full text] [doi: 10.1016/S2213-2600(17)30293-X] [Medline: 28822787]
17. GBD Results Tool. Institute for Health Metrics and Evaluation. 2019. URL: http://ghdx.healthdata.org/gbd-results-tool [accessed 2019-09-10]
18. Tran US, Andel R, Niederkrotenthaler T, Till B, Ajdacic-Gross V, Voracek M. Low validity of Google Trends for behavioral forecasting of national suicide rates. PLoS One 2017;12(8):e0183149 [FREE Full text] [doi: 10.1371/journal.pone.0183149] [Medline: 28813490]

19. Savin NE, White KJ. The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors. Econometrica 1977 Nov;45(8):1989. [doi: 10.2307/1914122]
20. Sadasivam RS, Kinney RL, Lemon SC, Shimada SL, Allison JJ, Houston TK. Internet health information seeking is a team sport: analysis of the Pew Internet Survey. Int J Med Inform 2013 Mar;82(3):193-200. [doi: 10.1016/j.ijmedinf.2012.09.008] [Medline: 23149121]
21. Langsetmo L, Platt RW, Ernst P, Bourbeau J. Underreporting exacerbation of chronic obstructive pulmonary disease in a longitudinal cohort. Am J Respir Crit Care Med 2008 Feb 15;177(4):396-401. [doi: 10.1164/rccm.200708-1290OC] [Medline: 18048806]
22. Benady S, Canadian Thoracic Society. The human and economic burden of COPD: a leading cause of hospital admission in Canada. Ottawa, Ontario: Canadian Thoracic Society; 2010.

Abbreviations

COPD: chronic obstructive pulmonary disease
DALY: disability-adjusted life years
GBD: Global Burden of Disease Study
ICC: intraclass correlation coefficient
OHISB: online health information–seeking behavior
RSI: relative search interest
TARIPH: Academic Respiratory Initiative for Pulmonary Health
YLD: years living with disability
YLL: years of life lost

Edited by R Kukafka; submitted 13.04.20; peer-reviewed by I Yang, T Cruvinel; comments to author 25.05.20; revised version received 20.10.20; accepted 04.06.21; published 18.10.21.

Please cite as:
Fang Y, Shepherd TA, Smith HE
Examining the Trends in Online Health Information–Seeking Behavior About Chronic Obstructive Pulmonary Disease in Singapore: Analysis of Data From Google Trends and the Global Burden of Disease Study
J Med Internet Res 2021;23(10):e19307
URL: https://www.jmir.org/2021/10/e19307
doi:10.2196/19307
PMID:34661539

Original Paper

Quantifying the Severity of Adverse Drug Reactions Using Social Media: Network Analysis

Adam Lavertu[1,2], PhD; Tymor Hamamsy[3], MS; Russ B Altman[2,4,5], MD, PhD

[1]Biomedical Informatics Training Program, Stanford University, Stanford, CA, United States

[2]Department of Biomedical Data Science, Stanford University, Stanford, CA, United States

[3]Center for Data Science, New York University, New York, NY, United States

[4]Department of Bioengineering, Stanford University, Stanford, CA, United States

[5]Department of Genetics, Stanford University, Stanford, CA, United States

Corresponding Author:
Russ B Altman, MD, PhD
Department of Bioengineering
Stanford University
Shriram Center Room 213
443 Via Ortega MC 4245
Stanford, CA, 94305
United States
Phone: 1 650 725 0659
Email: rbaltman@stanford.edu

Abstract

Background: Adverse drug reactions (ADRs) affect the health of hundreds of thousands of individuals annually in the United States, with associated costs of hundreds of billions of dollars. The monitoring and analysis of the severity of ADRs is limited by the current qualitative and categorical systems of severity classification. Previous efforts have generated quantitative estimates for a subset of ADRs but were limited in scope because of the time and costs associated with the efforts.

Objective: The aim of this study is to increase the number of ADRs for which there are quantitative severity estimates while improving the quality of these severity estimates.

Methods: We present a semisupervised approach that estimates ADR severity by using social media word embeddings to construct a lexical network of ADRs and perform label propagation. We used this method to estimate the severity of 28,113 ADRs, representing 12,198 unique ADR concepts from the Medical Dictionary for Regulatory Activities.

Results: Our Severity of Adverse Events Derived from Reddit (SAEDR) scores have good correlations with real-world outcomes. The SAEDR scores had Spearman correlations of 0.595, 0.633, and −0.748 for death, serious outcome, and no outcome, respectively, with ADR case outcomes in the Food and Drug Administration Adverse Event Reporting System. We investigated different methods for defining initial seed term sets and evaluated their impact on the severity estimates. We analyzed severity distributions for ADRs based on their appearance in boxed warning drug label sections, as well as for ADRs with sex-specific associations. We found that ADRs discovered in the postmarketing period had significantly greater severity than those discovered during the clinical trial (P<.001). We created quantitative drug-risk profile (DRIP) scores for 968 drugs that had a Spearman correlation of 0.377 with drugs ranked by the Food and Drug Administration Adverse Event Reporting System cases resulting in death, where the given drug was the primary suspect.

Conclusions: Our SAEDR and DRIP scores are well correlated with the real-world outcomes of the entities they represent and have demonstrated utility in pharmacovigilance research. We make the SAEDR scores for 12,198 ADRs and the DRIP scores for 968 drugs publicly available to enable more quantitative analysis of pharmacovigilance data.

(J Med Internet Res 2021;23(10):e27714) doi:10.2196/27714

KEYWORDS

social media for health; pharmacovigilance; adverse drug reactions; machine learning; network analysis; word embeddings; drug safety; social media

Introduction

Background

Adverse drug reactions (ADRs) are among the leading causes of mortality and morbidity in the United States, affecting hundreds of thousands of people and costing more than US $500 billion every year in the United States alone [1,2]. An ADR is characterized as "an appreciably harmful or unpleasant reaction resulting from an intervention related to the use of a medicinal product" [3]. A drug's ADRs are primarily derived from clinical trial data and augmented through postmarketing surveillance [4]. The ADR labeling process for each drug is based on the frequency of the ADRs in the treated populations and the severity of the outcomes associated with each ADR. The severity of an ADR is traditionally classified into one of three categories characterized by where it appears on the label: boxed warning, warnings and precautions, or adverse reactions, listed in decreasing order of associated severity [5]. Boxed warnings refer to "serious warnings, particularly those that lead to death or serious injury." Under the existing system, death is severe enough to warrant a boxed warning, but so is restlessness; therefore, it would be useful for prescribers, patients, and researchers to have systems for recognizing that death is more severe than restlessness [6]. Similarly, warnings, precautions, and adverse reactions include a diverse spectrum of ADRs, and it can be difficult for patients, prescribers, and researchers to compare the risk profiles of different drugs.

The existing categorical definitions of ADR severity limit the ability of researchers and regulators to apply quantitative methods to regulatory and pharmacovigilance efforts. For instance, tracking the regulatory performance of drug safety efforts is primarily achieved through a categorical analysis of ADR case outcomes and drug label changes. The creation of a quantitative numerical scale that provides a relative severity score for each ADR would enable these analyses to leverage the powerful tools of quantitative decision and utility theory.

The pioneering study by Tallarida et al [7] examined the ability to create a continuous ADR severity scale. A total of 53 physicians were interviewed and asked to estimate the acceptable probabilities in a set of scenarios specifying risk-benefit trade-offs. These probabilities were then used to define a set of equations and the expected benefit to the patient in each scenario. Solving the system of equations resulted in numerical severity weights that were relative to the lowest ADR severity. The effort took approximately 45 minutes per interview and resulted in relative scores for seven ordinal categories of ADR severity.

The study by Gottlieb et al [8] expanded this ADR severity ranking task using the crowdsourcing platform Amazon Mechanical Turk. Individual workers were presented with a pair of ADRs and a link to more information on these ADRs and asked to select the ADR that they perceived to be more severe. Over 146 person-days, 2589 workers produced severity comparisons for 126,512 ADR pairs composed from 2929 unique ADRs. A linear programming algorithm was used to create a unified ADR severity ranking. To validate the ranking, the ADR severities were correlated with ADR case outcomes from the Food and Drug Administration Adverse Event Reporting System (FAERS). The highest Spearman correlation, $\rho=0.53$, was associated with the proportion of cases resulting in death. This effort drastically increased the number of ADRs included in the severity ranking but was costly in terms of human time, labor, and money.

Recent work in natural language processing has resulted in numerous methods for creating vectorized word representations (also known as *embeddings*) that capture semantic meaning in a dense numerical representation [9,10]. These methods rely on the distributional hypothesis that word meaning is captured by the contexts in which a word appears [11]. The practical implication of the distributional hypothesis is that training a model to predict word context (ie, co-occurring word pairs) results in model weights that capture the meaning of the word. These model weights can then be used as a numerical representation of a given word.

Word embeddings learned on social media data sets have been deployed for pharmacovigilance previously but not for the purpose of exploring ADR severity [12]. By using a social media corpus generated by the general public, especially on pseudoanonymous social media platforms such as Reddit, researchers can capture meanings that reflect people's unfiltered experiences of, and opinions about, health and disease [13,14]. Investigating the utility, benefits, and challenges of different social media platforms and methods for pharmacovigilance has been an area of active research [15]. Previous research on Twitter data annotated samples of tweets for personal medication intake versus individuals simply mentioning a drug [16,17]. Together, the 2 studies found that approximately 40% of the tweets mentioning a drug indicated that the individual tweeting was possibly personally taking that medication. Here, we focused on ADRs, but it is likely that the individuals discussing ADRs on social media have more direct experience with the ADRs, either by being directly affected by the ADRs or being informed about the ADR experience by a close relation, than individuals selected from a pool of crowdworkers. Word embeddings trained on a corpus generated by the general public can then be leveraged as a metric for public opinion in a similar fashion to previous crowdsourced approaches. We used the RedMed embeddings trained on more than 580 million health-enriched Reddit comments from more than 10 million users [18]. These numbers dwarf the number of votes gathered in a typical crowdsourcing experiment and are therefore potentially more indicative of a representative population's perception of ADRs than traditional survey-based methods.

Objective

In this study, we used publicly available word embeddings and a network-based label propagation method to estimate the severity of 12,198 ADR concepts from Medical Dictionary for Regulatory Activities (MedDRA) terminology. The resulting Severity of Adverse Events Derived from Reddit (SAEDR) scores were validated against human rankings as well as FAERS case outcomes. We used System Organ Classes (SOCs) and other groupings within the MedDRA to examine how the SAEDR severity scores differ at various levels of abstraction and within ADR categories. We used the SAEDR scores to

compare the severity among ADRs in different sections of drug labels, ADRs with disproportionate rates between sexes, and ADRs discovered at different stages of drug development. We combined the SAEDR scores with frequency information from Side Effect Resource (SIDER)–derived drug labels to generate drug-specific aggregate side effect severity scores [19]. The SAEDR scores enabled new analyses that were not possible with the existing categorical classifications.

Methods

Data Sources and Preparation

ADR Terms and Phrases

We sourced our ADR phrases and their synonyms from the *lowest-level terms* within the terminology of the MedDRA, version 22 [20]. We filtered terms based on their semantic types within the Unified Medical Language System Metathesaurus [21] and retained terms of the following semantic types: *Disease or Syndrome* (T047), *Finding* (T033), *Neoplastic Process* (T191), *Injury or Poisoning* (T037), *Pathologic Function* (T046), *Sign or Symptom* (T184), *Mental or Behavioral Dysfunction* (T048), and *Congenital Abnormality* (T019).

Word-Embedding Model

We used the word-embedding model from the RedMed project [18]. This model was selected because it was directly optimized for medical term similarity and was trained on a corpus of Reddit comments generated by the public and preprocessed to maximize the inclusion of ADR terms. ADRs that were a phrase such as "abdominal pain" were represented using the average embedding of all terms within the phrase. In previous work, not presented here, we found that the use of average embeddings, for example, "abdominal" + "pain"/2, produced better results than the use of embedded phrases, for example, "abdominal pain."

Gottlieb Severity Data

We used the crowdsourced severity estimates included within the supplement of the study by Gottlieb et al [8], which included 2929 ADRs with rank scores.

FAERS Data

All adverse drug event data were downloaded in JSON format from the openFDA website, which includes data from both the Legacy Adverse Event Reporting System and the FAERS. Data were retrieved on August 8, 2020, and included adverse event case reports up to June 30, 2020. Drug reactions were normalized to the MedDRA, version 22. We filtered to unique case IDs using the provided duplicated flag to remove duplicate reports and case reports that did not originate in the United States. Case outcomes were normalized according to the following schema: "Death": {"Death"}, "Serious Outcome": {"Life-Threatening," "Hospitalization," "Other Serious," "Required Intervention"}, and "Disability": {"Congenital Anomaly," "Disability"}. We created an additional outcome category denoted *No Outcome* for cases with no reported outcome. We felt the need to create this category because the FAERS only allows for the reporting of serious case outcomes, and there is information in the absence of a serious outcome being reported. The outcome proportions for each ADR were calculated by dividing the number of cases reporting that ADR for a particular outcome by the total number of cases reporting that ADR.

FAERS Severity Rankings

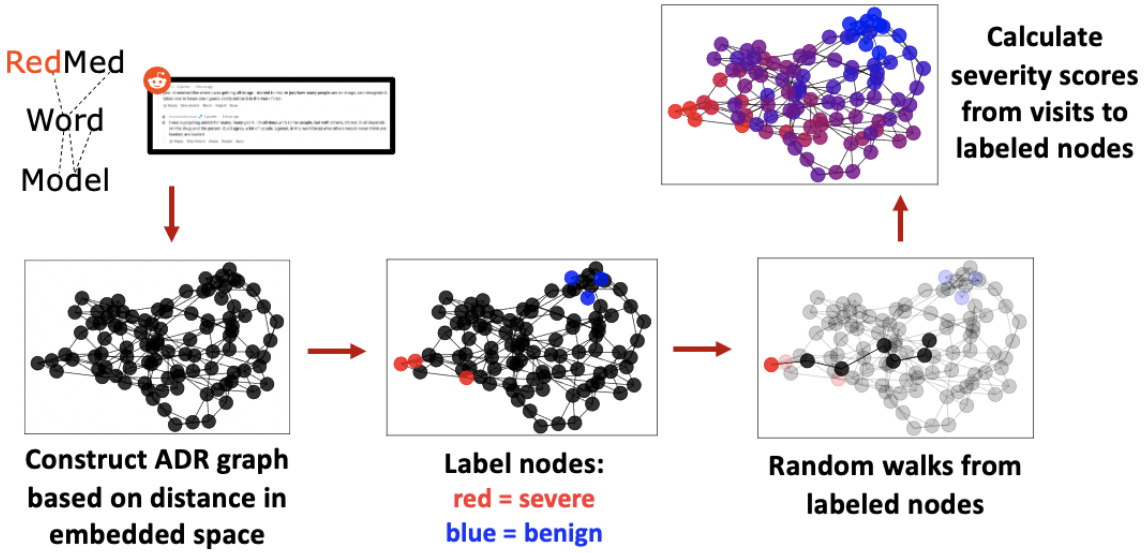
For the ranking of ADRs based on the FAERS data, we ranked ADRs based on their marginal likelihood to be included in a case with death or serious outcome as the reported case outcome. This was calculated by dividing the number of cases with the ADR that resulted in death or serious outcome by the number of cases without the ADR that resulted in these outcomes.

Semisupervised Severity Propagation

Overview

Given a word-embedding model and a set of initial, potentially noisy, seed-word labels for the severe and benign categories, we sought to propagate severity information over the rest of the vocabulary, similar to the sentiment propagation methods described in the study by Hamilton et al [22]. A graphical overview of this method is presented in Figure 1.

Figure 1. Overview of the network method for estimating ADR severity from word embeddings. Word embeddings for 28,113 ADR terms and phrases were extracted from the RedMed word-embedding model. A network is constructed based on nearest-neighbors in the embedding space. A subset of nodes is labeled as severe or benign ADRs, and random walks from these labeled nodes are conducted. The Severity of Adverse Events Derived from Reddit scores for each ADR are calculated based on the relative number of encounters in random walks initiated from severe versus benign nodes. ADR: adverse drug reaction.



Lexical Network Creation

The network was constructed by connecting words (nodes) with edges to their *k-nearest neighbors* based on the cosine distance between their embeddings. The corresponding edge weight was set to the cosine similarity of the two words (ie, the edge weight increases with term similarity).

Seed Selection From Terminology

An initial set of ranked seed terms was labeled. Biomedical terminologies and ontologies often contain many terms that are highly similar to each other. To reduce the effect of lexical similarities on label propagation, we filtered the benign and severe seed terms based on the edit ratio among the terms in each list, as defined by the ratio function in the python-Levenshtein package [23]. Only terms that had an edit ratio >0.5 from a higher-ranking term were included in the list, resulting in a reduced list of lexically distinct seed terms. Both benign and severe seed term lists were truncated to the minimum seed term list size of both seed sets.

For instance, if the severe seed term list had [“death,” “cardiac arrest acute,” “cardiac arrest”] in rank order and the benign seed term list had [“dry skin,” “yawning,” “cold sweat”] in rank order, then, after lexical filtering, the severe seed term list would be [“death,” “cardiac arrest acute”], and the benign seed term list would remain unchanged. The final seed term lists would be truncated to the minimum seed term list size, resulting in a severe seed term list [“death,” “cardiac arrest acute”] and a benign seed term list [“dry skin,” “yawning”].

Random Walks From Seed Nodes

Given a network and a set of seed nodes, we modified code from *node2vec* [24] to perform 5000 weighted random walks of length 200 from each seed node. We selected the number of random walks to ensure that all nodes within the graph were visited a nonzero number of times. We noted that this value needs to be empirically determined and will likely change based on the size and structure of the network.

SAEDR Score Calculation

The SAEDR score of a given node *u* is calculated using the following formula:

$$u_{SAEDR} = s_u / s_u + b_u$$

In the formula, s_u and b_u are the respective number of times that node *u* is encountered in a random walk from a severe node or a benign node. If node *u* was contained within one of our seed node sets, self-visits during the random walks were excluded from the calculation. To increase the robustness of these score estimates, we performed 10,000 iterations of bootstrap sampling [25] of the random walks, and the average of these bootstrap estimates was calculated. The SAEDR scores from multiple seed sets were averaged at the preferred term (PT) level, and the final combined scores were normalized to a zero to one range.

Hyperparameter Tuning for Severity Propagation

We randomly split the crowdsourced severity data into training (1778/2369, 75.05%) and test (591/2369, 24.95%) sets, with 284 ADRs dropped because of mapping. We performed a grid search over the number of neighbors used to construct the lexical network {2, 5, 10, 15, 20, 25, 30} and the percentage of nodes to label for the severe and benign seed nodes {2, 5, 10,15, 20, 25}. For example, an individual run would use the 25 nearest neighbors and 5% at each end of the training severity rank (ie, the top 5% most severe and bottom 5% least severe ADRs). We ultimately selected the parameters with the highest Spearman correlation with the training data.

ADR Discovery Group Analysis

We analyzed the SAEDR score distributions of several different categories of ADRs.

ADRs Included in a Boxed Warning

We downloaded the counts for the number of appearances of ADRs in the boxed warning section of drug labels from the supplement of the study by Wu et al [26]. ADRs that appear on a drug label based on their appearance in the SIDER but were not included in the list provided in the study by Wu et al [26] were considered to not have appeared in a boxed warning section [19].

ADRs With Disproportionate Reporting Between Sexes

A set of ADRs shown to have sex differences was identified from the supplement of the study by Chandak et al [27]. We filtered out ADRs considered borderline (mean log(ROR)<0.4 and mean log(ROR)>−0.4), as defined in the original study by Chandak et al [27]. Their study reported ADRs at the high-level group term (HLGT) level within the MedDRA, whereas our severity estimates were generated at the PT level. There are many PT ADRs per HLGT; therefore, all PTs within the specified HLGT terms were averaged to create a single SAEDR score for that HLGT. We did not use the drug information except to note distinct instances of sex-specific HLGTs (ie, HLGTs associated multiple times with different drugs each time).

ADRs Discovered at Different Stages of Drug Development

ADRs discovered at the clinical trial stage were identified based on their inclusion in a drug label in the SIDER with a non-postmarketing ADR frequency [19]. Postmarketing ADRs were identified as those included on drug labels with postmarketing frequency information and no clinical trial frequency information. OFFSIDES and TWOSIDES ADRs were discovered from the FAERS data by looking at disproportionate reporting after correcting for the effects of demographics and other case information [28]. The OFFSIDES and TWOSIDES sets are the results of a study by Tatonetti et al [28] in which propensity score matching was used to identify matched cases and controls among the FAERS cases and to also identify potentially novel ADRs based on disproportional reporting metrics. This was done to discover ADRs that were associated with a single drug, OFFSIDES, as well as ADRs that were potentially the result of drug-drug interactions, TWOSIDES. We considered both OFFSIDES and TWOSIDES ADRs discovery-stage postmarketing ADRs because they have not yet been included in the drug labeling.

Drug-Risk Profile Scores

The drug-risk profile (DRIP) scores were calculated based on our SAEDR score and the frequency of each ADR for a given drug. The DRIP score for a given drug is the sum of the severity multiplied by the frequency of each ADR on the drug label. Drug ADR frequency information was retrieved from the SIDER [19]. In situations where multiple ADR frequencies were indicated, 1000 frequency estimates were sampled from a uniform distribution with a lower bound at the minimum-reported frequency and an upper bound at the maximum-reported frequency. When there was no frequency information, we sampled 1000 estimates from a uniform distribution over the interval of 0.001 to 0.01. The final DRIP score was the average score across 1000 samples.

Results

Network Statistics

We used the ADR terms and phrases from the MedDRA, version 22.0, lowest-level terms [20] as the initial lexicon for the lexical ADR network. Although the FAERS uses the MedDRA to encode ADRs, versioning differences over the years have resulted in some FAERS ADRs not being included in the MedDRA. We were able to generate embeddings for 92.35% (2450/2653) of the unique crowdsourced ADRs, 100% (14,045/14,045) of the unique FAERS ADRs, and 68.72% (28,113/40,905) of all filtered MedDRA terms. This resulted in a final lexical network with 28,113 nodes, representing 12,198 MedDRA PTs.

SAEDR Score Performance

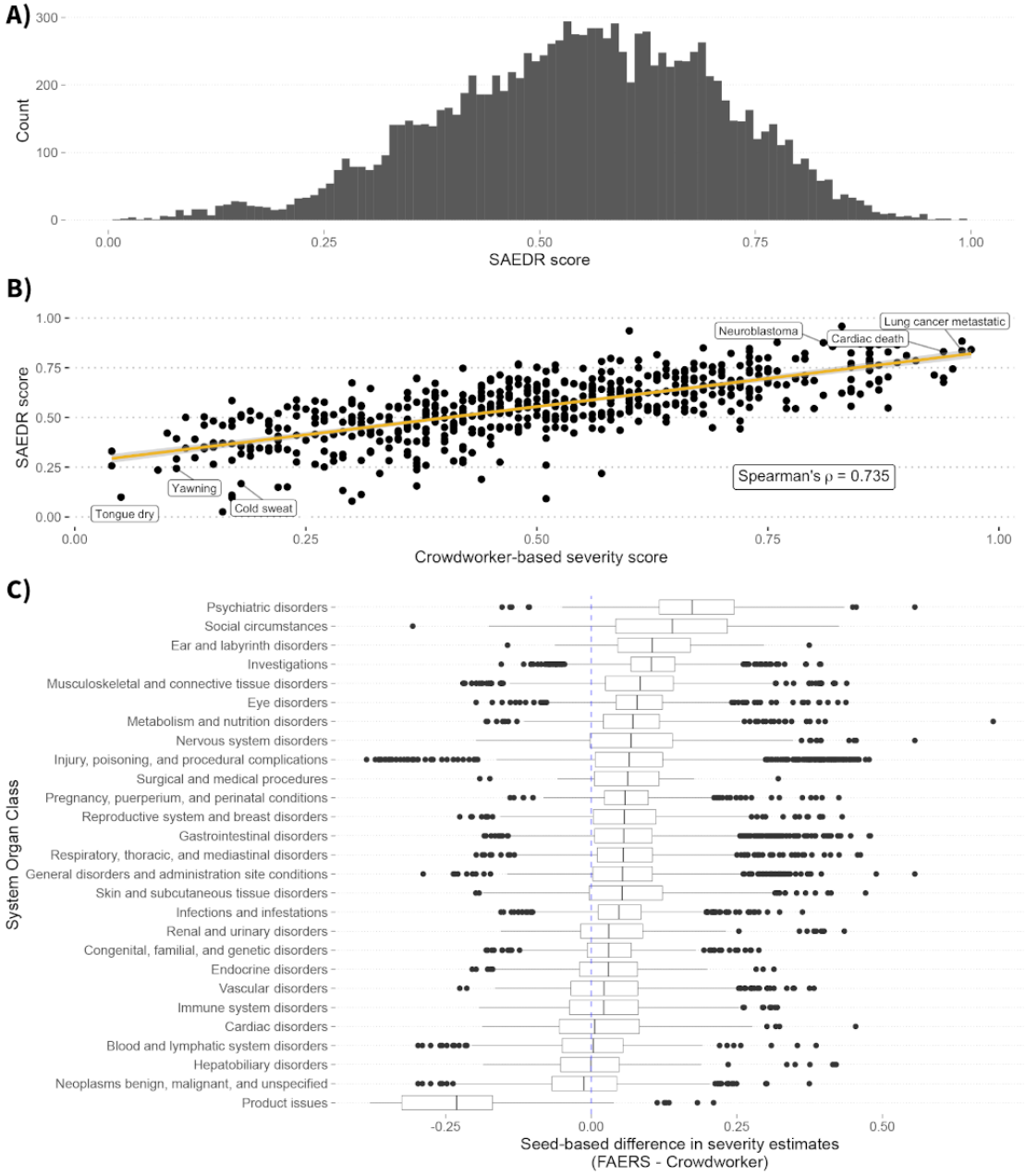
We compared our severity estimates using crowdworker-ranked seeds, FAERS-ranked seeds, and the average severity estimate across the two seed sets with two different ADR rankings: (1) a held-out test set of crowdworker-ranked ADRs (n=591) and (2) ADRs ranked by case outcome statistics in the FAERS database.

The crowdworker-seeded severity estimates had the highest training performance with a 25 nearest-neighbors graph and using 10% of the most and least severe ADRs as seeds. The Spearman correlation with the crowdworker test set was 0.747 (P<.001). The Spearman correlations for events with at least 100 reports in the FAERS were 0.595, 0.616, and −0.732 for death, serious outcome, and no outcome, respectively, with P<.001 for all (Figures S1 and S2 of Multimedia Appendix 1).

The FAERS-seeded severity estimates had the highest training performance with a 10 nearest-neighbors graph and using 10% of the most and least severe ADRs as seeds. The Spearman correlation with the crowdworker test set was 0.587. As no information from the crowdworker rankings was used to select these seeds, we can also report the Spearman correlation with the entire set of crowdworker ADRs: 0.765. The Spearman correlations for events with at least 100 reports in the FAERS were 0.509, 0.557, and −0.656 for death, serious outcome, and no outcome, respectively, with P<.001 for all (Figures S1 and S2 of Multimedia Appendix 1).

The SAEDR score—the average of the two severity estimates—had a Spearman correlation with the crowdworker test set of 0.735 (Figure 2). The Spearman correlations for events with at least 100 reports in the FAERS were 0.595, 0.633, and −0.748 for death, serious outcome, and no outcome, respectively (Figures S1 and S2 of Multimedia Appendix 1). All these correlations were statistically significant (P<.001).

Figure 2. Comparison of SAEDR scores and crowdworker severity estimates. (A) Histogram of SAEDR scores for 12,198 adverse drug reactions (ADRs). (B) Crowdworker severity estimates (x-axis) versus SAEDR scores (y-axis) for a test set of 591 ADRs. The SAEDR scores showed a strong Spearman correlation, ρ=0.735, with the human crowdworker rankings. We noted that this correlation is greater than the interrater correlation, ρ=0.71, reported in the original crowdsource study by Gottlieb et al [8]. A select set of the least and most severe ADRs based on the SAEDR score has been annotated. (C) Differences between severity estimates seeded with FAERS rankings and those seeded with crowdworker rankings (x-axis) for different System Organ Class groups (y-axis). The dashed blue line indicates where the severity would be the same for both estimates. FAERS: Food and Drug Administration Adverse Event Reporting System; SAEDR: Severity of Adverse Events Derived from Reddit.



Seed-Ranking Comparisons

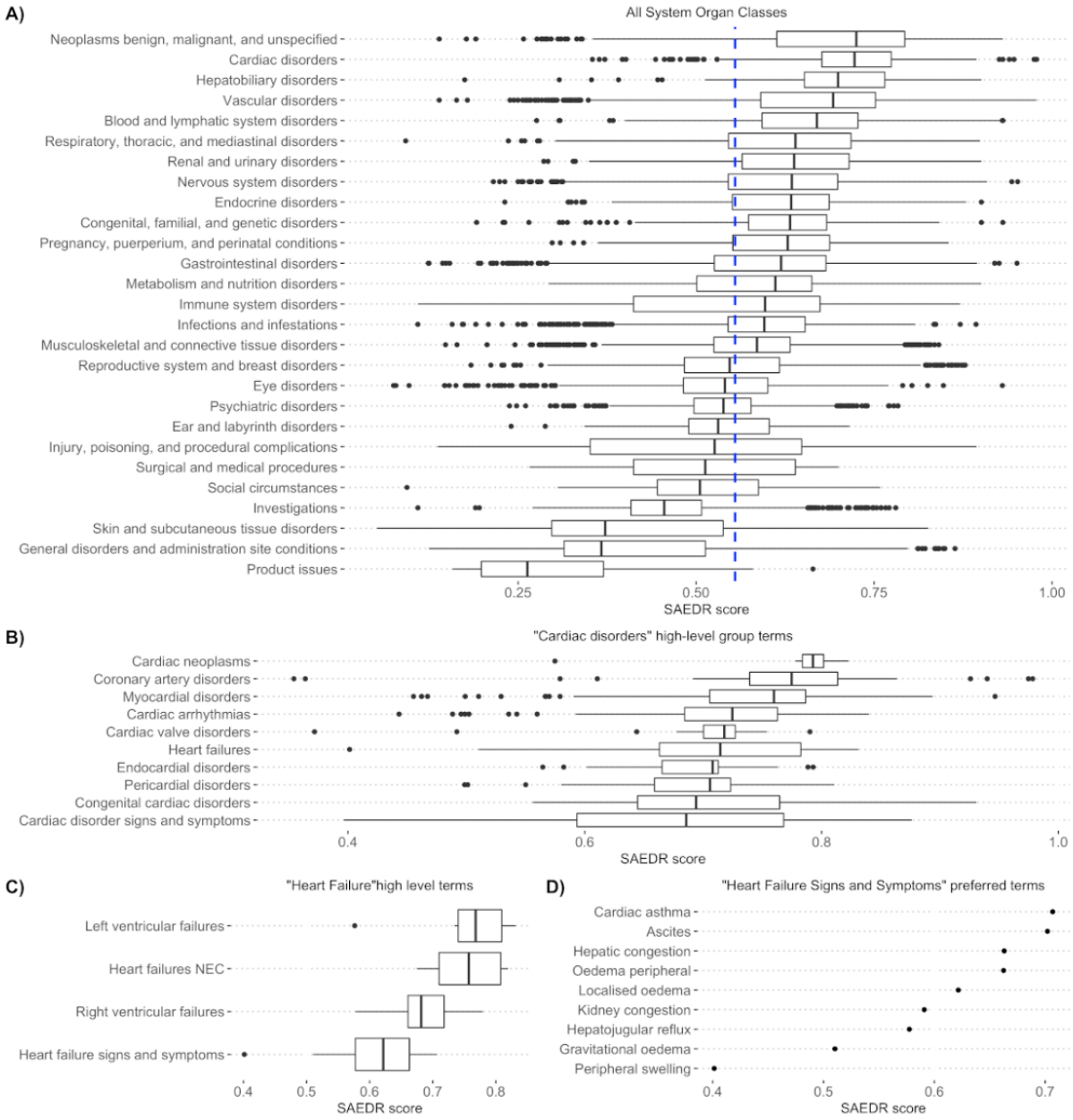
We examined the effects of seeding on the severity estimates of individual ADRs. Most ADRs were not substantially changed by the use of different seeds. We aggregated ADR severity differences at the MedDRA SOC level and found that the FAERS seeds led to increased severity estimates for *Psychiatric Disorders* and *Social Circumstances* (Figure 2). The most drastic shift in values was the decrease in severity of *Product Issue*–related ADRs using FAERS seeds compared with their severity when using crowdworker-based seeding.

SOC Severity Rankings

We compared the severity of ADRs based on the MedDRA SOC groupings as a qualitative evaluation (Figure 3). We found the ADRs related to cancer (neoplasms), cardiac, and liver (hepatobiliary) to be the most severe, with high SAEDR scores, whereas skin, general disorders, and product issues were considered the least severe, with low SAEDR scores. We performed a qualitative examination of ADRs at various other levels. We compared the severity distributions between the ADRs describing benign neoplasms and those describing malignant neoplasms (Figure S3 of Multimedia Appendix 1). A one-sided *t* test found malignant neoplasms to be significantly

more severe than benign neoplasms based on their SAEDR scores (P<.001). We found a wide range in terms of severity within the different levels of the MedDRA ADR term groupings (Figure 3). We noted that among ADRs in the *Cardiac Disorders* HLGT, cardiac neoplasms have the highest SAEDR scores, whereas signs and symptoms of cardiac disorder have the lowest relative SAEDR scores (Figure 3). When examining the level of individual ADRs within the *Heart Failure Signs and Symptoms* HLGT, we found that the ADRs that we would judge to be more severe were ranked higher than those that are more general and considered less severe.

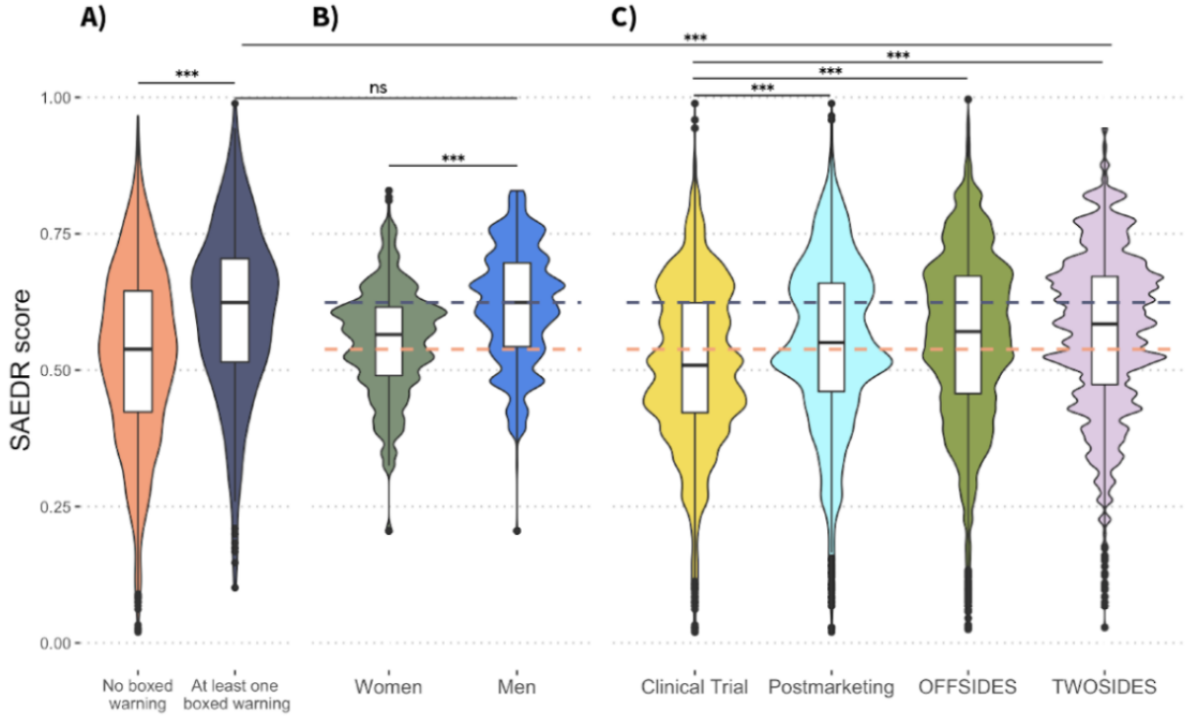
Figure 3. Severity of adverse drug reactions (ADRs) at different group resolutions. (A) SAEDR score (x-axis) for 11,981 ADRs with System Organ Class (SOC) groups (y-axis). The dashed blue line indicates an overall median SAEDR score of 0.554. The SAEDR score distributions indicate that cancers and cardiac-related side effects are considered the most severe, with product issues being the least severe. (B) Severity distributions of the Medical Dictionary for Regulatory Activities high-level group terms (HLGTs) within the Cardiac Disorders SOC. There are large differences in severity within these term groups. (C) High-level terms within the Heart Failure HLGTs show a tighter distribution of severity, with "Heart Failure Signs and Symptoms" being the least severe within this group. (D) Individual side effects within the signs and symptoms of heart failure suggest cardiac asthma and ascites (accumulation of fluid in the peritoneal cavity) are the most severe symptoms, and peripheral swelling is the least severe symptom NEC: necrotizing enterocolitis; SAEDR: Severity of Adverse Events Derived from Reddit.



Severity of ADRs Grouped by Labeling Section, Sex Effects, and Time Point of Discovery

On the basis of our estimate, we calculated that the ADRs that have never been included in a boxed warning had a median SAEDR score of 0.538, whereas the ADRs that had been included in at least one boxed warning had a median SAEDR score of 0.624. We found that among the ADRs that have appeared on a drug label, those that have been included in at least one boxed warning were significantly more severe (P<.001) than the ADRs that have never been included in a boxed warning section based on a one-sided t test (Figure 4).

Figure 4. Differences in adverse drug reaction (ADR) severity between ADR groupings and discovery periods: ADR groups (x-axis) versus SAEDR scores (y-axis). The grey dashed line indicates the median severity, 0.624, of the ADRs that have been included in a boxed warning. The orange boxed line indicates the median severity, 0.538, of the ADRs that appear on a drug label but have not been included in a boxed warning. (A) The ADRs that were listed as a black box warning at least once (n=356) were significantly more severe than those that have not appeared in a black box warning (n=3305). (B) The ADRs that are disproportionately reported for men (n=56,405) are significantly more severe than those disproportionately reported for women (n=50,801). There was no significant difference (ns) in severity between the ADRs included in black box warnings and those disproportionately reported for men. (C) The ADRs discovered in the postmarketing period (n=11,506) are significantly more severe (*** indicate *P*<.001) than those discovered in the clinical trials (n=35,450). The ADRs identified in the postmarketing period through OFFSIDES (n=350,631) and the postmarketing polypharmacy ADRs identified through TWOSIDES (n=4,210,513) are significantly more severe than those discovered in the clinical trials. The severity of all postmarketing ADR groups is significantly less than the severity of the ADRs that have appeared in a black box warning. SAEDR: Severity of Adverse Events Derived from Reddit.



We found that the ADR HLGTs that are disproportionately reported for men were significantly more severe (*P*<.001) than those disproportionately reported for women (Figure 4). The HLGTs disproportionately experienced by men were not significantly different from the severity of the ADRs that have been included in at least one boxed warning, *P*=.13, based on a one-sided *t* test.

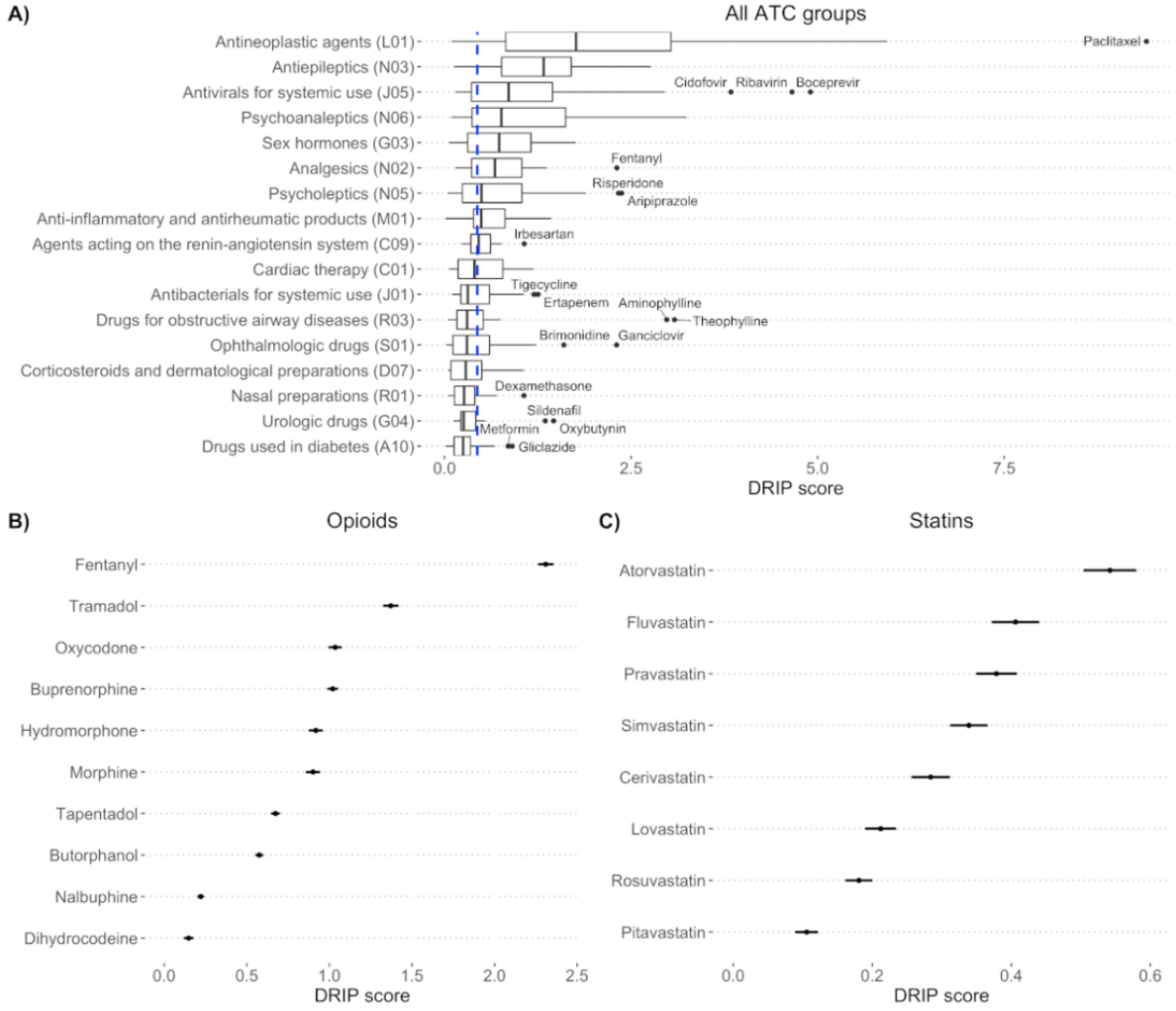
We examined the severity of the ADRs that were found during the clinical trial versus the severity of those discovered in the postmarketing period based on the SIDER label annotations. We found that the ADRs discovered in the postmarketing period had significantly higher severity (*P*<.001) than the ADRs discovered in the clinical trials based on a one-sided *t* test (Figure 4). We compared the severity of the clinical trial ADRs with the severity of those discovered in the postmarketing period using FAERS data. The study by Tatonetti et al [28] identified two different sets of ADRs: OFFSIDES is a set of ADRs disproportionately reported for a drug, whereas TWOSIDES is a set of ADRs disproportionately reported for a pair of drugs being taken concurrently, after correcting for case demographics and other information. The severity distribution of ADRs in both OFFSIDES and TWOSIDES was significantly higher (*P*<.001) than the severity distribution discovered in the clinical trials based on a one-sided *t* test (Figure 4). We found that the ADRs associated with polypharmacy, through TWOSIDES, had the highest severity of the postmarketing ADRs. All postmarketing ADRs were significantly lower in severity than the boxed warning ADRs based on a one-sided *t* test.

DRIP Score Analysis

We calculated the DRIP scores for 968 drugs using SIDER label data, with a resulting median DRIP score of 0.439. The Spearman correlation between the drugs ranked by the proportion of FAERS cases, with that drug as the primary suspect that resulted in death, and the drugs ranked by our DRIP scores was 0.377, *P*<.001.

We analyzed DRIP score distributions by the Anatomical Therapeutic Chemical (ATC) classification system for the subset of drugs with an ATC designation (Figure 5). Antineoplastic and antiepileptic drugs were the drug classes with the highest DRIP scores, indicating that the drugs in these classes have severe ADR profiles. Drugs used in the management and treatment of diabetes and urologic issues had the lowest DRIP scores, indicating that these drugs are relatively safe and have less-severe side effects. We examined individual drugs within the opioid class and found that fentanyl had a markedly higher DRIP score than other drugs in the same class (Figure 5). Drugs within the statin class were primarily below the overall median DRIP score (0.439), with only atorvastatin being markedly above the median DRIP score (Figure 5).

Figure 5. DRIP scores calculated using side effect severity and frequency. (A) DRIP scores (x-axis) for ATC classification system groups (y-axis) ordered by median severity with outliers labeled. The dashed blue line indicates the median DRIP score of 0.439 across all drugs. (B) DRIP scores (x-axis) for opioid class drugs (y-axis), as denoted by the N02A ATC group. Fentanyl, an extremely potent synthetic opioid, is ranked first as having the most severe side effect. (C) DRIP scores (x-axis) for statins (y-axis), as denoted by the C10AA ATC group. Atorvastatin, the strongest statin in broad use, is ranked to have the most severe side effect, whereas lower-efficacy statins are ranked as having less-severe side effects. ATC: Anatomical Therapeutic Chemical; DRIP: drug-risk profile.



Discussion

Principal Findings

In this study, we present the SAEDR scores and quantitative estimates of the severity of 28,113 MedDRA ADR terms. The SAEDR scores have strong correlations with both crowdworker-based severity estimates and real-world ADR case outcome statistics. We generated these estimates using a network-based label propagation approach that required only a small percentage of the terms to be labeled. This study demonstrated the feasibility of using distantly supervised techniques such as label propagation to generate quantitative values for medical concepts.

Our approach enabled us to increase the number of ADRs with quantitative severity estimates from 2929 to 28,113, an almost tenfold increase, while minimizing the human time involved in generating the estimates. Our severity estimates can be updated routinely based on the recomputation of the embeddings and reseeding of the reference severe or benign ADRs. We noted that our severity estimates had a higher correlation with the aggregate crowdworker rankings (ρ=0.735) than the crowdworker rankings had among the three replicates in the original study (ρ=0.710). In addition, our estimates had a greater correlation with real-world outcomes such as FAERS cases resulting in death (ρ=0.595) than the crowdworker rankings (ρ=0.53). Thus, our severity estimates track with both human judgments of ADR severity and real-world outcomes of consequence to patient health.

We found that our SAEDR scores had a strong negative correlation, $\rho=-0.748$, with no case outcome being reported in the FAERS. We assigned a *No Outcome* label to the FAERS cases for which none of the available but relatively severe case outcomes had been reported. We took this as an indication that the FAERS cases lacking outcomes may be a result of cases having an outcome below the level of seriousness that is reportable in the FAERS. FAERS case outcome completeness could potentially be improved by creating a lower-severity category for case outcome documentation. We also noted that

the SAEDR scores have a stronger correlation with serious outcomes in FAERS cases than with death outcomes. This can be partially explained by the fact that many more ADRs result in serious outcomes than in death, leading to poor differentiation of severity based on death for many ADRs where death is a rare or unobserved outcome.

In an effort to compare the impacts of using different methods to select the initial severe and benign ADR seed terms for label propagation, we compared crowdworker and real-world outcomes–based seed selections. We found that both methods resulted in consistently high severity scores for ADRs in the highest-severity SOC groups such as *Neoplasms* and *Cardiac Disorders*. We observed the largest differences in the *Product Issues* SOC group. Real-world outcomes–based seeds increased the severity of psychiatric disorders and social circumstances and decreased the severity of product issues. The importance and impact of psychiatric disorders may not be apparent to those not in the medical profession. Thus, it is possible that the crowdworkers did not perceive mental illness to be as severe as physical illnesses, leading to a bias in the rankings. The use of real-world outcomes–based rankings to select seeds led to more psychiatric disorders being included in the seed terms for the severity propagation. Similarly, the shifts in social circumstances and product issues are likely due to their limited presence or absence within the original crowdworker rankings because only seven ADRs in the social circumstances category were included in the 2929 ADRs ranked by crowdworkers, and no *Product Issues* ADRs were included in that ranking. This absence resulted in no ADRs from these groups being included in the initial seed set of labeled ADRs for the crowdworker-seeded method run, likely limiting the ability of the label propagation to accurately estimate the severity of these ADRs.

We found that the overall severity ordering of the SOC classes based on our SAEDR scores made intuitive sense. We were surprised by the ability of the label propagation approach to adjust the severity estimates based on particular key terms. For instance, the ADR terms for cancer (*Neoplasm*) ADRs are relatively similar on a lexical basis, but individual modifiers such as malignant and benign resulted in significantly different SAEDR scores for the respective groups of ADRs (Figure S3 of Multimedia Appendix 1). Similarly, examining groups of ADRs at different MedDRA hierarchy levels demonstrated relative severity estimates that were sensible to us, such as *Heart Failure Signs and Symptoms* having lower severity than the different types of actual heart failure.

We found a significantly higher estimate of severity among the ADRs that had been included in a boxed warning section than among those that had not. This offers further validation of the SAEDR scores because they are in agreement with past regulatory and drug labeling decisions. Our comparison of HLGTs with disproportionate rates between sexes revealed that male-associated ADRs were more severe. Notably, the SAEDR scores of the male-associated ADRs were not significantly different from those of the ADRs that have been included in a boxed warning. This is in agreement with previous work that highlighted the relatively higher severity of ADRs experienced by men [29,30].

We found that ADRs discovered in the postmarketing period are generally more severe than those discovered in the clinical trials based on our SAEDR scores. This indicates that postmarketing surveillance and ongoing regulatory discussions of risk-benefit trade-offs for particular drugs are necessary to keep the public safe [31]. These findings are not new because a recent study examining internal Food and Drug Administration data on drug labeling changes found that 35% of the ADRs added to drug labels were added to the boxed warnings and warnings and precautions sections [32]. Of the postmarketing ADRs, those involving more than one drug, as identified by TWOSIDES, were among those with the highest severity, indicating the increased risks associated with polypharmacy. This finding highlights the need for further research into the safety of polypharmacy because approximately half of the individuals prescribed a prescription drug are prescribed more than one concurrent medication [33].

One of the original aims of the pioneering study by Tallarida et al [7] was to create quantitative risk scores for individual drugs. Our DRIP scores are an attempt to do so, and we combined ADR severity with frequency information to generate a numerical estimate of a drug's risk profile. We were able to generate DRIP scores for 968 drugs using ADR frequency data from the SIDER. Our DRIP scores have a modest Spearman correlation, $\rho=0.377$, with FAERS cases that resulted in death, where that drug was the primary suspect. Generally, correlations greater than 0.8 are considered strong, but this generally assumes that the comparison is with a gold standard ground truth. Because of a lack of an objective gold standard of DRIPs, we compared with FAERS case outcomes for the drugs in question, although the case outcomes are imperfect because of reporting bias and other issues. In addition, the DRIP scores were created using estimates of ADR frequencies from the SIDER that are primarily derived from clinical trial data. These ADR frequency estimates are affected by the dosage and underlying health of the clinical trial population, both of which are only sometimes reported. High-quality frequency estimates for postmarketing ADRs are essentially nonexistent, and many challenges exist to accurately estimating their frequency [29]. These issues limit the ability of our DRIP scores to accurately quantify risk. However, despite the limitations of the benchmark data and the ADR frequency information, the DRIP scores still demonstrate a modest correlation with real-world outcomes. We interpret this as an indication that our DRIP scores track with real-world outcomes and capture signals related to a drug's safety profile that could be useful for downstream application. Further evidence of their validity is provided by the qualitative evaluation of the DRIP score distributions for the ATC groups. The relative distributions of the ATC groups make intuitive sense, with drugs used to treat cancer—antineoplastic drugs—having the most severe side effect profiles, whereas drugs used to treat diseases that are known to have safe and effective treatment options, such as diabetes, are ranked among the safest drug categories.

We present ADR severity estimates that track with both real-world outcomes and human perception, but these estimates are still limited. We used word embeddings trained on Reddit data, but other social media data could improve the severity

prediction model. We used RedMed because the model was publicly available and contained many ADR terms of interest to this study. However, Reddit as a social media platform is skewed toward men and young people, who may discuss ADRs in different ways from the general population. Learning new embeddings from social media corpora generated by the patient groups most at risk for a given set of ADR experiences or from biomedical literature might improve model performance and address issues with demographic model biases.

FAERS case outcome proportions as a benchmark for severity are limited because of the outcomes being reported at the case level by the FAERS. Because of this case-level aggregation, the ADRs in the severe category may have inflated severity. For instance, symptoms of cardiac arrest may often be reported as separate ADRs for a FAERS case of cardiac arrest, resulting in symptoms that seem to be more severe than they are when occurring independently of the severe ADR with which they are associated.

Our results indicate that men reported more severe ADRs based on sex-specific ADRs identified in the study by Chandak et al [27]. Other research efforts have reported similar findings based on ADR-reporting databases [29], but there are potential sources of bias that could have affected this result. The data underlying the sex-specific ADRs identified in the study by Chandak et al [27] were derived from the FAERS. The study method focused on the analysis of a sex-balanced cohort of case reports, with cohort creation through propensity score matching of individuals. Propensity score matching can correct for factors included in the model, but it is unable to correct for underlying biases in reporting. It has been documented that there are higher rates of ADR reporting for female patients than for male patients, whereas male patients tend to have more severe outcomes reported [30]. It is likely that male patients are enriched for severe ADRs because of this reporting bias, and ADR reports should not be conflated with all ADRs experienced by individuals because of the low rate of ADR reporting [34]. Another limitation is that sex-specific ADRs were only reported at the HLGT level, resulting in SAEDR scores that were an average of all PTs within the HLGT. Figure 3 highlights the large range in terms of severity, even within a particular HLGT. It is possible that an analysis of sex-specific ADRs at the PT level would result in a different finding. Overall, although ADRs disproportionately reported in men are more severe than those disproportionately reported in women, more research is needed to reach a complete understanding of sex-specific differences in ADR severity.

Although we captured some nuances in ADR severity, such as benign versus malignant neoplasms, we only created SAEDR scores for the ADR terms contained in the MedDRA. However, while exploring the RedMed word embeddings, we observed that individual ADR terms were often contained in phrases indicating modified severity. For instance, a report of stomach pain might be modified with an adjective such as excruciating or mild. Although the MedDRA does not contain an intensity scale component for its terms, capturing this information from patients might further enable risk-benefit trade-off calculations. The approaches for general word sentiment assignment, similar to the one used in the study by Hamilton et al [22], could be repurposed here to assign severity to ADR modifier terms.

Conclusions

In summary, we demonstrated that lexical networks and label propagation can be used quantitatively to estimate the severity of medical conditions. We showed the distributions of ADR severity among different groups of conditions, different groups of patients, and different ADR discovery time points. We combined our SAEDR scores with available ADR frequency data to generate quantitative DRIP scores and examined the distribution of the resulting DRIP scores. Our results (and future improved estimates) enable new quantitative analyses within the field of pharmacovigilance. To this end, we provide the complete set of SAEDR scores and DRIP scores in Multimedia Appendix 2 and Multimedia Appendix 3, respectively.

Acknowledgments

The authors would like to thank the editor and the reviewers for their helpful and thoughtful feedback during the review process. Some of the computations for this project were performed on the Sherlock cluster. The authors would like to thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to the research results. AL is supported by the National Science Foundation (NSF) Graduate Research Fellowship (DGE 1656518). TH is supported by the National Institutes of Health (NIH; R01DK103358); Simons Foundation and NSF (IOS-1546218, R35GM122515, and NSF CBET 1728858); and the NIH (R01AI130945). RBA is supported by the NIH (GM102365 and HG010615) and the Chan Zuckerberg Biohub.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary figures and associated captions.
[PPTX File , 885 KB - jmir_v23i10e27714_app1.pptx]

Multimedia Appendix 2
Severity of Adverse Events Derived from Reddit scores for 12,198 adverse drug reactions.

[ZIP File (Zip Archive), 391 KB - jmir_v23i10e27714_app2.zip]

Multimedia Appendix 3
Drug-risk profile scores.
[ZIP File (Zip Archive), 30 KB - jmir_v23i10e27714_app3.zip]

References

1. Bourgeois F, Shannon M, Valim C, Mandl KD. Adverse drug events in the outpatient setting: an 11-year national analysis. Pharmacoepidemiol Drug Saf 2010 Sep;19(9):901-910 [FREE Full text] [doi: 10.1002/pds.1984] [Medline: 20623513]
2. Watanabe JH, McInnis T, Hirsch JD. Cost of prescription drug-related morbidity and mortality. Ann Pharmacother 2018 Sep;52(9):829-837. [doi: 10.1177/1060028018765159] [Medline: 29577766]
3. Aronson JK, Ferner RE. Clarification of terminology in drug safety. Drug Saf 2005;28(10):851-870. [doi: 10.2165/00002018-200528100-00003] [Medline: 16180936]
4. Ahmad SR. Adverse drug event monitoring at the Food and Drug Administration. J Gen Intern Med 2003 Jan;18(1):57-60 [FREE Full text] [doi: 10.1046/j.1525-1497.2003.20130.x] [Medline: 12534765]
5. Gunderson F, Gunderson H, Ferguson Jr E. Code of Federal Regulations. In: Food Standards and Definitions in the United States. Cambridge, Massachusetts, United States: Academic Press; 1963:147-154.
6. Shaffer ML, Watterberg KL. Joint distribution approaches to simultaneously quantifying benefit and risk. BMC Med Res Methodol 2006 Oct 12;6:48 [FREE Full text] [doi: 10.1186/1471-2288-6-48] [Medline: 17038184]
7. Tallarida RJ, Murray RB, Eiben C. A scale for assessing the severity of diseases and adverse drug reactions. Application to drug benefit and risk. Clin Pharmacol Ther 1979 Apr;25(4):381-390. [doi: 10.1002/cpt1979254381] [Medline: 428184]
8. Gottlieb A, Hoehndorf R, Dumontier M, Altman RB. Ranking adverse drug reactions with crowdsourcing. J Med Internet Res 2015;17(3):e80 [FREE Full text] [doi: 10.2196/jmir.3962] [Medline: 25800813]
9. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. arXiv - Computer Science, Computation and Language. URL: http://arxiv.org/abs/1301.3781 [accessed 2021-07-19]
10. Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014 Presented at: Conference on Empirical Methods in Natural Language Processing (EMNLP); October, 2014; Doha, Qatar p. 1532-1543. [doi: 10.3115/v1/d14-1162]
11. Firth J. A Synopsis of Linguistic Theory 1930-1955, Selected Papers of JR Firth 1952 - 1959. In: Studies in Linguistic Analysis. Oxford, United Kingdom: Philological Society; 1957:168-205.
12. Nikfarjam A, Sarker A, O'Connor K, Ginn R, Gonzalez G. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. J Am Med Inform Assoc 2015 May;22(3):671-681 [FREE Full text] [doi: 10.1093/jamia/ocu041] [Medline: 25755127]
13. Hanson CL, Cannon B, Burton S, Giraud-Carrier C. An exploration of social circles and prescription drug abuse through Twitter. J Med Internet Res 2013;15(9):e189 [FREE Full text] [doi: 10.2196/jmir.2741] [Medline: 24014109]
14. Yan L, Yan X, Tan Y, Sun S. Shared Minds: How patients use collaborative information sharing via social media platforms. Prod Oper Manag 2018 Jun 26;28(1):9-26. [doi: 10.1111/poms.12895]
15. Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance: A review. J Biomed Inform 2015 Apr;54:202-212 [FREE Full text] [doi: 10.1016/j.jbi.2015.02.004] [Medline: 25720841]
16. Alvaro N, Conway M, Doan S, Lofi C, Overington J, Collier N. Crowdsourcing Twitter annotations to identify first-hand experiences of prescription drug use. J Biomed Inform 2015 Dec;58:280-287 [FREE Full text] [doi: 10.1016/j.jbi.2015.11.004] [Medline: 26556646]
17. Klein A, Sarker A, Rouhizadeh M, O'Connor K, Gonzalez G. Detecting personal medication intake in Twitter: An annotated corpus and baseline classification system. BioNLP 2017:136-142. [doi: 10.18653/v1/W17-2316]
18. Lavertu A, Altman RB. RedMed: Extending drug lexicons for social media applications. J Biomed Inform 2019 Nov;99:103307 [FREE Full text] [doi: 10.1016/j.jbi.2019.103307] [Medline: 31627020]
19. Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. Nucleic Acids Res 2015 Oct 19:1075-1079 [FREE Full text] [doi: 10.1093/nar/gkv1075] [Medline: 26481350]
20. Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). Drug Saf 1999 Feb;20(2):109-117 [FREE Full text] [doi: 10.2165/00002018-199920020-00002] [Medline: 10082069]
21. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004 Jan 1;32(Database issue):267-270 [FREE Full text] [doi: 10.1093/nar/gkh061] [Medline: 14681409]
22. Hamilton WL, Clark K, Leskovec J, Jurafsky D. Inducing domain-specific sentiment lexicons from unlabeled corpora. Proc Conf Empir Methods Nat Lang Process 2016 Nov;2016:595-605 [FREE Full text] [doi: 10.18653/v1/D16-1057] [Medline: 28660257]
23. Määttä E, Haapala A. Python-Levenshtein. Python Software Foundation. URL: https://pypi.org/project/python-Levenshtein/ [accessed 2020-05-25]

24. Grover A, Leskovec J. node2vec: Scalable feature learning for networks. KDD 2016 Aug;2016:855-864 [FREE Full text] [doi: 10.1145/2939672.2939754] [Medline: 27853626]
25. Efron B. Bootstrap Methods: Another look at the jackknife. Ann Statist 1979 Jan 1;7(1):1-26. [doi: 10.1214/aos/1176344552]
26. Wu L, Ingle T, Liu Z, Zhao-Wong A, Harris S, Thakkar S, et al. Study of serious adverse drug reactions using FDA-approved drug labeling and MedDRA. BMC Bioinformatics 2019 Mar 14;20(Suppl 2):97 [FREE Full text] [doi: 10.1186/s12859-019-2628-5] [Medline: 30871458]
27. Chandak P, Tatonetti NP. Using machine learning to identify adverse drug effects posing increased risk to women. Patterns (N Y) 2020 Oct 09;1(7):100108 [FREE Full text] [doi: 10.1016/j.patter.2020.100108] [Medline: 33179017]
28. Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. Sci Transl Med 2012 Mar 14;4(125):125ra31 [FREE Full text] [doi: 10.1126/scitranslmed.3003377] [Medline: 22422992]
29. Watson S, Caster O, Rochon PA, den Ruijter H. Reported adverse drug reactions in women and men: Aggregated evidence from globally collected individual case reports during half a century. EClinicalMedicine 2019 Dec;17:100188 [FREE Full text] [doi: 10.1016/j.eclinm.2019.10.001] [Medline: 31891132]
30. Holm L, Ekman E, Jorsäter Blomgren K. Influence of age, sex and seriousness on reporting of adverse drug reactions in Sweden. Pharmacoepidemiol Drug Saf 2017 Mar;26(3):335-343. [doi: 10.1002/pds.4155] [Medline: 28071845]
31. Lavertu A, Vora B, Giacomini KM, Altman R, Rensi S. A new era in pharmacovigilance: Toward real-world data and digital monitoring. Clin Pharmacol Ther 2021 May 25;109(5):1197-1202 [FREE Full text] [doi: 10.1002/cpt.2172] [Medline: 33492663]
32. Pinnow E, Amr S, Bentzen SM, Brajovic S, Hungerford L, St George DM, et al. Postmarket safety outcomes for New Molecular Entity (NME) drugs approved by the Food and Drug Administration between 2002 and 2014. Clin Pharmacol Ther 2018 Aug;104(2):390-400. [doi: 10.1002/cpt.944] [Medline: 29266187]
33. Quinn KJ, Shah NH. A dataset quantifying polypharmacy in the United States. Sci Data 2017 Oct 31;4:170167 [FREE Full text] [doi: 10.1038/sdata.2017.167] [Medline: 29087369]
34. Hazell L, Shakir SA. Under-reporting of adverse drug reactions : a systematic review. Drug Saf 2006;29(5):385-396. [doi: 10.2165/00002018-200629050-00003] [Medline: 16689555]

Abbreviations

ADR: adverse drug reaction
ATC: Anatomical Therapeutic Chemical classification system
DRIP: drug-risk profile
FAERS: Food and Drug Administration Adverse Event Reporting System
HLGT: high-level group term
MedDRA: Medical Dictionary for Regulatory Activities
PT: preferred term
SAEDR: Severity of Adverse Events Derived from Reddit
SIDER: Side Effect Resource
SOC: System Organ Class

Edited by R Kukafka; submitted 03.02.21; peer-reviewed by A Sarker, S Matsuda; comments to author 03.04.21; revised version received 25.05.21; accepted 14.06.21; published 21.10.21.

Please cite as:
Lavertu A, Hamamsy T, Altman RB
Quantifying the Severity of Adverse Drug Reactions Using Social Media: Network Analysis
J Med Internet Res 2021;23(10):e27714
URL: https://www.jmir.org/2021/10/e27714
doi:10.2196/27714
PMID:34673524

Original Paper

Areas of Interest and Attitudes Toward Antiobesity Drugs: Thematic and Quantitative Analysis Using Twitter

Miguel Angel Alvarez-Mon[1,2], MD, PhD; Maria Llavero-Valero[3,4], MD, PhD; Angel Asunsolo del Barco[5,6], MD, PhD; Cristina Zaragozá[7], PhD; Miguel A Ortega[2,6], PhD; Guillermo Lahera[2,6,8,9], MD, PhD; Javier Quintero[1,10], MD, PhD; Melchor Alvarez-Mon[2,6,11,12], MD, PhD

[1]Department of Psychiatry and Mental Health, Hospital Universitario Infanta Leonor, Madrid, Spain
[2]Department of Medicine and Medical Specialities, Faculty of Medicine and Health Sciences, University of Alcala, Alcalá de Henares, Spain
[3]Department of Endocrinology and Clinical Nutrition, Hospital Universitario Infanta Leonor, Madrid, Spain
[4]Department of Preventive Medicine and Public Health, School of Medicine, University of Navarra, Pamplona, Spain
[5]Department of Surgery, Medical and Social Sciences, Faculty of Medicine and Health Sciences, University of Alcala, Madrid, Spain
[6]Ramón y Cajal Institute of Sanitary Research, Madrid, Spain
[7]Pharmacology Unit, Biomedical Sciences Department, University of Alcalá, Madrid, Spain
[8]Centro de Investigación Biomédica en Red de Salud Mental, Madrid, Spain
[9]Psychiatry Service, Hospital Universitario Principe de Asturias, Alcalá de Henares, Spain
[10]Department of Legal Medicine and Psychiatry, Complutense University, Madrid, Spain
[11]Internal Medicine and Autoimmunity/Rheumatology Service, Hospital Universitario Principe de Asturias, Madrid, Spain
[12]Centro de Investigación Biomédica en Red en Enfermedades Hepáticas y Digestivas, Madrid, Spain

Corresponding Author:
Miguel Angel Alvarez-Mon, MD, PhD
Department of Psychiatry and Mental Health
Hospital Universitario Infanta Leonor
Av Gran Vía del Este 80
Madrid, 28031
Spain
Phone: 34 911 91 80 00
Email: maalvarezdemon@icloud.com

Abstract

Background: Antiobesity drugs are prescribed for the treatment of obesity in conjunction with healthy eating, physical activity, and behavior modification. However, poor adherence rates have been reported. Attitudes or beliefs toward medications are important to ascertain because they may be associated with patient behavior. The analysis of tweets has become a tool for health research.

Objective: The aim of this study is to investigate the content and key metrics of tweets referring to antiobesity drugs.

Methods: In this observational quantitative and qualitative study, we focused on tweets containing hashtags related to antiobesity drugs between September 20, 2019, and October 31, 2019. Tweets were first classified according to whether they described medical issues or not. Tweets with medical content were classified according to the topic they referred to: side effects, efficacy, or adherence. We additionally rated it as positive or negative. Furthermore, we classified any links included within a tweet as either scientific or nonscientific. Finally, the number of retweets generated as well as the dissemination and sentiment score obtained by the antiobesity drugs analyzed were also measured.

Results: We analyzed a total of 2045 tweets, 945 of which were excluded according to the criteria of the study. Finally, 320 out of the 1,100 remaining tweets were also excluded because their content, although related to drugs for obesity treatment, did not address the efficacy, side effects, or adherence to medication. Liraglutide and semaglutide accumulated the majority of tweets (682/780, 87.4%). Notably, the content that generated the highest frequency of tweets was related to treatment efficacy, with liraglutide-, semaglutide-, and lorcaserin-related tweets accumulating the highest proportion of positive consideration. We found the highest percentages of tweets with scientific links in those posts related to liraglutide and semaglutide. Semaglutide-related tweets obtained the highest probability of likes and were the most disseminated within the Twitter community.

Conclusions: This analysis of posted tweets related to antiobesity drugs shows that the interest, beliefs, and experiences regarding these pharmacological treatments are heterogeneous. The efficacy of the treatment accounts for the majority of interest among Twitter users.

(J Med Internet Res 2021;23(10):e24336) doi:10.2196/24336

KEYWORDS

obesity; social media; Twitter; drug therapy; pharmacotherapy; attitude; thematic analysis; quantitative analysis; drug

Introduction

Obesity is an increasingly prevalent disease, with high rates of associated morbidity and mortality [1]. The treatment of obesity remains only partially effective [2]. Moreover, the pharmacological treatment of obesity is becoming a more significant tool in the management of the disease [2]. However, frequency in the use of antiobesity drug treatments among those patients that could potentially benefit from them is minimal [3]. Moreover, both personal and social attitudes, combined with accessible information on available treatments, have been shown to be relevant for obtaining the expected clinical outcomes of pharmacological intervention [4]. Furthermore, it has been shown that social support is a potential beneficial component of weight loss programs [5,6].

In recent years, social media has become a pivotal instrument for disseminating knowledge [7]. Accordingly, the internet has modified how people communicate and how they share and seek out information regarding health [8]. Social networks are extensively used for the study of obesity, including the analysis of public attitudes, the social support of patients, patients' behavior, and treatment efficacy [9-11]. Being that information pertaining to health posted over social media is oftentimes more spontaneous in nature, it serves more as a complementary perspective to data collected from medical surveys, clinical trials, and consultancies with medical professionals [12-14].

Twitter, one of the most popular and widely used social media platforms, is currently considered to be an effective channel of communication [15]. Within this context, different agents in the health sector have realized Twitter´s potential for acquiring and distributing medical information [16,17]. In addition, Twitter users demonstrate a great interest in obesity and eating disorders [18,19]. Moreover, it has been shown that Twitter can be an effective platform for delivering interventions aimed at treating obesity [20]. The analysis of tweets about obesity, diet, and treatments is a recent relevant area of study for understanding the actual sentiments of society, patients, and health providers [21]. The trivialization, stigmatization, and mockery directed at obesity and other disorders by Twitter users have been reported [22]. Until now, however, topics of interest among Twitter users regarding the pharmacological treatment of obesity have not been identified. Furthermore, the dissemination of medication-related tweets tied to obesity remains unknown. The analysis of the feelings and experiences toward pharmacological treatment is relevant for the understanding of patients' attitudes to these drugs and the identification of concerns and needs potentially related to treatment adherence [23,24].

In this study, we performed an analysis of the content and key metrics of all the tweets generated concerning medications approved by the US Food and Drug Administration (FDA) for obesity treatment over a period of 6 weeks. We also investigated the areas of interest of those tweets containing medical content and the inclusion of links to related informative resources. Finally, we investigated the tweets' dissemination and overall sentiment.

Methods

Data Collection

In this observational quantitative and qualitative study, we focused on searching for tweets that referred to medications approved by the FDA for the treatment of overweight or obese status: Xenical, orlistat, Alli, Belviq, lorcaserin, Qsymia, phentermine-topiramate, Contrave, bupropion-naltrexone, Saxenda, Victoza, liraglutide, Ozempic, and semaglutide. The inclusion criteria for tweets were the following: being public; using any of the previously mentioned hashtags; being posted between September 20, 2019, and October 31, 2019; and being posted in English language. This 6-week period was chosen to avoid any potential bias within the content of the tweets. In addition, we obtained the number of likes each tweet generated, the date and time of each tweet, and the potential reach and impact of each hashtag.

Search Tool

We used the Twitter Firehose data stream, which is managed by Gnip and allows access to 100% of all public tweets that match a certain criteria (query) [25]. In our study, the search criteria were the previously mentioned hashtags. Tweet Binder, the search engine we employed, uses node.js and PHP language, which enabled us to analyze tweets in JavaScript Object Notation (JSON) format (used by Gnip).

Content Analysis Process

All 2045 retrieved tweets were directly inspected by 2 raters (MAAM and MLV). First, we scanned all of the tweets, excluding 945 tweets that provided information that was too limited (eg, tweets consisting mainly of hashtags), that contained only pictures, or that included hashtags of more than 1 treatment. All the remaining tweets were considered for thematic content analysis. Second, we created a codebook based on our research questions, our previous experience in analyzing tweets, and what we determined to be the most common tweet themes. Third, 2 raters analyzed 150 tweets separately to test the suitability of the codebook. Discrepancies were discussed between the raters and with another author (MAM). After the codebook was revised, the interrater reliability was reassessed

with a different set of 150 tweets. As this resulted in adequate κ values (range 0.68-0.99), the raters then proceeded to analyze 1100 tweets equally distributed among both. Each tweet, depending on its content, was categorized under side effects, efficacy, or adherence. In addition, users were classified into 3 categories: patients and relatives, health professionals, or health institutions. These categorizations were defined based on the description of user profiles and the content of user tweets. In those cases in which the nature of the user was not possible to know, they were classified as indeterminate. Finally, we analyzed any links included within a tweet, classifying them as either scientific or nonscientific. More specifically, those links attributed to a scientific source, including medical journals, academic institutions, hospitals, and official websites, were classified as scientific. The classification criteria we used and examples of tweets are shown in Textbox 1.

Textbox 1. Examples of tweets related to efficacy, side effects, and adherence (usernames and personal names have been removed).

Efficacy (the ability or inability of a treatment to provide a beneficial effect)

- "Oral semaglutide can effectively and safely reduce blood glucose, body weight and systolic blood pressure: A systematic review and meta-analysis."
- "Ozempic is superior to Invokana in reducing A1c and body weight".
- "More than just #weightloss ? Liraglutide improves hepatic steatosis and metabolic dysfunctions in a 3-week dietary mouse model of nonalcoholic steatohepatitis."
- "The PIONEER 4 trial showed that oral semaglutide is noninferior to injectable liraglutide and superior to placebo in improving glycemic control and weight loss at 26 weeks among patients with type 2 #diabetes.".

Side effects (any effect that is secondary to the one intended either adverse or beneficial; tweets discussing tolerability of the drug were also included)

- "No side effects with the Ozempic and I've been on it since July of last year. Now the metformin is a whole different story. First week or so the sight of food made me sick and it made my stomach act up. I'm actually glad to be off of it."
- "Does Contrave Make Anyone Else's Hands Shake?"
- "I saw the Saxenda results on people I know. It is fantastic. However, it comes with its challenges. Nausea, headaches and terrible moods. glucagon-like peptide 1 (GLP-1) analogs, such as liraglutide, is the possibility of developing pancreatitis. #usmle"
- "A bit personal, but I'm curious to hear others' stories. I've been taking Ozempic for a year or so, but the last 6-8 weeks I've started throwing up quite a bit. I don't have a history of this at all. Anyone else had this problem with Ozempic? Thank you!"
- "Those embalming leaves seem to have less side effects than ozempic."
- "Gallstone Disease in Patients Treated with Liraglutide: In a large randomized trial a small but statistically significant rise in this adverse event was noted."

Adherence (the degree of conformity to the recommendations about the treatment with respect to the timing, dosage, or frequency)

- "I'm a quibbler, I can't help it. There is a medicine called Qsymia that seems effective for long term weight loss. Its not without side effects or risks."
- "Morning Twitter nightmare week over. Week 21 liraglutide. Despite everything total adherence. Adherence to duloxetine not good. Quorn burger for breakfast & fruit. Happy today."

Measuring Interest and Influence on Twitter

We analyzed the number of likes generated by each tweet as an indicator of user interest on a given topic. We also measured the potential reach and impact of all analyzed hashtags in order to best assess tendencies in the dissemination of tweets. In this study, impact was defined as a numerical value representing the potential views a tweet may receive, while reach was defined as a numerical value measuring the potential audience of the hashtag.

In addition, we measured how positive or negative a hashtag was on a scale from 1 (negative) to 100 (positive). Sentiment analysis tools were used to analyze all words contained in a tweet, with each word having its own score that could vary depending on the context. The average score of all the tweets with a certain hashtag determined that hashtag's overall sentiment score. According to this score, we classified each hashtag into 3 categories: negative (0-40), neutral (>40-60), and positive (>60-100).

Ethical Considerations

This study was approved by the Research Ethics Committee at the University of Alcala.

Statistical Analysis

A descriptive study of the sample was performed. The qualitative variables are described as absolute frequency (number) and relative frequency (percentage). The percentages found were compared using the chi-square test. The numbers of retweets and likes per original tweet about the different hashtags were verified by graphs and statistical test (Kolmogorov-Smirnof test), and they did not follow a normal distribution. The differences found between the treatment groups were compared using the Kruskal-Wallis test. All statistical analyses were performed using SPSS version 25 (IBM Corp).

Results

Liraglutide and Semaglutide Accumulated the Most Interest Among Twitter Users

Our search tool provided a total of 2045 tweets, 945 of which were excluded according to the criteria of the study. Next, 320 out of the 1100 remaining tweets were also excluded because their content, although related to drugs for obesity treatment, did not address the efficacy, side effects, or adherence to the medication. Finally, we classified the content of the remaining 780 tweets.

The number of tweets with hashtags referencing the 6 drug groups approved for obesity treatment were significantly different, with the incidence of tweets related to liraglutide and semaglutide at least being 10 times higher than that related to the other 5 drug groups (Table 1).

Next, we found significant differences in the distribution of the content. Notably, the content that generated the highest frequency of tweets was that related to treatment efficacy, with liraglutide-, semaglutide-, and lorcaserin-related tweets accumulating the highest proportion of positive consideration (P<.001). In contrast, the highest percentage of tweets with a negative valuation towards efficacy was found in those posts related to bupropion-naltrexone (3/30, 10%), while those containing a mention of liraglutide (14/319, 4.4%) and semaglutide (2/363, 0.6%) had a much lower negative percentage.

Tweets with a negative valuation of side effects were mainly observed in those related to orlistat (4/15, 26.7%) and bupropion-naltrexone (4/30, 13.3%) but rarely in those mentioning liraglutide, semaglutide, and phentermine-topiramate. On the other hand, tweets with a positive valuation of side effects were minimal and were found mainly in those posts related to bupropion-naltrexone, semaglutide, and liraglutide (P=.04). Finally, the frequency of tweets with content related to adherence to treatment was low, with negative considerations predominant among liraglutide, bupropion-naltrexone, and semaglutide (P<.001). On the other hand, positive valuations of adherence to treatment were observed in those tweets related to orlistat, semaglutide, and liraglutide.

Table 1. Descriptive characteristics of the original tweets included in the analysis categorized by total amount per drug and category (side effects, efficacy, adherence, and link).

Category	Orlistat, n (%) (N=15)	Lorcaserin, n (%) (N=27)	Phentermine-topiramate, n (%) (N=26)	Bupropion-naltrexone, n (%) (N=30)	Liraglutide, n (%) (N=319)	Semaglutide, n (%) (N=363)	P value[a]
Side effects							.04
No mention	11 (73.3)	27 (100)	25 (96.2)	25 (83.3)	299 (93.7)	337 (92.8)	
Positive	0 (0)	0 (0)	0 (0)	1 (3.3)	2 (0.6)	8 (2.2)	
Negative	4 (26.7)	0 (0)	1 (3.8)	4 (13.3)	18 (5.6)	18 (5.0)	
Efficacy							<.001
No mention	9 (60.0)	8 (29.6)	15 (57.7)	14 (46.7)	62 (19.4)	131 (36.1)	
Positive	6 (40.0)	19 (70.4)	11 (42.3)	13 (43.3)	243 (76.2)	230 (63.4)	
Negative	0 (0)	0 (0)	0 (0)	3 (10.0)	14 (4.4)	2 (0.6)	
Adherence							<.001
No mention	14 (93.3)	27 (100)	26 (100)	28 (93.3)	288 (90.3)	336 (92.6)	
Positive	1 (6.7)	0 (0)	0 (0)	0 (0)	3 (0.9)	21 (5.8)	
Negative	0 (0)	0 (0)	0 (0)	2 (6.7)	28 (8.8)	6 (1.7)	
Link							<.001
None	6 (40.0)	2 (7.4)	4 (15.4)	18 (60.0)	93 (29.2)	91 (25.1)	
Scientific	1 (6.7)	11 (40.7)	5 (19.2)	0 (0)	196 (61.4)	270 (74.4)	
Nonscientific	8 (53.3)	14 (51.9)	17 (65.4)	12 (40)	30 (9.4)	2 (0.6)	

[a]Chi-square tests were conducted to assess statistical differences.

Scientific Links Were Mainly Found Within Liraglutide- and Semaglutide-Related Tweets

We found significant differences between the distribution of those tweets including a link, whether scientific or nonscientific, among the 6 different drug groups (P<.001 Table 1). More specifically, we found the highest percentages of tweets with scientific links in those posts related to liraglutide (196/319, 61.4%) and semaglutide (270/363, 74.4%), followed by those tweets referencing lorcaserin (11/27, 40.7%) and phentermine-topiramate (5/26, 19.2%). The frequency of tweets with a nonscientific link was very low in those related to semaglutide and liraglutide; on the other hand, more than half of the tweets referencing phentermine-topiramate, lorcaserin, and orlistat included a nonscientific link.

As liraglutide and semaglutide accumulated the majority of tweets (682/780, 87.4%), we decided to investigate the use of links in these tweets depending on their content (Table 2). The use of links was mainly focused on those tweets with a positive consideration towards the efficacy of the treatment, whereas in those tweets referencing side effects and adherence to treatment, the use of links was marginal.

Table 2. Descriptive characteristics of the original tweets included in the analysis categorized by total amount per drug and category.

Category	Liraglutide, n (%)		*P* value[a]	Semaglutide, n (%)		*P* value[a]
	Without link (N=93)	With link (N=226)		Without link (N=91)	With link (N=272)	
Side effects			.006			<.001
No mention	81 (87.1)	218 (96.5)		77 (84.6)	260 (95.6)	
Positive	1 (1.1)	0 (0)		2 (2.2)	6 (2.2)	
Negative	11 (11.8)	8 (3.5)		12 (13.2)	6 (2.2)	
Efficacy			<.001			<.001
No mention	24 (25.8)	38 (16.8)		17 (18.7)	114 (41.9)	
Positive	55 (59.1)	188 (83.2)		72 (79.1)	158 (58.1)	
Negative	14 (15.1)	0 (0)		2 (2.2)	0 (0)	
Adherence			<.001			<.001
No mention	73 (78.5)	215 (95.1)		80 (87.9)	256 (94.1)	
Positive	3 (3.2)	0 (0)		5 (5.5)	16 (5.9)	
Negative	17 (18.3)	11 (4.9)		6 (6.6)	0 (0)	

[a]Chi-square tests were conducted to assess statistical differences.

Semaglutide-Related Tweets Obtained the Highest Probability of Likes and Were the Most Disseminated Within the Twitter Community

We found that the probabilities of a tweet being liked between the groups were significantly different (*P*<.001). Semaglutide-related tweets accumulated the highest number of likes per tweet (median 3; 95% CI 1-12). In addition, we analyzed the number of likes received per tweet as classified by the inclusion or absence of a link. We found no differences in the median of likes per tweet between those posts including or not including a link (*P*=.27).

We found that semaglutide-related tweets had the highest potential reach and impact (2,522,621 and 4,676,763 , respectively), which was double that of liraglutide (719,382 and 1,631,062, respectively). On the other hand, both parameters were markedly lower for bupropion-naltrexone (996,398 and 1,603,556, respectively), orlistat (486,533 and 697,956 , respectively), phentermine-topiramate (183,919 and 187,094, respectively), and lorcaserin (29,420 and 30,341, respectively).

Regarding the sentiment analyses of the content of the tweets, we found that those posts related to semaglutide (mean 79.67), liraglutide (mean 61.46), lorcaserin (mean 75.14), and phentermine-topiramate (mean 60.06) received positive sentiment. However, the sentiment was neutral for orlistat (mean 43.9) and bupropion-naltrexone (mean 53.8).

Health Institutions Were the Most Active Twitter Users

We investigated the type of users that posted the tweets. Of the total number of tweets, 7.9% (62/780) were posted by users identified as patients or relatives, 16% (125/780) by health institutions, and 27.1% (211/780) by health care professionals. Of the remaining 49% (382/780) of tweets, the users were considered indeterminate.

Next, we investigated those tweets related to side effects according to the different types of users and found significant differences in the frequency and content of the postings (Table 3; *P*<.001). Patients were the users that posted most about the presence of side effects, whereas health institutions mentioned the presence of side effects the least. Moreover, we also found significant differences between users in regards to tweets about efficacy (*P*<.001) and adherence (*P*<.001). Interestingly, patients were also those who most frequently expressed a lack of efficacy or adherence (Table 4 and Table 5). On the other hand, users classified as health institutions were those that posted most frequently on the efficacy of treatment and promoted adherence to it. Additionally, we assessed who the users were that most frequently included a link within their tweets (Table 6). We found that health institutions included a link, either designated as scientific or nonscientific, more frequently in their posts than did users classified as health professionals or patients (*P*<.001). Finally, we assessed the frequency of user postings according to the different antiobesity drugs analyzed, finding significant differences among them. In particular, health institutions generated the majority of tweets concerning the latest antiobesity drugs.

Table 3. Descriptive characteristics of those tweets referencing side effects in which the user was categorized as a patient, health professional, or health institution. Tweets were further classified as either not mentioning side effects or mentioning them positively or negatively.

Side effects	Patients, n (%)	Health professionals, n (%)	Health institutions, n (%)	Total, n (%)
No mention	40 (11.59)	114 (33.04)	191 (55.36)	345 (100)
Positive	2 (66.67)	0 (0)	1 (33.33)	3 (100)
Negative	11 (39.29)	8 (28.57)	9 (32.14)	28 (100)
Total	53 (14.10)	122 (32.45)	201 (53.46)	376 (100)

Table 4. Descriptive characteristics of those tweets referencing efficacy in which the user was categorized as a patient, health professional, or health institution. Tweets were further classified as either not mentioning efficacy or mentioning it positively or negatively.

Efficacy	Patients, n (%)	Health professionals, n (%)	Health institutions, n (%)	Total, n (%)
No mention	27 (20.45)	34 (25.76)	71 (53.79)	132 (100)
Positive	19 (8.19)	83 (35.78	130 (56.03)	232 (100)
Negative	7 (58.33)	5 (41.67)	0 (0)	12 (100)
Total	53 (14.10)	122 (32.45)	201 (53.46)	376 (100)

Table 5. Descriptive characteristics of those tweets referencing adherence in which the user was categorized as a patient, health professional, or health institution. Tweets were further classified as either not mentioning adherence or mentioning it positively or negatively.

Adherence	Patients, n (%)	Health professionals, n (%)	Health institutions, n (%)	Total, n (%)
No mention	37 (10.60)	117 (33.52)	195 (55.87)	349 (100)
Positive	1 (12.5)	1 (12.5)	6 (75)	8 (100)
Negative	15 (78.95)	4 (21.05)	0 (0)	19 (100)
Total	53 (14.10)	122 (32.45)	201 (53.46)	376 (100)

Table 6. Descriptive characteristics of those tweets in which the user was categorized as a patient, health professional, or health institutions, and further designated as either not including a link or including a scientific or a nonscientific link.

Link	Patients, n (%)	Health professionals, n (%)	Health institutions, n (%)	Total, n (%)
None	52 (50)	39 (37.50)	13 (12.50)	104 (100)
Scientific	0 (0)	70 (31.25)	154 (68.75)	224 (100)
Nonscientific	1 (2.1)	13 (27.08)	34 (70.83)	48 (100)
Total	53 (14.10)	122 (32.45)	201 (53.46)	376 (100)

Discussion

Principal Findings

In this study, we have found that Twitter users show an interest in antiobesity drugs and mainly focus on semaglutide and liraglutide. These tweets are centered on the efficacy of the treatment and principally refer to liraglutide, semaglutide, and lorcaserin. Tweet content containing a negative consideration of side effects was mainly observed in those tweets related to orlistat and bupropion-naltrexone. The frequency of tweets with content related to adherence to treatment was marginal. The highest percentages of tweets with scientific links were observed in those related to liraglutide and semaglutide. Furthermore, those tweets referencing semaglutide obtained the highest potential reach and impact.

Diet, exercise, and lifestyle are considered relevant elements for maintaining a weight within the recommended range [26]. The prevention and treatment of overweight and obese status are considered public health priorities [27]. Currently, the use of pharmacological treatment is becoming pivotal in obesity management [28].

The outcomes of pharmacological treatments for chronic diseases are conditioned by different elements [29]. The efficacy and side effects of antiobesity drugs are critical for the success of these treatments [30]. However, the results of real-world pharmacological interventions are also dependent on treatment adherence [30]. Different factors, such as access to drug information and social considerations, modulate patients' attitudes toward treatment [31,32]. Therefore, identifying patients' areas of concern and the sources of information used are relevant for improving the clinical outcomes. Additionally, patients with health behaviors that are frequently disapproved of by society are oftentimes reluctant to disclose to health providers their noncompliance with treatment and medical advice [33]. In this context, the anonymity of social media may provide greater insight into the beliefs, interests, and experiences

of patients with regard to antiobesity drugs. Furthermore, family members of the patients, doctors, and health care providers can also participate in the social media community and post their comments related to the pharmacological treatment of obesity. The identification of the needs, concerns, and feelings of the patients related to their treatment may improve their adherence and contribute positively to achieving therapeutic objectives [34].

Interest in Antiobesity Drugs on Twitter

Our data show that antiobesity drugs are areas of interest within the Twitter community. The attention paid to antiobesity drugs is reflected in the number of tweets posted with content related to these drugs, which was higher than that reported on other medications employed to treat chronic diseases [35,36]. Furthermore, it is also significant that the majority of the posted tweets were related to the medical aspects of antiobesity drugs in contrast to the reported results of other medications in which the interest generated was nonmedical in nature [37]. In addition, the interest of Twitter users was mainly centered on liraglutide and semaglutide, which accumulated nearly 90% of the tweets. Likewise, differences in interest shown by social media users towards drugs with similar clinical indications have been previously observed, for example, in the case of statins [38].

In addition, we also studied users' areas of interest regarding antiobesity drugs. Our findings show that the one clearly predominant area was drug efficacy, but with different levels of positive consideration being present, as liraglutide, semaglutide, and lorcaserin achieved the highest valuations. With the exception of the tweets related to orlistat, the frequency of references to the side effects of the antiobesity was very low. Considering our study was conducted in the period from September 2019 to October 2019, it is notable that lorcaserin was soon after withdrawn from the market (February 2020) due to its potentially severe side effects [39].

There may be several reasons for these findings. First, differences in the efficacy could explain the different frequencies found between tweets posted about different drugs [30]. However, this factor is unlikely to prove fully conclusive as the results obtained in the clinical trials and metanalysis referencing these drugs do not wholly support the differences observed. Second, there were different patterns of side effects [30]. However, Twitter users have shown little interest in the side effects of the drugs, and it has only been a focus in relation to orlistat and bupropion-naltrexone. Thus, the references to side effects do not seem to explain the differences in interest. Third, it is possible that Twitter users might show a special interest in treatments suppressing appetite. However, this mechanism of action is not only characteristic of glucagon-like peptide-1 inhibitors but also lorcaserin, which only obtained a small number of tweets. The fourth potential factor is availability of scientific information. Liraglutide and semaglutide, the latest approved drugs, have been subject to most of the recent clinical trials involving antiobesity drugs. Furthermore, both drugs have the majority of tweets containing scientific links with content focused mainly on efficacy. Thus, it is possible that recent scientific publications and reports on clinical trials involving liraglutide and semaglutide might in part explain the significant interest toward these drugs within the Twitter community. The fifth reason may involve accessible information in the press and social media. It is known that newly launched drugs or recently approved indications receive enhanced interest from pharmaceutical companies, health providers, and the media [40]. Consequently, the fact that liraglutide and semaglutide have been the most recently approved drugs might explain the special interest towards them.

The dissemination of tweets referencing the 6 different drugs was also heterogenous, with the potential reach and impact of semaglutide clearly being the highest, demonstrating total numbers similar to those of the other 5 drugs combined. This finding could be explained by the fact that semaglutide has been the most recently approved antiobesity drug. In addition, clinical trials concerning semaglutide have been an area of interest for companies and prestigious scientific journals, which have published results in support of the approval of this treatment [41,42].

Furthermore, our data show that most of the tweets were focused on efficacy and rarely mention side effects or discuss issues related to personal adherence to treatment. Thus, there might have been a bias in the information related to these drugs. Moreover, in contrast to previous reports, most of the tweets included a link to sources supporting their content [43]. Interestingly, concerns about the efficacy or tolerability of antiobesity drugs were identified mostly in those tweets not containing a link. The sharing of personal experiences is unlikely to be associated with a link. The analysis of these tweets reveals relevant information for health care providers because many patients that question the efficacy of treatments or abandon their treatments entirely tend to withhold this information from doctors due to feelings of shame or guilt [33]. Indeed, social media has been found to identify side effects not always uncovered via traditional surveys [44]. In semaglutide- and liraglutide-related tweets, most links were scientific, whereas with the rest of the drugs, the majority of links contained a nonscientific source. This may indicate that tweets discussing issues related to semaglutide or liraglutide may be based on medical articles reporting on efficacy. This may thus imply that pharmaceutical companies, scientific journals, and researchers play a key role in Twitter conversations related to obesity medications. Therefore, it is possible that a potential increase in the investigation of the adherence to obesity treatment might increase the dissemination and relevance given by social media users to this pivotal aspect of obesity management.

The important role of social media in generating popular opinion and emotions via information distribution has been established [45], and social media has become a pivotal instrument for sharing knowledge and news [46]. The interest shown by Twitter users in antiobesity drugs support the relevance of social media in the diffusion of medical information. In addition, social media is used to carry out medical interventions, promote preventive health campaigns, and recruit participants for medical research [47,48].

Finally, we studied the sentiment of tweets and found that most drugs obtained a high score. In contrast, bupropion-naltrexone obtained a very low score. This low sentiment toward

bupropion-naltrexone may be reflective of the other indications this drug has: smoking cessation or depression. In this regard, it is worth noting the mockery of psychiatric conditions in Twitter [49]. However, this is unlikely to be the only cause of the low score obtained in the sentiment analysis because phentermine-topiramate is frequently prescribed for the treatment of psychiatric conditions and obtained an average score. Thus, it is possible to suggest that this low sentiment is due to the poor effect bupropion has on weight loss [30].

Understanding the public view of the pharmacological treatment of obesity is useful to better appraising the perceived demands for clinical care related to this condition. It could also help designing better promotional health initiatives and awareness strategies that include content of interest to social media users. In addition, this information can facilitate open conversations about a patient's most common concerns during the medical consultancy. Although this study focused on antiobesity drugs, these results provide relevant information which more than likely can be applied to other pharmacological treatments. The involvement of health institutions in related conversations over social media appears to be desirable given the interest raised by the medical content posted on Twitter.

Strengths and Limitations

First, the relevance of Twitter as a marker of patient´s voice is a matter of controversy. In addition, tweets do not necessarily reflect the overall experience of patients. Second, regarding the collection of tweets, there is the risk that some were not detected since they might have used different hashtags. However, including brand names and the active pharmaceutical ingredients minimized possible bias related to the choice of hashtags. Third, we did not determine whether the date of FDA approval affected Twitter activity differently in more recent versus less recent medications. Fourth, the codebook design and text analysis involved a degree of subjectivity. Nevertheless, this methodology is consistent with previous medical research studies using Twitter. Although computerized machine learning methods have been tested to automatically identify and classify topics in medical research in social media, we used an analysis strategy based on the raters' clinical expertise in obesity, which constitutes a qualitative advantage compared to automated strategies [50]. Finally, the inability to verify the precise identity of the majority of Twitter users posting about antiobesity drugs, in addition to a lack of geolocational data, may constitute a limitation in our capacity to interpret results.

Conclusions

This study demonstrates that Twitter users show an interest in antiobesity pharmacological treatment. The positive consideration of the efficacy of antiobesity drugs accounted for the majority of tweets. In contrast, the side effects of these treatments was only marginally described. Adherence to treatment received little interest from the Twitter community. The nature of the links included in the tweets was heterogenous between the different antiobesity drugs. Thus, this study highlights the opportunity for sharing scientific information, especially that aimed at promoting adherence to pharmacological treatment, which we have detected as being overlooked.

Acknowledgments

This work was partially supported by grants from the Fondo de Investigación de la Seguridad Social, Instituto de Salud Carlos III (PI18/01726), Spain; and the Programa de Actividades de I+D de la Comunidad de Madrid en Biomedicina (B2017/BMD-3804), Madrid, Spain.

Conflicts of Interest

None declared.

References

1. González-Muniesa P, Mártinez-González MA, Hu F, Després JP, Matsuzawa Y, Loos R, et al. Obesity. Nat Rev Dis Primers 2017 Jun 15;3:17034. [doi: 10.1038/nrdp.2017.34] [Medline: 28617414]
2. Bray GA, Frühbeck G, Ryan DH, Wilding JPH. Management of obesity. Lancet 2016 May 07;387(10031):1947-1956. [doi: 10.1016/S0140-6736(16)00271-3] [Medline: 26868660]
3. Thomas CE, Mauer EA, Shukla AP, Rathi S, Aronne LJ. Low adoption of weight loss medications: A comparison of prescribing patterns of antiobesity pharmacotherapies and SGLT2s. Obesity (Silver Spring) 2016 Sep;24(9):1955-1961 [FREE Full text] [doi: 10.1002/oby.21533] [Medline: 27569120]
4. Cassell B, Gyawali CP, Kushnir VM, Gott BM, Nix BD, Sayuk GS. Beliefs about GI medications and adherence to pharmacotherapy in functional GI disorder outpatients. Am J Gastroenterol 2015 Oct;110(10):1382-1387 [FREE Full text] [doi: 10.1038/ajg.2015.132] [Medline: 25916226]
5. Greaves CJ, Sheppard KE, Abraham C, Hardeman W, Roden M, Evans PH, et al. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health 2011;11:119 [FREE Full text] [doi: 10.1186/1471-2458-11-119] [Medline: 21333011]
6. Verheijden MW, Bakx JC, van Weel C, Koelen MA, van Staveren WA. Role of social support in lifestyle-focused weight management interventions. Eur J Clin Nutr 2005 Aug;59 Suppl 1:S179-S186. [doi: 10.1038/sj.ejcn.1602194] [Medline: 16052189]

7. McGowan BS, Wasko M, Vartabedian BS, Miller RS, Freiherr DD, Abdolrasulnia M. Understanding the factors that influence the adoption and meaningful use of social media by physicians to share medical information. J Med Internet Res 2012;14(5):e117 [FREE Full text] [doi: 10.2196/jmir.2138] [Medline: 23006336]
8. Lee K, Hoti K, Hughes J, Emmerton L. Dr Google and the consumer: a qualitative study exploring the navigational needs and online health information-seeking behaviors of consumers with chronic health conditions. J Med Internet Res 2014;16(12):e262 [FREE Full text] [doi: 10.2196/jmir.3706] [Medline: 25470306]
9. Waring ME, Jake-Schoffman DE, Holovatska MM, Mejia C, Williams JC, Pagoto SL. Social media and obesity in adults: a review of recent research and future directions. Curr Diab Rep 2018 Apr 18;18(6):34. [doi: 10.1007/s11892-018-1001-9] [Medline: 29671135]
10. Pagoto S, Schneider KL, Evans M, Waring ME, Appelhans B, Busch AM, et al. Tweeting it off: characteristics of adults who tweet about a weight loss attempt. J Am Med Inform Assoc 2014;21(6):1032-1037. [doi: 10.1136/amiajnl-2014-002652] [Medline: 24928175]
11. Chou WS, Prestin A, Kunath S. Obesity in social media: a mixed methods analysis. Transl Behav Med 2014 Sep;4(3):314-323 [FREE Full text] [doi: 10.1007/s13142-014-0256-1] [Medline: 25264470]
12. Cavazos-Rehg PA, Krauss MJ, Costello SJ, Kaiser N, Cahn ES, Fitzsimmons-Craft EE, et al. "I just want to be skinny.": A content analysis of tweets expressing eating disorder symptoms. PLoS One 2019;14(1):e0207506 [FREE Full text] [doi: 10.1371/journal.pone.0207506] [Medline: 30650072]
13. Golder S, Norman G, Loke YK. Systematic review on the prevalence, frequency and comparative value of adverse events data in social media. Br J Clin Pharmacol 2015 Oct;80(4):878-888. [doi: 10.1111/bcp.12746] [Medline: 26271492]
14. Joseph AJ, Tandon N, Yang LH, Duckworth K, Torous J, Seidman LJ, et al. #Schizophrenia: Use and misuse on Twitter. Schizophr Res 2015 Jul;165(2-3):111-115. [doi: 10.1016/j.schres.2015.04.009] [Medline: 25937459]
15. Attai D, Cowher M, Al-Hamadani M, Schoger J, Staley A, Landercasper J. Twitter social media is an effective tool for breast cancer patient education and support: patient-reported outcomes by survey. J Med Internet Res 2015 Jul 30;17(7):e188 [FREE Full text] [doi: 10.2196/jmir.4721] [Medline: 26228234]
16. Thangasamy IA, Leveridge M, Davies BJ, Finelli A, Stork B, Woo HH. International Urology Journal Club via Twitter: 12-month experience. Eur Urol 2014 Jul;66(1):112-117. [doi: 10.1016/j.eururo.2014.01.034] [Medline: 24548686]
17. Nikiphorou E, Studenic P, Ammitzbøll CG, Canavan M, Jani M, Ospelt C, EMEUNET. Social media use among young rheumatologists and basic scientists: results of an international survey by the Emerging EULAR Network (EMEUNET). Ann Rheum Dis 2017 Apr;76(4):712-715. [doi: 10.1136/annrheumdis-2016-209718] [Medline: 27797750]
18. Turner-McGrievy GM, Beets MW. Tweet for health: using an online social network to examine temporal trends in weight loss-related posts. Transl Behav Med 2015 Jun;5(2):160-166 [FREE Full text] [doi: 10.1007/s13142-015-0308-1] [Medline: 26029278]
19. Viguria I, Alvarez-Mon MA, Llavero-Valero M, Asunsolo Del Barco A, Ortuño F, Alvarez-Mon M. Eating disorder awareness campaigns: thematic and quantitative analysis using Twitter. J Med Internet Res 2020 Jul 14;22(7):e17626 [FREE Full text] [doi: 10.2196/17626] [Medline: 32673225]
20. Turner-McGrievy GM, Tate DF. Weight loss social support in 140 characters or less: use of an online social network in a remotely delivered weight loss intervention. Transl Behav Med 2013 Sep;3(3):287-294 [FREE Full text] [doi: 10.1007/s13142-012-0183-y] [Medline: 24073180]
21. Sinnenberg L, Buttenheim A, Padrez K, Mancheno C, Ungar L, Merchant R. Twitter as a tool for health research: a systematic review. Am J Public Health 2017 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
22. Lydecker J, Cotter E, Palmberg A, Simpson C, Kwitowski M, White K. Does this Tweet make me look fat? A content analysis of weight stigma on Twitter. Eat Weight Disord Internet 2016:229 [FREE Full text] [doi: 10.1007/s40519-016-0272-x] [Medline: 27068174]
23. Mohammed MA, Moles RJ, Chen TF. Medication-related burden and patients' lived experience with medicine: a systematic review and metasynthesis of qualitative studies. BMJ Open 2016 Feb 02;6(2):e010035 [FREE Full text] [doi: 10.1136/bmjopen-2015-010035] [Medline: 26839015]
24. Gagnon MD, Waltermaurer E, Martin A, Friedenson C, Gayle E, Hauser DL. Patient beliefs have a greater impact than barriers on medication adherence in a community health center. J Am Board Fam Med 2017;30(3):331-336 [FREE Full text] [doi: 10.3122/jabfm.2017.03.160129] [Medline: 28484065]
25. Joseph K, Landwehr P, Carley K. Two 1%s don't make a whole: comparing simultaneous samples from Twitter's streaming API. ISPRS International Journal of Geo-Information Internet 2014:75 [FREE Full text] [doi: 10.1007/978-3-319-05579-4_10]
26. Patnode CD, Evans CV, Senger CA, Redmond N, Lin JS. behavioral counseling to promote a healthful diet and physical activity for cardiovascular disease prevention in adults without known cardiovascular disease risk factors: updated evidence report and systematic review for the US preventive services task force. JAMA 2017 Jul 11;318(2):175-193. [doi: 10.1001/jama.2017.3303] [Medline: 28697259]
27. Lloyd-Jones D, Hong Y, Labarthe D, Mozaffarian D, Appel L, Van Horn L, et al. Defining and setting national goals for cardiovascular health promotion and disease reduction. Circulation 2010 Feb 02;121(4):586-613 [FREE Full text] [doi: 10.1161/circulationaha.109.192703]

28. Jensen M, Ryan D, Apovian C, Ard J, Comuzzie A, Donato K, American College of Cardiology/American Heart Association Task Force on Practice Guidelines, Obesity Society. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. Circulation 2014 Jun 24;129(25 Suppl 2):S102-S138 [FREE Full text] [doi: 10.1161/01.cir.0000437739.71477.ee] [Medline: 24222017]
29. Middleton KMR, Patidar SM, Perri MG. The impact of extended care on the long-term maintenance of weight loss: a systematic review and meta-analysis. Obes Rev 2012 Jun;13(6):509-517. [doi: 10.1111/j.1467-789X.2011.00972.x] [Medline: 22212682]
30. Khera R, Murad MH, Chandar AK, Dulai PS, Wang Z, Prokop LJ, et al. Association of pharmacological treatments for obesity with weight loss and adverse events: a systematic review and meta-analysis. JAMA 2016;315(22):2424-2434. [doi: 10.1001/jama.2016.7602] [Medline: 27299618]
31. Nanna MG, Navar AM, Zakroysky P, Xiang Q, Goldberg AC, Robinson J, et al. Association of patient perceptions of cardiovascular risk and beliefs on statin drugs with racial differences in statin use: insights from the patient and provider assessment of lipid management registry. JAMA Cardiol 2018 Aug 01;3(8):739-748 [FREE Full text] [doi: 10.1001/jamacardio.2018.1511] [Medline: 29898219]
32. Bradley CK, Wang TY, Li S, Robinson JG, Roger VL, Goldberg AC, et al. Patient-reported reasons for declining or discontinuing statin therapy: insights from the PALM registry. J Am Heart Assoc 2019 Apr 02;8(7):e011765 [FREE Full text] [doi: 10.1161/JAHA.118.011765] [Medline: 30913959]
33. Levy AG, Scherer AM, Zikmund-Fisher BJ, Larkin K, Barnes GD, Fagerlin A. Prevalence of and factors associated with patient nondisclosure of medically relevant information to clinicians. JAMA Netw Open 2018 Nov 02;1(7):e185293 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.5293] [Medline: 30646397]
34. Barbosa CD, Balp M, Kulich K, Germain N, Rofail D. A literature review to explore the link between treatment satisfaction and adherence, compliance, and persistence. Patient Prefer Adherence 2012;6:39-48 [FREE Full text] [doi: 10.2147/PPA.S24752] [Medline: 22272068]
35. Giunti G, Claes M, Dorronzoro Zubiete E, Rivera-Romero O, Gabarron E. Analysing sentiment and topics related to multiple sclerosis on Twitter. Stud Health Technol Inform 2020 Jun 16;270:911-915. [doi: 10.3233/SHTI200294] [Medline: 32570514]
36. Cook N, Mullins A, Gautam R, Medi S, Prince C, Tyagi N, et al. Evaluating patient experiences in dry eye disease through social media listening research. Ophthalmol Ther 2019 Sep;8(3):407-420 [FREE Full text] [doi: 10.1007/s40123-019-0188-4] [Medline: 31161531]
37. Alghamdi A, Abumelha K, Allarakia J, Al-Shehri A. Conversations and misconceptions about chemotherapy in Arabic tweets: content analysis. J Med Internet Res 2020 Jul 29;22(7):e13979 [FREE Full text] [doi: 10.2196/13979] [Medline: 32723724]
38. Golder S, O'Connor K, Hennessy S, Gross R, Gonzalez-Hernandez G. Assessment of beliefs and attitudes about statins posted on twitter: a qualitative study. JAMA Netw Open 2020 Jun 01;3(6):e208953 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.8953] [Medline: 32584408]
39. Halpern B, Mancini MC. Should the same safety scrutiny of antiobesity medications be applied to other chronic usage drugs? Obesity (Silver Spring) 2020 Jul;28(7):1171-1172. [doi: 10.1002/oby.22810] [Medline: 32374528]
40. Carlos S, de Irala J, Hanley M, Martínez-González. The use of expensive technologies instead of simple, sound and effective lifestyle interventions: a perpetual delusion. J Epidemiol Community Health 2014 Sep;68(9):897-904 [FREE Full text] [doi: 10.1136/jech-2014-203884] [Medline: 24962820]
41. Husain M, Birkenfeld AL, Donsmark M, Dungan K, Eliaschewitz FG, Franco DR, PIONEER 6 Investigators. Oral semaglutide and cardiovascular outcomes in patients with type 2 diabetes. N Engl J Med 2019 Aug 29;381(9):841-851. [doi: 10.1056/NEJMoa1901118] [Medline: 31185157]
42. Marso SP, Bain SC, Consoli A, Eliaschewitz FG, Jódar E, Leiter LA, SUSTAIN-6 Investigators. Semaglutide and cardiovascular outcomes in patients with type 2 diabetes. N Engl J Med 2016 Dec 10;375(19):1834-1844. [doi: 10.1056/NEJMoa1607141] [Medline: 27633186]
43. Pereira-Sanchez V, Alvarez-Mon MA, Asunsolo Del Barco A, Alvarez-Mon M, Teo A. Exploring the extent of the Hikikomori phenomenon on Twitter: mixed methods study of Western language tweets. J Med Internet Res 2019 May 29;21(5):e14167 [FREE Full text] [doi: 10.2196/14167] [Medline: 31144665]
44. Kim MG, Kim J, Kim SC, Jeong J. Twitter analysis of the nonmedical use and side effects of methylphenidate: machine learning study. J Med Internet Res 2020 Feb 24;22(2):e16466 [FREE Full text] [doi: 10.2196/16466] [Medline: 32130160]
45. King G, Schneer B, White A. How the news media activate public expression and influence national agendas. Science 2017 Nov 10;358(6364):776-780 [FREE Full text] [doi: 10.1126/science.aao1100] [Medline: 29123065]
46. Alvarez-Mon M, Asunsolo Del Barco A, Lahera G, Quintero J, Ferre F, Pereira-Sanchez V, et al. Increasing interest of mass communication media and the general public in the distribution of tweets about mental disorders: observational study. J Med Internet Res 2018 May 28;20(5):e205 [FREE Full text] [doi: 10.2196/jmir.9582] [Medline: 29807880]
47. Teo AR, Liebow SB, Chan B, Dobscha SK, Graham AL. Reaching those at risk for psychiatric disorders and suicidal ideation: Facebook advertisements to recruit military veterans. JMIR Ment Health 2018 Jul 05;5(3):e10078 [FREE Full text] [doi: 10.2196/10078] [Medline: 29980498]

48. Hales SB, Davidson C, Turner-McGrievy GM. Varying social media post types differentially impacts engagement in a behavioral weight loss intervention. Transl Behav Med 2014 Dec;4(4):355-362 [FREE Full text] [doi: 10.1007/s13142-014-0274-z] [Medline: 25584084]
49. Alvarez-Mon M, Llavero-Valero M, Sánchez-Bayona R, Pereira-Sanchez V, Vallejo-Valdivielso M, Monserrat J, et al. Areas of interest and stigmatic attitudes of the general public in five relevant medical conditions: thematic and quantitative analysis using Twitter. J Med Internet Res 2019 May 28;21(5):e14110 [FREE Full text] [doi: 10.2196/14110] [Medline: 31140438]
50. Saha K, Torous J, Ernala SK, Rizuto C, Stafford A, De Choudhury M. A computational study of mental health awareness campaigns on social media. Transl Behav Med 2019 Nov 25;9(6):1197-1207 [FREE Full text] [doi: 10.1093/tbm/ibz028] [Medline: 30834942]

Abbreviations

FDA: US Food and Drug Administration
JSON: JavaScript Object Notation

Edited by R Kukafka; submitted 15.09.20; peer-reviewed by JA Giron-González, SP Nischwitz, D Valdez, M Claes; comments to author 05.10.20; revised version received 02.12.20; accepted 12.08.21; published 26.10.21.

Please cite as:
Alvarez-Mon MA, Llavero-Valero M, Asunsolo del Barco A, Zaragozá C, Ortega MA, Lahera G, Quintero J, Alvarez-Mon M
Areas of Interest and Attitudes Toward Antiobesity Drugs: Thematic and Quantitative Analysis Using Twitter
J Med Internet Res 2021;23(10):e24336
URL: https://www.jmir.org/2021/10/e24336
doi:10.2196/24336
PMID:34698653

Original Paper

Characterizing Vaping Industry Political Influence and Mobilization on Facebook: Social Network Analysis

Michael Robert Haupt[1,2], MA; Qing Xu[2,3,4], MAS; Joshua Yang[5], PhD; Mingxiang Cai[2,3,4], MS; Tim K Mackey[2,4,6], MAS, PhD

[1]Department of Cognitive Science, University of California, San Diego, La Jolla, CA, United States

[2]Global Health Policy and Data Institute, San Diego, CA, United States

[3]Department of Healthcare Research and Policy, University of California, San Diego Extension, La Jolla, CA, United States

[4]S-3 Research, San Diego, CA, United States

[5]Department of Public Health, California State University, Fullerton, CA, United States

[6]Global Health Program, Department of Anthropology, University of California, San Diego, La Jolla, CA, United States

Corresponding Author:
Tim K Mackey, MAS, PhD
Global Health Program
Department of Anthropology
University of California, San Diego
9500 Gilman Drive
La Jolla, CA, 92093
United States
Phone: 1 951 491 4161
Email: tmackey@ucsd.edu

Abstract

Background: In response to recent policy efforts to regulate tobacco and vaping products, the vaping industry has been aggressive in mobilizing opposition by using a network of manufacturers, trade associations, and tobacco user communities, and by appealing to the general public. One strategy the alternative tobacco industry uses to mobilize political action is coordinating on social media platforms, such as the social networking site Facebook. However, few studies have specifically assessed how platforms such as Facebook are used to influence public sentiment and attitudes towards tobacco control policy.

Objective: This study used social network analysis to examine how the alternative tobacco industry uses Facebook to mobilize online users to influence tobacco control policy outcomes with a focus on the state of California.

Methods: Data were collected from local and national alternative tobacco Facebook groups that had affiliations with activities in the state of California. Network ties were constructed based on users' reactions to posts (eg, "like" and "love") and comments to characterize political mobilization networks.

Results: Findings show that alternative tobacco industry employees were more likely to engage within these networks and that these employees were also more likely to be influential members (ie, be more active) in the network. Comparisons between subnetworks show that communication within the local alternative tobacco advocacy group network was less dense and more centralized in contrast to a national advocacy group that had overall higher levels of engagement among members. A timeline analysis found that a higher number of influential posts that disseminated widely across networks occurred during e-cigarette–related legislative events, suggesting strategic online engagement and increased mobilization of online activity for the purposes of influencing policy outcomes.

Conclusions: Results from this study provide important insights into how tobacco industry–related advocacy groups leverage the Facebook platform to mobilize their online constituents in an effort to influence public perceptions and coordinate to defeat tobacco control efforts at the local, state, and federal level. Study results reveal one part of a vast network of socially enabled alternative tobacco industry actors and constituents that use Facebook as a mobilization point to support goals of the alternative tobacco industry.

(J Med Internet Res 2021;23(10):e28069) doi:10.2196/28069

KEYWORDS

vaping; alternative tobacco industry; e-cigarettes; Facebook; social network analysis; social networks; ehealth; health policy

Introduction

Since the introduction of e-cigarettes—also known as electronic nicotine delivery systems (ENDS)—as a commercial product in the mid-2000s, their popularity has grown significantly. In 2018, 3.2% of US adults aged 18 years or older reported using e-cigarettes every day or some days [1]. Among middle and high school students in the US, e-cigarettes were the most commonly used tobacco product in 2019, with 5.4 million (20%) of youths reporting current use [2]. The global revenue from e-cigarettes in 2019 was US $15.7 billion, with a projected growth rate of 9.2% between 2020 and 2030 equating to an estimated US $39 billion in revenue by 2030 [3].This growth in sales is troubling from a public health standpoint, with evidence showing that e-cigarette usage is associated with breathing difficulty and cardiopulmonary health risks [4] as well as a higher likelihood to engage in risky behaviors among high school students [5]. In response to this growth in use and public health concerns, government agencies such as the US Food and Drug Administration have begun to assert their regulatory authority over e-cigarettes [6,7] along with a number of US states and municipalities that have instituted a variety of tobacco control policies (eg, sales bans of flavored tobacco products, tax regimes, license requirements, restrictions on youth access) to reduce the uptake, marketing, and sale of ENDS [8,9].

An abundance of evidence has shown that as tobacco control efforts expand, the tobacco industry has responded with a variety of messages and tactics to protect its business interests, including but not limited to lobbying, shaping the evidence base to support product use or harm reduction messages, policy substitution, and litigation [10-14]. As the overall reputation of the tobacco industry has declined over time, front groups and alliances with more reputable organizations have been especially important to its political activity because they shield such efforts from negative public perception about the industry [15-20]. Smokers' rights groups were a particular group the tobacco industry used that attempted to mobilize smokers as a force to lobby policy makers to resist efforts that infringed on their "right to smoke" [17,21-23]. Although smokers' rights groups were projected as being grassroots, they were a form of astroturfing, or "artificial grassroots campaigns created by public relations firms," to increase the number of contacts that are made with policy makers in addition to traditional lobbying approaches [24].

The e-cigarette or "vaping" industry is also distinct in its structure compared to the overall tobacco industry. Specifically, the tobacco industry is highly concentrated among a few multinational corporations (China National Tobacco Corporation, Altria/Philip Morris International, British American Tobacco, Japan Tobacco International, and Imperial Brands make up approximately 82% of the global market share for cigarettes) [25]. These multinational tobacco actors also have a stake in the vaping industry as evidenced by the introduction by these companies of noncombustible products, mergers and acquisition activity with ENDS manufacturers, and other ownership of ENDS companies [26,27]. However, despite involvement from the tobacco industry, the ENDS industry has its own unique construction, including its own independent brands with national footprints (eg, Juul and the 35% share acquisition by Altria), smaller manufacturers and retailers of ENDS, and retailers independent of larger tobacco manufacturers [28]. Therefore, understanding the political resistance to expanding ENDS legislation can be informed by prior lessons learned from tobacco industry interference but will also need to evolve due to the differing landscape of actors associated with the growing ENDS industry.

A proactive leader in state-based policy efforts to regulate tobacco and ENDS is the state of California. Specifically, the state has been one of the leading jurisdictions in the United States in implementing local- and state-level tobacco control policies, including those related to smoke-free air and housing and tobacco prevention programs, which also includes measures directed at regulating ENDS [29,30]. For example, recent tobacco control legislation has included subjecting ENDS to existing antitobacco laws (Senate Bill [SB] X2-5), raising the purchasing age to 21 (SB X2-7), and imposing additional taxes on ENDS products (Proposition 56). In response, the vaping industry has been aggressive in mobilizing opposition to these policies by using a network of manufacturers, trade associations, and tobacco user communities, and by appealing to the general public in order to advocate for "vaper rights" [31-35]. One of the strategies used by the alternative tobacco industry has been to mobilize political action through social media platforms, such as the popular social networking site Facebook, which can extend the scope and reach of these antipolicy and advocacy efforts.

Previous studies examining the impact of social media platforms on the ENDS industry have identified user attitudes towards alternative tobacco products and behaviors of users [36-38]; characterized marketing tactics, sales strategies, and pricing of ENDS [39-46]; and identified geographic locations where people use ENDS [47]. However, few studies have specifically assessed how social media can influence antitobacco public policy and methods of social media mobilization among digital constituents [48,49]. This study builds on prior research by describing the membership and network structure of interest groups for alternative tobacco products on Facebook by using social network analysis (SNA). In order to observe mobilization across different online contexts, we conducted an exploratory investigation characterizing network structures of an alternative tobacco industry trade association Facebook page and a consumer-focused ENDS Facebook group.

Methods

Overview

This study used SNA to identify and characterize influential members and vape industry employees engaged with California chapters of 2 alternative tobacco interest groups. The proportion of influential members were compared between industry

employees and nonemployees. An exponential random graph model (ERGM) was then used to detect statistically significant differences between vape employment status and likelihood to engage within the network, and a timeline analysis was used to investigate alignment with national- and state-level tobacco control policy events. All data collection and statistical and SNA analyses were completed in the computer programming languages Python (Python Software Foundation) and R (The R Foundation for Statistical Computing).

Data Collection and Processing

This study used membership information available from 2 Facebook groups—the California Consumer Advocates for Smoke-Free Alternatives Association (CCASAA) and the Northern California Chapter of Smoke-Free Alternatives Trade Association (NC-SFATA)—to conduct SNA that characterizes the communication networks and identifies the position of influential members within and between these online communities.

The CASAA was founded in 2009 and describes itself as an advocacy group with an all-volunteer board and grassroots membership. It distinguishes itself as claiming to operate as a consumer organization and not a trade organization [50].The CASAA has a national Facebook page (address in New York) as well as pages for state chapters. The CCASAA Facebook page is for CASAA members residing in California and characterizes CASAA as "formed by people concerned about the continued availability of safer alternatives to tobacco" [51]. Members can "support CASAA's goals by advocating in California for reasonable laws for products such as electronic cigarettes and smokeless tobacco products." The CCASAA Facebook page has 195 members and is currently active.

The SFATA is a trade organization with a mission to "advocate for a reasonably regulated U.S. marketplace which allows our member companies to provide smoke-free products to adult consumers that are attractive in choice, while promoting a positive public image for vapor products…" [52]. Its membership includes "manufacturers, distributors, wholesalers, retailers, and the various service providers involved in their business" with membership opportunities open to the academic community and nonprofit organizations. The SFATA extends its national work into state and local level policy through its different state chapters. The NC-SFATA Facebook page characterizes SFATA as "representing the interests of the small- to mid-sized businesses by engaging with political decision makers, with advocacy at the national and state levels…SFATA is run and founded by the companies that built the vapor industry; with no ties or alliance to 'Big Tobacco'" [53]. It discloses lobbying activity at the federal, state, and local levels as part of its range of services offered to members. The parent SFATA organization (address in Washington DC) operates an active Facebook page, and there are also numerous Facebook SFATA chapter pages in other states with varying levels of activity. The NC-SFATA Facebook page has 387 followers but has had limited activity since February 2017 (eg, only 2 recent posts in January 2021.) Despite its low recent activity, the page remains available and also provides an important digital record of how the group has mobilized to influence past California tobacco control legislation.

Data mining approaches using the Python programming language were used to collect publicly available data for all posts and comments from the Facebook pages of the 2 target groups over a 15-day period, from July 1, 2020, to July 14, 2020. This allowed us to collect all Facebook posts and comments retrospectively from users prior to July 1, 2020, and prospectively until July 14, 2020. Reposts were not removed as they were used for the purpose of analyzing interactions between posts and user interactions. Data were then restructured manually to a format suitable for conducting SNA. We note that public Facebook pages do not require request for joining or membership, and all posts, comments, and other information are publicly available to any online user.

Self-reported occupation data were also collected for public accounts and group members among these 2 target pages if published on a publicly available Facebook account profile. For users whose occupation data were unavailable on Facebook, additional searches on other online platforms with professional public profiles (eg, Twitter, LinkedIn) matched to public metadata profile information were conducted to ascertain possible work position status. These additional metadata were matched based on matching names with at least 1 other identifier before cross-referencing with the Facebook data. All data generated from occupational classification were in the public domain, and no individual identifiers of these accounts are reported in this study. Occupation data from users that did not have a link to their profile page in the metadata were marked as missing, as additional data could not be cross-referenced and matched.

Social Network Analysis

SNA was conducted to detect influential members among alternative tobacco Facebook group communication networks reviewed. In the model presented in this paper, each node is a Facebook user associated with a California alternative tobacco trade association page (eg, CCASAA and NC-SFATA), and each edge (ie, link) between nodes represents reactions (eg, "like," "love," and "angry") or comments on a post. As previous research has demonstrated the importance of distinguishing between active versus passive engagement with Facebook content [54-57], comments were weighted as twice the value of a reaction (eg, a like, emoji) in order to indicate higher engagement on posts. Generally, the source node is the user who produced the original post while the target nodes are users who reacted to or commented on a post. Within the context of this study, a node with a high in-degree centrality indicated that the user received a higher number of reactions on their posts while a high-outdegree centrality indicated users who were more active within their Facebook page.

Network visualizations for the individual CCASAA and NC-SFATA groups used a spherical layout, while the visualization that included both networks used the Fruchterman-Reingold layout algorithm [58] to emphasize differences in the communication structures between the 2 groups. Eigenvector centrality, which accounts for both the number of edges of each node and the level of connectivity of

each node's connections (ie, the extent to which a node's connections is connected to others within the network), was used to measure influence within the network. Previous research has demonstrated that network structure measures such as Eigenvector centrality are reliable for detecting influential members within a network [59-61], and this measure has been validated in studies using data from both surveys and social media platforms [62-65]. Eigenvector scores were assigned to each node, with higher scores indicating higher influence within the network.

Proportions of users who were vaping industry employees were compared among the top 25% of users with the highest Eigenvector scores and remaining users with chi-square and a 2-proportion *z* test to test for statistical significance. Due to its increased usage within social network research [66] and previous work modeling social influence processes [67], a valued EGRM was used to detect a statistically significant influence between vape employee status and engagement within the network (ie, the likelihood of a tie formation) [68]. An ERGM is a statistical model that simulates alternative configurations of the observed network in order to determine the likelihood of a given structural feature, such as connections between nodes, which is referred to as "degrees" in SNA. Within the context of Facebook data used for this study, degrees correspond to the number of reactions and comments exchanged between users.

In order to model the value of the edges within the networks (as ERGMs traditionally only model binary ties), we employed an ERGM that accounts for valued edges by specifying a Poisson distribution as a reference for the distribution of edge values [69]. This study also examined homophily effects among ENDS industry–affiliated employees within the network. Homophily in social network research describes how members that share similar attributes, such as age, race, and gender, are more likely to interact with one another within a network [70] and has also been shown in health behaviors such as smoking [71]. Higher homophily among ENDS employees would indicate that users associated with the alternative tobacco industry are more likely to engage with one another on the Facebook group platform, suggesting higher coordination among members. The sum of reactions and comments for each post were calculated in order to identify influential posts based on whether the post was in the 75th percentile of engagement. Nodes with missing data for the ENDS employee status were kept in the analysis and imputed as not employed (46 nodes in total) in order to maintain network structure. This makes it possible that the models could be underestimating the strength of the effect that ENDS employee status and homophily have on the likelihood of communication ties.

Policy Timeline Analysis

A timeline analysis was also conducted to detect associations between the number of influential posts and relevant federal and California state legislative events concerning e-cigarettes during the study period. This was done in order to observe how social media activity relates to policy outcomes, with histograms created to show the number of posts for each Facebook group by the month and year they were posted. Separate analyses for both groups were conducted with consideration to the different timeframes of activity between the organizations (the CCASSA activity was from February 2019 to June 2020, while NC-SFATA activity occurred between May 2015 and June 2017). Additionally, posts were categorized as either "Top 25% engagement" or "Remaining 75%" based on whether they were in the 75th percentile of posts that received the highest number of reactions and comments.

Ethics

This study only involved the use of publicly available information in the public domain and did not include any interaction with social media users or other human participants. No personally identifiable information was included in the results of this study, and all results have been aggregated to prevent inadvertent disclosure of identifiable information. Hence, this study was not subject to ethics review.

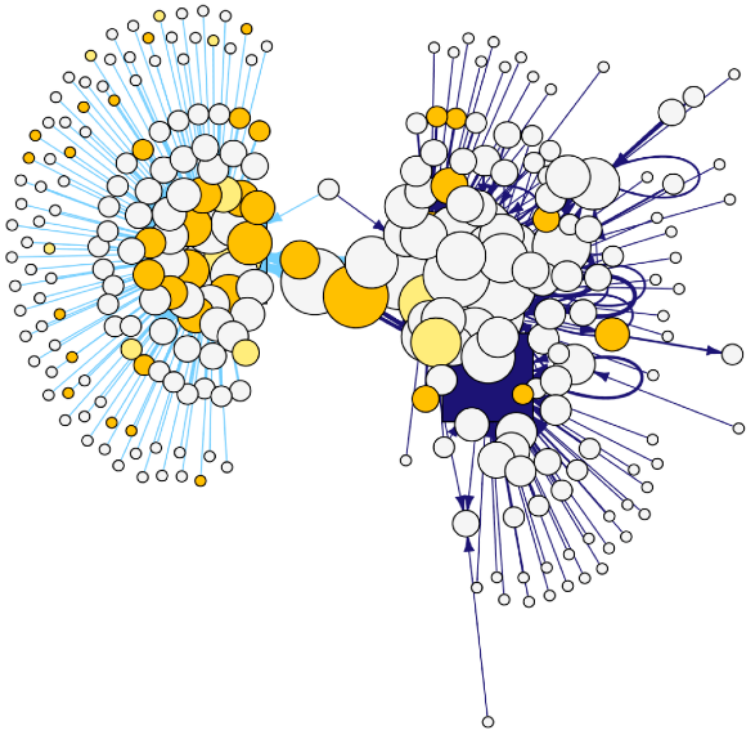
Results

Comparing CCASAA and NF-SFATA Networks

A total of 292 active users (ie, nodes) among both CCASAA and NC-SFATA Facebook groups were included in the SNA from interactions consisting of posts, comments, and reactions resulting in a total of 509 edges (ie, connections between users). Of the total users, 246 had data linking to their public Facebook account and were also reviewed for identifying occupation data. Among these 246 users, 53 (21.5%) self-reported being employed by the alternative tobacco industry, which consisted of 10 retail workers (4.1%), 24 vape shop owners (9.8%), and 19 alternative tobacco industry public relations or marketing employees (7.7%). The predominant themes of Facebook posts and comments of these users included event invitations for activities directly mobilizing against tobacco control–related policies, public mobilization messages to encourage users to take individual action against tobacco control legislation, information about a tobacco control policy or introduced bill, and negative comments and opinions on tobacco control policies from the perspective of the alternative tobacco industry and ENDS users. A more in-depth qualitative description of these themes and specific content is being developed for separate analysis.

SNA visualization of Facebook users in Figure 1 shows the overall network of both the CCASAA and NC-SFATA members while Figures 2 and 3 showcase how users are spread out among the Facebook pages. Light yellow nodes represent users who are alternative tobacco retail workers while dark yellow represents vape shop owners and alternative tobacco public relations or marketing employees. All remaining white nodes are nonalternative tobacco employee users. The source node is the user who produced the original post while the target nodes are users who reacted to or commented on a post. Edges representing online discourse on the CCASAA Facebook page are colored dark blue while activity on the NC-SFATA page is depicted by light blue edges. The size of nodes and edges depends on the weighted sum of the edge connections (ie, larger node and edge sizes equate to a higher number of reactions and comments).

Figure 1. Consumer Advocates for Smoke-Free Alternatives Association and Northern California Chapter of Smoke-Free Alternatives Trade Association networks.



Unsurprisingly, in an examination of the size of user nodes expressed in Figures 2 and 3, both the CCASAA and NC-SFATA are highly engaged within the networks as they are administrators of their respective Facebook pages. Most notably, the figures reveal that employees of the alternative tobacco industry are among some of the more engaged nodes. In Figure 1, both networks are shown to have mostly separate followings, with the CCASAA network having more engagement among nodes and the NC-SFATA network showing a more centralized, yet less engaged discourse. This is further supported when the degree distribution between networks is compared as shown in Table 1, where higher degree indicates more engagement. The mean degree for CCASSA users is 4.84 compared to 1.99 among NC-SFATA users, which shows that CCASAA users on average react, comment, or post more often. Additionally, CCASAA users have a higher degree across percentiles while NC-SFATA users consistently have a degree of 1 despite having a larger max value, which further indicates the centralized nature of the NC-SFATA communications. Network densities (ie, the ratio of the number of edges to the number of possible edges) were also calculated and show that the CCASAA network is denser (density 0.016) compared to that of the NC-SFATA (density 0.007), which is another indicator of greater activity levels among nodes within the CCASAA network.

Between the 2 groups, 8 Facebook users were identified as being active in both networks. Of these overlapping users, 6 were employed in the alternative tobacco industry, with 1 being a retail worker, 4 being vape shop owners, and 1 having served in a leadership role with the NC-SFATA. The involvement of a user with a leadership role in the SFATA chapter suggests the possibility of coordinated policy mobilization between these 2 California-based pro–alternative tobacco organizations and potential alignment of messaging and advocacy approaches. Additionally, Table 1 shows that alternative tobacco employees had a higher mean degree (mean 8.02) compared to the total network (mean 3.49), which indicates that these industry employees on average are more engaged within the groups compared to other users.

Figure 2. Consumer Advocates for Smoke-Free Alternatives Association network. The Consumer Advocates for Smoke-Free Alternatives Association is represented by the dark blue square.

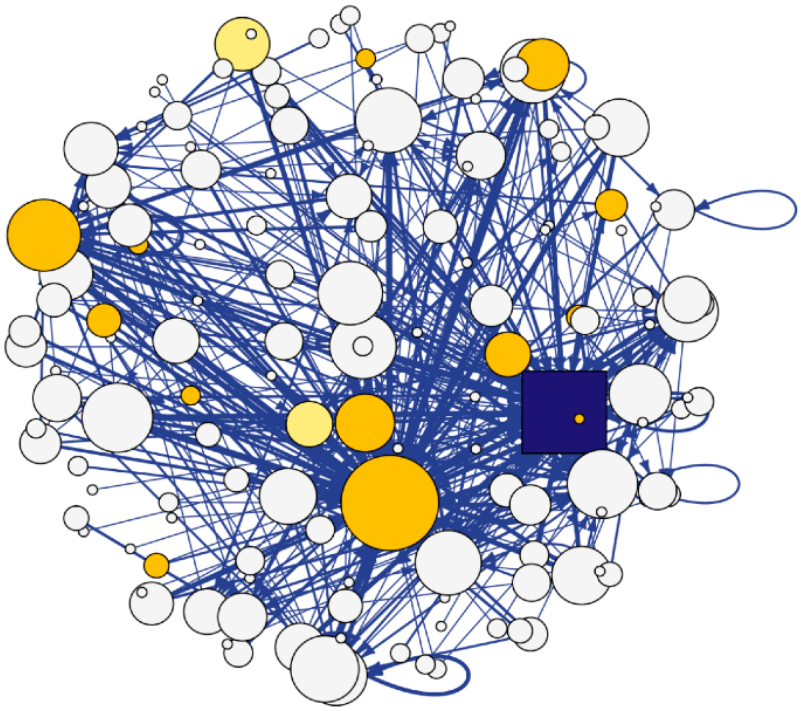


Figure 3. Northern California Chapter of Smoke-Free Alternatives Trade Association network. The Smoke-Free Alternatives Trade Association is represented by the light blue square.

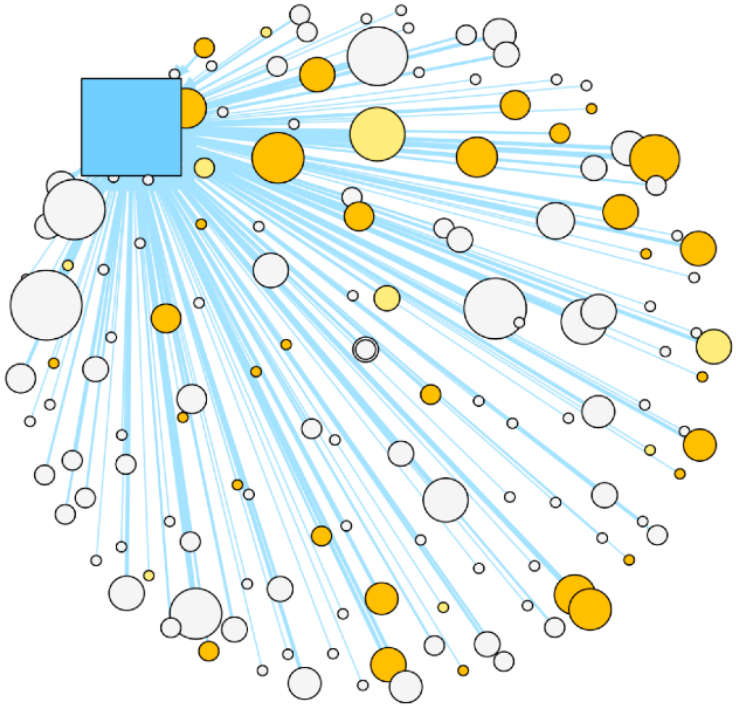


Table 1. Network statistics and degree distribution.

Statistics	Total network	CCASAA[a] Only	NC-SFATA[b] Only	Vape employee (CCASSA + SFATA)	Top 25% influential (CCASSA + SFATA)
Total nodes, n	292	148	152	53	76
Degree, mean (SD)	3.49 (11.52)	4.84 (10.28)	1.99 (12.17)	8.02 (25.18)	9.45 (21.43)
60th percentile (degree)	2	4	1	2	8
90th percentile (degree)	6	10	1	8	18
95th percentile (degree)	10	17	1	47	31
Max, n	151	83	151	151	151

[a]CCASAA: Consumer Advocates for Smoke-Free Alternatives Association.

[b]NC-SFATA: Northern California Chapter of Smoke-Free Alternatives Trade Association.

Table 2 shows that 32% (23/71) of the most influential members within the network were self-identified alternative tobacco employees compared to 17.1% (30/175) of other users, with this difference being statistically significant (*P*=.007). This indicates that alternative tobacco industry employees were more likely to be influential nodes within the communication network.

Results from the chi-square test provide further evidence of an association between alternative tobacco industry employment and network influence, which shows that the distribution of alternative tobacco employees across influential and noninfluential members were statistically different from expected probabilities based on random chance (*P*=.002).

Table 2. Chi-square and 2-proportion z test comparing vape employee status between influential and noninfluential users (N=264[a]).

Tests for detecting statistical significance	Top 25% most influential	Remaining 75%
Chi-square test[b]		
Expected probability	.25	.75
Observed count (vape employees), n	23	30
Two-proportion *z* test[c]		
Users, n	71	175
Vape employees, n (%)	23 (32)	30 (17.1)

[a]Only users with links to Facebook profiles are included.

[b]X^2_1=9.566; *P* value=.002.

[c]X^2_1=6.078; *P* value=.007.

As shown in Table 3, the ERGM was run on the networks examined to detect statistically significant effects between being an alternative tobacco industry employee and the likelihood that a tie would form within the network, which in this study would signify engagement via a reaction or comment. The variable labeled "Nodefactor: *Vape*" measures how alternative tobacco employee status (coded as a categorical variable) influences the likelihood of a tie within the network. Model 2 additionally tests for homophily among alternative tobacco employees, labeled as "Nodematch: *Vape*." The term "Nonzero" was included in the model to control for zero inflation of the network (ie, when a network is sparse but still has high interaction between nodes).

Results from Model 1 show that if a user was an alternative tobacco employee, then the log odds of sending a reaction or comment was 0.35 times greater than that of non–alternative tobacco employees when the density of the network is controlled for. This effect is statistically significant (*P*<.001) in both models. When homophily in Model 2 was tested for, there was a statistically significant negative effect (–0.27 log odds; *P*<.001) indicating that alternative tobacco users were less likely to engage with one another within the network. These findings indicate that while being an alternative tobacco employee increases the likelihood that a user engages within the CCASAA and NC-SFATA networks, these same users are more likely to interact with non–alternative tobacco industry employees within these online contexts.

Table 3. Exponential random graph model for valued edges to detect likelihood of being an alternative tobacco industry employee within the network.

Network parameter[a]	Model 1	*P* value	Model 2	*P* value
Sum (network density)	0.91	<.001	0.91	<.001
Nodefactor: Vape	0.35	<.001	0.42	<.001
Nonzero	–8.22	<.001	–8.08	<.001
Nodematch: Vape	—[b]	—	–0.27	<.001
AIC[c]	–163446	—	–163579	—

[a]Total nodes: 292; total edges: 500 (loop edges removed for the exponential random graph model); sample size per chain: 5000; thinning interval: 5000; reference distribution: Poisson.

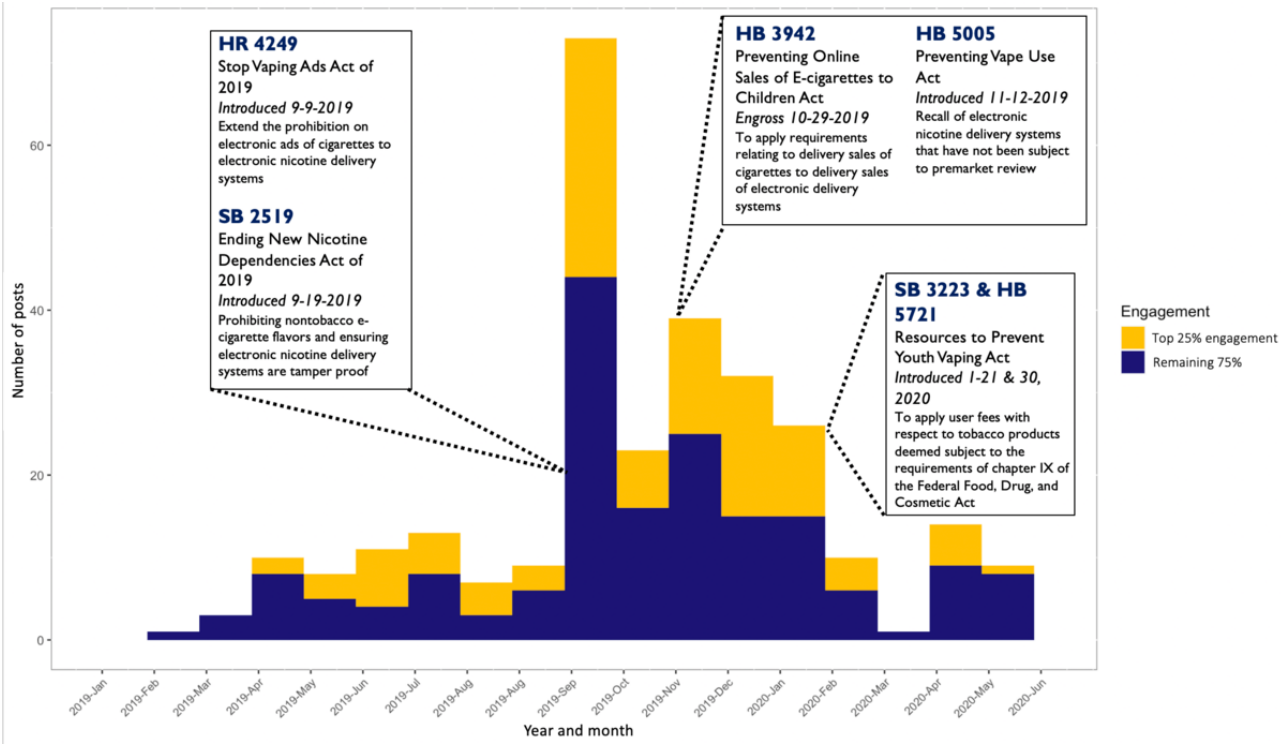
[b]Not included in the model.

[c]AIC: Akaike information criterion. AIC within the context of an exponential random graph model measures deviance based on log-likelihood, which is calculated by summing the differences between predicted probabilities and observed values.

Post Timeline Analysis

Both Figures 4 and 5 show the post timelines for CCASSA and NC-SFATA content, respectively, with the proportion of posts receiving the top 25% most engagement filled in yellow. Each timeline includes markers detailing the date and descriptions of important vaping legislation during the time period in which posts were analyzed. Although not every post explicitly mentions a tobacco control policy, it can be interpreted that different forms of provaping messaging posted around legislative events can influence the opinions and mobilization actions of users. For example, an advertisement for a vaping device that shows up on a group member's newsfeed still supports and communicates a provaping narrative whether or not it explicitly endorses a policy outcome. Users with higher exposure and engagement to pro–alternative tobacco messaging, which is the predominant theme within these Facebook groups, could be more likely to mobilize.

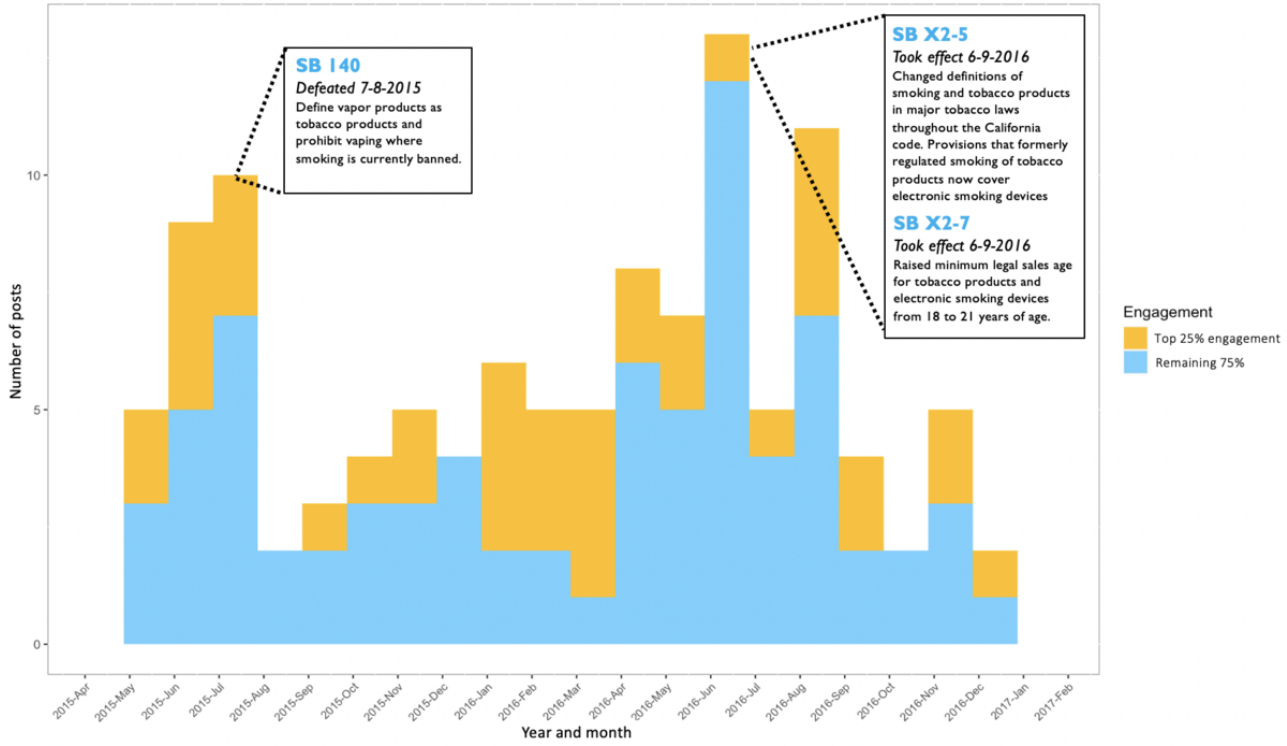
Figure 4. Consumer Advocates for Smoke-Free Alternatives Association posts timeline (2019-2020). HB: House Bill; HR: House of Representative; SB: Senate Bill.



As seen in Figure 4, the largest number and largest proportion of highly engaged posts within the CCASAA network were created in September 2019, around the same time that federal legislation regulating alternative tobacco ads and calling for prohibiting of nontobacco e-cigarette flavors were introduced (House of Representative [HR] 4249 and SB 2519). Although posts had notably dropped in the following month in October 2019, there was a modest increase in activity in November around the introduction of additional regulation to the delivery sales of ENDS (House Bill [HB] 3942 and HB 5005). Importantly, this timeline generally depicts periods of increased activity and posts among the CCASAA Facebook users during periods when federal anti–tobacco and alternative tobacco legislation was introduced, both in the US Senate and House of Representatives.

Figure 5 shows a bimodal distribution of posts within the NC-SFATA group. The first spike of posts begins on May 2015 and peaks on July 2015 around the same time that the law SB 140 was defeated in the California Senate, which would have classified ENDS products as tobacco and extended the prohibition of tobacco smoking in public places to ENDS products. The second peak of posts happened during June 2016, when the same legislation succeeded in a subsequent legislative session and began to take effect as did other tobacco control legislation that increased the legal sales age of ENDS from 18 to 21 (SB X2-5 and SB X2-7). Despite having the highest number of posts within the timeline, the proportion of high engagement posts during June 2016 was low, possibly suffering from lower user engagement as anti–alternative tobacco legislation had already been successfully passed.

Figure 5. Northern California Chapter of Smoke-Free Alternatives Trade Association posts timeline (2015-2017). SB: Senate Bill.



Discussion

Summary of Findings and Implications

Results from this study identify and characterize ways in which alternative tobacco interest groups act as virtual mobilization points across distinct online networks to influence tobacco control policy at both the federal and local level. This specifically includes our observation that alternative tobacco industry–affiliated actors are highly involved in antitobacco policy mobilization and advocacy activity with both an industry trade–focused group (NC-SFATA) that represents the interest of manufacturers, retailers, and distributors, but also a group that represents itself as a consumer organization for safer alternatives to tobacco. Specifically, our study found that there is a greater likelihood that alternative tobacco employees are within these Facebook user networks and that a higher proportion of industry employees are among the most influential members of the network. These findings are worrying as the presence of alternative tobacco employees and representatives may influence the factual narrative of tobacco and ENDS policy discussions due to financial interests and industry ties while not accurately conveying the concerns of consumers or consequences for public health outcomes.

Additionally, based on our timeline analysis, we observed that the highest number of messages with high engagement among these networks corresponded with dates tied to important tobacco control legislative events, evidencing increased activity around important policy decision-making windows. Some evidence of coordination between groups existed by examining network graphs with overlapping users in both organizations, and despite policy mobilization being active in both groups, they differed in their policy targets, with the CCASAA focusing on federal tobacco legislation and the NC-SFATA focusing on California state legislation. These results provide some indication of distinct mobilization and advocacy efforts at multiple policy-making levels. Although the purpose and messages of these activities clearly focused on disseminating information and influencing public perception for the purposes of defeating alternative tobacco legislation, levels of user engagement on these activities differed based on the group pages reviewed.

The network structure of the Facebook groups considered in this analysis differed in their density and post volume. The NC-SFATA Facebook group had lower density and more uneven distribution of engagement, while the NC-SFATA account was the main driver of posts to members. Additionally, the overall volume of posts was low compared to that of the CCASAA, and activity slowed after February 2017. The CCASAA had a

denser network structure, a larger volume of posts compared to the NC-SFATA, and a more even distribution of activity among users that continues to the present. This suggests that the sustainability of engagement levels across interest groups is highly variable and may also be influenced by the presence and interaction of other affiliated groups (eg, SFATA and CASAA parent organizations that are also both active on Facebook).

Determining the specific factors that influence levels of engagement was beyond the scope of the current analysis; differences may be due to the more formal nature of trade associations (which may lend themselves to more linear communication structure) versus that of consumer-focused groups or to other organizational changes that occurred within the CCASAA. The existing literature, however, suggests that emotion plays a strong role in the alternative tobacco market as demonstrated in previous work which found business strategies used within the alternative tobacco market to be highly contested, volatile, and interwoven with competition, emotion, and conflicting beliefs [72]. The network characteristics shown in this study may indicate a higher emotional commitment among CCASAA members, a factor important in countering their political actions.

Although independent manufacturers and retailers within the alternative tobacco industry attempt to distinguish themselves from Big Tobacco and the goals of these 2 segments of the provaping movement may differ [28,73] (eg, implications of changes in tobacco use behavior and product sales due to dual use of cigarettes and e-cigarettes vs transition from cigarettes to e-cigarettes), they nevertheless are working to the same policy ends: to minimize restrictions on the marketing, sale, and use of tobacco and alternative tobacco products. As multinational tobacco companies engage in efforts to increase their share of the alternative tobacco market and expend large sums of money to lobby and fight ENDS regulations and restrictions, the less politically powerful and poorly resourced independent entities within the alternative tobacco industry may serve the role that front groups traditionally have served for the tobacco industry [74]. Through consumer groups and trade associations, they may attempt to influence policy makers through power in grassroots-based advocacy: mobilizing numerous “legitimate” voices from consumers and small businesses against tobacco and alternative tobacco control measures.

Further, results from our study likely only represent a very small segment of political influence exercised by the tobacco and alternative tobacco industry on social media platforms. For example, our study only examined 2 relatively small California-specific group pages and chapters; although the CASAA and SFATA’s parent organizations had much higher levels of activity and engagement on Facebook and Twitter (@CASAAmedia on has 33,000 and 22,000 followers on Facebook and Twitter, respectively, while @sfataorg has 15,000 and 10,000 followers on Facebook and Twitter, respectively), there are CASAA and SFATA Facebook group chapters in Florida, Missouri, Colorado, Connecticut, Hawaii, Texas, Arizona, and Ohio, to name a few other state-specific examples. There are also other pro–alternative tobacco industry groups engaged on social media that actively advocate against tobacco control measures that were not included in this study as we only focused on those associated with the state of California. Hence, although our study is limited to a single US jurisdiction and 2 pages, it is highly probable that lobbying and digital mobilization against tobacco control legislation via social media is occurring across multiple jurisdictions and a diversity of user groups, likely with similar ties to tobacco industry–affiliated employees and actors. This socially enabled online environment represents an important public constituency that can be activated against federal, state, and local alternative tobacco control measures, necessitating further research into ties with the broader tobacco industry and other front groups.

In response to this strategic use of Facebook (and other social media platforms) by the alternative tobacco industry to mobilize efforts to influence public perception and the outcome of federal and state tobacco control policy, public health stakeholders should expand their own efforts to mitigate and counter provaping narratives, particularly if they originate from alternative tobacco employees or lobbyists and misrepresent or include misinformation about current or pending tobacco control legislation. Public health stakeholders should also make a concerted effort to engage in these almost exclusively provaping virtual communities by establishing their own counter narratives highlighting the health and addictive harms of alternative tobacco products. Specifically, the posts on these pages appear to have an echo chamber effect, in that all the posts exclusively present a provaping narrative that is disseminated among users who are members of these pages. For Facebook pages that are open to the public, these discussions could spill over to other online communities, and in the absence of effective counter marketing, fact-checking, and health promotion, could lead to protobacco messages influencing opinions of other users. Platforms should also consider requiring disclosure of industry affiliations by page users or administrators and any corresponding potential conflicts of interest that may influence the type of information presented, including on individual posts that relate to claims about tobacco control policy.

Limitations

As not every user was included in the *z* test comparing the proportion of alternative tobacco employees between influential and noninfluential users, it is possible that access to profile data could not be randomly distributed across alternative tobacco employee versus nonalternative tobacco users. If alternative tobacco employee status influences the likelihood of having available profile data, then this could have biased the results. Additionally, the results from this study only examined 2 California-affiliated organizations. More networks between local and national organizations would need to be analyzed in order to generalize findings about communication patterns in the overall alternative tobacco industry. As data were collected retrospectively, each network is also susceptible to data loss for posts collected before July 1, 2020. As mentioned in the methods section, the NC-SFATA is a Facebook page while the CCASSA is a Facebook community, which may account for many of the structural differences between communication networks. However, both formats share the same functions for interaction, so the main difference is between being listed as a group member versus only liking the page. Even if the format of the group on Facebook influences how users communicate with

each other, it is possible that choosing a particular format could be done intentionally to better align with the communication goals of the page creator. Finally, the timeline analysis was conducted to clarify the association between the number of posts and legislative events and cannot be used to establish a causal connection.

Conclusions

Future research should continue to characterize online communication strategies that differ between national and state or local alternative tobacco trade associations, consumer groups, and lobbying organizations in the social media sphere. This should include more in-depth characterization of formal coordination efforts on messaging, policy advocacy planning, and attempts to introduce misinformation about the impact of tobacco control legislation. For example, our study found that the NC-SFATA group had less engagement but a higher number of followers while the CCASSA group had fewer members but more engagement from these members. This may indicate specialization in the context of grass roots mobilization, advocacy, and policy substitution within the context of specific constituents, with SFATA focusing on mobilizing larger numbers of local and state actors and CCASSA focusing on smaller but more active engagement on federal issues. This potentially includes aligning with goals of each organization's respective national parent associations, with the aim of defeating both federal and state legislation in a strategic and coordinated fashion. Unfortunately, the same coordination that is needed among federal regulators, public health professionals, and state health agencies to promote tobacco control policies may be absent from these social networking service platforms, meaning that pro–alternative tobacco narratives may unduly influence the policy-making process, threatening future tobacco legislation and implementation of policy already in place.

Acknowledgments

This study was funded by the University of California Tobacco-Related Disease Research Program (award #T29IP0384).

Authors' Contributions

MN collected the data, while all authors designed the study, conducted the data analyses, wrote the manuscript, and approved the final version of the manuscript.

Conflicts of Interest

QX, MN, and TKM are employees of the startup company S-3 Research LLC. S-3 Research is a startup funded and currently supported by the National Institutes of Health-National Institute on Drug Abuse through a Small Business Innovation and Research contract for opioid-related social media research and technology commercialization. The authors have no other conflicts of interest to declare.

References

1. Creamer MR, Wang TW, Babb S, Cullen KA, Day H, Willis G, et al. Tobacco product use and cessation indicators among adults - United States, 2018. MMWR Morb Mortal Wkly Rep 2019 Nov 15;68(45):1013-1019 [FREE Full text] [doi: 10.15585/mmwr.mm6845a2] [Medline: 31725711]
2. Wang TW, Gentzke AS, Creamer MR, Cullen KA, Holder-Hayes E, Sawdey MD, et al. Tobacco product use and associated factors among middle and high school students - United States, 2019. MMWR Surveill Summ 2019 Nov 06;68(12):1-22 [FREE Full text] [doi: 10.15585/mmwr.ss6812a1] [Medline: 31805035]
3. E-cigarette industry to surpass $39 billion revenue by 2030, globally. Bloomberg. 2020. URL: https://www.bloomberg.com/press-releases/2020-10-12/e-cigarette-industry-to-surpass-39-billion-revenue-by-2030-globally-p-s-intelligence [accessed 2021-01-30]
4. Wang JB, Olgin JE, Nah G, Vittinghoff E, Cataldo JK, Pletcher MJ, et al. Cigarette and e-cigarette dual use and risk of cardiopulmonary symptoms in the Health eHeart Study. PLoS One 2018;13(7):e0198681 [FREE Full text] [doi: 10.1371/journal.pone.0198681] [Medline: 30044773]
5. Wills TA, Knight R, Williams RJ, Pagano I, Sargent JD. Risk factors for exclusive e-cigarette use and dual e-cigarette use and tobacco use in adolescents. Pediatrics 2015 Jan 15;135(1):e43-e51 [FREE Full text] [doi: 10.1542/peds.2014-0760] [Medline: 25511118]
6. Houston AR, Howard A, Sweanor D. Placing the legal vape market in the hands of big tobacco. Am J Public Health 2020 Jun;110(6):781-782. [doi: 10.2105/ajph.2020.305676]
7. McKee M. Will e-cigarette regulation keep pace with new products? Am J Public Health 2020 Jun;110(6):782-783. [doi: 10.2105/AJPH.2020.305668] [Medline: 32374689]
8. Electronic, or e-cigarette, and other electronic smoking devices policy adoption. American Nonsmokers' Rights Foundation. 2018. URL: https://no-smoke.org/electronic-e-cigarette-electronic-smoking-devices-policy-adoption/ [accessed 2021-01-30]
9. U.S. e-cigarette regulations - 50 state review (2021). Public Health Law Center. URL: https://www.publichealthlawcenter.org/resources/us-e-cigarette-regulations-50-state-review [accessed 2020-11-19]

10. Malone R. The tobacco industry. In: The Bottom Line Or Public Health: Tactics Corporations Use to Influence Health and Health Policy, and What We Can Do to Counter Them. New York: Oxford University Press; 2010:155-191.
11. McDaniel PA, Malone RE. Tobacco industry and public health responses to state and local efforts to end tobacco sales from 1969-2020. PLoS One 2020 May 22;15(5):e0233417 [FREE Full text] [doi: 10.1371/journal.pone.0233417] [Medline: 32442202]
12. Savell E, Gilmore AB, Fooks G. How does the tobacco industry attempt to influence marketing regulations? A systematic review. PLoS One 2014;9(2):e87389 [FREE Full text] [doi: 10.1371/journal.pone.0087389] [Medline: 24505286]
13. Smith KE, Savell E, Gilmore AB. What is known about tobacco industry efforts to influence tobacco tax? A systematic review of empirical studies. Tob Control 2013 Mar;22(2):144-153 [FREE Full text] [doi: 10.1136/tobaccocontrol-2011-050098] [Medline: 22887175]
14. World Health Organization. Tobacco Industry Interference With Tobacco Control. Geneva: WHO Press; 2008.
15. Alamar BC, Glantz SA. The tobacco industry's use of Wall Street analysts in shaping policy. Tob Control 2004 Sep;13(3):223-227. [doi: 10.1136/tc.2003.006643] [Medline: 15333876]
16. Apollonio DE, Bero LA. Industry front groups: a tobacco case study. J. Verbr. Lebensm 2007 Jul 30;2(3):341-348. [doi: 10.1007/s00003-007-0205-8]
17. Fallin A, Grana R, Glantz SA. 'To quarterback behind the scenes, third-party efforts': the tobacco industry and the Tea Party. Tob Control 2014 Jul 08;23(4):322-331 [FREE Full text] [doi: 10.1136/tobaccocontrol-2012-050815] [Medline: 23396417]
18. Muggli ME, Forster JL, Hurt RD, Repace JL. The smoke you don't see: uncovering tobacco industry scientific strategies aimed against environmental tobacco smoke policies. Am J Public Health 2001 Sep;91(9):1419-1423. [doi: 10.2105/ajph.91.9.1419] [Medline: 11527774]
19. Stauber JC. Smokers' hacks: the tobacco lobby's PR front groups. PR Watch. 1994. URL: https://www.prwatch.org/content/pr-watch-archives-1994 [accessed 2021-01-30]
20. Velicer C, Glantz SA. Hiding in the shadows: Philip Morris and the use of third parties to oppose ingredient disclosure regulations. PLoS One 2015;10(12):e0142032 [FREE Full text] [doi: 10.1371/journal.pone.0142032] [Medline: 26717245]
21. Samuels B, Glantz SA. The politics of local tobacco control. JAMA: The Journal of the American Medical Association 1991;266(15):2110-2117. [doi: 10.1001/jama.266.15.2110]
22. Smith EA, Malone RE. 'We will speak as the smoker': the tobacco industry's smokers' rights groups. Eur J Public Health 2007 Jun;17(3):306-313 [FREE Full text] [doi: 10.1093/eurpub/ckl244] [Medline: 17065174]
23. Traynor MP. New tobacco industry strategy to prevent local tobacco control. JAMA 1993 Jul 28;270(4):479. [doi: 10.1001/jama.1993.03510040083034]
24. Lyon TP, Maxwell JW. Astroturf: interest group lobbying and corporate strategy. J Economics Management Strategy 2004 Dec;13(4):561-597. [doi: 10.1111/j.1430-9134.2004.00023.x]
25. Cigarette market share worldwide as of 2019, by company. Statista. URL: https://www.statista.com/statistics/279873/global-cigarette-market-share-by-group/ [accessed 2021-04-30]
26. Levy DT, Lindblom EN, Sweanor DT, Chaloupka F, O'Connor RJ, Shang C, et al. An economic analysis of the pre-deeming US market for nicotine vaping products. Tob Regul Sci 2019 Mar;5(2):169-181 [FREE Full text] [doi: 10.18001/trs.5.2.8] [Medline: 32864395]
27. Levy DT, Sweanor D, Sanchez-Romero LM, O'Connor R, Goniewicz ML, Borland R. Altria-Juul Labs deal: why did it occur and what does it mean for the US nicotine delivery product market. Tob Control 2020 Dec;29(e1):e171-e174 [FREE Full text] [doi: 10.1136/tobaccocontrol-2019-055081] [Medline: 31484802]
28. Sussman S, Baezconde-Garbanati L, Garcia R, Barker DC, Samet JM, Leventhal A, et al. Forces that drive the vape shop industry and implications for the health professions. Eval Health Prof 2016 Sep;39(3):379-388 [FREE Full text] [doi: 10.1177/0163278715586295] [Medline: 25967071]
29. Smith EA, McDaniel PA, Malone RE. California advocates' perspectives on challenges and risks of advancing the tobacco endgame. J Public Health Policy 2020 Sep;41(3):321-333. [doi: 10.1057/s41271-020-00230-5] [Medline: 32461584]
30. State of tobacco control California 2020. American Lung Association. URL: https://www.lung.org/local-content/ca/state-of-tobacco-control/2020/sota-20 [accessed 2021-04-29]
31. Cressey D. California's anti-vaping bill goes up in smoke. Nature 2015 Jul 16;523(7560):267-267. [doi: 10.1038/523267a] [Medline: 26178943]
32. D'Souza D. JUUL spent millions as e-cigarette lobbying fired up this year. Investopedia. URL: https://www.investopedia.com/e-cigarette-lobbying-fires-up-as-juul-spends-millions-4771268 [accessed 2021-04-27]
33. Karlis N. The e-cig industry poured money into lobbying prior to Trump's reversal on vaping ban. Salon. 2019. URL: https://www.salon.com/2019/11/18/the-e-cig-industry-poured-money-into-lobbying-prior-to-trumps-reversal-on-vaping-ban/ [accessed 2021-04-27]
34. Thomas K, Kaplan S. E-cigarettes went unchecked in 10 years of federal inaction. The New York Times. 2019. URL: https://www.nytimes.com/2019/10/14/health/vaping-e-cigarettes-fda.html [accessed 2021-04-27]
35. Ramamurthi D, Fadadu RP, Jackler RK. Electronic cigarette marketers manipulate antitobacco advertisements to promote vaping. Tob Control 2016 Nov 06;25(6):720-722. [doi: 10.1136/tobaccocontrol-2015-052661] [Medline: 26546153]

36. Cole-Lewis H, Varghese A, Sanders A, Schwarz M, Pugatch J, Augustson E. Assessing electronic cigarette-related tweets for sentiment and content using supervised machine learning. J Med Internet Res 2015 Aug 25;17(8):e208 [FREE Full text] [doi: 10.2196/jmir.4392] [Medline: 26307512]
37. Cole-Lewis H, Pugatch J, Sanders A, Varghese A, Posada S, Yun C, et al. Social listening: a content analysis of e-cigarette discussions on Twitter. J Med Internet Res 2015 Oct 27;17(10):e243 [FREE Full text] [doi: 10.2196/jmir.4969] [Medline: 26508089]
38. Myslín M, Zhu S, Chapman W, Conway M. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. J Med Internet Res 2013 Aug;15(8):e174 [FREE Full text] [doi: 10.2196/jmir.2534] [Medline: 23989137]
39. Mackey TK, Miner A, Cuomo RE. Exploring the e-cigarette e-commerce marketplace: Identifying Internet e-cigarette marketing characteristics and regulatory gaps. Drug Alcohol Depend 2015 Nov 01;156:97-103. [doi: 10.1016/j.drugalcdep.2015.08.032] [Medline: 26431794]
40. Zhu S, Sun JY, Bonnevie E, Cummins SE, Gamst A, Yin L, et al. Four hundred and sixty brands of e-cigarettes and counting: implications for product regulation. Tob Control 2014 Jul 16;23 Suppl 3(suppl 3):iii3-iii9 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051670] [Medline: 24935895]
41. Huang J, Kornfield R, Szczypka G, Emery SL. A cross-sectional examination of marketing of electronic cigarettes on Twitter. Tob Control 2014 Jul;23 Suppl 3:iii26-iii30 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051551] [Medline: 24935894]
42. Liang Y, Zheng X, Zeng DD, Zhou X, Leischow SJ, Chung W. Exploring how the tobacco industry presents and promotes itself in social media. J Med Internet Res 2015 Jan;17(1):e24 [FREE Full text] [doi: 10.2196/jmir.3665] [Medline: 25608524]
43. Sowles SJ, Krauss MJ, Connolly S, Cavazos-Rehg PA. A content analysis of vaping advertisements on Twitter, November 2014. Prev Chronic Dis 2016 Sep 29;13:E139 [FREE Full text] [doi: 10.5888/pcd13.160274] [Medline: 27685432]
44. Cuomo RE, Miner A, Mackey TK. Pricing and sales tax collection policies for e-cigarette starter kits and disposable products sold online. Drug Alcohol Rev 2016 Jan;35(1):110-114. [doi: 10.1111/dar.12353] [Medline: 26492956]
45. Jo CL, Kornfield R, Kim Y, Emery S, Ribisl KM. Price-related promotions for tobacco products on Twitter. Tob Control 2016 Jul;25(4):476-479. [doi: 10.1136/tobaccocontrol-2015-052260] [Medline: 26130449]
46. Kaplan S, Hoffman J. F.D.A. targets vaping, alarmed by teenage use. The New York Times. 2018. URL: https://www.nytimes.com/2018/09/12/health/juul-fda-vaping-ecigarettes.html [accessed 2021-01-30]
47. Kim AE, Hopper T, Simpson S, Nonnemaker J, Lieberman AJ, Hansen H, et al. Using Twitter data to gain insights into e-cigarette marketing and locations of use: an infoveillance study. J Med Internet Res 2015 Nov 06;17(11):e251 [FREE Full text] [doi: 10.2196/jmir.4466] [Medline: 26545927]
48. Watts C, Hefler M, Freeman B. 'We have a rich heritage and, we believe, a bright future': how transnational tobacco companies are using Twitter to oppose policy and shape their public identity. Tob Control 2019 Mar;28(2):227-232. [doi: 10.1136/tobaccocontrol-2017-054188] [Medline: 29666168]
49. Hefler M, Freeman B, Chapman S. Tobacco control advocacy in the age of social media: using Facebook, Twitter and change. Tob Control 2013 May;22(3):210-214. [doi: 10.1136/tobaccocontrol-2012-050721] [Medline: 23047890]
50. About us. CASAA. URL: https://www.casaa.org/gallery-with-sidebar/ [accessed 2020-11-19]
51. California Consumer Advocates for Smokefree Alternatives Assoc. Facebook Groups. URL: https://www.facebook.com/groups/299151773932781/ [accessed 2020-11-19]
52. SFATA Mission Statement. Smoke Free Alternatives Trade Association. URL: https://sfata.org/content.aspx?page_id=22&club_id=89995&module_id=260904 [accessed 2020-11-19]
53. Northern California Chapter of SFATA. Facebook. URL: https://m.facebook.com/NorCalSFATA/about/?ref=page_internal&mt_nav=0&_rdr [accessed 2020-11-19]
54. Gerson J, Plagnol AC, Corr PJ. Passive and Active Facebook Use Measure (PAUM): validation and relationship to the Reinforcement Sensitivity Theory. Personality and Individual Differences 2017 Oct;117:81-90. [doi: 10.1016/j.paid.2017.05.034]
55. Verduyn P, Lee DS, Park J, Shablack H, Orvell A, Bayer J, et al. Passive Facebook usage undermines affective well-being: experimental and longitudinal evidence. J Exp Psychol Gen 2015 Apr;144(2):480-488. [doi: 10.1037/xge0000057] [Medline: 25706656]
56. Krasnova H, Widjaja T, Buxmann P, Wenninger H, Benbasat I. Envy on Facebook: A hidden threat to users? life satisfaction? Information Systems Research 2015 Sep;26(3):585-605 [FREE Full text] [doi: 10.1287/isre.2015.0588]
57. Deters FG, Mehl MR. Does posting Facebook status updates increase or decrease loneliness? An online social networking experiment. Soc Psychol Personal Sci 2013 Sep 01;4(5):1-13 [FREE Full text] [doi: 10.1177/1948550612469233] [Medline: 24224070]
58. Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. Softw: Pract. Exper 1991 Nov;21(11):1129-1164. [doi: 10.1002/spe.4380211102]
59. Katz L. A new status index derived from sociometric analysis. Psychometrika 1953 Mar;18(1):39-43. [doi: 10.1007/bf02289026]
60. Bonacich P. Power and centrality: a family of measures. American Journal of Sociology 1987 Mar;92(5):1170-1182. [doi: 10.1086/228631]

61. Bonacich P. Some unique properties of eigenvector centrality. Social Networks 2007 Oct;29(4):555-564. [doi: 10.1016/j.socnet.2007.04.002]
62. Batool K, Niazi MA. Towards a methodology for validation of centrality measures in complex networks. PLoS One 2014;9(4):e90283 [FREE Full text] [doi: 10.1371/journal.pone.0090283] [Medline: 24709999]
63. Bonacich P, Lloyd P. Eigenvector-like measures of centrality for asymmetric relations. Social Networks 2001 Jul;23(3):191-201. [doi: 10.1016/s0378-8733(01)00038-7]
64. Dubois E, Gaffney D. The multiple facets of influence. American Behavioral Scientist 2014 Apr 08;58(10):1260-1277. [doi: 10.1177/0002764214527088]
65. Everett MG, Borgatti SP. Extending centrality. Models and methods in social network analysis 2005;35(1):57-76. [doi: 10.1017/cbo9780511811395.004]
66. Robins G, Pattison P, Elliott P. Network models for social influence processes. Psychometrika 2001 Jun;66(2):161-189. [doi: 10.1007/bf02294834]
67. van der Pol J. Introduction to network modeling using exponential random graph models (ERGM): theory and an application using R-project. Comput Econ 2018 Sep 17;54(3):845-875. [doi: 10.1007/s10614-018-9853-2]
68. Krivitsky P, Handcock M, Hunter D. ergm.count: fit, simulate and diagnose exponential-family models for networks with count edges. CRAN.R-project. URL: https://CRAN.R-project.org/package=ergm.count [accessed 2021-02-18]
69. Krivitsky PN. Exponential-family random graph models for valued networks. Electron. J. Statist 2012 Jan 1;6(none):1100-1128. [doi: 10.1214/12-ejs696]
70. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. Annu. Rev. Sociol 2001 Aug;27(1):415-444. [doi: 10.1146/annurev.soc.27.1.415]
71. Flatt J, Agimi Y, Albert S. Homophily and health behavior in social networks of older adults. Fam Community Health 2012;35(4):312-321. [doi: 10.1097/fch.0b013e3182666650]
72. de Andrade M, Spotswood F, Hastings G, Angus K, Angelova N. Emotion in the ANDS (alternative nicotine delivery systems) market: practice-theoretical insight into a volatile market. Social Business 2017 Dec 21;7(3):391-418. [doi: 10.1362/204440817x15108539431569]
73. de Andrade M, Angus K, Hastings G, Angelova N. Hostage to fortune: an empirical study of the tobacco industry's business strategies since the advent of e-cigarettes. Crit Public Health 2020 Dec 18;30(3):280-293 [FREE Full text] [doi: 10.1080/09581596.2018.1552778] [Medline: 32536808]
74. Levy DT, Sweanor D, Sanchez-Romero LM, O'Connor R, Goniewicz ML, Borland R. Altria-Juul Labs deal: why did it occur and what does it mean for the US nicotine delivery product market. Tob Control 2020 Dec 04;29(e1):e171-e174 [FREE Full text] [doi: 10.1136/tobaccocontrol-2019-055081] [Medline: 31484802]

Abbreviations

CCASAA: Consumer Advocates for Smoke-Free Alternatives Association
ENDS: electronic nicotine delivery systems
ERGM: exponential random graph model
HB: House Bill
HR: House of Representatives
NC-SFATA: Northern California Chapter of Smoke-Free Alternatives Trade Association
SB: Senate Bill
SNA: social network analysis

Edited by R Kukafka; submitted 19.02.21; peer-reviewed by C Watts, K Durant; comments to author 11.04.21; revised version received 01.05.21; accepted 04.08.21; published 29.10.21.

Please cite as:
Haupt MR, Xu Q, Yang J, Cai M, Mackey TK
Characterizing Vaping Industry Political Influence and Mobilization on Facebook: Social Network Analysis
J Med Internet Res 2021;23(10):e28069
URL: https://www.jmir.org/2021/10/e28069
doi:10.2196/28069
PMID:

Original Paper

Digital Tracking of Physical Activity, Heart Rate, and Inhalation Behavior in Patients With Pulmonary Arterial Hypertension Treated With Inhaled Iloprost: Observational Study (VENTASTEP)

Barbara Stollfuss[1], MD, PhD; Manuel Richter[2], MD; Daniel Drömann[3], MD; Hans Klose[4], MD; Martin Schwaiblmair[5], MD; Ekkehard Gruenig[6], MD; Ralf Ewert[7], MD; Martin C Kirchner[1], Dipl-Biol; Frank Kleinjung[8], PhD; Valeska Irrgang[1], MD; Christian Mueller[1], PhD

[1]Bayer Vital GmbH, Leverkusen, Germany

[2]Department of Internal Medicine, Justus-Liebig-University Giessen, Universities of Giessen and Marburg Lung Center, Member of the German Center for Lung Research, Giessen, Germany

[3]Department of Pneumology, Universitätsklinikum Schleswig-Holstein, Member of the German Center for Lung Research, Lübeck, Germany

[4]Department of Pneumology, Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany

[5]Department of Pneumology, I. Medizinische Klinik, Universitätsklinikum Augsburg, Augsburg, Germany

[6]Centre for Pulmonary Hypertension, Thoraxclinic Heidelberg GmbH at Heidelberg University Hospital and German Centre for Lung Research, Heidelberg, Germany

[7]Department of Internal Medicine, Universitätsmedizin Greifswald, Greifswald, Germany

[8]Bayer AG, Berlin, Germany

Corresponding Author:
Christian Mueller, PhD
Bayer Vital GmbH
Building K 56, 1D321
Leverkusen, 51368
Germany
Phone: 49 2143046587
Email: christian.mueller4@bayer.com

Abstract

Background: Pulmonary arterial hypertension restricts the ability of patients to perform routine physical activities. As part of pulmonary arterial hypertension treatment, inhaled iloprost can be administered via a nebulizer that tracks inhalation behavior. Pulmonary arterial hypertension treatment is guided by intermittent clinical measurements, such as 6-minute walk distance, assessed during regular physician visits. Continuous digital monitoring of physical activity may facilitate more complete assessment of the impact of pulmonary arterial hypertension on daily life. Physical activity tracking with a wearable has not yet been assessed with simultaneous tracking of pulmonary arterial hypertension medication intake.

Objective: We aimed to digitally track the physical parameters of patients with pulmonary arterial hypertension who were starting treatment with iloprost using a Breelib nebulizer. The primary objective was to investigate correlations between changes in digital physical activity measures and changes in traditional clinical measures and health-related quality of life over 3 months. Secondary objectives were to evaluate inhalation behavior, adverse events, and changes in heart rate and sleep quality.

Methods: We conducted a prospective, multicenter observational study of adults with pulmonary arterial hypertension in World Health Organization functional class III who were adding inhaled iloprost to existing pulmonary arterial hypertension therapy. Daily distance walked, step count, number of standing-up events, heart rate, and 6-minute walk distance were digitally captured using smartwatch (Apple Watch Series 2) and smartphone (iPhone 6S) apps during a 3-month observation period (which began when iloprost treatment began). Before and at the end of the observation period (within 2 weeks), we also evaluated 6-minute walk distance, Borg dyspnea, functional class, B-type natriuretic peptide (or N-terminal pro–B-type natriuretic peptide) levels, health-related quality of life (EQ-5D questionnaire), and sleep quality (Pittsburgh Sleep Quality Index).

Results: Of 31 patients, 18 were included in the full analysis (observation period: median 91.5 days, IQR 88.0 to 92.0). Changes from baseline in traditional and digital 6-minute walk distance were moderately correlated (r=0.57). Physical activity (daily distance walked: median 0.4 km, IQR –0.2 to 1.9; daily step count: median 591, IQR −509 to 2413) and clinical measures

(traditional 6-minute walk distance: median 26 m, IQR 0 to 40) changed concordantly from baseline to the end of the observation period. Health-related quality of life showed little change. Total sleep score and resting heart rate slightly decreased. Distance walked and step count showed short-term increases after each iloprost inhalation. No new safety signals were identified (safety analysis set: n=30).

Conclusions: Our results suggest that despite challenges, parallel monitoring of physical activity, heart rate, and iloprost inhalation is feasible in patients with pulmonary arterial hypertension and may complement traditional measures in guiding treatment; however, the sample size of this study limits generalizability.

Trial Registration: ClinicalTrials.gov NCT03293407; https://clinicaltrials.gov/ct2/show/NCT03293407

International Registered Report Identifier (IRRID): RR2-10.2196/12144

(J Med Internet Res 2021;23(10):e25163) doi:10.2196/25163

KEYWORDS

6-minute walk distance; 6MWD; Breelib; daily physical activity; digital monitoring; health-related quality of life; iloprost; Ventavis; inhalation behavior; mobile phone; pulmonary arterial hypertension; PAH; sleeping behavior; behavior; sleep; monitoring; physical activity; heart; cardiology

Introduction

Background

Remote patient monitoring can play an important role in disease management, supporting the conventional approach of face-to-face visits. In patients with chronic diseases such as left-sided heart failure, remote monitoring has been reported to have beneficial effects—reduced readmissions (possibly due to earlier treatment for symptoms detected before scheduled follow-up visits), reduced mortality, increased quality of life, increased participation in self-management of the disease, and improved knowledge of the disease [1-4].

Pulmonary arterial hypertension is a progressive, life-threatening disease that can lead to right-sided heart failure and death [5]. Dyspnea, a common symptom of pulmonary arterial hypertension, restricts physical activity, which in turn impairs quality of life. The impairment of quality of life in patients with pulmonary arterial hypertension is severe, with health-related quality of life scores comparable to those reported for patients with spinal cord injuries or cancer unresponsive to therapy [6]. However, the use of vasoactive treatment to address pulmonary arterial hypertension is usually guided by intermittent clinical measurements of the 6-minute walk distance (a routine assessment of exercise capacity in pulmonary arterial hypertension), World Health Organization (WHO) functional class, and levels of B-type natriuretic peptide (BNP) or N-terminal pro–B-type natriuretic peptide (NT-proBNP) [7]. These traditional clinical measures are assessed during regular out- or inpatient visits, usually at widely spaced intervals (eg, every 3-6 months), and therefore, provide only a snapshot of the patient's health status. In times of increased reliance on telemedicine, such as during the current COVID-19 pandemic, visits between patients and their treating physicians may not be possible. For example, patients may attempt to reduce the probability of infection by avoiding visits in hospitals, and hospitals may enact restrictions and cancel visits. Furthermore, traditional clinical parameters do not address the impact of pulmonary arterial hypertension on daily life, which may be more important to the patient than functional or laboratory measures [8,9]. Remote monitoring may provide a more complete picture by allowing continuous assessment of parameters (such as physical activity) that reflect the impact of pulmonary arterial hypertension on daily life. However, it is unclear whether digital measures of physical activity are associated with traditional clinical parameters during long-term treatment. Furthermore, physical activity has not yet been assessed simultaneously with medication intake tracking.

Iloprost, a prostacyclin-based inhaled therapy for patients with primary pulmonary hypertension (idiopathic pulmonary arterial hypertension) in WHO functional class III, has shown beneficial effects on 6-minute walk distance and WHO functional class in clinical trials [10-12]. Iloprost is administered via nebulizers such as the Breelib nebulizer (Vectura Group plc) [13], which automatically saves inhalation information (dates, times, and completeness of inhalations).

Objectives

The primary aim of our study was to track daily physical activity in patients with pulmonary arterial hypertension starting treatment with inhaled iloprost, using a commercially available smartwatch and smartphone, and to investigate the association between changes in digital physical activity measures and changes in traditional clinical measures of disease severity and health-related quality of life over a 12-week period.

Secondary aims were to assess acceptance of the smartwatch by patients and the feasibility of measuring 6-minute walk distance digitally and to identify digital measures linked to traditional 6-minute walk distance. Further aims were to digitally monitor heart rate and iloprost inhalation behavior, to explore temporal patterns in activity level and heart rate relative to iloprost inhalation periods, and to evaluate changes in clinical outcome measures, daily physical activity, heart rate, and sleep quality after initiation of inhaled iloprost therapy. The study was not designed to investigate or confirm the effectiveness and safety of iloprost (Ventavis).

Methods

Study Design

Additional details are available in a previous publication [14] and in Multimedia Appendix 1.

As part of a national, prospective, observational, multicenter, single-arm cohort study (VENTASTEP; Clinical trial.gov, NCT03293407) investigating the associations between clinical and device-reported outcomes in patients with pulmonary arterial hypertension (WHO functional class III) who were adding inhaled iloprost to existing pulmonary arterial hypertension therapy, we collected digital measures of daily physical activity, heart rate, and 6-minute walk distance using smartwatch (Apple Watch Series 2) and smartphone (Apple iPhone 6S) apps (xbird GmbH). The apps continuously captured data from the smartwatch for the purpose of analyzing physical activity and behavior patterns; the 6-minute walk distance app captured data from the smartwatch during the 6-minute walk distance test. Raw sensor data from the smartphone were not used because it was impossible to ascertain whether patients always carried the smartphone with them. Device functionality was reduced to the minimum requirements for the study. To reduce bias, the apps did not provide feedback on physical activity to patients or physicians. A nebulizer (Breelib, Bayer AG) was used to record inhalation behavior [13,15,16].

The study period consisted of baseline data collection during the period from the initial visit and decision to start iloprost treatment until the actual start of iloprost treatment (≤2 weeks), an observation period of 3 months (±2 weeks) from the start of iloprost treatment, and final data collection at discontinuation of therapy or at the end of the study (whichever was earliest). The decision to start treatment with inhaled iloprost was made at the discretion of the treating physician; the decision was made in advance and independently of inclusion in the study. Follow-up was conducted during routine patient visits to their pulmonary arterial hypertension centers.

The study was approved by the central ethics committee of the Justus Liebig University Giessen (AZ153/17) and the ethics committees of all participating sites.

Patients

In preparation for this study, a survey on acceptance of digital tracking was performed, which indicated that approximately 50% of patients were willing to participate [14].

Patients were enrolled from 7 specialized pulmonary arterial hypertension centers in Germany from February 2018 to July 2019. Patients aged ≥18 years with pulmonary arterial hypertension in WHO functional class III despite treatment were eligible for enrollment if they were willing to wear a smartwatch for the duration of the study, had no previous treatment with inhaled iloprost, and they and their treating physician had decided to initiate treatment with inhaled iloprost using a Breelib nebulizer. Patients were excluded if they were allergic to nickel and methacrylates, or if they were already participating in an investigational program that included an intervention outside of routine clinical practice. All patients who were enrolled provided signed informed consent.

Study participants were trained by site personnel or investigators to correctly use the smartphone and smartwatch. Study participants were given training on inhaler use through a patient support program (VENTAPLUS, ContraCare GmbH and Vitartis Medical-Services GmbH), which is routinely offered to patients.

Data Collection

Anamnestic data (demographic and clinical characteristics) were collected by an investigator from medical records, if available, and entered into an electronic case report form that formed part of an electronic data capture system developed and managed by a contract research organization (Institut Dr. Schauerte).

Traditional clinical measures (6-minute walk distance, Borg dyspnea score at end of 6-minute walk distance test, WHO functional class, and levels of BNP or NT-proBNP) and vital parameters were assessed at the initial and final visit to the study center; data were also recorded in the electronic case report form. Patient-reported outcomes—EuroQol 5-dimension questionnaire (EQ-5D) and Pittsburgh Sleep Quality Index (PSQI)—were documented on paper at the initial and final visits and sent to the contract research organization for analysis. Adverse events, from the periods between the first use of inhaled iloprost to 30 days after the last dose within the study period, were documented in the electronic case report form.

Digital measurements of daily physical activity and heart rate were captured continuously by the smartwatch and app. Digital 6-minute walk distance data (steps, distance walked, heart rate, and raw data from motion sensors) were collected at the initial and final visit using the smartwatch and 6-minute walk distance app (which uses step count and a stride length algorithm trained on healthy volunteer data). Inhalation data (frequency of inhalations, completeness of inhalations, and duration of inhalations) were captured by the nebulizer, transferred to the smartphone using the BreeConnect App (Bayer AG), and then transferred to the contract research organization for analysis.

All data collected by the smartwatch were stored pseudonymized on a secure cloud server (xbird GmbH) for further processing. Processed data points for all variables were transferred as pseudonymized data to the electronic data capture system and matched to other collected data after the last visit of the last patient.

Outcome Measures

The primary endpoint was correlation between changes from baseline to the final visit in digital measures of physical activity, traditional clinical measures, and in health-related quality of life. Changes from baseline were evaluated as a first step. The traditional clinical measures included in the primary analysis were the 6-minute walk distance, Borg dyspnea score after the 6-minute walk distance test, WHO functional class, and BNP or NT-proBNP levels. Health-related quality of life was assessed using the EQ-5D. The digital measures were distance walked, number of steps, number of standing-up events, and digital 6-minute walk distance (Multimedia Appendix 2).

Secondary endpoints included inhalation behavior (mean daily inhalation duration per session, mean daily number of inhalations, and mean daily proportion of complete and

incomplete inhalations) and the mean association between physical activity level (device-based) and time since the last inhalation. The change from baseline to the end of the observation period in sleep quality (PSQI) and heart rate (at rest and during the 6-minute walk distance test), the association between heart rate and other device-based measures during the whole study period, and the incidence of treatment-emergent adverse events and serious treatment-emergent adverse events were also evaluated as secondary endpoints. Other endpoints included acceptance of the smartwatch by patients, correlation of traditional and digital 6-minute walk distance measurements, and association of digital measures with 6-minute walk distance.

Statistical Analysis

Statistical analyses were exploratory and descriptive.

Based on a feasibility study conducted in 2018 and ongoing study experience, the final planned sample size was 25 to 50 patients. The minimum number of 25 patients was considered sufficient to obtain reasonably precise correlation coefficients (even with 40% missing data).

The safety analysis set included all patients who received at least 1 dose of inhaled iloprost. If a patient withdrew consent without agreeing to further use of their data, they were excluded from the safety analysis set. The full analysis set included all patients from the safety analysis set who had ≥3 days of activity measurements during the baseline period, ≥3 days of activity measurements during the last 2 weeks of the observation period, and data at the initial and final visits for at least 1 of the following clinical outcome measures: 6-minute walk distance, BNP/NT-proBNP levels, and WHO functional class. Patients identified as screening failures after enrollment (included by mistake) were excluded from the full analysis set.

Measurements per day were normalized, for an assumed 18 hours (6 AM to midnight) of activity per day if parts of the observational time period of that day were missing, based on the percentage of time spent wearing the smartwatch. If the watch was worn for less than 10% of a given day, data for that day were treated as missing.

Device-measured daily physical activity at baseline was calculated as the median of ≤14 daily assessments before the first intake of inhaled iloprost. Device-measured daily physical activity at the end of the observation period was calculated as the median of daily assessments in the last 14 days of the observation period.

Pearson correlations were calculated for complete, pairwise observations of traditional and digital measures included in the primary analysis. Missing data for these parameters were not imputed.

Regression analyses were performed with change from baseline in a traditional clinical measure as the dependent variable and changes from baseline in digital measures (distance walked, number of standing up events, and 6-minute walk distance) as independent variables.

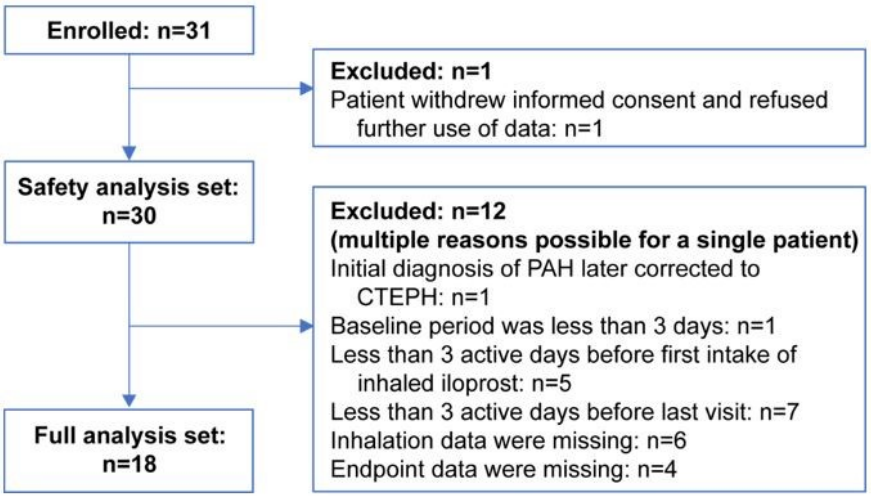
Results

Patients

We screened and enrolled 31 patients; 30 patients were included in the safety analysis set, and 18 patients were included in the full analysis set (Figure 1). Most patients in the full analysis set (Multimedia Appendix 3) were older adults (aged ≥65 years; n=12), female (n=12), and White (n=17). All patients were taking oral pulmonary arterial hypertension therapies at baseline.

The most common iloprost dose at the start and end of the observation period was 2.5 µg; 3 patients changed doses (Multimedia Appendix 4). The median number of inhalation sessions per day was 5.0 (IQR 4.3 to 5.7), and the inhalation duration per session was 5.5 minutes (IQR 3.8 to 6.3). More than 75% of the inhalations were complete—the median daily percentage of complete inhalations was 100% (IQR 100% to 100%; range 50%-100%).

Figure 1. Patient disposition. CTEPH: chronic thromboembolic pulmonary hypertension; PAH: pulmonary arterial hypertension.



Use of the Smartwatch During the Study

The median observation period in the full analysis set was 91.5 days (IQR 88.0 to 92.0; range: 34-113 days). Participants wore the smartwatch between 30 and 113 days overall (baseline period: range 4-14 days; end of the observation period: 8-14 days), for a mean time ranging from 4.8 hours per day to 11.5

hours per day (baseline period: range 4.1-14.6 hours per day; end of the observation period: range 4.7-10.5 hours per day). During the baseline period, 12 of the 18 patients (67%) wore the smartwatch for less than 7 days, and 6 of the 18 patients (33%) wore their smartwatch for at least 7 days. During the end of the observation period, almost three-quarters of the patients (13/18, 72%) wore the smartwatch every day, and all patients (18/18, 100%) wore it for at least 7 days. (Table 1). The number of hours spent wearing the watch per day decreased over the study period, but 89% of the patients (16/18) still wore the smartwatch for ≥6 hours per day on average during the end of the observation period (Table 1).

Table 1. Smartwatch use during the study (full analysis set).

Use characteristic	Patients, n (%)	
	Baseline period	Last 14 days of observation period
Number of days worn		
<7 days	12 (67)	0 (0)
≥7 days	6 (33)	18 (100)
≥10 days	3 (17)	17 (94)
14 days	2 (11)	13 (72)
Average daily hours worn		
<6 hours	2 (11)	2 (11)
≥6 hours	16 (89)	16 (89)
≥10 hours	9 (50)	4 (22)
≥14 hours	1 (6)	0 (0)
18 hours	0 (0)	0 (0)

Association Between Changes in Digital Physical Activity Measures and Changes in Traditional Clinical Measures

Comparison of Traditional and Digital Measures for 6-Minute Walk Distance

The median increase in traditional 6-minute walk distance was 26 m (Table 2). Although traditional and digital 6-minute walk distances were measured in parallel, digital 6-minute walk distance showed no substantial change from baseline (median −4.3 m, IQR −33.5 to 35.8), despite the observed change in traditionally measured 6-minute walk distance.

Table 2. Traditional and digital measures for 6-minute walk distance before and after 12 weeks of inhaled iloprost therapy.

Measure	Patients	Values		
	n	Median	IQR	Range
Traditional 6-minute walk distance (meters)				
First visit	18	339	250 to 420	196 to 546
Last visit	16[a]	366	282 to 418	207 to 585
Change[b]	16[a]	26	0 to 40	−43 to 87
Digital 6-minute walk distance[c] (meters)				
First visit	14	423.1	364.2 to 460.6	301.3 to 556.8
Last visit	14	433.9	329.2 to 486.3	267.8 to 592.6
Change	14	−4.3	−33.5 to 35.8	−69.2 to 90.0

[a]Two of the 18 patients withdrew consent during the observation period and did not perform a 6-minute walk distance test at their final visit (they agreed to the use of data that had already been collected).

[b]Patients analyzed at the first and last visit were used for the calculation of change. The median change as a robust center estimate is not necessarily the difference of the median of the total distance walked.

[c]Only for patients with available values for both baseline and final visit.

Further Traditional Measures and Health-Related Quality of Life

Borg dyspnea index increased from baseline in 7/18 patients (39%), remained stable in 6/18 patients (33%), and decreased in 3/18 patients (17%) (Table 3). WHO functional class changed from III to II in 4 patients and remained unchanged in the other 14 patients. The EQ-5D weighted index did not change considerably.

Table 3. Traditional clinical measures and health-related quality of life before and after 12 weeks of inhaled iloprost therapy.

Measure	Patients	Values		
	n	Median	IQR	Range
Borg dyspnea				
First visit	18	5.0	4.0 to 6.0	1.0 to 10.0
Last visit	16[a]	3.5	2.5 to 6.5	1.0 to 8.0
Change	16[a]	−0.5	−2.0 to 0.0	−3.0 to 3.0
EQ-5D[b]				
First visit	17	0.88	0.74 to 0.96	0.38 to 1.00
Last visit	16	0.87	0.74 to 0.97	−0.26 to 1.00
Change	15	0.02	0.00 to 0.08	−0.21 to 0.18
B-type natriuretic peptide level (ng/L)				
First visit	6	181	75 to 352	17 to 430
Last visit	6	136	44 to 166	22 to 304
Change	6	−16	−236 to 5	−264 to 99
N-terminal pro–B-type natriuretic peptide level (ng/L)				
First visit	10	1836	692 to 4676	78 to 15,749
Last visit	8	2005	759 to 3726	92 to 9859
Change	8	−749	−3415 to 212	−5890 to 1888

[a]Two of the 18 patients withdrew consent during the observation period and did not perform a 6-minute walk distance test at their final visit (they agreed to the use of data that had already been collected).

[b]EQ-5D: EuroQol 5-dimension questionnaire.

Digital Measures of Physical Activity

Digital measures of daily physical activity (daily distance walked, number of steps per day, and number of standing-up events per day) increased from baseline to the end of the observation period (Table 4).

Table 4. Digital measures before and after 12 weeks of inhaled iloprost therapy.

Measure	Patients	Values		
	n	Median	IQR	Range
Distance walked per day[a] (km)				
Baseline	18	5.2	3.1 to 7.6	1.7 to 14.9
End of observation	18	6.5	4.0 to 7.8	2.6 to 15.7
Change	18	0.4	−0.2 to 1.9	−3.6 to 4.7
Number of steps per day[a]				
Baseline	18	6721	4073 to 10,258	2268 to 18,398
End of observation	18	8332	5330 to 10,101	3389 to 19,451
Change	18	591	−509 to 2413	−4605 to 6075
Number of standing-up events per day[a]				
Baseline	18	24.2	20.4 to 25.6	0.0 to 56.5
End of observation	18	24.8	23.1 to 26.7	14.5 to 55.0
Change	18	1.7	−1.1 to 5.4	−9.9 to 25.5

[a]Average during baseline and last 14 days of observation period.

Correlation Between Changes From Baseline in Traditional and Digital Measures

Changes from baseline in traditional 6-minute walk distance and digital 6-minute walk distance were moderately correlated (Table 5). The numbers of patients with available data for BNP and NT-proBNP levels (n=6 and n=8, respectively) were too low for a meaningful analysis of correlation. A sensitivity analysis that had been planned (calculation using the complete data set instead of pairwise complete observations) was not performed because the volume of data was insufficient for further analysis.

Table 5. Pearson correlation between changes from baseline in traditional clinical parameters, health-related quality of life, and digital parameters after starting inhaled iloprost therapy.

Measure	Patients	Digital parameters, *r*			
	n	Average distance walked per day[a]	Average number of steps per day[a]	Average number of standing-up events per day[a]	6-minute walk distance[b]
Traditional parameters					
6-minute walk distance (meters)[b]	16	0.35	0.35	−0.30	0.57
Borg dyspnea[b]	16	−0.33	−0.33	−0.08	0.37
B-type natriuretic peptide level (ng/L)	6	−0.78	−0.79	0.27	−0.13
N-terminal pro–B-type natriuretic peptide level (ng/L)	8	−0.01	−0.02	0.11	−0.29
EQ-5D[c] weighted index	15	0.14	0.14	0.05	0.52

[a]Calculation based on a daily period from 6 AM to midnight.

[b]Two of the 18 patients withdrew consent during the observation period and did not perform a 6-minute walk distance test at their final visit (they agreed to the use of data that had already been collected).

[c]EQ-5D: EuroQol 5-dimension questionnaire.

Heart Rate

The median average heart rate at rest was 75.9 beats per minute (IQR 70.4 to 88.5) at baseline and 76.5 beats per minute (IQR 69.5 to 85.4) at the final visit. From baseline to the final visit, the average heart rate at rest showed a median decrease of −1.3 beats per minute (IQR −8.0 to 3.5).

Heart rates measured during the 6-minute walk distance test were slightly higher at the last visit compared with those at the baseline visit (Figure 2). The median changes in heart rate from 1 minute before the 6-minute walk distance test to each minute during the 6-minute walk distance test were also higher at the last visit (4.6 beats per minute to 23.9 beats per minute) than those at the baseline visit (3.0 beats per minute to 20.5 beats per minute) (Multimedia Appendix 5). Heart rate recovery immediately after the 6-minute walk distance test was slower at the last visit than that at the baseline visit (Figure 3).

Figure 2. Median heart rate Data collected by smartwatch during 6-minute walk distance test before (baseline) and under (last visit) inhaled iloprost therapy. Error bars show IQR.

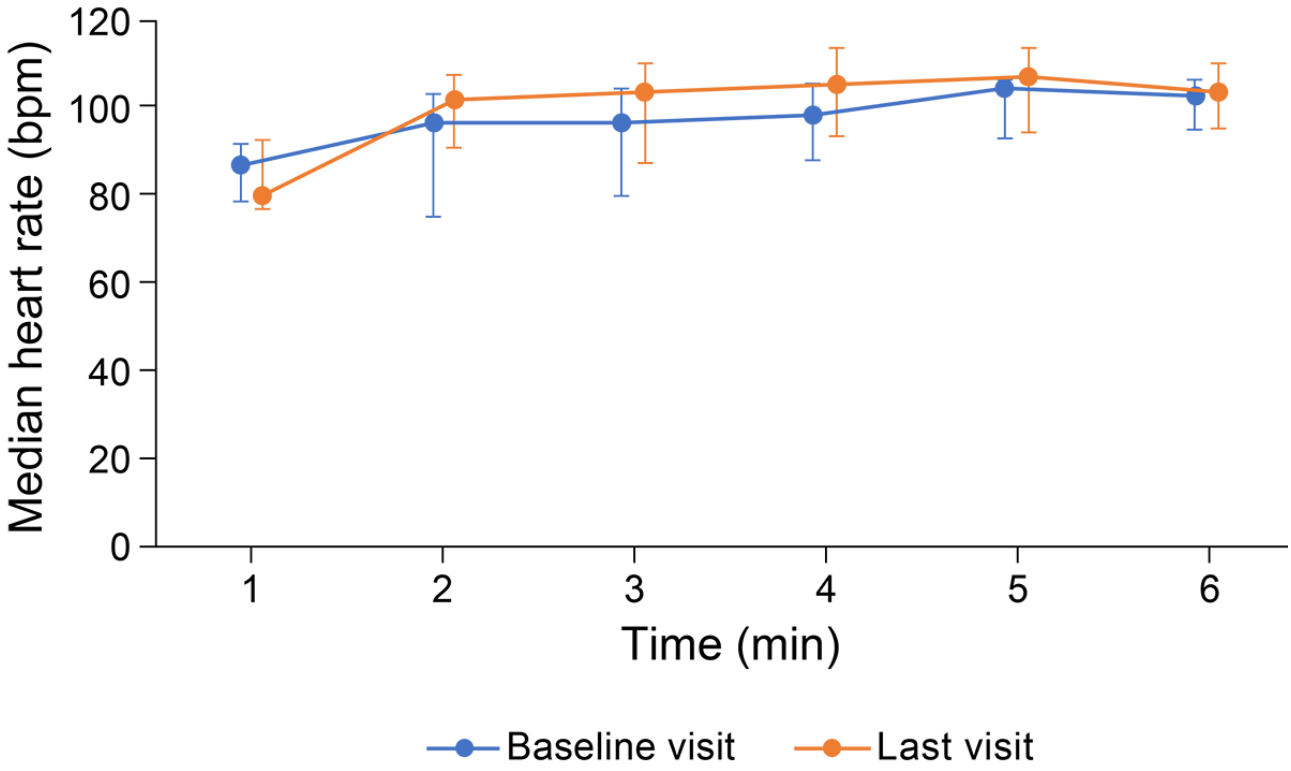
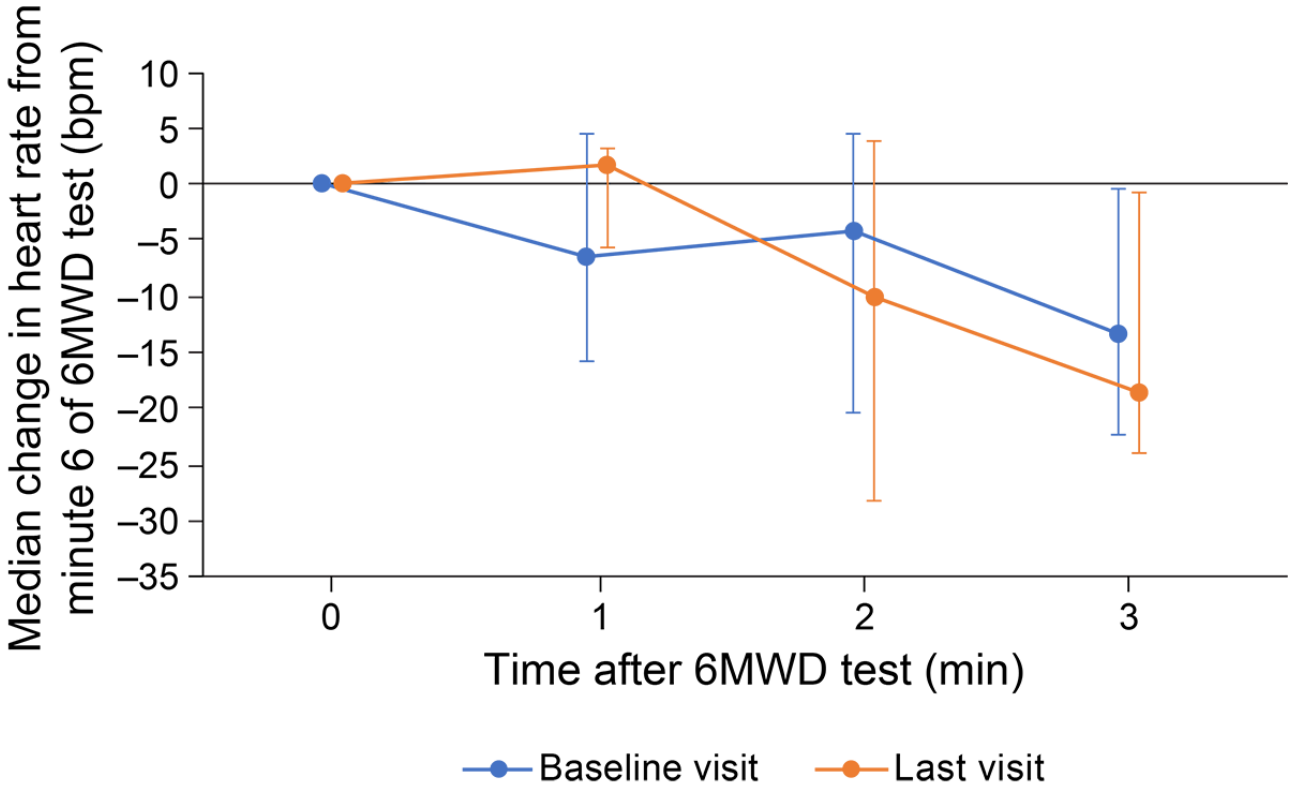


Figure 3. Heart rate recovery data collected by smartwatch at the end of the 6-minute walk distance (6MWD) test before (baseline) and under (last visit) inhaled iloprost therapy. Error bars show IQR.



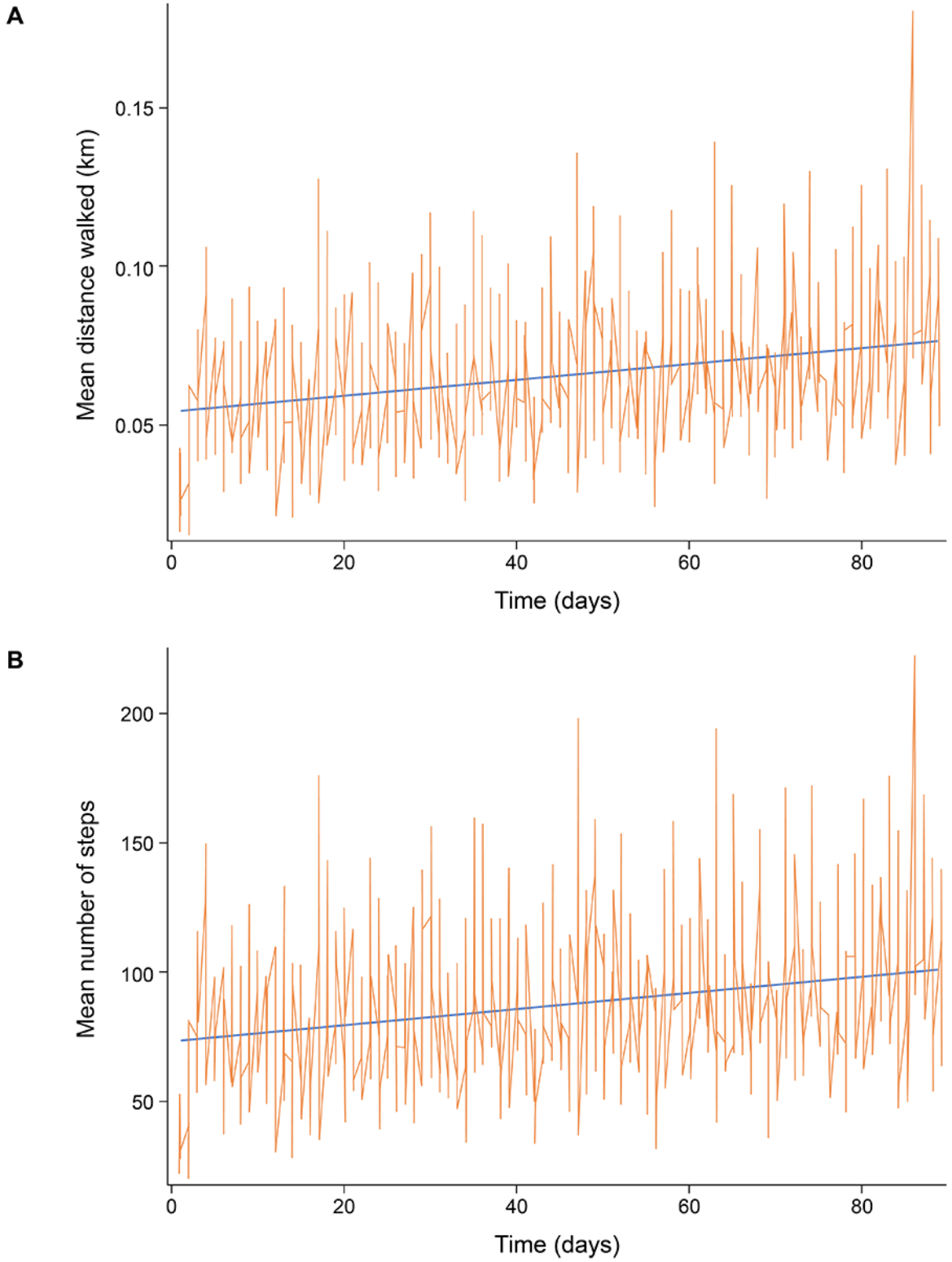
Inhalation and Changes in Heart Rate and Physical Activity

The study design enabled parallel digital tracking of iloprost inhalation, heart rate, and physical activity. An analysis of patients stratified by the median duration of iloprost inhalation and treatment month suggested that the average heart rate during iloprost inhalation showed a slight increase with duration of inhalation (Multimedia Appendix 6). Owing to large differences

in subgroup size, no clear trend for the change in average heart rate during treatment could be detected.

Distance walked (measured digitally) showed a short-term increase in activity shortly after inhalation and a long-term increase in activity over the entire observation period (Figure 4A). A similar association was observed for the number of steps (Figure 4B) but was not observed for the number of standing-up events (not shown).

Figure 4. (A) Distance walked and (B) number of steps (measured at 15-minute intervals for 3 hours after the end of each iloprost inhalation session) over the whole observation period.



For each iloprost inhalation session per day, 12 intervals (of 15 minutes) beginning at the end of the inhalation session were defined, and the distance walked per interval was recorded. The corresponding intervals for each inhalation per day were averaged for each patient (ie, the first intervals after the inhalations were averaged, then the second intervals after the

inhalations, etc) The resulting 12 average values for each patient each day were then averaged over all patients. Complete and incomplete inhalations were included. Approximately 3700 inhalations were excluded from the analysis because the watch was not worn over the entire period of 3 hours after the end of inhalation. Some of the 3-hour intervals were overlapping. The coverage for the first analyzed 15-minute interval of each inhalation could be less than 100%, which occurred about 40 times. Within 1 day, the intervals after the inhalations were not distributed equally over the day.

Additional Analyses

In regression analyses, the change from baseline in traditional 6-minute walk distance was associated with changes in digital outcomes (distance walked, number of standing-up events, and digital 6-minute walk distance). Absolute measurements of traditional and digital 6-minute walk distance were strongly correlated but not congruent at baseline (r=0.86) and at the final visit (r=0.87). Further details of these and other additional analyses are presented in Multimedia Appendix 7.

Discussion

Feasibility of Digital Monitoring

To the best of our knowledge, this study is the first to monitor physical activity and heart rate using iPhones and Apple Watches in patients with pulmonary arterial hypertension. The results demonstrate the feasibility of using digital parameters as digital biomarkers for long-term activity levels in patients with pulmonary arterial hypertension. Despite the fact that patients were blinded to smartwatch-collected data, acceptance was good—almost 90% wore the smartwatch for at least 6 hours per day at baseline (16/18, 89%) and for the end of the observation period (16/18, 89%). Although correlations between changes in digital and traditional clinical outcome measures were few, the overall directions of those changes were generally concordant. The results also demonstrated the feasibility of continuously measuring multiple digital parameters (eg, inhalations, heart rate, and physical activity) in parallel with different devices and matching the data. We were able to detect short-term increases in physical activity after each iloprost inhalation session with fine granularity. The results also highlighted some important challenges and limitations that should be considered for future digital studies in patients with pulmonary arterial hypertension.

Recruitment was limited by a lack of available patients, which resulted in a small study population. This could be partly due to the rarity of pulmonary arterial hypertension [17], the willingness of patients to undergo digital monitoring, and the suitability of the study devices for the patients [14]. Nevertheless, data were fully available in the full analysis set for most outcome variables. Thus, reasonable precision was obtained for most correlation coefficients in the primary endpoint analysis, with the exceptions of BNP and NT-proBNP levels; as no central laboratory facility was used, clinicians generally measured either BNP or NT-proBNP (rather than both) based on local availability, which meant that few patients had data for each individual laboratory parameter.

Comparison Between Clinical and Digital Parameters

Although there was general agreement between changes in digital and traditional parameters from baseline, parallel digital and traditional measurements of 6-minute walk distance were not congruent. The proprietary digital 6-minute walk distance algorithm was based on step count and the step length of healthy individuals, which in general is not valid for patients with pulmonary arterial hypertension. Substitution of the healthy step length with a patient's individualized step length could provide better estimation of distance. The current series of Apple devices includes individual and repeated measurement of step length, which may be used in future studies to better estimate distance. This feature may also enable researchers to characterize other changes in health status (eg, dyspnea) more precisely. Alternatively, measurement of distance by GPS may be suitable if the 6-minute walk distance test can be performed outdoors. Wi-Fi could be a potential alternative to GPS for 6-minute walk distance tests conducted indoors. However, measurement of 6-minute walk distance by GPS or Wi-Fi would require precise location tracking. A feasibility survey suggested that precise tracking on a long-term basis would not be accepted by patients [14]; such tracking would, therefore, need to be restricted to during the 6-minute walk distance test (eg, by designing the 6-minute walk distance app to automatically switch GPS and Wi-Fi tracking on at the start of the 6-minute walk distance test and off at the end of the test).

Previous digital studies [18-22] with patients with pulmonary arterial hypertension have shown mean step counts in the range of 3234 to 5041 and mean traditional 6-minute walk distances in the range of 343 m to 458 m. In comparison, this study population had higher step counts (median 6721 at baseline and 8332 at the end of the observation period) but broadly similar traditional 6-minute walk distance (median 339 m at baseline and 366 m at the final visit). The reasons for this are unclear but may include methodological differences and the small sample size in our study. Patients in previous studies were monitored for only short periods of time (≥3 days [18], ≥3 weekdays [20], 6 days on average [21,22], or 7 days [19]), whereas patients in our study were monitored for 30 to 113 days (including 4-14 days in the baseline period and 8-14 days at the end of the observation period). Studies [23,24] with healthy adults have shown variations in physical activity over the course of a week, with the lowest levels of inactivity occurring on Saturdays. The results in patients with pulmonary arterial hypertension for shorter duration study periods may, therefore, have been influenced by the timing of physical activity monitoring. Mainguy et al [19] evaluated patients for 7 consecutive days but asked them to avoid performing unfamiliar activities during that time; this restriction was not used in our study.

The number of standing-up events is difficult to interpret. In a recent study with patients with pulmonary arterial hypertension, sit-to-stand test results (the number of sit-to-stand movements that patients were able to complete in 30 seconds) were associated with quadriceps muscle strength, which in turn was associated with 6-minute walk distance [25]. An official statement from the European Respiratory Society [26] noted that the sit-to-stand test depends substantially more on muscle

strength, equilibrium, and gait balance than on the mechanisms of oxygen and carbon dioxide transport that are the target of vasoactive pulmonary arterial hypertension treatments. During long-term monitoring, an increase in the number of standing-up events per day could reflect an increase in distance walked or an increase in sitting-down events (in the absence of an increase in activity, the latter would suggest a worsening of the patient's condition). The assessment of standing-up events in patients with pulmonary arterial hypertension is not yet well established, and more data are needed.

The physical activity level may also be influenced by factors other than the severity of pulmonary arterial hypertension (eg, patient preferences, patient behavior, or the season in which the patient was recruited). Nevertheless, daily activity has been linked to prognosis in patients with pulmonary hypertension [27].

The abovementioned limitations provide examples of some of the many criteria that must be fulfilled to ensure a successful digital study. The criteria can be broadly categorized into 6 themes, collectively referred to as the digital real-world evidence matrix: patient (eg, patient willingness to use digital devices and preferences regarding design and features of digital devices), indication (eg, number of available patients and suitability of disease-relevant outcome measures for digital monitoring), biomarker (eg, connection of digital measure to a health-related outcome), type of sensor or device (eg, ease of use, personalization, and data security), quality (eg, accuracy of primary data capture), and endpoint (eg, transformation of sensor data into a meaningful, patient-centered endpoint). These themes were addressed in the design of the VENTASTEP study, which combined digital and traditional monitoring to provide new insights into the real-life treatment of patients with pulmonary arterial hypertension.

Changes in Digital and Traditional Measures After Initiation of Inhaled Iloprost

There was an increase of 26 m for traditional investigator-measured 6-minute walk distance from baseline after addition of inhaled iloprost to oral pulmonary arterial hypertension therapy, which is consistent with the results of previous clinical trials [10,11]. The increase in traditional 6-minute walk distance was accompanied by changes in other traditional parameters and digital measures of daily physical activity.

6-Minute walk distance assessed during an office visit may be affected by daily fluctuations in physical function and may not accurately capture long-term changes in symptoms and physical function. Furthermore, the 6-minute walk distance test reflects maximal activity, whereas daily physical activity parameters provide an average of overall activity. Digital monitoring provides new insights into patient activity, with a more complete picture of the patient's health status than traditional 6-minute walk distance alone. For example, the digital data showed that physical activity increased shortly after each iloprost inhalation session. This sheds light on the day-to-day experiences of patients with pulmonary arterial hypertension taking inhaled iloprost and may be helpful information for patients or physicians considering iloprost therapy.

The increase in heart rate during the 6-minute walk distance test was marginally greater at the final visit compared with that the baseline visit, and heart rate recovery time after the 6-minute walk distance test increased. A previous study [28] showed significant associations between the chronotropic response (peak walking heart rate minus resting heart rate) and 6-minute walk distance in patients with pulmonary arterial hypertension, which led Provencher et al [28] to suggest that a lack of chronotropic response may reflect a loss of normal physiological reserve in patients who were more unwell. Preservation of the chronotropic response may therefore be a positive sign. However, given the high variability of heart rate in our study, changes to this parameter cannot be clearly interpreted.

Despite increases in daily physical activity over the course of the study, the EQ-5D weighted index showed little change. The EQ-5D score has produced mixed results in key clinical trials of pulmonary arterial hypertension therapies [9], showing improvements in placebo-controlled trials of sildenafil [29] and tadalafil [30] but not in a placebo-controlled trial of riociguat [31], despite riociguat resulting in patients showing significant improvements in traditional measures such as 6-minute walk distance and WHO functional class. The EQ-5D is a generic instrument [9] and may therefore lack specificity to address the important health-related quality of life issues in pulmonary arterial hypertension [8]. Health-related quality of life may also be influenced by associated comorbidities such as scleroderma, liver cirrhosis, and HIV [32].

Sleep disturbance in patients with pulmonary arterial hypertension has been associated with symptom severity, psychological state, and health-related quality of life [33]. PSQI score showed slight changes in sleep quality from baseline to the final visit. A global PSQI score of 5 has been proposed as a cut-off to distinguish good and poor sleep [34]. Thus, sleep quality in the study group as a whole shifted from poor to good during the course of the study.

Strengths and Limitations

The strengths of this study include its continuous, parallel, digital measurement of inhalation behavior, heart rate, and physical activity over a long observation period. The limitations of this study include its small sample size, insufficient BNP or NT-proBNP data, and the fact that the digital 6-minute walk distance algorithm was based on the step length of healthy individuals rather than that of patients with pulmonary arterial hypertension. A substantial proportion of the study population took concomitant beta-blockers, which should be considered when interpreting the heart rate data.

Implications for Future Research

Further research is warranted to assess the association of iloprost inhalation behavior with changes in digital physical activity measures, traditional clinical measures, and health-related quality of life. The relationship between daily physical activity and traditional 6-minute walk distance requires further exploration. In addition, studies with longer-term follow-up are needed to determine if clinical outcomes (eg, mortality or time to clinical worsening) are better predicted by the change in

traditional 6-minute walk distance or the change in digital 6-minute walk distance over time.

The recruitment difficulties in this study (which focused on a rare disease) suggest that digital monitoring studies may be more feasible in indications with large patient populations, such as heart failure. In addition, most patients were older adults; recruitment may be higher among populations of younger patients who are more familiar with digital technology. Refinement of the digital 6-minute walk distance algorithm for the target patient population is also needed.

Conclusions

Although generalizability is limited because of the small sample size and other limitations, the study demonstrated the general feasibility of performing digital assessments with a commercially available smartwatch and smartphone in patients with pulmonary arterial hypertension. Digital measures of daily physical activity and traditional clinical measures showed concordant changes from baseline after addition of inhaled iloprost to oral pulmonary arterial hypertension therapies. However, the data indicate that activity tracking algorithms validated for healthy people require adaptation for patients with pulmonary arterial hypertension. Further investigations are therefore necessary.

Acknowledgments

This study was funded by Bayer AG and was conducted in collaboration with the suppliers of the patient support program VENTAPLUS (ContraCare GmbH and Vitartis Medical-Services GmbH) and xbird GmbH, which created the study app (for analysis of daily physical activity and heart rate) and the 6-minute walk distance app. Medical writing assistance was provided by Dr. Claire Mulligan (Beacon Medical Communications Ltd) and was funded by Bayer AG.

Conflicts of Interest

BS, MK, VI, and CM are employees of Bayer Vital GmbH, which is the local representative of the marketing authorization holder (Bayer AG) for inhaled iloprost in Europe. DD has received support from Bayer Vital, Actelion, Novartis, and Boehringer. DD also received payments for lectures/consulting honorarium/travel support from Actelion, AstraZeneca, Bayer Healthcare, Boehringer Ingelheim, GlaxoSmithKline, Novartis, Med. Clinic III, UKSH Campus Lübeck and received compensation for participation in clinical trials from AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Merck, Novartis, Pfizer, Bayer HealthCare, BMS, MSD, and Novartis Roche. EG has received fees for lectures or consultations from Actelion, Bayer, GlaxoSmithKline, Janssen, Merck Sharp & Dohme, Pfizer, and United Therapeutics. FK is an employee of Bayer AG. HK has received support from Bayer Vital, Actelion, Novartis, Boehringer, Janssen, Merck Sharp & Dohme, GlaxoSmithKline, and United Therapeutics. MR has received support from United Therapeutics and Bayer Pharma AG, speaker fees from Actelion, Bayer Pharma AG, Mundipharma, Roche, and OMT, and consultancy fees from Bayer 202. MS has received support from Bayer Vital GmbH, GlaxoSmithKline, Johnson & Johnson, Merck Sharp & Dohme, and Pfizer. RE has received honoraria or fees for consultancy from Janssen Pharmaceutical Companies of Johnson & Johnson, Bayer Vital, and OMT and as a corporate board member for Actelion Pharmaceuticals Ltd, OMT, and Novartis. RE has received grants from Janssen Pharmaceutical Companies of Johnson & Johnson, Boehringer Ingelheim, OMT, Berlin Chemie, and Novartis.

Multimedia Appendix 1
Amendment to protocol.
[DOC File , 55 KB - jmir_v23i10e25163_app1.doc]

Multimedia Appendix 2
Additional information on digital parameters.
[DOC File , 50 KB - jmir_v23i10e25163_app2.doc]

Multimedia Appendix 3
Patient characteristics at baseline.
[DOC File , 81 KB - jmir_v23i10e25163_app3.doc]

Multimedia Appendix 4
Iloprost dose and inhalation characteristics.
[DOC File , 58 KB - jmir_v23i10e25163_app4.doc]

Multimedia Appendix 5
Change in heart rate from 1 minute before the 6MWD test to each minute during the 6MWD test (full analysis set).
[DOC File , 74 KB - jmir_v23i10e25163_app5.doc]

Multimedia Appendix 6
Average heart rate during iloprost inhalation per month of observation, stratified by median length of inhalation (full analysis set).
[DOC File , 74 KB - jmir_v23i10e25163_app6.doc]

Multimedia Appendix 7
Additional results – secondary objectives.
[DOC File , 94 KB - jmir_v23i10e25163_app7.doc]

References

1. Benatar D, Bondmass M, Ghitelman J, Avitall B. Outcomes of chronic heart failure. Arch Intern Med 2003 Mar 10;163(3):347-352. [doi: 10.1001/archinte.163.3.347] [Medline: 12578516]
2. Darkins A, Ryan P, Kobb R, Foster L, Edmonson E, Wakefield B, et al. Care Coordination/Home Telehealth: the systematic implementation of health informatics, home telehealth, and disease management to support the care of veteran patients with chronic conditions. Telemed J E Health 2008 Dec;14(10):1118-1126. [doi: 10.1089/tmj.2008.0021] [Medline: 19119835]
3. Inglis SC, Clark RA, Dierckx R, Prieto-Merino D, Cleland JGF. Structured telephone support or non-invasive telemonitoring for patients with heart failure. Cochrane Database Syst Rev 2015(10):CD007228. [doi: 10.1002/14651858.CD007228.pub3] [Medline: 26517969]
4. Bashi N, Karunanithi M, Fatehi F, Ding H, Walters D. Remote monitoring of patients with heart failure: an overview of systematic reviews. J Med Internet Res 2017 Jan 20;19(1):e18 [FREE Full text] [doi: 10.2196/jmir.6571] [Medline: 28108430]
5. Montani D, Günther S, Dorfmüller P, Perros F, Girerd B, Garcia G, et al. Pulmonary arterial hypertension. Orphanet J Rare Dis 2013 Jul 06;8:97 [FREE Full text] [doi: 10.1186/1750-1172-8-97] [Medline: 23829793]
6. Taichman DB, Shin J, Hud L, Archer-Chicko C, Kaplan S, Sager JS, et al. Health-related quality of life in patients with pulmonary arterial hypertension. Respir Res 2005 Aug 10;6:92 [FREE Full text] [doi: 10.1186/1465-9921-6-92] [Medline: 16092961]
7. Galiè N, Humbert M, Vachiery J, Gibbs S, Lang I, Torbicki A, ESC Scientific Document Group. 2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension: the Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS): Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). Eur Heart J 2016 Jan 01;37(1):67-119. [doi: 10.1093/eurheartj/ehv317] [Medline: 26320113]
8. Howard LS, Ferrari P, Mehta S. Physicians' and patients' expectations of therapies for pulmonary arterial hypertension: where do they meet? Eur Respir Rev 2014 Dec;23(134):458-468 [FREE Full text] [doi: 10.1183/09059180.00007514] [Medline: 25445944]
9. McGoon MD, Ferrari P, Armstrong I, Denis M, Howard LS, Lowe G, et al. The importance of patient perspectives in pulmonary hypertension. Eur Respir J 2019 Jan;53(1):1-15 [FREE Full text] [doi: 10.1183/13993003.01919-2018] [Medline: 30545977]
10. McLaughlin VV, Oudiz RJ, Frost A, Tapson VF, Murali S, Channick RN, et al. Randomized study of adding inhaled iloprost to existing bosentan in pulmonary arterial hypertension. Am J Respir Crit Care Med 2006 Dec 01;174(11):1257-1263. [doi: 10.1164/rccm.200603-358OC] [Medline: 16946127]
11. Saji T, Myoishi M, Sugimura K, Tahara N, Takeda Y, Fukuda K, et al. Efficacy and safety of inhaled Iloprost in Japanese patients with pulmonary arterial hypertension - insights from the IBUKI and AIR studies. Circ J 2016;80(4):835-842 [FREE Full text] [doi: 10.1253/circj.CJ-16-0097] [Medline: 27001191]
12. Olschewski H, Simonneau G, Galiè N, Higenbottam T, Naeije R, Rubin LJ, Aerosolized Iloprost Randomized Study Group. Inhaled iloprost for severe pulmonary hypertension. N Engl J Med 2002 Aug 01;347(5):322-329. [doi: 10.1056/NEJMoa020204] [Medline: 12151469]
13. Gessler T. lloprost delivered via the BREELIB nebulizer: a review of the clinical evidence for efficacy and safety. Ther Adv Respir Dis 2019;13:1753466619835497 [FREE Full text] [doi: 10.1177/1753466619835497] [Medline: 30874487]
14. Mueller C, Stollfuss B, Roitenberg A, Harder J, Richter MJ. Evaluation of clinical outcomes and simultaneous digital tracking of daily physical activity, heart rate, and inhalation behavior in patients with pulmonary arterial hypertension treated with inhaled iloprost: protocol for the observational VENTASTEP study. JMIR Res Protoc 2019 Apr 15;8(4):e12144 [FREE Full text] [doi: 10.2196/12144] [Medline: 30985279]
15. Gessler T, Ghofrani H, Held M, Klose H, Leuchte H, Olschewski H, et al. The safety and pharmacokinetics of rapid iloprost aerosol delivery via the BREELIB nebulizer in pulmonary arterial hypertension. Pulm Circ 2017;7(2):505-513 [FREE Full text] [doi: 10.1177/2045893217706691] [Medline: 28597762]
16. Richter MJ, Wan J, Ghofrani HA, Seeger W, Gall H, Rieth A, et al. Acute response to rapid iloprost inhalation using the Breelib™ nebulizer in pulmonary arterial hypertension: the Breelib™ acute study. Pulm Circ 2019;9(3):2045894019875342 [FREE Full text] [doi: 10.1177/2045894019875342] [Medline: 31588353]

17. Swinnen K, Quarck R, Godinas L, Belge C, Delcroix M. Learning from registries in pulmonary arterial hypertension: pitfalls and recommendations. Eur Respir Rev 2019 Dec 31;28(154):1-12 [FREE Full text] [doi: 10.1183/16000617.0050-2019] [Medline: 31852746]
18. Aslan GK, Akinci B, Yeldan I, Okumus G. Respiratory muscle strength in patients with pulmonary hypertension: the relationship with exercise capacity, physical activity level, and quality of life. Clin Respir J 2018 Feb;12(2):699-705. [doi: 10.1111/crj.12582] [Medline: 27860259]
19. Mainguy V, Provencher S, Maltais F, Malenfant S, Saey D. Assessment of daily life physical activities in pulmonary arterial hypertension. PLoS One 2011;6(11):e27993 [FREE Full text] [doi: 10.1371/journal.pone.0027993] [Medline: 22110770]
20. Okumus G, Aslan GK, Arseven O, Ongen G, Issever H, Kiyan E. The role of an activity monitor in the objective evaluation of patients with pulmonary hypertension. Clin Respir J 2018 Jan;12(1):119-125. [doi: 10.1111/crj.12495] [Medline: 27149246]
21. Panagiotou M, Johnson MK, Louvaris Z, Baker JS, Church AC, Peacock AJ, et al. A study of clinical and physiological relations of daily physical activity in precapillary pulmonary hypertension. J Appl Physiol (1985) 2017 Oct 01;123(4):851-859. [doi: 10.1152/japplphysiol.00986.2016] [Medline: 28663381]
22. Saxer S, Lichtblau M, Berlier C, Hasler ED, Schwarz EI, Ulrich S. Physical activity in incident patients with pulmonary arterial and chronic thromboembolic hypertension. Lung 2019 Oct;197(5):617-625. [doi: 10.1007/s00408-019-00248-x] [Medline: 31263960]
23. Matthews CE, Ainsworth BE, Thompson RW, Bassett DR. Sources of variance in daily physical activity levels as measured by an accelerometer. Med Sci Sports Exerc 2002 Aug;34(8):1376-1381. [Medline: 12165695]
24. Scheers T, Philippaerts R, Lefevre J. Variability in physical activity patterns as measured by the SenseWear Armband: how many days are needed? Eur J Appl Physiol 2012 May;112(5):1653-1662. [doi: 10.1007/s00421-011-2131-9] [Medline: 21874552]
25. Ozcan Kahraman B, Ozsoy I, Akdeniz B, Ozpelit E, Sevinc C, Acar S, et al. Test-retest reliability and validity of the timed up and go test and 30-second sit to stand test in patients with pulmonary hypertension. Int J Cardiol 2020 Apr 01;304:159-163. [doi: 10.1016/j.ijcard.2020.01.028] [Medline: 31980271]
26. Puente-Maestu L, Palange P, Casaburi R, Laveneziana P, Maltais F, Neder JA, et al. Use of exercise testing in the evaluation of interventional efficacy: an official ERS statement. Eur Respir J 2016 Feb;47(2):429-460 [FREE Full text] [doi: 10.1183/13993003.00745-2015] [Medline: 26797036]
27. Ulrich S, Fischler M, Speich R, Bloch KE. Wrist actigraphy predicts outcome in patients with pulmonary hypertension. Respiration 2013;86(1):45-51 [FREE Full text] [doi: 10.1159/000342351] [Medline: 23234873]
28. Provencher S, Chemla D, Hervé P, Sitbon O, Humbert M, Simonneau G. Heart rate responses during the 6-minute walk test in pulmonary arterial hypertension. Eur Respir J 2006 Jan;27(1):114-120 [FREE Full text] [doi: 10.1183/09031936.06.00042705] [Medline: 16387943]
29. Pepke-Zaba J, Gilbert C, Collings L, Brown MCJ. Sildenafil improves health-related quality of life in patients with pulmonary arterial hypertension. Chest 2008 Jan;133(1):183-189. [doi: 10.1378/chest.07-0592] [Medline: 18187744]
30. Pepke-Zaba J, Beardsworth A, Chan M, Angalakuditi M. Tadalafil therapy and health-related quality of life in pulmonary arterial hypertension. Curr Med Res Opin 2009 Oct;25(10):2479-2485. [doi: 10.1185/03007990903210066] [Medline: 19686085]
31. Ghofrani H, Galiè N, Grimminger F, Grünig E, Humbert M, Jing Z, PATENT-1 Study Group. Riociguat for the treatment of pulmonary arterial hypertension. N Engl J Med 2013 Jul 25;369(4):330-340. [doi: 10.1056/NEJMoa1209655] [Medline: 23883378]
32. Rubenfire M, Lippo G, Bodini BD, Blasi F, Allegra L, Bossone E. Evaluating health-related quality of life, work ability, and disability in pulmonary arterial hypertension: an unmet need. Chest 2009 Aug;136(2):597-603. [doi: 10.1378/chest.08-1260] [Medline: 19666759]
33. Matura LA, McDonough A, Hanlon AL, Carroll DL, Riegel B. Sleep disturbance, symptoms, psychological distress, and health-related quality of life in pulmonary arterial hypertension. Eur J Cardiovasc Nurs 2015 Oct;14(5):423-430. [doi: 10.1177/1474515114537951] [Medline: 24867878]
34. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]

Abbreviations

BNP: B-type natriuretic peptide
EQ-5D: EuroQol 5-dimension questionnaire
NT-proBNP: N-terminal pro–B-type natriuretic peptide
PSQI: Pittsburgh Sleep Quality Index
WHO: World Health Organization

Edited by G Eysenbach; submitted 04.11.20; peer-reviewed by JB Park; comments to author 25.11.20; revised version received 01.12.20; accepted 16.01.21; published 08.10.21.

Please cite as:
Stollfuss B, Richter M, Drömann D, Klose H, Schwaiblmair M, Gruenig E, Ewert R, Kirchner MC, Kleinjung F, Irrgang V, Mueller C
Digital Tracking of Physical Activity, Heart Rate, and Inhalation Behavior in Patients With Pulmonary Arterial Hypertension Treated With Inhaled Iloprost: Observational Study (VENTASTEP)
J Med Internet Res 2021;23(10):e25163
URL: https://www.jmir.org/2021/10/e25163
doi:10.2196/25163
PMID:34623313

Original Paper

Management and Treatment of Patients With Obstructive Sleep Apnea Using an Intelligent Monitoring System Based on Machine Learning Aiming to Improve Continuous Positive Airway Pressure Treatment Compliance: Randomized Controlled Trial

Cecilia Turino[1,2*], MD; Ivan D Benítez[1,2*], BSc; Xavier Rafael-Palou[3,4], PhD; Ana Mayoral[5], MD; Alejandro Lopera[5], MInf; Lydia Pascual[1,2], BSc; Rafaela Vaca[1,2], BSc; Anunciación Cortijo[1,2], MD; Anna Moncusí-Moix[1,2], MBioInf; Mireia Dalmases[1,2], MD; Eloisa Vargiu[3], PhD; Jordi Blanco[5], MBus; Ferran Barbé[1,2], MD; Jordi de Batlle[1,2], PhD

[1]Group of Translational Research in Respiratory Medicine, Institut de Recerca Biomèdica de Lleida (IRBLleida), Hospital Universitari Arnau de Vilanova and Santa Maria, Lleida, Spain

[2]Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES), Madrid, Spain

[3]eHealth Unit, Eurecat Centre Tecnològic de Catalunya, Barcelona, Spain

[4]Barcelona Centre for New Medical Technologies (BCN Medtech), Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain

[5]Oxigen salud, Barcelona, Spain

[*]these authors contributed equally

Corresponding Author:
Jordi de Batlle, PhD
Group of Translational Research in Respiratory Medicine
Institut de Recerca Biomèdica de Lleida (IRBLleida)
Hospital Universitari Arnau de Vilanova and Santa Maria
Rovira Roure 80
Lleida, 25198
Spain
Phone: 34 645624734
Email: jordidebatlle@gmail.com

Abstract

Background: Continuous positive airway pressure (CPAP) is an effective treatment for obstructive sleep apnea (OSA), but treatment compliance is often unsatisfactory.

Objective: The aim of this study was to assess the effectiveness and cost-effectiveness of an intelligent monitoring system for improving CPAP compliance.

Methods: This is a prospective, open label, parallel, randomized controlled trial including 60 newly diagnosed patients with OSA requiring CPAP (Apnea–Hypopnea Index [AHI] >15) from Lleida, Spain. Participants were randomized (1:1) to standard management or the MiSAOS intelligent monitoring system, involving (1) early compliance detection, thus providing measures of patient's CPAP compliance from the very first days of usage; (2) machine learning–based prediction of midterm future CPAP compliance; and (3) rule-based recommendations for the patient (app) and care team. Clinical and anthropometric variables, daytime sleepiness, and quality of life were recorded at baseline and after 6 months, together with patient's compliance, satisfaction, and health care costs.

Results: Randomized patients had a mean age of 57 (SD 11) years, mean AHI of 50 (SD 27), and 13% (8/60) were women. Patients in the intervention arm had a mean (95% CI) of 1.14 (0.04-2.23) hours/day higher adjusted CPAP compliance than controls (P=.047). Patients' satisfaction was excellent in both arms, and up to 88% (15/17) of intervention patients reported willingness to keep using the MiSAOS app in the future. No significant differences were found in costs (control: mean €90.2 (SD 53.14) (US $105.76 [SD 62.31]); intervention: mean €96.2 (SD 62.13) (US $112.70 [SD 72.85]); P=.70; €1=US $1.17 was considered throughout). Overall costs combined with results on compliance demonstrated cost-effectiveness in a bootstrap-based simulation analysis.

Conclusions: A machine learning–based intelligent monitoring system increased daily compliance, reported excellent patient satisfaction similar to that reported in usual care, and did not incur in a substantial increase in costs, thus proving cost-effectiveness. This study supports the implementation of intelligent eHealth frameworks for the management of patients with CPAP-treated OSA and confirms the value of patients' empowerment in the management of chronic diseases.

Trial Registration: ClinicalTrials.gov NCT03116958; https://clinicaltrials.gov/ct2/show/NCT03116958

(J Med Internet Res 2021;23(10):e24072) doi:10.2196/24072

KEYWORDS

obstructive sleep apnea; continuous positive airway pressure; patient compliance; remote monitoring; machine learning

Introduction

Obstructive sleep apnea (OSA) is the most prevalent sleep-disordered breathing condition, affecting 15%-30% of adults in Western countries [1]. It is characterized by repetitive episodes of airways collapse during sleep, causing sleep fragmentation, intermittent hypoxia, and daytime somnolence. OSA has been associated with increased morbidity and mortality, and has an impact on quality of life [2]. In this sense, increased inflammation, oxidative stress, sympathetic activation, and hypercoagulability are the main mechanisms associating OSA with hypertension; cancer; and cardiovascular, cerebrovascular, and metabolic diseases [2].

Nocturnal continuous positive airway pressure (CPAP), preventing upper airway collapse during sleep, is the treatment of choice for patients with symptomatic OSA [3]. A satisfactory CPAP compliance (≥4 hours/day) improves daytime sleepiness and overall quality of life; reduces OSA severity markers, such as the Apnea–Hypopnea Index (AHI); moderately decreases arterial blood pressure (BP), mainly in patients with resistant hypertension [3,4]; and contributes to preventing the onset of newly diagnosed hypertension [5]. Compliance is, therefore, essential for the efficacy of CPAP treatment and its optimization is an important aspect of patient management. However, up to one-third of patients underuse or even discontinue CPAP [6-8], mostly because of treatment-related side effects such as machine noise, pressure intolerance, mask displacement, or claustrophobia [9]. In this sense, issues hampering CPAP compliance during the first months of treatment are likely to have a significant impact on long-term CPAP compliance [10]. Therefore, there is a need to implement effective strategies for the promotion of CPAP compliance, especially during the first months of treatment.

So far, interventions tackling CPAP compliance, including novel educational and supportive or therapeutic strategies, have reported low to moderate evidence of success [11,12]. By contrast, when these strategies are wrapped up in comprehensive packages making use of information and communication technologies (eHealth) and targeting the initial months after CPAP prescription, the potential for success can be significantly enhanced [12-14]. In this scenario, and within the frame of the MiSAOS project, an internet of things–based intelligent monitoring system relying on machine learning [15] was developed in Catalonia, Spain, with a fourfold goal: (1) predicting patient's potential early CPAP compliance; (2) providing real-time monitoring of patient's CPAP compliance, informing both the patient and the care team, and granting decision support; (3) empowering the patient by means of feedback and recommendations; and (4) reducing patient's overall management costs. This study compares, in terms of effectiveness and cost-effectiveness, the MiSAOS intelligent monitoring system model, based on early compliance detection, compliance prediction, and rule-based recommendations, with the usual care provided to patients using CPAP in the region of Lleida, Catalonia.

Methods

Study Design

This is a prospective, open label, parallel, randomized controlled trial comparing the MiSAOS management model with care as usual for a duration of 6 months after CPAP prescription (ClinicalTrials.gov NCT03116958). The study was conducted from November 2016 to December 2017 in Lleida, Catalonia. Multimedia Appendix 1 shows the CONSORT checklist of this study.

Target Population

Eligible population included patients with OSA (AHI ≥15) being newly diagnosed in the sleep unit of University Hospital Santa Maria, Lleida, and requiring CPAP treatment according to the Spanish Respiratory Society (SEPAR) guidelines [16]. The specific eligibility criteria were aged over 18 years; having a sufficient competence in the use of smartphones; not having been previously treated with CPAP; not having impaired lung function (overlap syndrome, obesity hypoventilation syndrome, and restrictive disorders), severe heart failure, severe chronic pathologies, psychiatric disorders, or periodic leg movements or other dyssomnias or parasomnias; and not being pregnant.

Sample Size

Accepting an α risk of .05 and a β risk of .2 in a 2-sided test, 29 patients per study arm were needed to recognize as statistically significant a difference in compliance greater than or equal to 1 hour/day. The common SD was assumed to be 1.35, based on previous research of the group.

Recruitment, Randomization, and Intervention

Patients were recruited in the sleep unit and randomized (1:1) to receive 6 months of either MiSAOS or usual care management. Patients in the usual care arm were managed according to the SEPAR guidelines [16]. Randomization was based on a permuted-block design with a computer random number generator and a fixed block size of 4. Patients were fitted with a mask and given a CPAP device (AirSense 10;

ResMed) and a leaflet explaining how to use it. A short training session on how to use the CPAP device was also given to patients and partners in the sleep unit by a trained nurse with experience in the follow-up of CPAP-treated patients. This included a practical demonstration of how to put on the mask, and the correct management and cleaning of the tubes, mask, and humidifier. Information on how to turn the CPAP device on and off was provided by the homecare provider at the time of machine delivery. According to SEPAR recommendations, patients were visited after 1 month of treatment by the specialist nurse at the sleep unit. Information about CPAP, compliance (use of CPAP for ≥4 hours/day), residual respiratory events, and leaks was downloaded from the device. CPAP-related side effects, CPAP machine care and maintenance actions (ie, changes of mask), and the number of required additional visits or calls were recorded by the nurse.

Similarly, patients in the MiSAOS arm were fitted with a mask, a CPAP device (AirSense 10; ResMed), and given a leaflet explaining its use. Patients received the same training sessions from the same personnel as in the usual care arm. However, these patients' CPAP devices were equipped with mobile 2G (global systems for mobile/general packet radio service [GSM/GPRS]) technology capable of sending daily information on CPAP compliance, CPAPs, mask leaks, and residual respiratory events to the MiSAOS–Oxigen salud web database. In addition, patients in the MiSAOS arm had access to an integrated platform including a website [17] and a mobile app (MiSAOS; available for both Android and iOS), benefitting from continuous monitoring and personalized feedback. Sample screenshots of the MiSAOS app showing its main functionalities and features are presented in Multimedia Appendix 2. Similarly, sample screenshots of the MiSAOS website showing some of its functionalities and features are presented in Multimedia Appendix 3. Hospital lung specialists managing these patients and the CPAP provider (Oxigen salud) also had access to the MiSAOS website that provided relevant information and decision support according to the specific role and access rights of each professional user. Finally, the cloud-based MiSAOS platform connected all the devices for data exchange and hosted an intelligent monitoring system, based on machine learning, capable of predicting the expected compliance with the therapy by a given patient, thus providing adequate feedback and proposing personalized interventions to increase compliance [15,18]. Predictions of patient's midterm compliance were based on patient's characteristics, such as anthropometric data and clinical information, and early compliance data. Examples of the needed information and provided outcome can be found in Multimedia Appendix 3. Based on these predictions, patients were classified into 2 groups: low compliance and medium/high compliance, and recommendations were provided based on these classes. In brief, recommendations included warnings and exhortation to do better in case of low compliance, or positive reinforcement messages in case of good compliance, highlighting the key areas to be improved regardless of the compliance. This platform was also used for the monitoring of patient compliance, prompting actions when compliance was too low.

Data Collection

Baseline information was collected by sleep unit personnel during recruitment, regardless of the study arm. This included age; gender; socioeconomic level; Epworth Sleepiness Scale (ESS) score; EuroQoL-5D quality of life (EQ-5D); lifestyle habits (tobacco and alcohol consumption); comorbidities; use of medications; weight; height; BMI; neck, waist, and hip circumference; and BP. Variables of the sleep study were also recorded and included registration time, sleep duration, AHI, and percentage of nighttime spent with an oxygen saturation less than 90%.

At 3 and 6 months all patients, regardless of the study arm, were visited at the sleep unit. Patients were checked about progress and compliance with therapy and any problems with their CPAP machine. During these visits we collected data on treatment compliance (number of hours/day), ESS score, OSA-related symptoms, EQ-5D, BP, and anthropometric variables. Additionally, data on CPAP, residual respiratory events and leaks, CPAP-related side effects (mask allergies and skin irritations, dry mouth, congestion, runny nose, sneezing, sinusitis, nosebleeds, and discomfort), overall satisfaction with the therapy (questionnaire), CPAP machine care and maintenance actions (ie, changes of mask), and the number of any additional visits and calls required by the patient during the follow-up were collected. Finally, costs for each component, use of services, and visits were computed based on standard prices of the CPAP provider and on Catalan Health Department official data (CVE-DOGC-A-13051031-2013) [19]. Only direct costs were considered.

Statistical Analyses

A *t* test, or an equivalent nonparametric test, or chi-square test was used for baseline bivariate analyses, depending on variables' characteristics. Differences in the primary and secondary outcomes between the intervention and control groups at 6 months were assessed using ordinary least-squares linear models. All models were adjusted by age, and models for secondary outcomes were further adjusted by the baseline values. A 2-sided *P* value and 95% CI were used. The cost-effectiveness analysis was performed using the total costs for each arm based on intervention effectiveness (CPAP treatment compliance). A probabilistic sensitivity analysis was performed using the bootstrap method, which was represented in a cost-effectiveness plan.

The primary and secondary analyses were performed on both the intention-to-treat (ITT) and per-protocol (PP) samples. The ITT sample included all the patients who were randomized. The PP sample excluded the patients who were lost during the follow-up period. Missing data were imputed using multiple imputation consisting of chained equations, for which 10 complete databases were checked. The R package "mice" was used for these calculations. All statistical analyses and data processing procedures were performed using R software, version 3.4.4 (The R Foundation).

Ethical Considerations

This study was approved by the Ethics Committee of Hospital Arnau de Vilanova (CEIC-1283) and all patients provided

written informed consent. This project was registered in ClinicalTrials.gov (registration number NCT03116958).

Results

A total of 60 patients were randomized to receive either MiSAOS (intervention; n=30) or usual care (control; n=30) management, and up to 53 patients completed the study (Figure 1). Patients' baseline characteristics in both study arms are presented in Table 1, which show that only age was statistically different among groups.

Table 2 shows the primary and secondary outcomes of the trial according to an ITT analysis. After 6 months, the mean (95% CI) CPAP compliance was 4.89 (4.05-5.72) hours/day in the control group and 5.79 (5.20-6.38) hours/day in the intervention group, with an adjusted difference of 1.14 (0.04-2.23; *P*=.047) hours/day in benefit of intervention. Furthermore, the intervention arm had a higher proportion of patients with good compliance (use of CPAP for ≥4 hours/day) than the control arm (88.5% [23/26] vs 70.4% [19/27], respectively; *P*=.20), although this did not achieve statistical significance. Regarding secondary outcomes, ESS, BP, and the EQ-5D visual analog scale scores improved after 6 months of CPAP treatment in both arms, although the change in systolic BP was significantly higher in patients in the control arm than in those in the intervention arm (adjusted *P*=.04). Results on a PP approach were similar and are shown in Multimedia Appendix 4.

Figure 1. Study flowchart. CPAP: continuous positive airway pressure, ITT: intention-to-treat

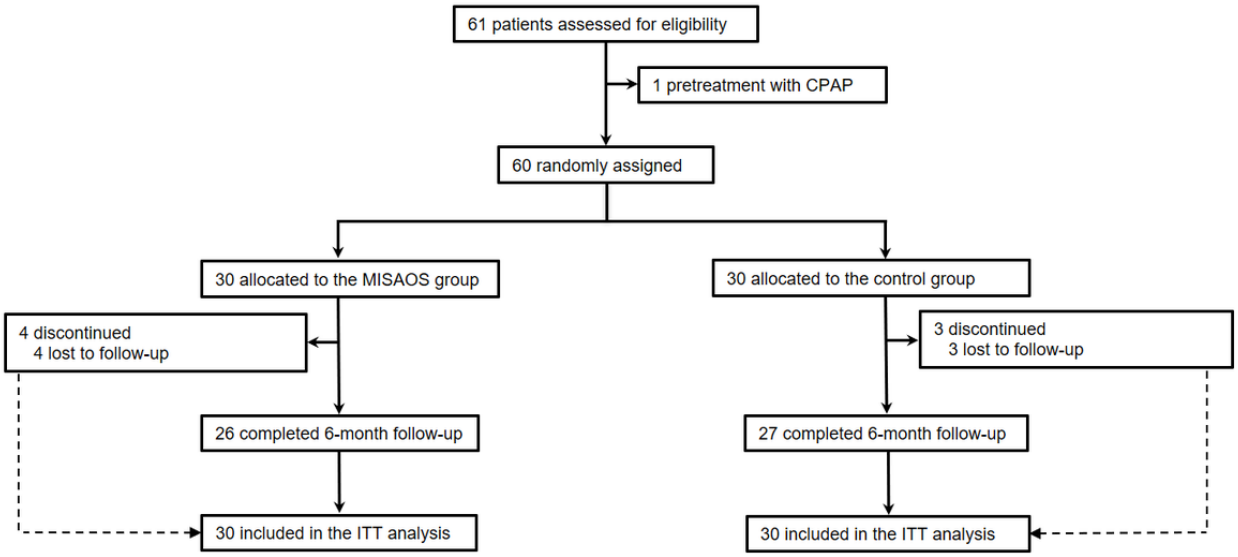


Table 1. Patients' baseline characteristics (N=60).[a]

Characteristic	Control (n=30)	Intervention (n=30)	*P* value
Gender (male), n (%)	26 (87)	26 (87)	.99
Age (years), mean (SD)	58 (10)	52 (12)	.04
Weight (kg), mean (SD)	97 (19)	101 (23)	.42
BMI (kg/m^2), mean (SD)	33.1 (6.4)	34.7 (7.3)	.38
Systolic blood pressure (mmHg), mean (SD)	138 (17)	142 (20)	.42
Diastolic blood pressure (mmHg), median (range)	87 (79-96)	88 (81-95)	.74
Apnea–Hypopnea Index (events/hour)	39 (25-71)	53 (35-65)	.22

[a]Data as per *t* test, or an equivalent nonparametric test, or chi-square test depending on variables' characteristics.

Table 2. Differences in primary and secondary outcomes of the trial according to an intention-to-treat analysis (N=60).[a]

Differences	Control (n=30), mean (SD)	Intervention (n=30), mean (SD)	Difference, mean (95% CI)
Primary outcome			
Compliance (hours/day)	4.89 (2.30)	5.79 (1.60)	
Crude difference			0.90 (–0.16 to 1.96)
Adjusted difference			1.14 (0.04 to 2.23)
Secondary outcomes			
ESS[b] score (0-24)			
Baseline	10.9 (5.35)	11.1 (5.35)	
6 months	4.90 (2.41)	5.85 (3.91)	
Change	–5.98 (4.42)	–5.22 (4.78)	
Crude difference			0.76 (–1.64 to 3.16)
Adjusted difference			1.05 (–0.51 to 2.61)
Weight (kg)			
Baseline	97.0 (18.6)	101 (22.5)	
6 months	98.2 (20.2)	100 (20.7)	
Change	1.26 (7.86)	–0.95 (7.91)	
Crude difference			–2.21 (–6.98 to 2.56)
Adjusted difference			–2.55 (–7.41 to 2.32)
BMI (kg/m^2)			
Baseline	33.3 (6.20)	34.7 (7.17)	
6 months	34.2 (6.80)	34.8 (6.32)	
Change	0.98 (3.26)	0.14 (3.16)	
Crude difference			–0.84 (–2.95 to 1.27)
Adjusted difference			–0.82 (–2.97 to 1.32)
Systolic blood pressure (mmHg)			
Baseline	138 (17.0)	142 (19.4)	
6 months	131 (12.7)	138 (17.2)	
Change	–7.02 (15.2)	–3.80 (12.7)	
Crude difference			3.22 (–5.03 to 11.47)
Adjusted difference			7.81 (0.57 to 15.05)
Diastolic blood pressure (mmHg)			
Baseline	87.7 (13.5)	90.3 (12.5)	
6 months	81.6 (8.84)	86.8 (9.23)	
Change	–6.13 (11.4)	–3.52 (10.6)	
Crude difference			2.61 (–4.04 to 9.27)
Adjusted difference			4.52 (–0.65 to 9.69)
EQ-5D[c] HUI[d] (0-1)			
Baseline	0.84 (0.22)	0.85 (0.17)	
6 months	0.80 (0.19)	0.86 (0.20)	
Change	–0.04 (0.17)	0.00 (0.18)	
Crude difference			0.05 (–0.05 to 0.15)
Adjusted difference			0.03 (–0.06 to 0.13)

Differences	Control (n=30), mean (SD)	Intervention (n=30), mean (SD)	Difference, mean (95% CI)
EQ-5D VAS[e] (0-10)			
Baseline	4.93 (3.41)	4.63 (3.55)	
6 months	7.35 (1.71)	8.03 (1.32)	
Change	2.42 (2.87)	3.40 (3.65)	
Crude difference			0.98 (–0.72 to 2.69)
Adjusted difference			0.51 (–0.3 to 1.33)

[a]Ordinary least-squares linear models adjusted by age and baseline value.

[b]ESS: Epworth Sleepiness Scale.

[c]EQ-5D: EuroQoL-5D quality of life.

[d]HUI: health utility index

[e]VAS: visual analog scale.

Patients' satisfaction with the management of their illness was excellent in both study groups (Table 3). Similarly, no differences were found in the occurrence of CPAP-related side effects. Finally, all patients in the intervention group reported the MiSAOS app to be useful; 94% (16/17) of patients reported that it was easy to use, and 88% (15/17) of patients reported the willingness to continue to use it in the future (Table 3).

Table 3. Overall patients' satisfaction and satisfaction with MiSAOS (N=45).[a]

Users' satisfaction	Control (n=26)	Intervention (n=19)	*P* value
Overall satisfaction			
The follow-up I received was sufficient to manage my health and medical needs			
Agrees/strongly agrees, n (%)	26 (100)	19 (100)	
Overall score (1-7), mean (SD)	6.38 (0.80)	6.53 (0.61)	.51
In general I am satisfied with the management of my illness			
Agrees/strongly agrees, n (%)	26 (100)	19 (100)	
Overall score (1-7), mean (SD)	6.62 (0.57)	6.53 (0.61)	.62
My contact with the hospital was sufficient			
Agrees/strongly agrees, n (%)	26 (100)	18 (95)	
Overall score (1-7), mean (SD)	6.62 (0.64)	6.21 (1.62)	.31
Satisfaction with MiSAOS (intervention only) (n=17)			
The app was useful			
Agrees/strongly agrees, n (%)		17 (100)	
Overall score (1-7), mean (SD)		6.53 (0.72)	
The app was easy to use			
Agrees/strongly agrees, n (%)		16 (94)	
Overall score (1-7), mean (SD)		6.41 (0.87)	
I would like to use the app every day in the future			
Agrees/strongly agrees, n (%)		15 (88)	
Overall score (1-7), mean (SD)		6.35 (1.17)	

[a]The overall satisfaction questionnaire was answered by 26 controls and 19 intervention participants. The satisfaction with MiSAOS questionnaire was answered by 17 participants.

Table 4 shows the costs of the intervention, costs of contacts with the CPAP provider and health system (not including the baseline, 3-month, and 6-month visits, as all patients regardless of study arm did them), and costs of any CPAP machine care and maintenance intervention actions (ie, changes of mask) during the 6 months of follow-up. The overall mean cost per patient was €90.2 (SD 53.14) (US $105.76 [SD 62.31]) in the control group and €96.2 (SD 62.13) (US $112.70 [SD 72.85]) in the intervention group, resulting in a nonsignificant cost difference between arms (*P*=.70). The main differences between arms were €49.5 (US $58.04) of the intervention costs (2G [GSM/GPRS] daily data transfer and activation and maintenance

costs) in the MiSAOS arm, and the €41 (US $48.07) on sleep unit visits in the usual care arm. This overall cost, combined with the results on CPAP treatment compliance (primary outcome), demonstrated cost-effectiveness in a bootstrap-based simulation analysis (Figure 2).

Table 4. Within-trial intervention and follow-up costs (average cost per randomized patient; N=60).

Concept	Control (n=30), €[a]/patient, mean (SD)	Intervention (n=30), €/patient, mean (SD)	Difference, mean (95% CI)
Intervention costs[b]			
2G (GSM[c]/GPRS[d]) daily data transfer	0 (0)	41.5 (0)	–41.5 (—)
Activation and maintenance	0 (0)	8 (0)	–8 (—)
Follow-up costs			
Sleep unit visits and consultations[e]	41 (0)	0 (0)	41 (—)
CPAP[f] provider visits and consultations[g]	9.7 (8.9)	10.0 (10.8)	–0.33 (–5.5 to 4.8)
Changes in CPAP device components[h]			
ResMed Mirage Quattro	12.5 (28.4)	10.0 (32.6)	2.5 (–13.3 to 18.3)
ResMed Mirage FX	0.8 (4.4)	3.2 (8.3)	–2.4 (–5.9 to 1.1)
ResMed Mirage Micro	0 (0)	1.1 (4.1)	–1.07 (–2.6 to 0.4)
ResMed Swift FX	1.5 (8.2)	4.5 (13.7)	–3 (–8.9 to 2.9)
ResMed Airfit P10 (without head-gear)	0 (0)	1.3 (7.3)	–1.33 (–4.1 to 1.4)
Philips Respironics Nuance gel	0 (0)	2.45 (13.4)	–2.45 (–7.5 to 2.6)
ResMed Airfit F10	15.0 (30.5)	10.0 (25.9)	5 (–9.6 to 19.6)
ResMed Airfit P10 (with head-gear)	1.50 (8.22)	0 (0)	1.5 (–1.6 to 4.6)
SleepNet IQ	2.20 (8.86)	0 (0)	2.2 (–1.1 to 5.5)
SleepNet Ascend	0 (0)	0.8 (4.4)	–0.8 (–2.4 to 0.8)
Philips Respironics Amara View	5.00 (15.3)	3.3 (18.3)	1.66 (–7.0 to 10.4)
Philips Respironics Comfort Gel Blue	1.00 (5.48)	0 (0)	1 (–1.0 to 3.0)
Total	90.2 (53.1)	96.2 (62.1)	–6.0 (–35.9 to 23.9)

[a]€1 = US $1.17.

[b]Estimated costs supplied by the CPAP provider: 2G (GSM/GPRS) daily data transfer (€83 [US $97.32]/year); activation and maintenance (€16 [US $18.76]/year).

[c]GSM: global systems for mobile.

[d]GPRS: general packet radio service.

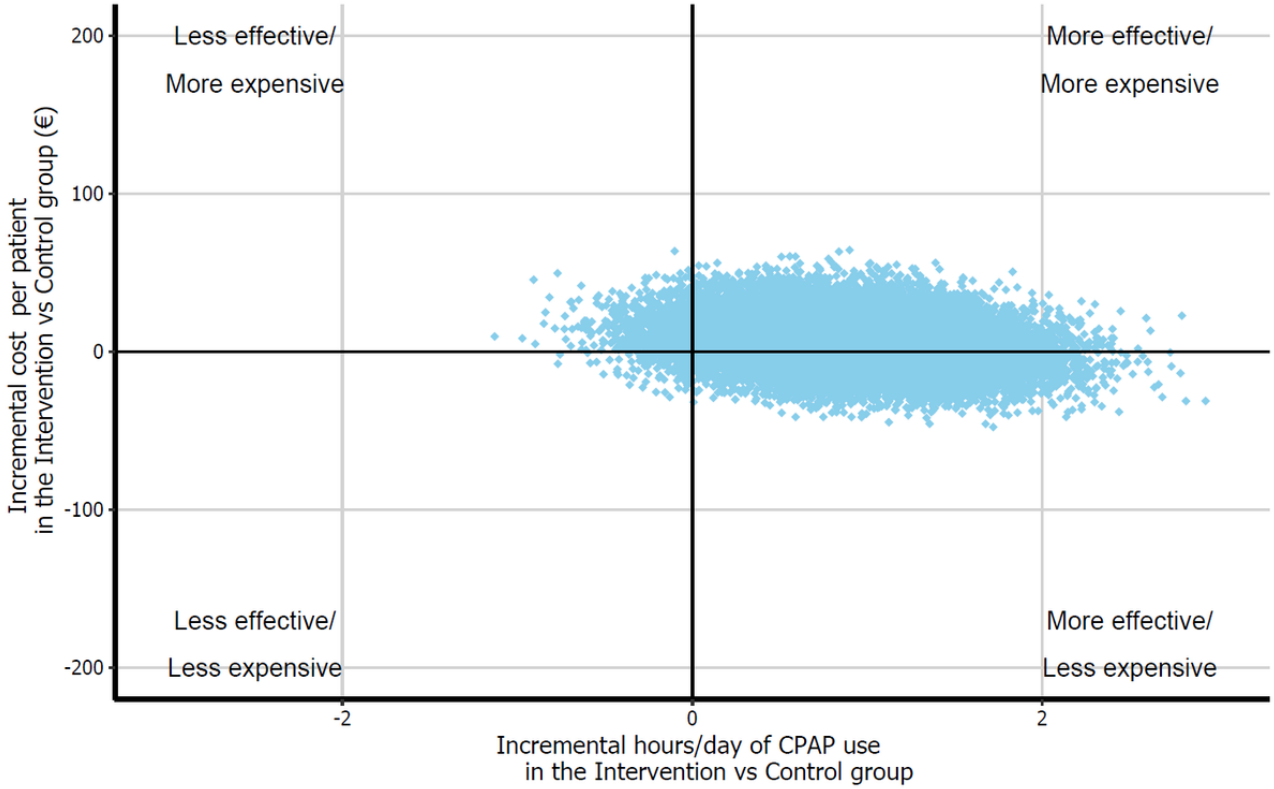
[e]Not including the baseline, 3-month, and 6-month visits, as all patients did them regardless of study arm. Costs based on the Catalan Institute of Health (CVE-DOGC-A-13051031-2013): sleep unit visits and consultations (€41 [US $48.07]/contact).

[f]CPAP: continuous positive airway pressure.

[g]Commercial costs supplied by the CPAP provider: CPAP provider visits and consultations (€10 [US $11.73]/contact).

[h]Commercial costs supplied by the CPAP provider: ResMed Mirage Quattro (€75 [US $87.94]/unit); ResMed Mirage FX (€24 [US $28.14]/unit); ResMed Mirage Micro (€16 [US $18.76]/unit); ResMed Swift FX (€45 [US $52.76]/unit); ResMed Airfit P10 (without headgear) (€40 [US $46.90]/unit); Philips Respironics Nuance gel (€73.5 [US $86.18]/unit); ResMed Airfit F10 (€75 [US $87.94]/unit); ResMed Airfit P10 (with head-gear) (€45 [US $52.76]/unit); SleepNet IQ (€22 [US $25.80]/unit); SleepNet Ascend (€24 [US $28.14]/unit); Philips Respironics Amara View (€50 [US $58.63]/unit); Philips Respironics Comfort Gel Blue (€30 [US $35.18]/unit).

Figure 2. Cost-effectiveness analysis based on treatment compliance (CPAP hours/day) and total costs for each arm, performed using a bootstrap probabilistic sensitivity analysis. CPAP: continuous positive airway pressure.



Discussion

Principal Findings

This study is the first randomized controlled clinical trial assessing the effectiveness and cost-effectiveness of a machine learning–based intelligent monitoring system aiming to improve CPAP compliance in patients with OSA. The MiSAOS intelligent monitoring system, based on early compliance detection, compliance prediction, and rule-based recommendations, was compared with usual care in the region of Lleida, showing a mean increase of 1.14 hours in daily compliance with no substantial differences in direct costs and an excellent patient satisfaction. This novel management system proved to be cost-effective and thus a viable option for the management of patients with OSA treated with CPAP.

Strengths and Limitations

This study has several strengths, including the (1) use of the same CPAP devices in both study arms; (2) use of an intelligent monitoring system model, based on early compliance detection, machine learning–based compliance prediction, and rule-based recommendations; (3) inclusion of continuous patient feedback through an app; (4) measurement of a broad range of effect measures (ie, compliance, changes in symptoms, and changes in quality of life); (5) assessment of patient comfort and satisfaction; and (6) inclusion of cost and cost-effectiveness analyses. Nevertheless, there are also some limitations to be acknowledged: (1) the slight infraestimation of the required number of study participants limited the statistical power of some of the between-arm comparisons, although this did not affect the results on the primary outcome and cost-effectiveness analysis; (2) the assessment of patient satisfaction was performed using a nonvalidated questionnaire; (3) the exclusion of patients with severe chronic pathologies and other dyssomnias or parasomnias could limit the generalizability of our results, although the included patients would be the ideal target for eHealth interventions as more complex patients could require a close follow-up in the sleep units; (4) the results of cost analyses are highly dependent on the characteristics of the health care setting in which they are conducted and, thus, extrapolation of the results to different settings should be done cautiously; and (5) the follow-up period does not allow the extrapolation of results to the long term.

Comparison With Existing Literature

Patients experiencing the MiSAOS intelligent monitoring system showed a mean increase of 1.14 hours in daily CPAP compliance when compared with patients in usual care. This result is more positive than the mean (95% CI) increase of 0.54 (0.29-0.79) hours reported by Aardoom et al [14] in a 2020 meta-analysis including 18 studies with eHealth interventions. Other recent studies exploring advanced monitoring systems have shown similarly inferior results, for instance, Pépin et al [20] reported a 0.53 hours' increase in CPAP compliance in patients with OSA with high cardiovascular risk using a multimodal telemonitoring intervention. Interestingly, granting patients an easy access to their compliance data has shown successful results in terms of increasing patient compliance [21-23]. Therefore, the combination of an intelligent machine learning–based monitoring system with the empowerment of patients, based on access to daily compliance and personalized feedback through an app, could represent a significant step forward in the promotion of CPAP compliance.

The impact of the CPAP treatment on secondary outcomes in the MiSAOS intervention was very similar to that achieved in usual care and reports in previous literature [12,20]. Sleepiness, overall quality of life, and BP improvements after 6 months of follow-up were similar in both study arms. The only difference between study arms was the significantly higher decrease in systolic BP in patients having usual care, which could be easily explained by the baseline differences in BP between study arms.

Patient's comfort and satisfaction are key drivers of compliance with CPAP treatment in the long term [9]. On the one hand, regarding comfort, the number of side effects in both arms was very similar. On the other hand, it must be noticed that in telemedicine interventions patients' satisfaction is usually similar to or lower than that of usual care [12], with privacy concerns being the main reported issue [24,25]. In this trial, all patients reported excellent satisfaction with their management, regardless of the study arm. Moreover, patients in the intervention group considered the MiSAOS app as useful and easy to use and reported their willingness to keep using it in the future. These results are better than those obtained in telemonitoring interventions in the same setting (Lleida, Spain) but not providing any direct feedback to the patients [25] and confirm that patient empowerment has a direct impact on patient satisfaction. Finally, potential issues on data privacy had no impact on the current trial results, in contrast to previous research [25], and in line with other interventions providing the patients with regular feedback on compliance [21].

A key aspect of any new management strategy is the cost of the intervention and its cost-effectiveness. In this study, the analysis of costs and cost-effectiveness showed that the MiSAOS intervention had an overall cost similar to that of usual care while providing better results in terms of treatment compliance, thus demonstrating cost-effectiveness. This result is in contrast to previous cost-effectiveness trials of telemonitoring interventions for CPAP-treated patients in Spain, where cost-effectiveness was demonstrated because of an overall reduction in costs and no significant differences in effectiveness were found [25,26]. Similarly, telemedicine platforms with automated functions to provide education or accountability have already shown cost-effectiveness in sleep medicine [23]. This suggests that the addition of machine learning data–processing functionalities together with the empowerment of patients by means of direct feedback could tip the scales toward significant increases in compliance and boost the cost-effectiveness of already existing telemonitoring interventions. Moreover, it should be noticed that a key factor of telemonitoring is the reduction in the number or duration of follow-up visits, which was quantified by Anttalainen et al [27], reporting a saving of 19 minutes in nursing time when comparing telemonitoring with usual care in the habituation phase of CPAP treatment (4 weeks) [27], and should be sufficient to mitigate the costs of telemonitoring.

Implications for Future Clinical Practice

As previously stated, the main barriers for the large-scale implementation of a novel management intervention are its costs and cost-effectiveness. In the optimal scenario, a novel management strategy should be cheaper than usual care while providing better results. The MiSAOS model has shown the potential to generate better results than usual care in terms of compliance. However, it was not cheaper than usual care. It is worth mentioning that a big proportion of the intervention's cost corresponded to the use of a 2G (GSM/GPRS) system for daily CPAP compliance data transfer. This technology could be easily replaced by a secure wireless connection to the patients' home Wi-Fi network, which would represent a huge saving and further boost cost-effectiveness. Even in rural areas such as Lleida, this scenario is rapidly becoming a reality and most homes have a suitable Wi-Fi network.

Conclusion

The use of a machine learning–based intelligent monitoring system increased daily compliance, reported an excellent patient satisfaction similar to that reported in usual care, and did not incur in a substantial increase in costs, thus proving cost-effectiveness. This study supports the implementation of intelligent eHealth frameworks for the management of patients with OSA treated by CPAP and confirms the value of patients' empowerment in the management of chronic diseases.

Acknowledgments

This work is part of the myOSA project (RTC-2014-3138-1), funded by the Spanish Ministry of Economy, Industry and Competitiveness (Ministerio de Economía, Industria y Competitividad) and Agencia Estatal de Investigación, under the framework "Retos-Colaboración", State Scientific and Technical Research and Innovation Plan 2013-2016. The work was cofunded by the European Regional Development Fund (ERDF), "A way to make Europe". JdB acknowledges receiving financial support from the Catalan Health Department (Pla Estratègic de Recerca i Innovació en Salut [PERIS] 2016: SLT002/16/00364) and Instituto de Salud Carlos III (ISCIII; Miguel Servet 2019: CP19/00108), co-funded by the European Social Fund (ESF), "Investing in your future". Funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Authors' Contributions

CT, EV, JB, FB, and JdB participated in the conceptualization of project. CT, AL, LP, RV, and AC conducted data collection. CT, IB and AM-M participated in data curation. IB conducted all statistical analyses. CT, IB, and JdB wrote the original draft of the manuscript. All authors reviewed the final manuscript. JB and FB secured funding for the project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHealth checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1751 KB - jmir_v23i10e24072_app1.pdf]

Multimedia Appendix 2
Screenshots of the MiSAOS app showing from left to right and top to bottom: (i) loading screen with the app's name; (ii) last week's overall CPAP treatment performance summary, including an overall numeric score (0-10) together with a summary smiley, average compliance in hours, average air leaks, average residual apnea-hypopnea index (AHI), specific compliances for the last 7 days, and a summary text including reinforcements and tips to improve the overall rating in upcoming weeks (in this case, a positive reinforcement message for and a tip regarding air-leaks and hours of use); (iii) last week's overall CPAP treatment performance summary with focus on air-leaks, including last 14 days air-leak information and additional details; (iv) patient and device's summary information including technical details on the CPAP treatment characteristics; (v) main achievements summary, including a ranking of the patient's performance compared to other CPAP users in the region and challenges' progression such as good compliance streaks; and, (vi) sample information and training screen, in this case explaining the basics of obstructive sleep apnea pathophysiology.
[DOCX File , 825 KB - jmir_v23i10e24072_app2.docx]

Multimedia Appendix 3
Screenshots of the MiSAOS website showing from top to bottom: (i) registry of events and actions taken in relation to a given patient, including date, action details and comments; (ii) example of data collection for the feed of the predictive algorithms; (iii) sample prediction provided by MiSAOS intelligent algorithms, in this case predicting 6-month compliance based on early compliance and information such as the one collected in the previous screenshot; and, (iv) example of available training material tackling the most common issues and doubts of patients (in addition to videos there are also manuals, FAQs, and tips).
[DOCX File , 582 KB - jmir_v23i10e24072_app3.docx]

Multimedia Appendix 4
Differences in primary and secondary outcomes of the trial according to a per-protocol analysis.
[DOCX File , 18 KB - jmir_v23i10e24072_app4.docx]

References

1. Senaratna CV, Perret JL, Lodge CJ, Lowe AJ, Campbell BE, Matheson MC, et al. Prevalence of obstructive sleep apnea in the general population: A systematic review. Sleep Medicine Reviews 2017 Aug;34:70-81. [doi: 10.1016/j.smrv.2016.07.002] [Medline: 27568340]
2. Sánchez-de-la-Torre M, Campos-Rodriguez F, Barbé F. Obstructive sleep apnoea and cardiovascular disease. The Lancet Respiratory Medicine 2013 Mar;1(1):61-72. [doi: 10.1016/s2213-2600(12)70051-6]
3. Iftikhar IH, Bittencourt L, Youngstedt SD, Ayas N, Cistulli P, Schwab R, et al. Comparative efficacy of CPAP, MADs, exercise-training, and dietary weight loss for sleep apnea: a network meta-analysis. Sleep Med 2017 Feb;30:7-14. [doi: 10.1016/j.sleep.2016.06.001] [Medline: 28215266]
4. Martínez-García MA, Capote F, Campos-Rodríguez F, Lloberes P, Díaz de Atauri MJ, Somoza M, Spanish Sleep Network. Effect of CPAP on blood pressure in patients with obstructive sleep apnea and resistant hypertension: the HIPARCO randomized clinical trial. JAMA 2013 Dec 11;310(22):2407-2415. [doi: 10.1001/jama.2013.281250] [Medline: 24327037]
5. Barbé F, Durán-Cantolla J, Sánchez-de-la-Torre M, Martínez-Alonso M, Carmona C, Barceló A, Spanish Sleep And Breathing Network. Effect of continuous positive airway pressure on the incidence of hypertension and cardiovascular events in nonsleepy patients with obstructive sleep apnea: a randomized controlled trial. JAMA 2012 May 23;307(20):2161-2168. [doi: 10.1001/jama.2012.4366] [Medline: 22618923]
6. Sawyer AM, Gooneratne NS, Marcus CL, Ofer D, Richards KC, Weaver TE. A systematic review of CPAP adherence across age groups: clinical and empiric insights for developing CPAP adherence interventions. Sleep Med Rev 2011 Dec;15(6):343-356 [FREE Full text] [doi: 10.1016/j.smrv.2011.01.003] [Medline: 21652236]
7. Hussain SF, Irfan M, Waheed Z, Alam N, Mansoor S, Islam M. Compliance with continuous positive airway pressure (CPAP) therapy for obstructive sleep apnea among privately paying patients- a cross sectional study. BMC Pulm Med 2014 Nov 29;14(1):188 [FREE Full text] [doi: 10.1186/1471-2466-14-188] [Medline: 25433468]
8. Campos-Rodriguez F, Martinez-Alonso M, Sanchez-de-la-Torre M, Barbe F. Long-term adherence to continuous positive airway pressure therapy in non-sleepy sleep apnea patients. Sleep Medicine 2016 Jan;17:1-6. [doi: 10.1016/j.sleep.2015.07.038]

9. Rolfe I, Olson LG, Saunders NA. Long-term acceptance of continuous positive airway pressure in obstructive sleep apnea. Am Rev Respir Dis 1991 Nov;144(5):1130-1133. [doi: 10.1164/ajrccm/144.5.1130] [Medline: 1952444]
10. McArdle N, Devereux G, Heidarnejad H, Engleman HM, Mackay TW, Douglas NJ. Long-term use of CPAP therapy for sleep apnea/hypopnea syndrome. Am J Respir Crit Care Med 1999 Apr;159(4 Pt 1):1108-1114. [doi: 10.1164/ajrccm.159.4.9807111] [Medline: 10194153]
11. Wozniak D, Lasserson T, Smith I. Educational, supportive and behavioural interventions to improve usage of continuous positive airway pressure machines in adults with obstructive sleep apnoea. Cochrane Database Syst Rev 2014 Jan 08(1):CD007736. [doi: 10.1002/14651858.CD007736.pub2] [Medline: 24399660]
12. Bruyneel M. Telemedicine in the diagnosis and treatment of sleep apnoea. Eur Respir Rev 2019 Mar 31;28(151):180093 [FREE Full text] [doi: 10.1183/16000617.0093-2018] [Medline: 30872397]
13. Pépin JL, Tamisier R, Hwang D, Mereddy S, Parthasarathy S. Does remote monitoring change OSA management and CPAP adherence? Respirology 2017 Nov 12;22(8):1508-1517. [doi: 10.1111/resp.13183] [Medline: 29024308]
14. Aardoom JJ, Loheide-Niesmann L, Ossebaard HC, Riper H. Effectiveness of eHealth Interventions in Improving Treatment Adherence for Adults With Obstructive Sleep Apnea: Meta-Analytic Review. J Med Internet Res 2020 Feb 18;22(2):e16972 [FREE Full text] [doi: 10.2196/16972] [Medline: 32130137]
15. Rafael-Palou X, Vargiu E, Turino C, Steblin A, Sánchez-de-la-Torre M, Barbe F. Towards an Intelligent Monitoring System for Patients with Obstrusive Sleep Apnea. ICST Transactions on Ambient Systems 2017 Dec 19;4(16):153481. [doi: 10.4108/eai.19-12-2017.153481]
16. Lloberes P, Durán-Cantolla J, Martínez-García, Marín JM, Ferrer A, Corral J, et al. Diagnosis and treatment of sleep apnea-hypopnea syndrome. Archivos de Bronconeumología ((English Edition)) 2011 Jan;47(3):143-156. [doi: 10.1016/s1579-2129(11)70034-9]
17. MiSAOS. URL: https://www.misaos.com [accessed 2021-09-19]
18. Rafael-Palou X, Turino C, Steblin A, Sánchez-de-la-Torre M, Barbé F, Vargiu E. Comparative analysis of predictive methods for early assessment of compliance with continuous positive airway pressure therapy. BMC Med Inform Decis Mak 2018 Sep 18;18(1):81 [FREE Full text] [doi: 10.1186/s12911-018-0657-z] [Medline: 30227856]
19. CVE-DOGC-A-13051031-2013: Resolution SLT/353/2013, of February 13, on the review of public prices for the services provided by the Catalan Health Institute. URL: https://www.comb.cat/Upload/Documents/4635.pdf [accessed 2021-09-19]
20. Pépin JL, Jullian-Desayes I, Sapène M, Treptow E, Joyeux-Faure M, Benmerad M, et al. Multimodal Remote Monitoring of High Cardiovascular Risk Patients With OSA Initiating CPAP: A Randomized Trial. Chest 2019 Apr;155(4):730-739. [doi: 10.1016/j.chest.2018.11.007] [Medline: 30472022]
21. Stepnowsky C, Edwards C, Zamora T, Barker R, Agha Z. Patient perspective on use of an interactive website for sleep apnea. Int J Telemed Appl 2013;2013:239382-239310 [FREE Full text] [doi: 10.1155/2013/239382] [Medline: 23573081]
22. Kuna ST, Shuttleworth D, Chi L, Schutte-Rodin S, Friedman E, Guo H, et al. Web-Based Access to Positive Airway Pressure Usage with or without an Initial Financial Incentive Improves Treatment Use in Patients with Obstructive Sleep Apnea. Sleep 2015 Aug 01;38(8):1229-1236 [FREE Full text] [doi: 10.5665/sleep.4898] [Medline: 25581921]
23. Hwang D, Chang JW, Benjafield AV, Crocker ME, Kelly C, Becker KA, et al. Effect of Telemedicine Education and Telemonitoring on Continuous Positive Airway Pressure Adherence. The Tele-OSA Randomized Trial. Am J Respir Crit Care Med 2018 Jan;197(1):117-126. [doi: 10.1164/rccm.201703-0582oc]
24. Frasnelli M, Baty F, Niedermann J, Brutsche MH, Schoch OD. Effect of telemetric monitoring in the first 30 days of continuous positive airway pressure adaptation for obstructive sleep apnoea syndrome – a controlled pilot study. J Telemed Telecare 2015 Aug 06;22(4):209-214. [doi: 10.1177/1357633x15598053]
25. Turino C, de Batlle J, Woehrle H, Mayoral A, Castro-Grattoni A, Gómez S, et al. Management of continuous positive airway pressure treatment compliance using telemonitoring in obstructive sleep apnoea. Eur Respir J 2017 Feb 08;49(2):1601128 [FREE Full text] [doi: 10.1183/13993003.01128-2016] [Medline: 28179438]
26. Isetta V, Negrín MA, Monasterio C, Masa JF, Feu N, Álvarez A, Spanish Sleep Network. A Bayesian cost-effectiveness analysis of a telemedicine-based strategy for the management of sleep apnoea: a multicentre randomised controlled trial. Thorax 2015 Nov 26;70(11):1054-1061. [doi: 10.1136/thoraxjnl-2015-207032] [Medline: 26310452]
27. Anttalainen U, Melkko S, Hakko S, Laitinen T, Saaresranta T. Telemonitoring of CPAP therapy may save nursing time. Sleep Breath 2016 Dec 4;20(4):1209-1215. [doi: 10.1007/s11325-016-1337-9] [Medline: 27043327]

Abbreviations

AHI: Apnea–Hypopnea Index
BP: blood pressure
CPAP: continuous positive airway pressure
ESS: Epworth Sleepiness Scale
ITT: intention-to-treat
OSA: obstructive sleep apnea
PP: per protocol

SEPAR: Spanish Respiratory Society

Edited by R Kukafka; submitted 02.09.20; peer-reviewed by I Cano, R Halkes; comments to author 21.12.20; revised version received 02.03.21; accepted 12.08.21; published 18.10.21.

Please cite as:
Turino C, Benítez ID, Rafael-Palou X, Mayoral A, Lopera A, Pascual L, Vaca R, Cortijo A, Moncusí-Moix A, Dalmases M, Vargiu E, Blanco J, Barbé F, de Batlle J
Management and Treatment of Patients With Obstructive Sleep Apnea Using an Intelligent Monitoring System Based on Machine Learning Aiming to Improve Continuous Positive Airway Pressure Treatment Compliance: Randomized Controlled Trial
J Med Internet Res 2021;23(10):e24072
URL: https://www.jmir.org/2021/10/e24072
doi:10.2196/24072
PMID:34661550

Original Paper

A Wi-Fi–Based Mask-Type Laryngoscope for Telediagnosis During the COVID-19 Pandemic: Instrument Validation Study

Youngjin Moon[1,2*], PhD; Jaeho Hyun[3*], MSc; Jeongmin Oh[1*], BSc; Kwanhee Lee[2], BSc; Yoon Se Lee[4], MD, PhD; Jun Ki Kim[1,2], PhD

[1]Biomedical Engineering Research Center, Asan Institute for Life Sciences, Asan Medical Center, Seoul, Republic of Korea

[2]Department of Convergence Medicine, University of Ulsan College of Medicine, Seoul, Republic of Korea

[3]Department of Biomedical Engineering, University of Ulsan College of Medicine, Seoul, Republic of Korea

[4]Department of Otorhinolaryngology - Head and Neck Surgery, Asan Medical Center, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Jun Ki Kim, PhD
Biomedical Engineering Research Center
Asan Institute for Life Sciences
Asan Medical Center
88, Olympic-ro 43-gil, Songpa-Gu
Seoul
Republic of Korea
Phone: 82 230108619
Email: kim@amc.seoul.kr

Abstract

Background: Owing to the COVID-19 pandemic, social distancing has become mandatory. Wireless endoscopy in contactless examinations promises to protect health care workers and reduce viral spread.

Objective: This study aimed to introduce a contactless endoscopic diagnosis system using a wireless endoscope resembling a mask.

Methods: The Wi-Fi–based contactless mask endoscopy system comprises a disposable endoscope and a controller. First, the effective force applied by the tip during insertion was evaluated in a simple transoral model consisting of a force sensor on a simulated oropharynx wall. Second, the delay in video streaming was evaluated by comparing the frame rate and delays between a movement and its image over direct and Wi-Fi connections. Third, the system was applied to a detailed laryngopharyngeal tract phantom.

Results: The smartphone-controlled wireless endoscopy system was successfully evaluated. The mean, maximum, and minimum collision forces against the wall of the transoral model were 296 mN (30 gf), 363 mN (37 gf), and 235 mN (24 gf), respectively. The delay resulting from the wireless connection was 0.72 seconds. Using the phantom, an inexperienced user took around 1 minute to orient the endoscope to a desired area via the app.

Conclusions: Device articulation does not pose a significant risk of laryngopharyngeal wall penetration, and latency does not significantly impede its use. Contactless wireless video streaming was successful within the access point range regardless of the presence of walls. The mask endoscope can be controlled and articulated wirelessly, minimizing contact between patients and device operators. By minimizing contact, the device can protect health care workers from infectious viruses like the coronavirus.

(J Med Internet Res 2021;23(10):e31224) doi:10.2196/31224

KEYWORDS

smartphone-based endoscope; mobile health; telediagnosis; continuum segment; articulable endoscope; COVID-19; point-of-care diagnostics; validation; medical device; endoscope; sensor; innovation; video; transmission

Introduction

With the recent outbreak of COVID-19, the importance of quarantine and contact reduction has increased. Social distancing has become mandatory to lessen the impact of the outbreak and prevent the further spread of the virus, because increased contact with infected individuals drastically increases the possibility of infection. Those who work in the medical field are at particularly high risk of exposure. According to Nguyen et al, front-line health care workers are 4 to 10 times more likely to get the disease in the United States and the United Kingdom [1].

To protect health care workers and prevent further spread of the coronavirus, contactless examination has emerged. For contactless examination and treatment, governments worldwide have started implementing various contactless infrastructure and devices. For example, South Korean health care service workers use walk-through screening centers for examination [2]. Separated by a box, respiratory splash is completely prevented with a negative pressure applied inside the booth. Experts speculate that even after the pandemic ends, with COVID-19 being eradicated or neutralized like the common cold, there will always be a possibility of another pandemic [3]. Therefore, requirements for contactless devices will persistently rise after 2021 for all medical procedures that include the risk of possible viral infection, and endoscopy will not be an exception.

The endoscope is widely accepted as an examination device for a wide range of diseases. Starting as a rigid bronchoscope, over 100 years of development have enabled endoscopes to become flexible, therefore enhancing both patient experience and accessibility to internal organs [4,5]. However, in a pandemic, even a simple endoscopy can be a serious threat to patients. According to Kovaleva et al, several microorganisms can be transferred from one patient to another during endoscopy [6]. Most infections are via direct contact with incompletely disinfected devices, but viruses and other microorganisms can be transferred by not only an infected endoscope but also aerosols generated during the procedure [7]. Especially for highly infectious diseases like COVID-19, the entire facility involved in endoscopy can be at risk. To minimize contact with patients who undergo endoscopy, several prevention methods have been suggested, from minimizing the practice of endoscopy itself [8] to completely covering patients with protective sheets. However, since endoscopy has to be performed right beside the patient, it does not afford the same degree of safety to health care workers.

Wireless endoscopy has been around for a while, but has never been practically used. Most wireless endoscopes are capsule endoscopes, whose purpose is to specifically record the gastrointestinal tract, usually with control of the magnetic force applied from the outer belt patients are asked to wear [9]. There are several reasons wireless endoscopy is not favored over traditional, wired endoscopy, such as battery depletion and a low data transfer rate. If the endoscope is flexible, flexibility is usually manifested by tension applied to the string attached to a motor, transforming the position of individual segments of the continuum robot. The endoscope can be controlled directly using the controller attached to the motor. There is no reason to split the flexible endoscope into 2 parts and, therefore, discard the wired endoscope over the wireless endoscope, unless the desired area is too deep for the cable to reach. However, in dire situations when a contactless procedure should be the new norm, a wireless endoscope with articulation can be a huge advantage over a wired endoscope.

In this study, we introduced a novel, contactless endoscopic diagnosis system resembling a mask, enabling users to put it on easily themselves. The newly designed continuum part comprises 2 articulating units rolled into the device and sliding out when needed. While the distal tip unit freely articulates in the direction the user desires, the proximal tip holds down the user's tongue, providing a better endoscopy experience. The endoscope unit is controlled by a mobile app, over a Wi-Fi connection. Once the patient wears the device, the system enables the operator to perform endoscopy without direct contact with the patient.

Methods

Development of a Wi-Fi–Based Contactless Mask Endoscope

We built a Wi-Fi–based contactless mask endoscopic system. Figure 1 depicts the overall scheme of how a Wi-Fi–based contactless mask endoscope works in the field. The device comprises 2 parts, a disposable endoscope and a controller. The disposable endoscope captures and transmits digital visual data to the controller. The controller comprises a continuum segment, motors, and a microprocessor. The controller and the operator's Android mobile device are connected over a wireless network. The controller is responsible for articulation of the disposable endoscope and transmits visual data obtained from the endoscope to the connected mobile device. On the operator's side, a specific app for the endoscope helps the operator interface with the device using a wireless connection. The app receives real-time visual data from the device, while it sends control data to the device. The communication was based on the connection between electrical components. Figure 2 describes the connection and general flow of data and electricity in the Wi-Fi–based contactless mask endoscope.

Being a portable device, the contactless mask endoscope draws electricity from a rechargeable lithium battery. Raspberry Pi Zero W, a microprocessor containing a Wi-Fi module and memory sufficient enough to serve as a simple Wi-Fi server, is the center of the device. The device is connected to 5 motors, which also draw electricity from the battery. The endoscope module is a separate pluggable module that can be unplugged for maintenance and cleaning. Connected to the endoscope control board, Raspberry Pi Zero W receives real-time video signals from the endoscope module. To communicate with each other, Raspberry Pi Zero W and the operator's Android tablet are connected within the same network. The touch-based input of the app is converted into a directional signal in the app and sent to Raspberry Pi Zero W to manipulate the continuum segment.

The detailed Wi-Fi communication flow is described in Figure 3. Both the smartphone app and the device are booted, and they are manually connected to the same network. When the Wi-Fi connection is secured, the server is automatically started in the endoscope, while the app waits for the server connection. When the server successfully starts, it immediately starts to stream video information, while listening for directional button input from the client app. The client app starts to receive the video stream and readies to send a motor signal to the endoscope server. At this point, the user can easily send a direction-steering signal to the server by pressing the directional button on the app user interface. The app saves a snapshot and the video stream, without affecting the video stream. If the Wi-Fi connection is severed, the app and server revert to the waiting stage. If the user stops the session, the app is safe to turn off.

Figure 1. Overall scheme of the Wi-Fi–based contactless mask endoscopic system.

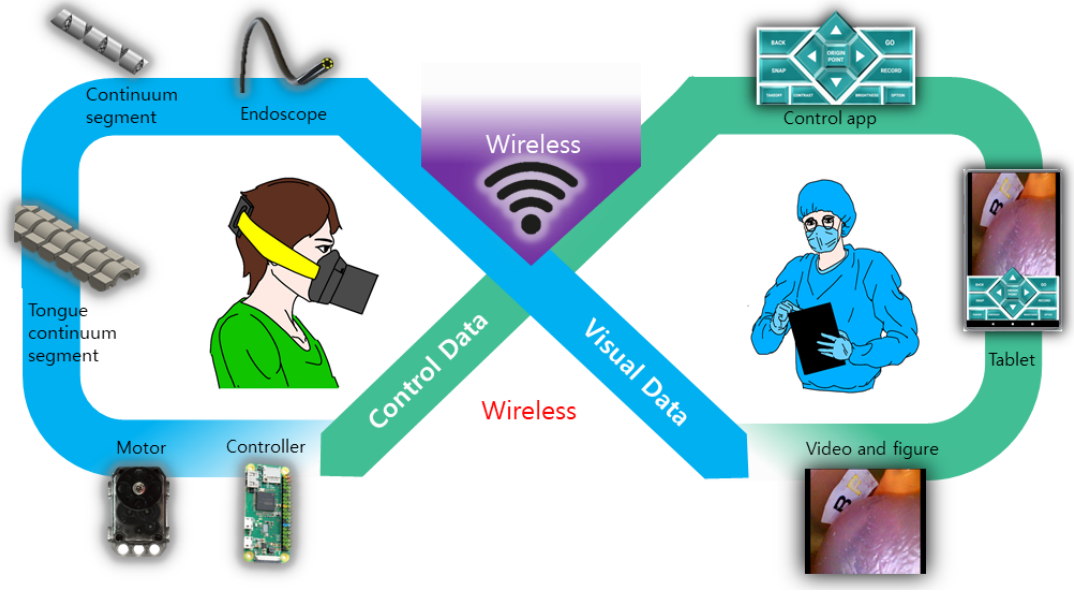


Figure 2. Electric wiring and key function of the Wi-Fi–based contactless mask endoscopic system. GND: ground; IDC: insulation-displacement contact.

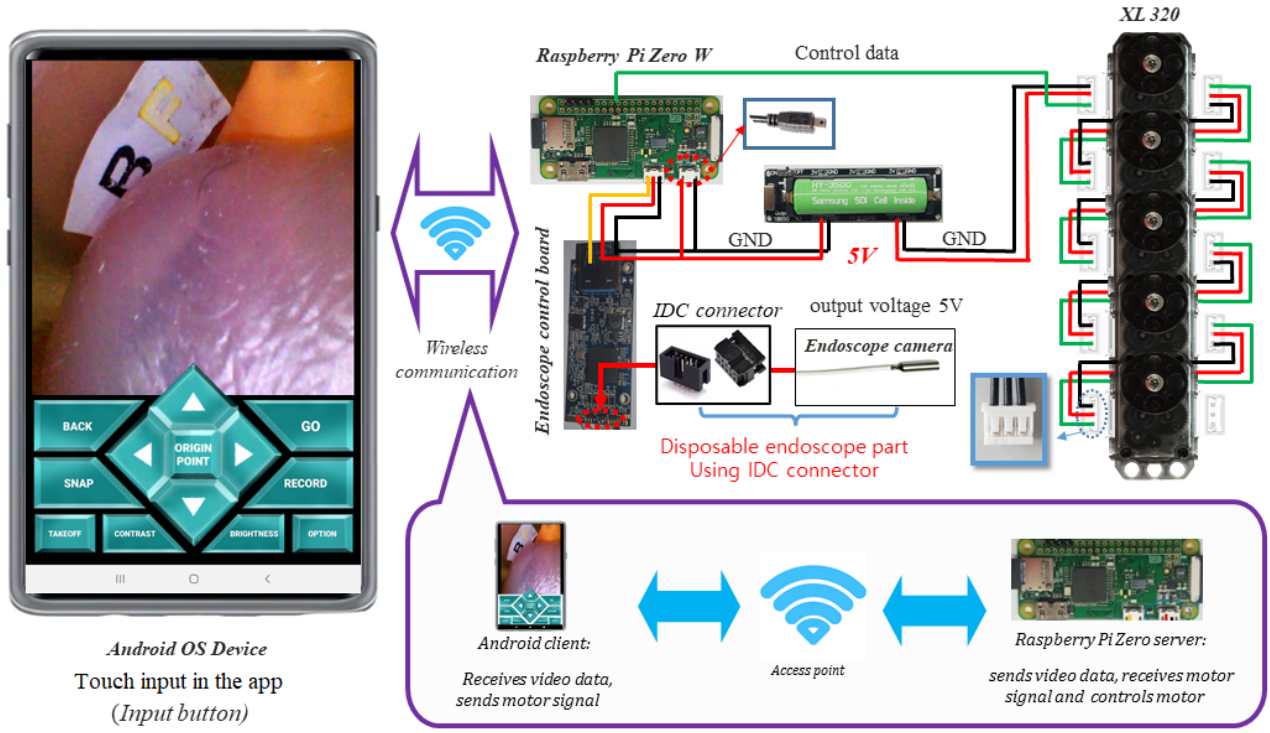
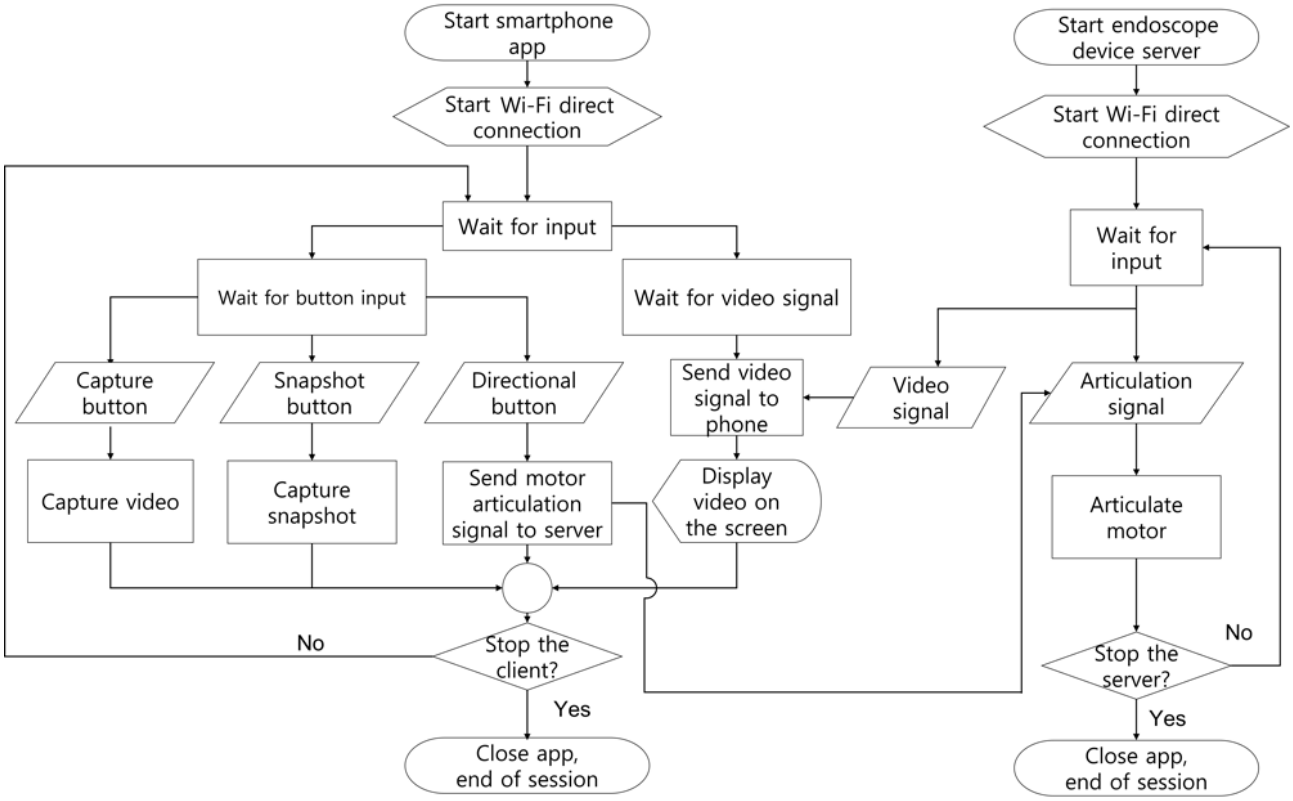
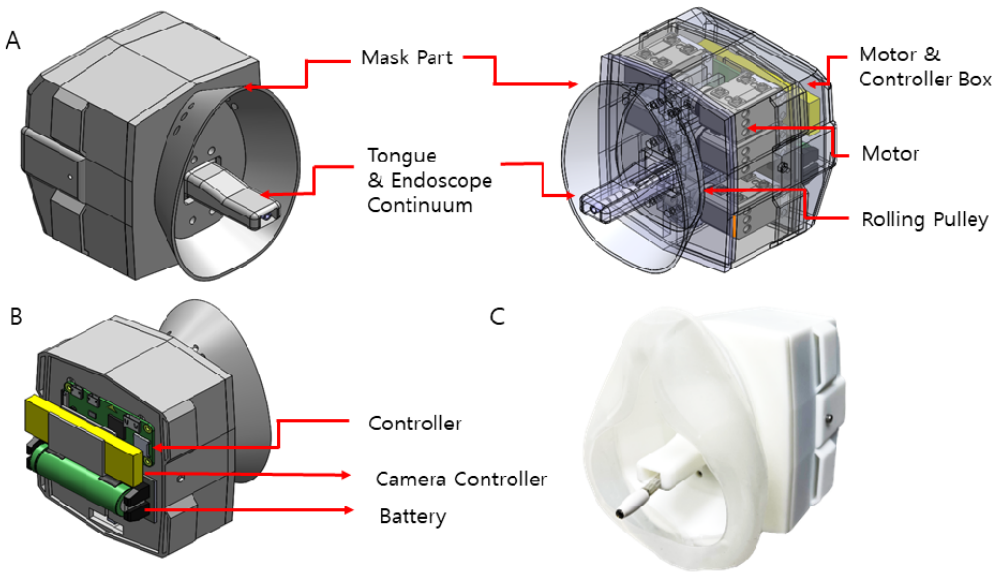


Figure 3. Flowchart of the software part of the Wi-Fi–based contactless mask endoscopic system in action.



The mechanical parts of the mask endoscope are roughly divided into a mask, a motor and controller box, and a continuum. The mask is where crucial systemic components are present. The lid of the mask can be removed to place a rechargeable battery and perform maintenance of the Raspberry server. Figure 4 shows the internal and external structure of the 3D mask endoscope in 3D space. The arrangement of the motor and controller box, pulley, and controlling parts are shown on the right of Figure 4A. The motors are separated from the controller by a partition. As shown in Figure 4B, the controllers and the battery are in the outermost side of the motor and controller box, easily accessible by removing the lid. Figure 4C shows a finished prototype in 3D space. In addition to the mask, a soft support is added to improve patient experience by enhancing texture.

Figure 4. Assembled hardware parts in 3D space. (A) Rendered 3D structure of the outside shell and internal articulating parts of the hardware. (B) Position of the electronic parts of the hardware, rendered in 3D space. (C) Printed and assembled final product in real space.



The endoscopic continuum consists of individual continuum segments connected by stranded wires. As shown in Figure 5A, each segment has 4 wire holes for the articulation wire and a big central cable hole for the endoscope camera cable. The tension of the wires is adjusted by winding, and rewinding narrows the lateral gap between segments, therefore generating 2-degree-of-freedom (2DOF) omnidirectional motion [10], as depicted in Figure 5B. Two opposite-sided wire holes in the individual segments are paired to each other, making 4 wire holes in total. Each wire hole pair is responsible for the rotational steering movement between 2 segments. The collectible movements of segments create the steering motion.

The tongue continuum is quite different. Only having a slope on the front and the back side of the single continuum segment, the tongue continuum has 2 wire holes for a supportive wire on each wing of the individual segment, a central cable hole for the endoscope cable, and 2 slits at the bottom for the articulating wire, as shown in Figure 5C. The tongue continuum can bend up and down (bidirectional), as shown in Figure 5D. In addition, the wire connecting all the tongue continuum parts is not skewed in a specific direction, let alone being connected to a separate motor. Lacking an active driving force of motion, it can only be passively rolled in and out. The tongue continuum flattens the tongue to get the endoscope safely in position. Coordinated with the motor movement, the tongue continuum module can be retracted and extended by rolling it in and out of the mask, as shown in Figure 6.

As shown in Figure 6, the 4 upper motors aligned in rectangular corners are responsible for articulation of the endoscope continuum. The motor positioned at the downmost location in the mask is responsible for the extending and retracting motions of the tongue continuum. The tongue continuum is directly connected to the winder wheel, and the wheel itself serves as the winder for the endoscope module and the tongue continuum's supporting wire inside the continuum. The wire connected to the tongue continuum is flexible enough to be kept rolled up in the mask when the endoscope is not in use. As the motor turns clockwise, the tongue continuum unit unrolls from the winder, extending forward. While the tongue continuum extends forward, the other 4 motors also correspond to the motion, unrolling the wires to match the length. This results in only the tongue continuum extending forward, without any bending of the endoscope continuum. This motion corresponds with the GO button on the app controller.

When the tongue continuum is extended to the desired length, the operator can manipulate buttons in the app to control the endoscope continuum. Rolled out with the tongue continuum, the endoscope continuum is connected to separate wires that lead to the winder connected to the 4 upper motors. Each motor is in charge of unidirectional bending of 2DOF motion, that is, a pair of motors corresponds to bidirectional bending motion, as shown in Figure 7.

After imaging, the operator can press the BACK button to turn the motor counterclockwise to make the tongue continuum roll back into the mask. For the pulling motion, while the winder of the bottom motor rotates, like the device articulated in the extending motion, the other 4 motors rotate to wind the wire back up to match the retracting length of the endoscope.

Figure 5. Configuration of motions in the continuum. (A) Structure of the endoscope continuum. (B) Bending motion of the endoscope continuum in action. (C) Structure of the tongue continuum. (D) Bending motion of the tongue continuum in action.

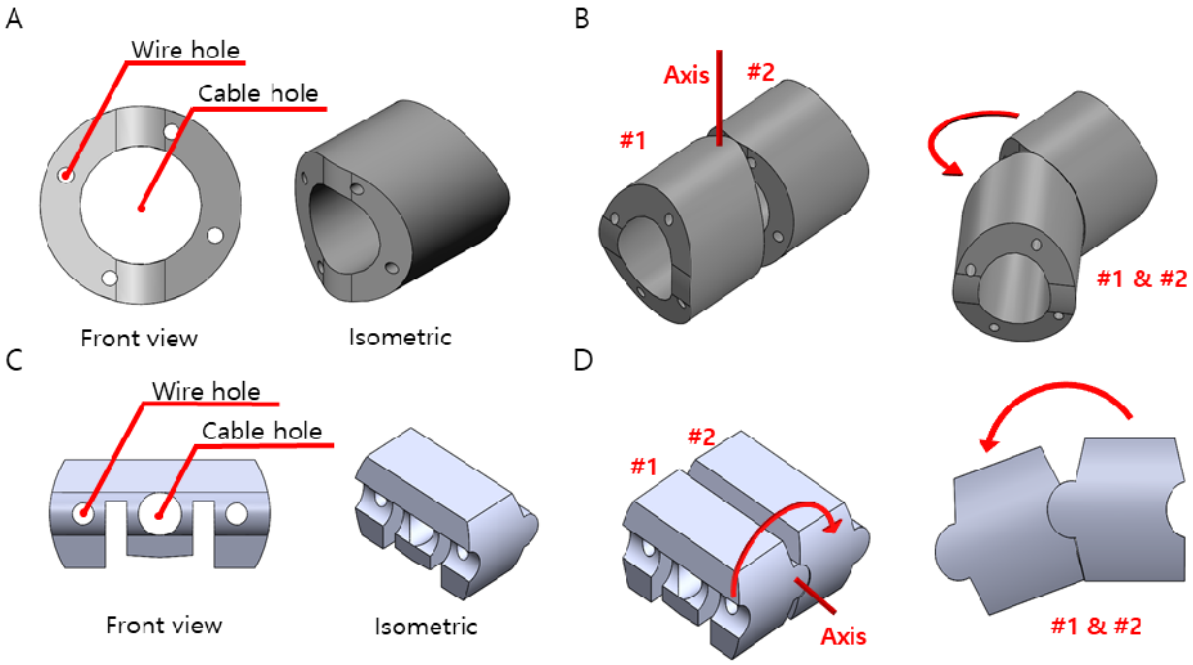


Figure 6. Articulation principle of continuums. (A) Extending and retracting motion resulting from the rolling motion of the tongue continuum winder motor. (B) Arrangement of the inner wire connection of the continuum. (C) Endoscope continuum in motion after extension.

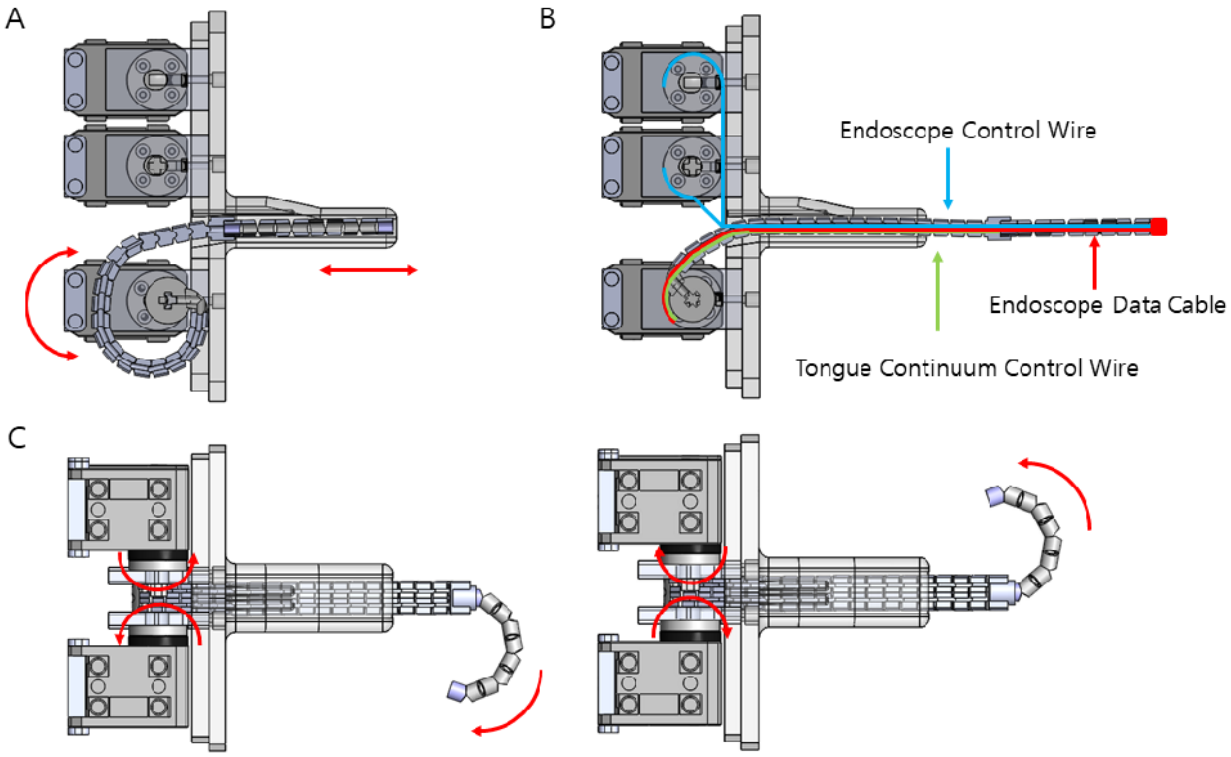
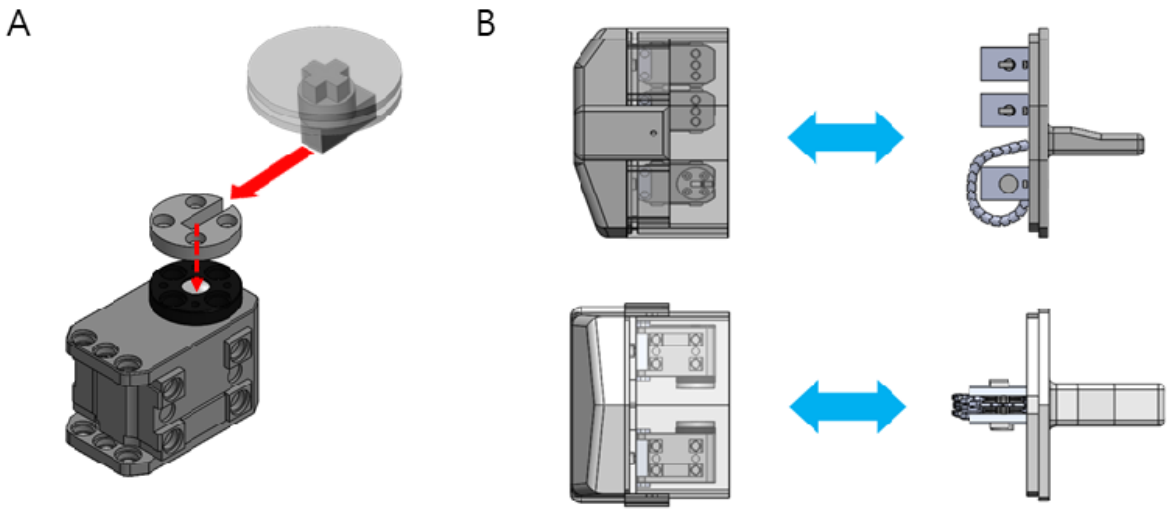


Figure 7. Principle of the key-and-pulley parts, rendered in 3D simulation. (A) Assembly of the key-and-pulley parts. (B) Combining the pulley and motor parts, and key and continuum parts, rendered in a 3D model.



Pulley and Key

Medical devices in contact with patients must be sterilized and reprocessed by thorough cleaning. However, complex medical devices, especially those with electronic parts, are difficult to sterilize without damaging the electronic parts. The contactless wireless mask endoscope is no exception. Many medical electronic devices overcome this problem by separating the parts in contact with the patient from the noncontact parts using disposable covers like a medical drape. As shown in Figure 7, our device is made of 2 detachable assemblies, a continuum module and a motor assembly module.

For separation, the device needs to connect and convey the force between parts. By employing a key on the pulley, the force from the motors can be transmitted to the continuum module directly. The pulley is designed as a simple disk with a slit that goes from the center of the pulley to the outermost rim, which appears U-shaped, as depicted in Figure 7A. The key is a small rectangular metal rod-like part, which has a small cross on the

disk attached to the tip of the long side of the rod. When fully assembled, the key is inserted into the slit of the pulley. Figure 7B shows the key-and-pulley assembly in action. The pulley is attached to the motor wheel by screws (4 holes). The pulley can be easily disassembled from the motor wheel by unscrewing the screws, but when it is screwed on, the pulley and the motor wheel are securely attached. On the other hand, the key is attached directly to the wire winder wheel, which is directly connected to each wire that is responsible for articulation. By sliding the assembled key into the slit of the pulley, the assembled motor unit and the endoscope modules are safely connected.

Results

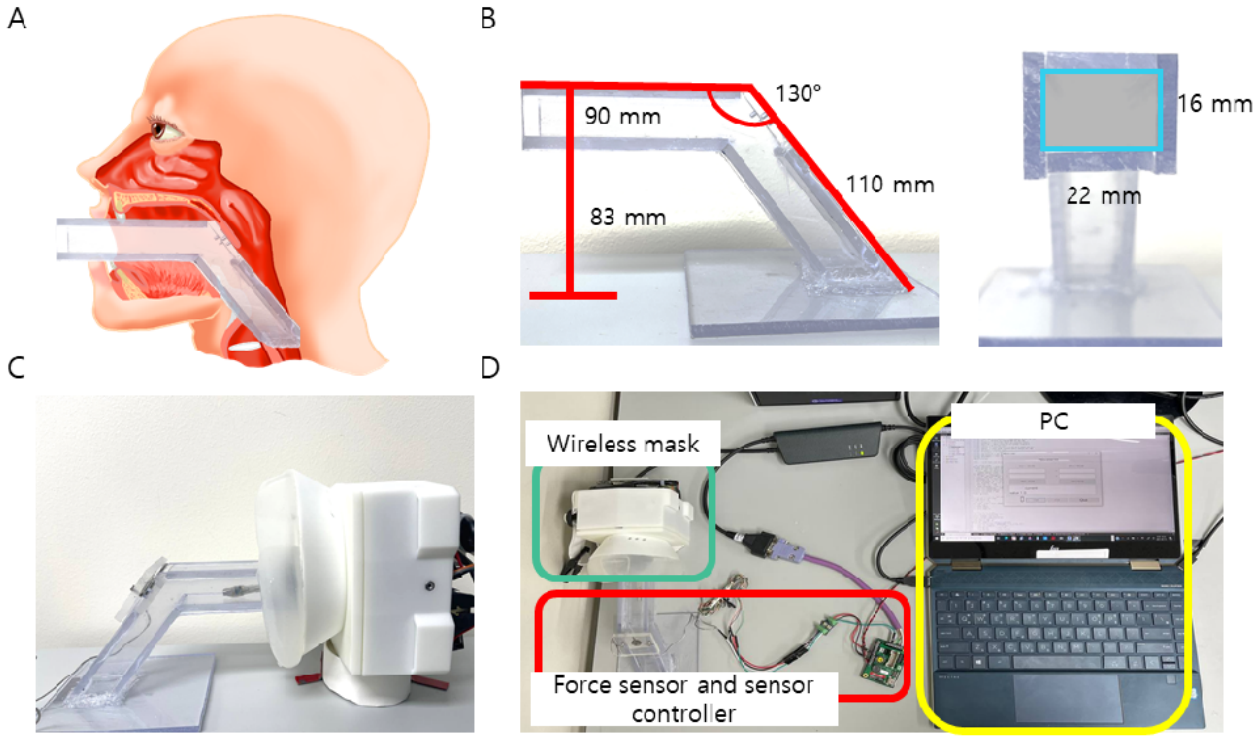
Continuum Contact Force Experiment

The contactless mask endoscope is designed as a laryngoscope, which means the endoscope part has to go over the patient's tongue, hard palate, and soft palate to finally arrive at the vocal cord. If the flexible continuum is too rigid, it can injure oropharyngeal tissues, resulting in secondary infection. A simple transoral model was created to evaluate the end-effect force when the continuum articulates.

Assuming the continuum always collides with the oral tract, a force sensor (Honeywell International Inc) was placed where the soft palate is present, since the continuum is most likely to collide with the soft palate. The force against the surface was measured 10 times. The overall experimental setup is presented in Figure 8D. According to studies on oral anatomical structures, the hard palate length is 46.65 to 51.89 mm for adult males and females [11-16], and the soft palate length is 25.3 mm on average. The radius of the esophagus at rest varies from 12.87 to 17 mm. The angle between the soft palate and the esophagus is 130.1° to 137° [17].

Figures 8A and 8B present the oral structures and transoral model constructed from the information about oral structures. The length of the simulated oral part of the transoral model was set to 90 mm, the angle between the oral part and the esophagus was 130°, the cross-sectional area of the simulated esophagus in the transoral model was 16 mm × 22 mm, and the height was set to 83 mm. The contactless wireless mask endoscope was set in the transoral model, as shown in Figure 8C. Without articulating the endoscope continuum, the tip of the endoscope was collided with the force sensor located on the simulated oropharynx wall of the transoral model only using back-and-forth articulation of the tongue continuum. The collision was repeated 11 times with 10-second intervals.

Figure 8. Continuum contact force measurement experiment and structure of the transoral model for the experiment. (A) Comparison of human oropharyngeal anatomy and the transoral model structure. (B) Dimension of the transoral model. (C) Wi-Fi–based contactless wireless mask endoscope positioned with the transoral model for the continuum force measurement experiment. (D) Experimental setting of the continuum tip force measurement.



Wireless Video Transfer Rate Evaluation

The contactless wireless mask endoscope processes raw video data in the controller and sends them to the Android tablet through an access point. There is a small delay due to the limited hardware specification and the nature of wireless communication. To evaluate the delay in the video broadcast from the endoscope, a comparison experiment was conducted. Figure 9 shows the experimental setup of the wireless and wired endoscope video delay test. From the same camera module, the video was transferred directly from the camera and broadcast from the Raspberry Pi module. Using forward articulating

movement as a standard, we compared the frame rate and delays between the 2 videos.

Figure 9. Wireless video transfer rate evaluation experiment. (A) Wireless and (B) wired articulation and video transfer between a contactless wireless mask endoscope and laptop computer. The transferred videos are displayed on the screens of both laptops.

Wireless Articulation Experiment Using a Phantom

Using a phantom resembling the real-life structure of a patient's laryngopharyngeal tract, a wireless articulation experiment was conducted to evaluate the practicality of the contactless mask endoscope.

The conjoined experiment involved a contactless mask endoscope, 2 cameras, a phantom, and an operator with an Android tablet. The abstract experimental setup is shown in Figure 10 and Multimedia Appendix 1. The laryngopharyngeal phantom was set at the center of the meeting room. The contactless mask endoscope was applied to the phantom, and the operator was outside the meeting room. The operator began to control the endoscope with the tablet through a wireless connection. The camera simultaneously recorded both sides as the operator controlled the endoscope to take snapshots and videos of stickers attached to the desired parts of the phantom. The process was repeated after changing the operator's location as follows: in the front of the meeting room, in the corridor, and in the laboratory right next to the meeting room.

The results of the continuum contact force experiment are shown in Figure 11. The magnitude of the collision force varied from a minimum of 235 mN (24 gf) to a maximum of 363 mN (37 gf). The average collision force was 296 mN (30.2 gf), which is just 6.5% of that of a commercial endoscope.

The results of the wireless video transfer rate evaluation experiment are summarized in Figure 12. Comparing wired connection and wireless communication, wireless video streaming was delayed by 0.72 seconds with the contactless mask endoscope.

In the wireless articulation experiment, we placed marker stickers in the desired areas as shown in Figure 10C and the operator of the mask endoscope controlled the device and examined the video data streamed from it while being spatially separated completely from the patient. The measured time to get the stickers in the video stream was around 1 minute in the desired separated location. From the experiment, we tried to evaluate the articulating range of the endoscope continuum and the capability of the device's wireless communication that enables users to control the device from a different space. Contactless wireless video streaming was successful within the range of the access point regardless of the presence of any material blocking the way.

In summary, the Wi-Fi–based contactless mask endoscope was controlled and articulated wirelessly, minimizing or completely avoiding possible contact between patients and device operators. By minimizing contact, contactless medical devices can protect health care workers from infectious diseases like COVID-19.

Figure 10. Wireless articulation experiment. Compilation of videos resulting from the experiment conducted outside of the office (A) over a glass partition and (B) behind a concrete wall. Each video is synced to the timestamp on the bottom right of the video. (C) Snapshot taken from each marked position by articulating the endoscope continuum to deliver the tip to the desired position. (D) Position of the mask endoscope and operator for each experiment.

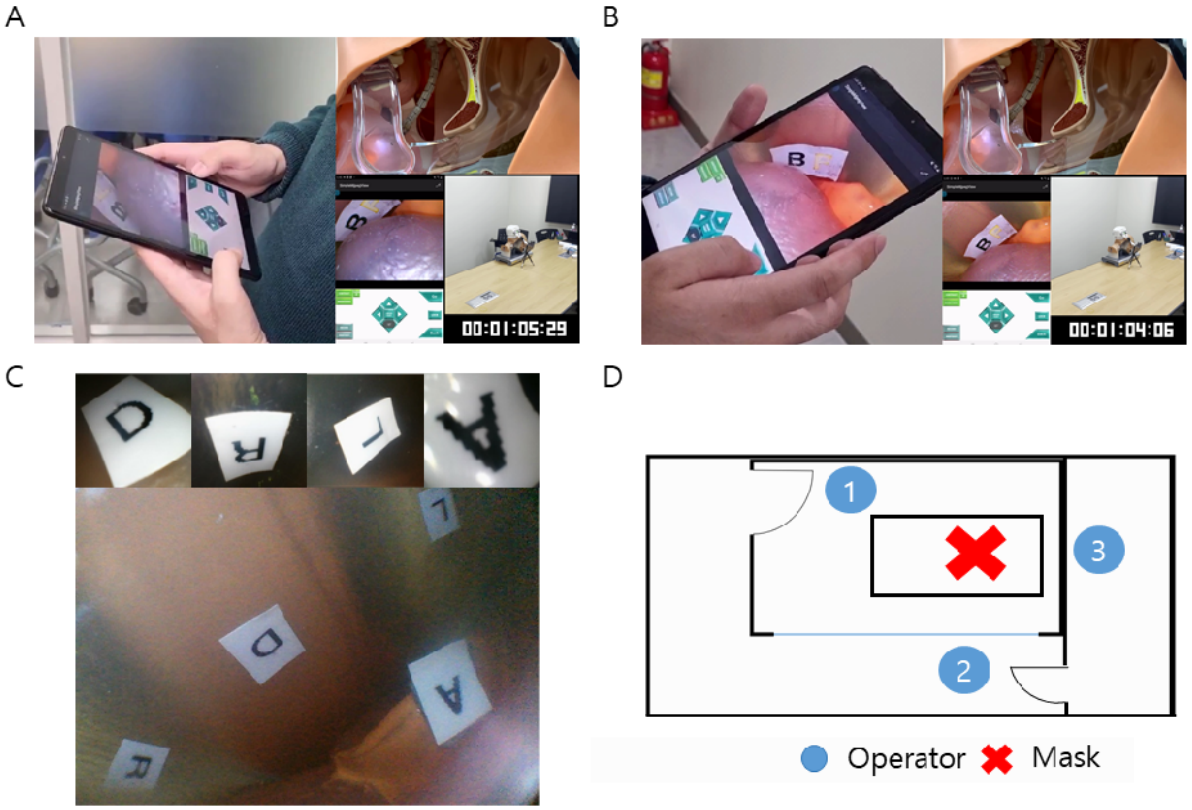


Figure 11. Measured force of the endoscope continuum colliding against the wall of the transoral model.

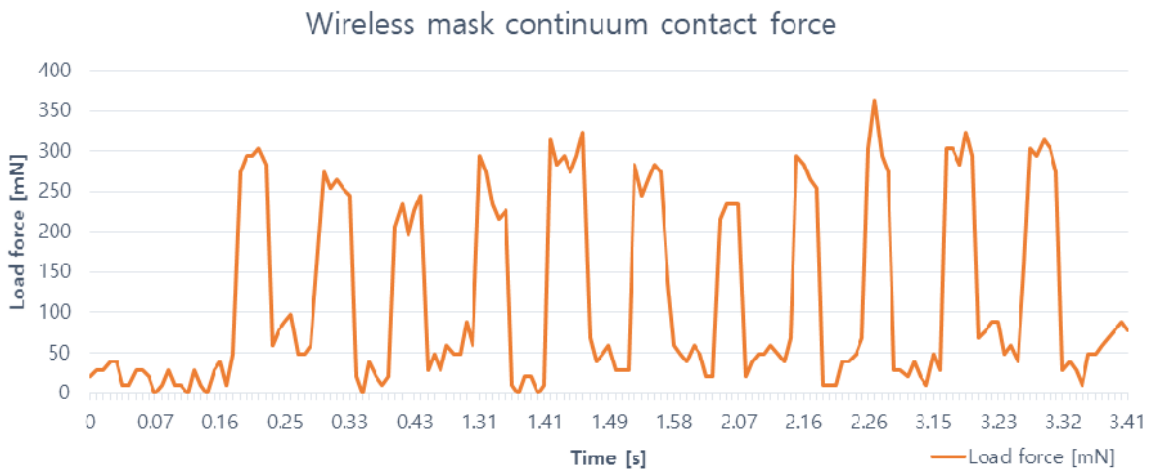
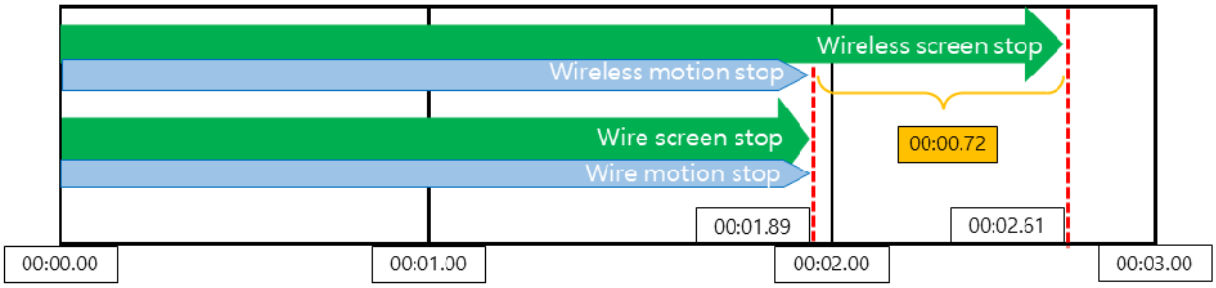


Figure 12. Time lapse of the wireless video transfer rate evaluation experiment.



Discussion

Principal Findings

In the current COVID-19 pandemic era, physicians run risks when they evaluate a patient with fever of unknown origin or with laryngeal symptoms, such as dyspnea, hoarseness, and dysphagia. However, these symptoms are associated with laryngeal obstruction and could be exacerbated when adequate evaluation and treatment are lacking. The proposed wireless device would help protect physicians from possible contagious diseases while allowing them to promptly evaluate patients.

Previously, our team developed a mobile-based USB on-the-go (OTG) endoscope that can be articulated using an Android mobile app [10]. The Wi-Fi–based contactless mask endoscope has several advantages over the USB OTG endoscope. First, it can be controlled over a wireless network, enabling the operator to position himself/herself freely, unlike the USB OTG endoscope, for which the operator has to be close to the patient. Second, the operator of the Wi-Fi–based mask endoscope can attach the endoscope via a strap, unlike the USB OTG endoscope, which has to be held with the cell phone. This has the secondary benefit of decreasing the load on the operator's wrist and hand, providing a better user experience. Finally, modulization of the Wi-Fi–based contactless mask endoscope makes it easier to replace parts, simplifying maintenance.

As individual parts can be replaced rather than replacing the whole module, maintenance costs are reduced. The price of the USB OTG endoscope was US $473, which was one of the main advantages of the device, as it was intended to be used in underdeveloped countries. Table 1 shows the price of the Wi-Fi–based contactless mask endoscope. Considering that the cost of the 3D-printed modules can be reduced with mass production, in addition to the reduced maintenance costs, the price may be better than that of the USB OTG endoscope [10].

Table 1. Prices of hardware parts of the Wi-Fi–based contactless wireless mask endoscope.

Hardware	Price (US $)
XL320 * 5ea	97.8
3.9-mm camera module	170
3D print of the case	68.96
3D print of the tongue and endoscope continuum	152.88
Controller	60.02
Wires, etc	49.27
Total	598.93

A snake-like continuum is commonly used in flexible endoscope articulation, enabling multiple degrees of motion for clinical needs. However, complex articulation was historically reserved for larger endoscopes. Endoscopes with multiple kinds of articulating units for flexible parts have some advantages. First, they can perform different functions, such as adjusting the mechanical force applied to the articulating units to ensure patient safety [18]. Second, the multiple kinds of articulating units can exert different kinds of motion for each kind of unit, enabling users to control dynamic motions [19]. While the endoscope continuum of this contactless wireless mask endoscope serves as the conventional endoscope articulating unit, the tongue continuum has 2 other functions. First, its broad shape and stiff mechanical property flattens the tongue. During endoscopy, especially of the stomach and upper respiratory organs, the tongue gets in the way of the endoscope, hindering its movement. Therefore, it is recommended that the examiner use anesthesia before endoscopy to keep the tongue out of the way. A mouthpiece with a tongue presser can be used [20]. The tongue continuum of the contactless wireless mask endoscope also serves as a tongue presser, making it easier to enter the desired area. Second, the tongue continuum itself works as a support for the regular continuum. If the snake-like continuum is too long, due to capstan friction, forces on the end point cannot drive the end point as intended [18]. Adding a tongue continuum shortens the length of the snake-like continuum, making it easier to relay force to the tip of the endoscope and enabling precise control. In conclusion, having a semirigid

tongue continuum has the advantage of articulation control and safe delivery of the endoscope tip to the desired area.

The system's modularity also reduces contamination concerns. Endoscopes are more prone to cross-infection than other noninvasive examination devices because endoscopy involves mucosal contact. The endoscope module cannot be sterilized by autoclaving because the electronics within it are sensitive to heat, pressure, and moisture. Given that autoclaving is one of the easiest and most reliable methods of disinfection, the endoscope module must be changed regularly regardless of the cost.

To solve the contamination problem, 2 major innovations were incorporated. First, a detachable endoscope module was used. With the key-and-pulley part inside the mask, the endoscope module is fixed to the motor. When the operator decides to separate the mask and the endoscope, the 2 parts can be separated by disassembling the key-and-pulley part. The inner space of the mask can be easily accessed by opening the outer lid. Separation of parts also creates a gap between the parts, where an anticontaminant cover sheet can be applied. Therefore, by replacing the wheel, the possibly contaminated continuum and endoscope modules can be separated from the mask module, and the mask can be kept clean. The second innovation is that of a simplified inner structure. For a larger endoscope, the sheath has to have some kind of contact point because of the channel in the endoscope. However, the contactless wireless mask endoscopic system does not have such a channel, and therefore, the sheath can be simple, like a condom. Without a channel, it is much easier to sterilize the smooth surface, and there is much less chance of contamination because of the small contact point.

For conventional endoscopes, portable devices with a wireless connection are not preferred for multiple reasons. First, a wireless connection has a limited data transfer rate. The Bluetooth 5.0 connection has a 1 Mbps data transfer rate for low-energy Bluetooth [21]. The 1 Mbps rate of Bluetooth 5.0 can barely cover a standard-frame-rate 360p video stream [22]. A Wi-Fi connection is more generous in terms of the bit rate, and 802.11n families can go up to 200 Mbps, enough to stream 4K videos [22,23]. Second, since endoscopy is a medical procedure for which many trained health care experts are needed, there is less need to separate the endoscope and the operating system. However, due to the COVID-19 pandemic, there is a possibility of infecting entire medical facilities during endoscopy. Therefore, the need for separation of the operator and patient has emerged.

The current focus mainly revolves around contactless endoscopy that is performed in close proximity but remains in the clinic. While the practitioner is in a separate room, a patient can follow instructions to wear the mask. When the mask is properly worn, endoscopy can be performed remotely using the smartphone app. Should an emergency occur, practitioners waiting outside the room can interrupt the procedure and ensure the patient's safety. Otherwise, the practitioner can utilize the internet to extend the range of communication. The endoscope streams the video from the patient to the practitioner. Theoretically, by setting up a public IP address and encrypted data transfer, users can safely control the endoscope over the internet from many kilometers away. Subsequent research will involve safety and clinical trials performed at Asan Medical Center using an improved version of the articulating endoscope system.

The designed laryngoscope goes over soft areas that can trigger a gag reflex. The oropharyngeal area is sensitive, so it can detect down to 2.07 mmHg of pressure [24,25]. A gag reflex can be triggered by a slight touch of the triggering area [25]. Generally, laryngoscopy involves carefully avoiding the areas that would trigger a gag reflex [26] or using local anesthesia [27-29]. In particular, noninvasive local anesthesia can be applied easily with fewer side effects and fewer tissue injuries. Therefore, in practice, the device can be accompanied with local anesthetic application by hand or machine. Lidocaine nebulizer, spray, or ingestion would help ameliorate the gag reflex.

Limitations

While making the Wi-Fi–based contactless mask endoscope, we faced a few problems mainly due to hardware limitations. The 3D-printed parts did not achieve a tight seal and allowed air to flow through the motor box. For clinical application, improved seals combined with a filtered ventilation port would reduce the risk of aerosol transmission. Further hardware limitations mandated a tradeoff between device size and capabilities. Because the mask needs to be worn like a conventional respiratory mask, the dimensions of the attached controller and motor parts were minimized to enhance the patient's experience (Table 2). The lack of space and need to downsize the controller also lowered the performance of the control system, resulting in delays and reduced video-processing performance, which led to a 0.72-second delay in wireless video streaming.

For future research, we will compensate the space limitation in hardware with better performance, while maintaining the portable point of care–like nature of the device. In particular, the device may be widened to incorporate an additional channel for biopsy or to increase patient comfort.

Table 2. Dimensions of the Wi-Fi–based contactless mask endoscope.

Part and dimension	Value (mm)
Mask	
Width (with cushion part)	99
Height (with cushion part)	120
Depth (with cushion part)	47
Width (without cushion part)	90
Height (without cushion part)	118
Depth (without cushion part)	26
Controller and motor box	
Width	97
Height	114
Depth	77
Hardware	
Width	99
Height	122
Depth	123

Conclusions

We developed a contactless wireless mask endoscopy and diagnosis system to be used during the COVID-19 pandemic. Using Wi-Fi and an Android mobile device, we successfully articulated the endoscope and took snapshots and videos of the patient's vocal cord. The device can separate health care workers from patients, while enabling diagnosis of internal parts of the body that cannot be reached without endoscopy. In conclusion, the contactless wireless mask endoscope can help minimize chances of contamination, further minimizing the impact of the pandemic in the medical field.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) (2019R1A2C2084122 and NRF-2021R1A2C3005763); a Medical Research Center grant (2018R1A5A2020732) by the Ministry of Science & ICT; the Ministry of Trade, Industry, and Energy under Industrial Technology Innovation Program (20000843); and a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute, funded by the Ministry of Health and Welfare, Republic of Korea (HI18C2391 and HI17C2410).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Contactless wireless endoscope mask in action. Left: operation of the endoscope from an adjacent hallway. Right top: close-up view of endoscope articulation in the laryngopharyngeal phantom. Right bottom left: screen capture of the mobile app in action. Right bottom right: view of the endoscope mask installed on the phantom.
[MP4 File (MP4 Video), 114697 KB - jmir_v23i10e31224_app1.mp4]

References

1. Nguyen L, Drew D, Graham MS, Joshi AD, Guo C, Ma W, et al. Risk of COVID-19 among front-line health-care workers and the general community: a prospective cohort study. The Lancet Public Health 2020 Sep 25;5(9):e475-e483 [FREE Full text] [doi: 10.1016/S2468-2667(20)30164-X]
2. Kim SI, Lee JY. Walk-Through Screening Center for COVID-19: an Accessible and Efficient Screening System in a Pandemic Situation. J Korean Med Sci 2020 Apr 20;35(15):e154 [FREE Full text] [doi: 10.3346/jkms.2020.35.e154] [Medline: 32301300]
3. Lavine JS, Bjornstad ON, Antia R. Immunological characteristics govern the transition of COVID-19 to endemicity. Science 2021 Feb 12;371(6530):741-745 [FREE Full text] [doi: 10.1126/science.abe6522] [Medline: 33436525]
4. Frangenheim H. History of Endoscopy. In: Gynaecological Endoscopy. Boston, MA: Springer; 1988:1-5.

5. Kaunitz JD. The fruits of fiber: the invention of the flexible fiberoptic gastroscope. Dig Dis Sci 2014 Nov;59(11):2616-2618 [FREE Full text] [doi: 10.1007/s10620-014-3339-4] [Medline: 25193388]
6. Kovaleva J, Peters FTM, van der Mei HC, Degener JE. Transmission of infection by flexible gastrointestinal endoscopy and bronchoscopy. Clin Microbiol Rev 2013 Apr;26(2):231-254 [FREE Full text] [doi: 10.1128/CMR.00085-12] [Medline: 23554415]
7. Soetikno R, Teoh AY, Kaltenbach T, Lau JY, Asokkumar R, Cabral-Prodigalidad P, et al. Considerations in performing endoscopy during the COVID-19 pandemic. Gastrointest Endosc 2020 Jul;92(1):176-183 [FREE Full text] [doi: 10.1016/j.gie.2020.03.3758] [Medline: 32229131]
8. Minelli Grazioli L, Milluzzo SM, Spada C. Safe endoscopy during the COVID-19 pandemic. Gastrointest Endosc 2020 Dec;92(6):1274-1275 [FREE Full text] [doi: 10.1016/j.gie.2020.07.058] [Medline: 32750322]
9. Marlicz W, Ren X, Robertson A, Skonieczna-Żydecka K, Łoniewski I, Dario P, et al. Frontiers of Robotic Gastroscopy: A Comprehensive Review of Robotic Gastroscopes and Technologies. Cancers (Basel) 2020 Sep 28;12(10):2775 [FREE Full text] [doi: 10.3390/cancers12102775] [Medline: 32998213]
10. Moon Y, Oh J, Hyun J, Kim Y, Choi J, Namgoong J, et al. Cost-Effective Smartphone-Based Articulable Endoscope Systems for Developing Countries: Instrument Validation Study. JMIR Mhealth Uhealth 2020 Sep 10;8(9):e17057 [FREE Full text] [doi: 10.2196/17057] [Medline: 32909951]
11. Mustafa AG, Tashtoush AA, Alshboul OA, Allouh MZ, Altarifi AA. Morphometric Study of the Hard Palate and Its Relevance to Dental and Forensic Sciences. Int J Dent 2019 Jan 28;2019:1687345 [FREE Full text] [doi: 10.1155/2019/1687345] [Medline: 30809259]
12. Mahendran S, Thenmozhi MS. Sexual Dimorphism of Adult Human Palate by its Dimensions in South Indian Dry Skulls. International Journal of Pharmaceutical Sciences Review and Research 2017;45(1):56-59 [FREE Full text]
13. Jotania B, Patel SV, Patel S, Patel PM, Patel SA, Patel K. Morphometric analysis of hard palate. International Journal of Research in Medicine 2013;2(2):72-75 [FREE Full text]
14. Fauzi NQBA, Eapen B. Patient management education for dentists. International Journal of Current Research 2017;9(9):57136-57139 [FREE Full text]
15. Fitch WT, Giedd J. Morphology and development of the human vocal tract: a study using magnetic resonance imaging. J Acoust Soc Am 1999 Sep;106(3 Pt 1):1511-1522. [doi: 10.1121/1.427148] [Medline: 10489707]
16. Badshah M, Soames R, Khan MJ, Hasnain J. Morphology of the human hard palate: a study on dry skulls. Italian Journal of Anatomy and Embryology 2018;123(1):55-63 [FREE Full text]
17. Joseph AA, Elbaum J, Cisneros GJ, Eisig SB. A cephalometric comparative study of the soft tissue airway dimensions in persons with hyperdivergent and normodivergent facial patterns. J Oral Maxillofac Surg 1998 Feb;56(2):135-9; discussion 139. [doi: 10.1016/s0278-2391(98)90850-3] [Medline: 9461134]
18. Chen Y, Tanaka S, Hunter IW. Disposable endoscope tip actuation design and robotic platform. 2010 Presented at: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology; August 31-September 4, 2010; Buenos Aires, Argentina p. 2279-2282. [doi: 10.1109/IEMBS.2010.5627677]
19. Yoon HS, Yi BJ. A 4-DOF flexible continuum robot using a spring backbone. 2009 Presented at: 2009 International Conference on Mechatronics and Automation; August 9-12, 2009; Changchun, China. [doi: 10.1109/ICMA.2009.5246612]
20. Lee S, Park Y, Cho S, Kang J, Lee D. Technical skills and training of upper gastrointestinal endoscopy for new beginners. World J Gastroenterol 2015 Jan 21;21(3):759-785 [FREE Full text] [doi: 10.3748/wjg.v21.i3.759] [Medline: 25624710]
21. Daudov IM, Orobey MN, Ignatev IV. Bluetooth based technology for industrial personnel local positioning. IOP Conf. Ser.: Mater. Sci. Eng 2021 Mar 01;1111(1):012029. [doi: 10.1088/1757-899x/1111/1/012029]
22. Recommended upload encoding settings. Google Support. URL: https://support.google.com/youtube/answer/1722171?hl=en [accessed 2021-04-04]
23. Chan H. Comparing wireless data network standards. 2007 Presented at: AFRICON; September 26-28, 2007; Windhoek, South Africa. [doi: 10.1109/afrcon.2007.4401623]
24. Bearelly S, Cheung SW. Sensory Topography of Oral Structures. JAMA Otolaryngol Head Neck Surg 2017 Jan 01;143(1):73-80. [doi: 10.1001/jamaoto.2016.2772] [Medline: 27684535]
25. Aviv JE, Martin JH, Jones ME, Wee TA, Diamond B, Keen MS, et al. Age-related changes in pharyngeal and supraglottic sensation. Ann Otol Rhinol Laryngol 1994 Oct 28;103(10):749-752. [doi: 10.1177/000348949410301001] [Medline: 7944164]
26. Beales PH, Al-Khaled MJ. Procedures in practice. Laryngoscopy. Br Med J 1980 Mar 01;280(6214):629-631 [FREE Full text] [doi: 10.1136/bmj.280.6214.629] [Medline: 7370612]
27. Kamran M, Qamar R. An easy and effective way to reduce gag during orthodontic impression recording. Pakistan Orthodontic Journal 2016;8(1):17-20 [FREE Full text]
28. Means CR, Flenniken IE. Gagging--a problem in prosthetic dentistry. J Prosthet Dent 1970 Jun;23(6):614-620. [doi: 10.1016/0022-3913(70)90224-6] [Medline: 4911058]
29. Singh S, Ali FM, Nazirkar G, Dole VK, Gaikwad B. Gag- Etiology and Its Skillfull Management- A Review. Journal of Evolution of Medical and Dental Sciences 2013 Mar 11;2(10):1509-1516. [doi: 10.14260/jemds/416]

Abbreviations

2DOF: 2-degree-of-freedom
OTG: on-the-go

Edited by G Eysenbach; submitted 14.06.21; peer-reviewed by W Choi, Y Man; comments to author 08.08.21; revised version received 25.08.21; accepted 13.09.21; published 18.10.21.

Please cite as:
Moon Y, Hyun J, Oh J, Lee K, Lee YS, Kim JK
A Wi-Fi–Based Mask-Type Laryngoscope for Telediagnosis During the COVID-19 Pandemic: Instrument Validation Study
J Med Internet Res 2021;23(10):e31224
URL: https://www.jmir.org/2021/10/e31224
doi:10.2196/31224
PMID:34518154

Original Paper

Revealing the Nature of Chronic Obstructive Pulmonary Disease Using Self-tracking and Analysis of Contact Patterns: Longitudinal Study

Klaus Phanareth[1], MD, PhD; Astrid Laura Dam[2], BSc; Martin A B C Hansen[1]; Signe Lindskrog[2], MSc; Søren Vingtoft[3], MD; Lars Kayser[2], MD, PhD

[1]Epital Health, Gentofte, Denmark

[2]Department of Public Health, University of Copenhagen, Copenhagen, Denmark

[3]Region Zealand, Sorø, Denmark

Corresponding Author:
Lars Kayser, MD, PhD
Department of Public Health
University of Copenhagen
Øster Farimagsgade 5
Copenhagen, 1353
Denmark
Phone: 45 26822657
Email: lk@sund.ku.dk

Abstract

Background: Chronic obstructive pulmonary disease (COPD) is the fourth leading cause of death and is characterized by a progressive loss of pulmonary function over time with intermittent episodes of exacerbations. Rapid and proactive interventions may reduce the burden of the condition for the patients. Telehealth solutions involving self-tracking of vital parameters such as pulmonary function, oxygen saturation, heart rate, and temperature with synchronous communication of health data may become a powerful solution as they enable health care professionals to react with a proactive and adequate response. We have taken this idea to the next level in the Epital Care Model and organized a person-centered technology-assisted ecosystem to provide health services to COPD patients.

Objective: The objective is to reveal the nature of COPD by combining technology with a person-centered design aimed to benefit from interactions based on patient-reported outcome data and to assess the needed kind of contacts to best treat exacerbations. We wanted to know the following: (1) What are the incidences of mild, moderate, and severe exacerbations in a mixed population of COPD patients? (2) What are the courses of mild, moderate, and severe exacerbations? And (3) How is the activity and pattern of contacts with health professionals related to the participant conditions?

Methods: Participants were recruited by convenience sampling from November 2013 to December 2015. The participants' sex, age, forced expiratory volume during the first second, pulse rate, and oxygen saturation were registered at entry. During the study, we registered number of days, number of exacerbations, and number of contact notes coded into care and treatment notes. Each participant was classified according to GOLD I-IV and risk factor group A-D. Participants reported their clinical status using a tablet by answering 4 questions and sending 3 semiautomated measurements.

Results: Of the 87 participants, 11 were in risk factor group A, 24 in B, 13 in C, and 39 in D. The number of observed days was 31,801 days with 12,470 measurements, 1397 care notes, and 1704 treatment notes. A total of 254 exacerbations were treated and only 18 caused hospitalization. Those in risk factor group D had the highest number of hospitalizations (16), exacerbations (151), and contacts (1910). The initial contacts during the first month declined within 3 months to one-third for care contacts and one-half for treatment contacts and reached a plateau after 4 months.

Conclusions: The majority of COPD patients in risk factor group D can be managed virtually, and only 13% of those with severe exacerbations required hospitalization. Contact to the health care professionals decreases markedly within the first months after enrollment. These results provide a new and detailed insight into the course of COPD. We propose a resilience index for virtual clinical management making it easier to compare results across settings.

(J Med Internet Res 2021;23(10):e22567) doi:10.2196/22567

KEYWORDS

COPD; exacerbations; patient-reported outcomes; Epital Care Model; early interventions

Introduction

Chronic disease poses one of the biggest challenges facing Western health care systems, and with its increasing prevalence in the population, it is a major driver of health care utilization and the leading cause of death [1]. Chronic obstructive pulmonary disease (COPD) is one of the chronic diseases that weighs heavily in this scenario. Currently the fourth leading cause of death in the world, COPD is projected to be the third leading cause of death by 2020 [2] with an increasing socioeconomic burden and growing strain on health care budgets [3]. COPD is characterized by a progressive loss of pulmonary function over time with intermittent episodes of exacerbations as part of its natural history [2,4]. The long-term course of the disease depends on how well individuals can understand, manage, and act upon risk factors such as smoking, nutritional state, exercise activity, and compliance to medical treatment plans [5]. In the short-term, the prevention and treatment of exacerbations are key elements in the management of COPD according to the Global Initiative on Chronic Obstructive Lung Disease (GOLD) as the number of exacerbation is closely associated with an increased rate of hospitalization, increased mortality, decline in lung function, increased use of health care resources, and decrease in quality of life [2,4]. According to the GOLD guidelines, COPD exacerbation is defined as “an acute worsening of respiratory symptoms that results in additional therapy.” These events are classified as mild, moderate, or severe, of which the first 2 categories are characterized by the type of medication given and the latter characterized by the patient’s need for hospitalization or visits to the emergency room [2].

Early detection of exacerbations and prompt interventions with antibiotics and corticosteroids can reduce hospital admissions, decrease use of health care services, and increase and improve quality of life [6,7]. Sixteen years ago, Wilkinson et al [8] showed that early recognition of exacerbations using symptom-based diaries completed by the patient followed by prompt treatment by a physician is beneficial to the recovery of the exacerbation and can improve the burden of exacerbation-related morbidity and mortality. The authors concluded a need for initiatives to encourage reporting of early presentation of exacerbations and enhanced symptom recognition tools that are easy to use and promptly managed by health professionals.

Telehealth solutions involving self-tracking of vital parameters such as pulmonary function, oxygen saturation, heart rate, and temperature with synchronous communication of health data may become a powerful solution in this context as it enables health care professionals to react with a proactive and adequate response. Over the last decade, data from several randomized controlled trials examining the effect of telehealth and how it may assist self-monitoring and management in COPD have been reported [9-12]. Some studies suggest a potential beneficial effect with a reduction of exacerbations [13], acute hospital admissions, and mortality [14] and increased health-related quality of life [15] whereas others have not been able to verify these results [16-18]. In COPD, no studies have yet demonstrated conclusively that telehealth is neither superior nor inferior to usual care; it is noteworthy that telehealth interventions have not caused any harm to COPD patients and the technology has been welcomed by the users [10,19,20].

One of the fundamental elements of a person-centered and value-based health care provision of services is the inclusion of the patients’ self-reported data including their experiences as patients. These data are often termed patient-reported outcomes (PROs) [21]. PROs are a natural component of telehealth as they can be used to report the condition dynamically over distance to service providers. PROs may also engage the patients and increase self-management due to a better insight into their own condition [21]. In the context of COPD, PROs are a key element for patients to communicate with their health providers, enabling them to understand their condition and proactively act upon potential deteriorations [22].

Telehealth and PROs may contribute to the fulfillment of the World Health Organization framework for integrated people-centered health services as they facilitate the engagement and empowerment of patients and the reorientation of how health care can be delivered [23].

In an attempt to demonstrate how these ideas and potentials can be realized, we have developed a new person-centered service model: the Epital Care Model (ECM). We have tested the ECM in a living lab [24] and evaluated potential candidates for PROs gathered by psychometric validated instruments focusing on self-management, health literacy, and physical and mental well-being as well as clinical parameters [25]. In this study, we report on the second part, in which we, based on a newly developed algorithm and setting (to be reported elsewhere), are able to evaluate how PROs relating to the clinical condition can be used to understand the actual needs of the patient and needs of resources required from the ECM living lab to help the patient return to their habitual condition.

For this purpose, we have followed the self-tracking activities of 87 patients with COPD and the corresponding standardized responses from health care professionals when responding to signs of deteriorations of the patients’ clinical condition or if the patients need support to manage their condition.

To understand the activities and required resources, we have addressed the following research questions:

- What is the incidence of mild, moderate, and severe exacerbations in a mixed population of COPD patients?
- What is the course of the mild, moderate, and severe exacerbations?
- How is the activity and pattern of contacts with health professionals related to the participant condition?

Methods

The setup and initial data of this study has been described previously by Phanareth [24]. In short, the data for this study was collected as part of the Epital living lab project which was established in a Danish municipality in 2012 [24].

Setting

The Epital Care Model is a model designed to help organize health care services both virtually and physically with a personalized health-centered approach and with the objective of enhancing empowerment and self-management. Services and interactions are organized according to 6 levels that can be divided into three 3 settings depending on the resources allocated. In the ECM1 level, participants live an independent and active life with their condition supported by technology. When participants need medical support in relation to their condition, they are classified as being at the ECM2 level. Both ECM1 and ECM2 interactions and services can be provided at any location and are performed virtually (setting 1). If the participant's condition deteriorates and a physical visit is necessary, nurses will go to the participant's physical location and assess the condition thereby placing the participant at the ECM3 level. If regular follow-ups by nurses and medical doctors are needed, this will be offered as a patient at home service termed "outmission" at the ECM4 level (setting 2). The outmission is equivalent to services offered when admitted to a hospital but in this case, they are provided at home by a virtual medical doctor (eDoctor) and a nurse (setting 3). The eDoctor performs virtual ward rounds but the patients can receive visits from community nurses at the participant's physical location if necessary. At the ECM5 and ECM6 levels, the participant is now considered a patient and is referred to a community or hospital bed depending on the need of care and monitoring.

Population

The study period took place from November 1, 2013, to December 31, 2015, and in this period, we consecutively enrolled participants recruited by a snowball-like method based on convenience sampling. This resulted in a random sampling over seasons and an uneven period for observations defined by either the time from enrollment to the planned end of the study or a period defined as the time from enrollment to the participant's death or choice to withdraw from the study.

A total of 93 participants with COPD were enrolled into the ECM living lab, of these 27 were transferred directly from the prior pilot study and 66 were recruited throughout the period of the study. For this analysis, 5 were excluded due to not having provided any measurements after the inclusion and one due to insufficient inclusion data, resulting in a total of 87 participants.

During the study, 29 patients dropped out: 9 of whom because they died, 8 for reasons not associated to the tracked exacerbations, and 12 for other reasons. In addition, 5% (4/87) of participants dropped out within 14 days, but data from these participants are included as an objective of this study was to assess the overall activity and event of the participants.

At entry, all participants were registered with baseline data including age, sex, heart rate, oxygen saturation, forced expiratory volume for 1 second (FEV_1), and a clinical examination [25]. Based on these data, participants were classified according to the previous GOLD guidelines from 2011 [26], as the study was planned in 2013. They were classified according to the air flow limitation severity in COPD (GOLD classes I-IV) and their risk factor group (GOLD risk factor groups A-D) assessed by the number of exacerbations per year and their Medical Research Council breathless test score. We recognize that these recommendations have changed in new edition of the GOLD guidelines and our data may therefore not be directly comparable to newer data sets [26].

Context and Data Collection

The ECM is a unique, personalized health-centered living lab project that consists of an environment and service ecosystem where it is possible to investigate everyday fluctuations when living with COPD. Simultaneously, the ECM makes it possible to estimate how many resources are needed to meet the needs of the participants both when the participants live an active and independent life in ECM1 and when their condition deteriorates and they are allocated resources as described in ECM2 to ECM6.

The ECM living lab provides the participants with 24/7 access to assistance from the response and coordination center (RCC) manned by nurses supported by eDoctors. The self-tracking activities are completely regulated by the participants, driven by either curiosity or needs in relation to the ongoing changes of their condition. As the participants register their data, they are simultaneously reported to the RCC. The participants are informed that their data will be visible for the RCC staff, they can expect the RCC to respond if their condition deteriorates, and self-tracking is digital assistance helping them to manage their condition. The participants are also able to contact the RCC through their tablet using video or voice communication if they feel insecure or have any questions [24].

A third kind of communication may be initiated by the RCC staff: when they register a deterioration, they will contact the participant to investigate whether there is any need of assistance. For the participants, this data-driven approach should result in a structured and less intrusive way to live with their COPD as they only need to report their PRO data when they feel like it and only contact the RCC if they need advice, care, or treatment.

The information technology (IT) support for processes and activities in the ECM living lab constitutes 2 IT systems specially tailored for this purpose [24]: a citizen-oriented monitoring system (Appinux Care, Appinux A/S) and a medical documentation system (EpiProcess, OCEAN Process A/S).

The COPD monitoring module in the Appinux platform allows participants to deliver real-time data to health care providers. Participants used a Samsung tablet with an app to report their data used to assess their condition. The participants themselves reported a measured body temperature, increased breathlessness, increased coughing, and changed color of sputum. These data were reported together with a guided semiautomatic measurement of heart rate, oxygen saturation, and FEV_1. The initial assessment of the data is performed by the system, and an algorithm presents this assessment to the participant as a

color (red, yellow, green) along with advice on the appropriate action to take.

Based on these data, the health care professionals can, together with the participant, make informed decisions on how to best manage the change in condition. This will be based on the previous measures evaluated with graphs, including plotted trends, which enable the health professionals to better predict the course of the participant. All participants were equipped with an acute medicine box at home, to be used only when exacerbations occurred to avoid delays in the initiation of a medical treatment. The content was prescribed by the eDoctor, provided by a local pharmacy, and used only in agreement with the RCC staff. The acute medicine box contained bronchodilators (short-acting β_2-agonist and long-acting β_2-agonist), inhaled corticosteroids, prednisolone tablets, and broad-spectrum antibiotic tablets.

The medical treatment was standardized to match 3 categories of exacerbations (mild, moderate, and severe) using a stepwise approach with increasing intensity of the medication to match the severity (Table 1). The initiation of medical treatment is based on symptoms and the algorithm system using the PRO data. Based on these and the history, the participant was treated with drugs according to Table 1. The pattern of used drugs was used afterward to classify the participant's severity of exacerbation for the purpose of this study.

Table 1. Description of severity based on medical prescription.

Corresponding medical treatment	Exacerbation type
Increase of ICS[a] and LABA[b] + SABA[c]	Mild
Increase of ICS and LABA + SABA + oral prednisolone[d]	Moderate
Increase of ICS and LABA + SABA + oral prednisolone + antibiotics[e]	Severe

[a]ICS: inhaled corticosteroid.

[b]LABA: long-acting β_2-agonist.

[c]SABA: short-acting β_2-agonist.

[d]Course of 37.5 mg daily for 5 to 10 days.

[e]Amoxcillin 500 mg × 3 daily for 10 days.

Our definition of severe exacerbation differs from the one purposed in the GOLD guidelines, as GOLD classifies this as a condition requiring hospitalization or emergency room visits. In our setting, we defined a severe exacerbation as a condition where an increase of inhaled corticosteroid, long-acting β_2-agonist, short-acting β_2-agonist, oral prednisolone, and antibiotic was needed (Table 1); this resembled more a moderate exacerbation compared to GOLD. In our case, however, if a severe exacerbation treatment was initiated it triggered a series of standardized health services including a follow-up program with daily virtual ward rounds, close monitoring of the condition, and medical adjustments if needed (ECM3 and ECM4).

We therefore argue that our approach in a combined virtual and physical setting is comparable to the procedures performed during a hospitalization or in the emergency room, and therefore may be covered by the definition of severe exacerbation as proposed in the GOLD guidelines.

Statistical Analysis

At entry, we registered sex, age, FEV_1, pulse rate, and oxygen saturation for each participant. During the study, we registered participant observation period, number of exacerbations, and number of contact notes, which were coded into care and treatment notes. Each participant was at entry classified according to the GOLD classes I through IV and placed in a risk factor group A through D. Categorical data were reported as numbers and percentages and continuous data were reported as mean values with standard deviations.

Age at entry; participant days; number of mild, moderate, severe, and total exacerbations; and number of measurements per participant year were compared between the risk factor groups using 1-way analysis of variance. All statistics were performed using Excel (version 2016, Microsoft Corp).

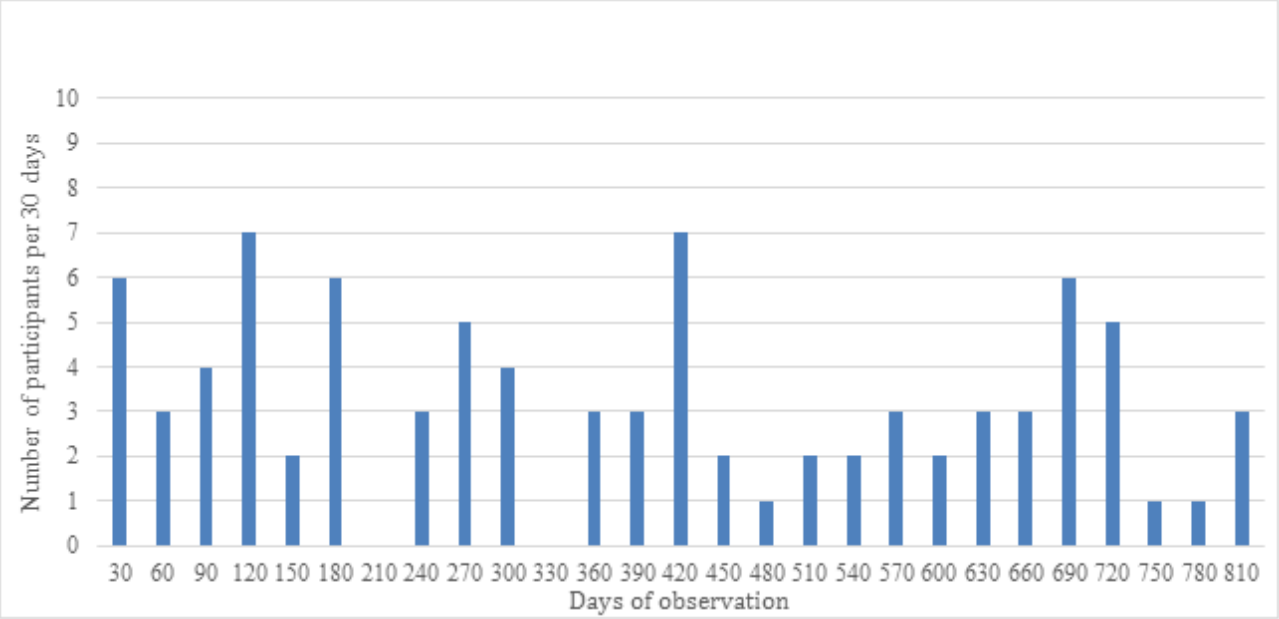
Ethics and Data Protection

As previously reported, the study was assessed and found not to need specific approval by the regional office of the National Danish Ethics Committee (H-3-2012-FSP31). The program was also registered with the National Danish Data Agency by first the University of Copenhagen (2012-41-0384) and since January 2014 by the municipality of Lyngby-Taarbæk, Denmark (20150910229). All participants were told about the study orally and with written material and informed that they could withdraw from the services at any time. Each participant signed a consent with a copy of this information. All data were handled according to Danish legislation and regulations.

Results

The study population consisted of 58 females and 29 males. The average age was 73.7 (range 47-91) years. The accumulated number of self-tracked measurements was 12,470. The accumulated number of days with self-tracking was 31,714, and the average of self-tracking days per participant was 368.8 (SD 248.7) with a range from 1 to 791 days per participant. As seen in Figure 1, only 7% (6/87) of the participants dropped out within the first month and 85% (74/87) participated for more than 3 months.

Figure 1. Observation period for the participants.



The number of participants classified both according to risk factor groups and severity is shown in Table 2. More than half of all patients were characterized with severe or very severe COPD. The mean FEV_1 and standard deviation at entry served as baseline for the evaluation of the participants during the observation period (also shown in Table 2).

Table 2. Forced expiratory volume during the first second at baseline of the study by participants' GOLD class and risk group.

GOLD[a]	A: MRC[b] score <2; 0-1 exacerbations per year, mean (SD)	B: MRC score ≥2; 0-1 exacerbations per year, mean (SD)	C: MRC score 0-1; ≥2 exacerbations per year, mean (SD)	D: MRC score ≥2; ≥2 exacerbations per year, mean (SD)	Total
I: FEV_1[c]≥80% predicted	—[d]	—	—	1.46, n=1	1
II: 50%≤FEV_1<80% predicted	1.47 (0.46), n=11	1.52 (0.46), n=23	1.49 (0.05), n=3	0.86 (0.26), n=4	41
III: 30%≤FEV_1<50% predicted	—	1.28, n=1	1.10 (0.08), n=8	0.74 (0.22), n=16	25
IV: FEV_1<30% predicted	—	—	0.97 (0.05), n=2	0.67 (0.3), n=18	20
Total	11	24	13	39	87

[a]GOLD: Global Initiative on Chronic Obstructive Lung Disease.

[b]MRC: Medical Research Council.

[c]FEV_1: Forced expiratory volume during the first second.

[d]Not applicable.

The number of participants were unevenly distributed across the severity classes and risk factor groups (Table 2). Therefore, we adjusted for observation period by normalizing data into participant years or days of observation (Table 3).

The mean number of measurements per participant days did not differ between the risk factor groups, although risk factor group D tended to have higher activity. The numbers are less than one per day and equivalent to a measurement every second to fifth day but with a wide range from 0.01 and up to 3.8 measurements per day.

The number of exacerbations divided into mild, moderate, and severe did not differ significantly among the groups, although the annual rate per patient tended to be higher in risk factor group D with risk factor group B having the second highest mean value. The mean numbers of care notes and treatment notes were similarly higher in risk factor group D, and risk factor group B had the second highest mean of care notes per participant year. The mean for treatment notes per participant year deviated from this pattern as risk factor group C had almost the same number as risk factor group B. This was primarily due to one participant in group C who had 93 treatment notes over a period of 219 days.

A total of 21 of the contact notes were related to other issues than living with COPD, 2 related to treatment of urinary tract infection, 2 to peripheral edema treated with diuretics, and 2 to acute need of hospitalization for other reasons. None of the contact notes was related to treatment or care of other somatic long-term conditions such as ischemic heart failure or diabetes.

In Table 3, the characteristics of exacerbations and contact notes distributed within risk groups are provided.

Table 3. Distribution of exacerbations and contacts within risk groups.

Characteristic	Risk factor				
	A (n=11)	B (n=24)	C (n=13)	D (n=39)	Total (n=87)
Participant years, mean (SD)	0.77 (0.57)	1.12 (0.68)	1.16 (0.72)	0.94 (0.66)	1.00 (0.67)
Measurements per participant day, mean (SD)	0.28 (0.18)	0.43 (0.36)	0.35 (0.27)	0.57 (0.55)	0.45 (0.46)
Total number of exacerbations	11	70	23	150	254
Exacerbations/year, mean (SD)	1 (1.49)	2.92 (3.36)	1.77 (2.35)	3.87 (4.76)	2.93 (3.89)
Mild exacerbations	4	29	9	34	76
Mild exacerbation/year (SD)	1.11 (2.91)	1.07 (1.77)	0.52 (1.02)	1.06 (1.71)	0.99 (1.81)
Moderate exacerbations	2	11	7	37	57
Moderate exacerbation/year, mean (SD)	0.18 (0.45)	0.38 (0.97)	0.64 (1.84)	1.06 (1.96)	0.70 (1.61)
Severe exacerbations	5	30	7	79	121
Severe exacerbations/year, mean (SD)	0.28 (0.53)	1.74 (3.50)	0.37 (0.58)	1.85 (2.17)	1.39 (2.41)
Care notes	102	382	124	789	1397
Care notes/year, mean (SD), (n=84)	23.33 (29.06)	21.94 (23.90)	12.33 (11.78)	31.13 (33.55)	24.80 (28.52)
Treatment notes	53	364	166	1121	1704
Treatment notes/year, mean (SD)	5.29 (6.90)	14.20 (22.00)	14.85 (42.40)	29.01 (32.27)	20.28 (30.60)
Total contact notes/year mean (SD)	28.62 (33.41)	37.43 (34.75)	27.18 (50.18)	57.71 (44.74)	45.31 (45.41)

The participants' total number of contacts were primarily related to ECM2 with the virtual clinical management. Less than 5.36% (165/3081) of the contact notes related to more resource-demanding activities such as evaluation in the participants' home by a mobile acute team (ECM3 and 4), referral to community bed care (ECM5) or hospitalization (ECM6). As seen with the number of exacerbations, the highest number of contacts were in risk factor group B and D (Table 4). This is partly due to the higher number of participants in these 2 groups, as the adjustment for participant years (Table 3) reduces the difference between group A and C versus group B and D. The differences, particularly in group C, may also be explained by a lower need for contacts not related to treatment as these contacts contribute to the largest difference between group C versus group B and group D (Table 3).

Table 4. Number of contact notes (%) within risk factor distributed at ECM levels.

Risk factor	ECM1[a]	ECM2[b]	ECM3[c]	ECM4[d]	ECM5[e]	ECM6[f]	Discharge notes	Total
A	24 (0.8)	129 (4.2)	1 (0.0)	2 (0.1)	0	0	__[g]	156 (5.1)
B	86 (2.8)	629 (20.4)	6 (0.2)	10 (0.3)	3 (0.1)	6 (0.2)	6 (0.2)	746 (24.2)
C	31 (1.0)	242 (7.9)	5 (0.2)	7 (0.2)	0	2 (0.1)	1 (0.0)	288 (9.6)
D	139 (4.5)	1615 (52.4)	35 (1.1)	35 (1.1)	28 (0.9)	25 (0.8)	14 (0.5)	1891 (61.3)
Total	280 (9.1)	2615 (84.9)	47 (1.5)	54 (1.8)	31 (1.0)	33 (1.1)	21 (0.7)	3081 (100)

[a]ECM1: Epital Care Model 1—Digitally facilitated active and independent living.

[b]ECM2: Epital Care Model 2—Virtual clinical management.

[c]ECM3: Epital Care Model 3—Clinical management @home, digitally assisted.

[d]ECM4: Epital Care Model 4—Outmitted @home.

[e]ECM5: Epital Care Model 5—Admitted to local community care facility.

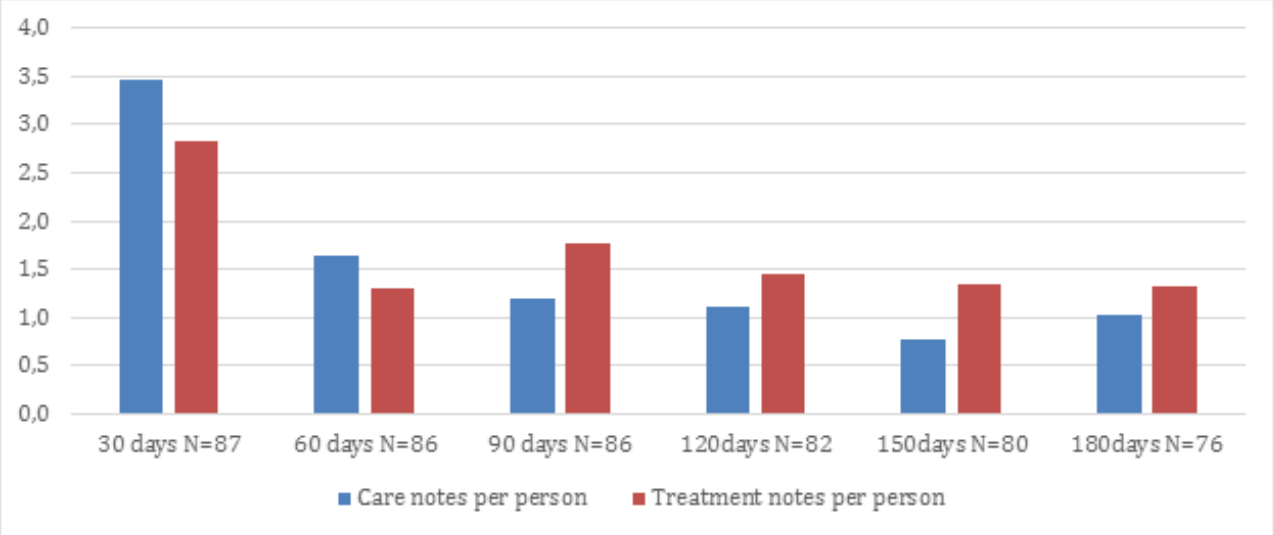
[f]ECM6: Epital Care Model 6—Admitted to hospital.

[g]Not applicable.

The numbers of care and treatment notes are highest in the first month after enrollment in the ECM (Figure 2). The care notes decline over time with a plateau after 3 months at approximately one-third of the initial number of care notes. The number of

treatment notes were initially lower than the care notes and reach a plateau after 4 months at approximately half of the initial number of treatment notes. Most of the care and treatment notes (2111/3060, 68.99%) were conducted in the daytime hours between 8:00 AM and 5:00 PM.

Figure 2. Number of care contact notes and treatment contact notes per person over the first 6 months after enrollment/on-boarding, reported in intervals of 30 days.



Most of the exacerbations were managed virtually in the context of the RCC (ECM2), and 69.4% (84/121) of the severe exacerbations were mitigated virtually (Table 5). Only 21.3% (54/254) of all exacerbations required more intensive diagnostics and follow-up with the majority of these belonging to the severe exacerbations. Only 7.1% (18/254) of the exacerbations required hospitalization and an equivalent number of exacerbations required hospitalization at home (ECM4).

Table 5. Number of exacerbations (%) in relation to severity of the exacerbations and the allocated resources (ECM level).

Exacerbation type	ECM2[a]	ECM3[b]	ECM4[c]	ECM5[d]	ECM6[e]	Total
Mild	72 (94.7)	2 (2.6)	2 (2.6)	—[f]	—	76 (100)
Moderate	44 (77.2)	6 (10.5)	5 (8.8)	—	2 (3.5)	57 (100)
Severe	84 (69.4)	4 (3.3)	13 (10.7)	4 (3.3)	16 (13.2)	121 (100)
Total	200 (78.7)	12 (4.7)	20 (7.9)	4 (1.6)	18 (7.1)	254 (100)

[a]ECM2: Epital Care Model 2—Virtual clinical management.

[b]ECM3: Epital Care Model 3—Clinical management @home, digitally assisted.

[c]ECM4: Epital Care Model 4—Outmitted @home.

[d]ECM5: Epital Care Model 5—Admitted to local community care facility.

[e]ECM6: Epital Care Model 6—Admitted to hospital.

[f]Not applicable.

Discussion

Principal Findings

This study is, to the best of our knowledge, the first to describe the activities and interactions that take place in an environment based on telehealth and a systematic use of PRO data. An environment was designed to meet the needs of patients and provide an active and independent living in accordance with the person-centered health service model, the ECM [24].

The ECM environment with its service ecosystem offers a unique opportunity to more thoroughly explore the nature of COPD in relation to the severity described as GOLD classes and risk factor groups and the required resources to assist the participants to adequately manage their own condition.

In our study, we detected a total of 254 exacerbations in 87 patients based on PRO measures with a subsequent clinical assessment that resulted in medical treatment corresponding to either mild, moderate, or severe exacerbation (Table 1). Overall, we found 59% (150/254) of the exacerbations occurred in the high-risk group (risk factor group D) with an annual rate of 3.9 exacerbations per participant year, and we found that approximately half of these were categorized as severe exacerbations with 1.86 severe exacerbations per participant year. These results are in the higher end compared to other studies that have also examined annual rates of exacerbations. Comparison of our data with other studies is difficult as it requires a standardization of the severity of the COPD condition and a classification of severity of exacerbations across studies. Our study design builds on the 2011 GOLD guidelines and a

pragmatic approach to classification of severity taking the strength of telehealth into consideration, as described in the methods section. Wilkinson et al [8] found, in accordance with our study, a mean of 2.5 exacerbations per patient year, but 40.1% of these were not detected by the clinicians and covered a broad spectrum of patients with respect to severity.

In a review by Seemungal et al [27], the authors found an estimated annual rate of COPD exacerbations as low as 0.5 to a high of 3.5 exacerbations per patient year and hospitalization rates with a range from 0.09 to 2.4 per patient year, the latter being categorized as severe exacerbation according to the GOLD definition [26].

It is surprising that the annual rate of exacerbations in risk factor group C tends to be lower, although not significantly, than in risk factor groups B and D. This may be due to the low number of participants in risk factor group C and is thereby a statistical coincidence and not real finding. In Denmark, patients with COPD will be referred to pulmonary specialists when they move from risk factor group B to C and this may also be a factor that can explain why group C has a lower annual rate of exacerbations than group B.

Our data, which build on a 24/7 self-tracking setup with immediate response from RCC staff, provide us with a detailed insight into the nature of the COPD condition and reports of deteriorations, which may have been missed in other settings. This may explain the higher number of total exacerbations (3.9 per participant year), whereas the number of severe exacerbations is closer to the other studies. Had we used a classification of severity similar to the GOLD guidelines by only including those in need of physical contact to health professionals in a setting equivalent to an emergency room or hospitalization, the number would only have been 37 of the detected 254 exacerbations, which is 31% of the 121 considered to be severe exacerbations based on our criteria or only 15% of all 254 exacerbations. It is also noteworthy that 2 of those initially classified as moderate exacerbations resulted in hospitalization and 16 of the 121 classified as severe exacerbations resulted in hospitalization. If the stricter approach was followed, by only classifying those in ECM3 to ECM6 as severe, 21 of 37 were not hospitalized but mitigated before hospitalization.

The number of exacerbations leading to hospitalizations is similar to the 6.4% found by Wilkinson et al [8]. Whether this is a result of a lack of effect of our setting or due to our inclusion of a higher number of participants in risk factor group D remains to be explored. In a study by Pinnock et al [19] comparing telehealth with conventional care they found an annual rate of number of hospitalizations per patient to be 1.2. This is considerably higher than our 18 hospitalizations out of 87 participants with 0.21 hospitalizations per participant year.

When comparing exacerbation rates, one must therefore be careful and very aware of the lack of consensus of clear definitions of exacerbations, grading systems for the severity of the disease, the high heterogeneity among studies, and the different treatment culture among clinicians which also have an influence.

The higher overall exacerbation rates in this study may be explained by the voluntary self-tracking setup with an easy way to report PRO data, which provides far more data than one would normally get in a traditional health system. This may provide a higher sensitivity for detecting exacerbations but may also decrease the specificity in comparison to traditional systems as the threshold for acting on changes in the pulmonary condition may be lowered. The higher occurrence of exacerbations in our study compared to others, may be explained by the notion that exacerbations go unreported [8,28,29].

Whether a lower threshold for detection and an apparent increased administration of drugs will be reasonable and be considered proactive by being able to prevent a deterioration of the COPD over time by a more aggressive handling of the acute exacerbations needs further studies.

Association Between Clinical Condition and RCC-Mediated Activities

When organizing a 24/7 service available for all individuals with COPD in a geographical area, it is important to understand whom will need which services and how these needs will evolve over time. As expected, most of the severe exacerbations and hospitalizations were associated with the participants in risk factor group D. This group also had the highest number of contacts, both related to care and treatment. Most of contacts were handled at the virtual clinical management level with 84.1% taken place in relation to ECM2 and with an additional 9.1% only 6% of all contacts were in relation to the more resource demanding activities in ECM3-ECM6.

We find that the activities as estimated by contact notes were more than twice as high during the first month of enrollment into the living lab than after 3 to 4 months. Even the number of treatments notes declined, indicating that the participants learned to better manage their condition or became less distressed. This is in accordance with Kargiannakis et al [30] who, in a service ecosystem comparable to ours, found that the initial contacts per participant dropped from 7 per day to 4 within the first week. The total number of contacts over a period of 42 days with 23 participants were considerable higher with a total of 451 or approximately 14 per participant within the first month [30].

Our results are in accordance with our previous finding that the participants' emotional distress decreases over time [25]. The reduced number of contacts to the RCC may also reflect the finding that the participants reported a perception of being less active in managing their condition [25]. Based on both the previous findings and the number of contacts, will we hypothesize that the ECM living lab environment and service ecosystem contributes to the participants ability to better manage the COPD and related challenges. It should also be noted that the majority of contacts were in the daytime, which indicates a relatively low need for urgent and often more costly contacts in the evening and at night.

Another important finding is that when an exacerbation is mild when detected, the likelihood of needing physical contact to the health professionals in the course is low (< 5%) and with moderate exacerbations this increases to approximately 23% whereas approximately 31% with severe exacerbations require

physical evaluation and approximately half of these needs to be hospitalized.

Virtual Clinical Management Resilience Index

The increasing digitalization and virtualization of clinical care and the integration of services across providers call for an indicator of the resilience of the ecosystem of services that can be provided by the circles of care to individuals in a way that supports an active and independent living assisted by technology and with assistance when their condition periodically deteriorates. The ECM1 and ECM2 is an example of such ecosystem, which is supported by the blended physical and virtual services provided in the participant's own home (ECM3 and ECM4) or referral to a bed outside the participant's home (ECM5 or ECM6) in case of severe deterioration. The capacity to manage the individual in ECM1 and ECM2 relates to the competence of the health professionals to manage acute episodes remotely, their access to technology that supports their decisions, equipment in the participants' homes that enables the handing of even more complex situations, and finally the participant's and their relatives' understanding of and ability to manage the fluctuations of their condition.

As our data points to not only the classification of COPD with respect to severity and type of exacerbations, the context and resources also need to be reported in comparable manners. Instead of comparing technology and capacity of all the involved persons, we suggest an index that describes the resilience of the ecosystem of health services provided virtually as an indicator of maturity and capacity to manage participants most efficiently with focus on their well-being.

We propose a Virtual Clinical Management Resilience Index (VCMRI) that reports on the success of keeping patients within the setting of the virtual clinical environment. The ratio should be calculated as the number of cases that can be managed in a virtual clinical environment divided by the total number of patients.

In the context of ECM, a VCMRI will be a ratio of the number of exacerbations that can be reversed at ECM2. We found 121 severe exacerbations among the participants in our study and by using the formula on VCMRI with ECM2 in the denominator, we calculated a VCMRI of 0.694 for ECM2 meaning that almost 70% of severe exacerbations mitigated in ECM2. This is of interest as all services in ECM2 are performed virtually without any physical contact to the surrounding health care system. The VCMRI may be an important tool for the development of future integrated systems as it can enable the documentation and comparison of the health care services' ability to keep the patient at home, which is a service goal for most patients, who prefer receiving treatment and care in their own home [31,32]. Therefore, the VCMRI may contribute to the development of more integrated and personalized health care services fulfilling the patient's wish to be at home.

With its model for describing activities, roles, contexts, and technologies, the ECM can be used to allocate resources both within the virtual ecosystem and in the physical activities from ECM3 to ECM6. In reporting the resilience index, the efforts and resources can be reported in accordance with the structure of the ECM with respect to roles of the actors, their activities, the technology involved, and geographical locations as well as roles and certification of the involved actors. Future studies reporting in this context may map themselves into this model and compare with our results.

Strengths and Limitations

The strength of our study is the close contact to the participants in the period they are connected to the living lab. This provides our RCC staff with an insight into the everyday life of participants and how they manage their condition. The use of technology to document all the events for this study imposes a limitation as we cannot be entirely sure that all communications and actions are documented. Also, we do not have access to other health care IT systems at the individual level, which may result in an underestimation of hospital admissions relating to other conditions.

Another limitation is the recruitment of participants as it is a convenience sampling from the local community and without direct referral from hospitals or general practitioners. We may therefore miss those who were not able to use technology or were disengaged from the services where we recruited. With a more vigorous inclusion, we might be able to include more COPD patients, who have difficulties in managing their health. This could result in more severe courses including a higher rate of hospitalization as Wilkinson et al [8] demonstrated that those not seeking their GP or hospital specialists are in a higher risk of deterioration.

The observation period varied from 1 day to more than 2 years, and the lack of exclusion of the 4.6% who participated for less than a 14-day period may contribute to a lack of statistical differences between the risk factor groups with respect to the number of exacerbations and contact notes, as normalization of these data to patient year may have increased the variance of the data set. We have kept all the participants regardless of days spent in the study to conform to our research protocol, and we also find it important to keep these data as information about how early withdrawal contributes to the full picture of the population. It should be noted that all participants have been through 2 visits in their home by both a medical doctor and a member of the service team who introduced them to the equipment for self-tracking. Thus, withdrawal after only a few days signifies that despite of the enrolled participants being well informed, there is still a small group of participants who find it too difficult when left to self-monitoring. Unfortunately, for ethical reasons, we were not permitted to ask for a reason for the withdrawal.

Our study is an observational study without any controls. We are therefore not able to document the effect of the ECM intervention compared to how the participants would behave in relation to their contact pattern to conventional health services.

Perspective

We recommend future studies focus on participants who have been enrolled at least 90 days prior to observations for 2 reasons. The first is to be able to obtain a steady state situation before activities and outcomes of the intervention are observed. The second is from a resource perspective, as our data indicate that

the resources and efforts during enrollment and the first 30 days can only be justified if the COPD patients are using the services for a longer period, as their need for resources declines after 90 days (Figure 2).

Future studies should also be more focused on a more appropriate number of participants in each of the GOLD classes and risk factor groups, and they should also include matched controls or include a randomized control group.

Conclusion

The establishment of a person- and health-centered and digitally supported environment to assist people living with COPD has offered us a unique opportunity to learn about the nature of COPD over a period of up to 2 years. We find a large variation in both to what extent the participants engage themselves in self-monitoring activities and their need for contact with health professionals. The participants who are in most need of interaction with the health professionals (both virtual and nonvirtual) and have with the highest number of exacerbations belong to risk factor group D. The participants' interactions with the RCC staff were highest during the first month of enrollment as also seen in other studies, indicating that the participants develop their capability to manage their own condition over time. Our results are difficult to compare to other technology-assisted settings and we therefore suggest a resilience index, the VCMRI. Together with the ECM, the resilience index could be used as a scaffold to report on the success with mitigating participants at the virtual clinical management level and at the same time report on involved capacity. We also find a need for a classification of severe exacerbation which can translate the current GOLD recommendation of hospitalization as a criterion with resources and efforts comparable to these in a telehealth care setting including the practice of referral to hospitalization at home or community care.

Acknowledgments

KP and SV have received support from the 3P project from the Norwegian Health Foundation during the period 2016-2018 while contributing to the manuscript. The authors thank the health professionals, patients, administrators, and politicians at the municipality of Lyngby-Taarbæk for their active participation in the project. They would also like to thank Charlotte Dorph Lyng and Jakob Sylvest Nielsen for their active engagement.

Conflicts of Interest

KP is a partner in the company Epital Health, which provides services based on the ECM.

References

1. Milani RV, Lavie CJ. Health care 2020: reengineering health care delivery to combat chronic disease. Am J Med 2015 Apr;128(4):337-343. [doi: 10.1016/j.amjmed.2014.10.047] [Medline: 25460529]
2. GOLD. Global strategy for the diagnosis, management, and preventien of chronic obstructive pulmonary disease (2018 report). 2018. URL: https://goldcopd.org/wp-content/uploads/2017/11/GOLD-2018-v6.0-FINAL-revised-20-Nov_WMS.pdf [accessed 2021-09-08]
3. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 2006 Nov;3(11):e442 [FREE Full text] [doi: 10.1371/journal.pmed.0030442] [Medline: 17132052]
4. Viniol C, Vogelmeier CF. Exacerbations of COPD. Eur Respir Rev 2018 Mar 31;27(147):170103 [FREE Full text] [doi: 10.1183/16000617.0103-2017] [Medline: 29540496]
5. GOLD. Chronic obstructive pulmonary disease in over 16s: diagnosis and management. National Institute for Health and Care Excellence. 2018. URL: https://www.nice.org.uk/guidance/ng115/resources/chronic-obstructive-pulmonary-disease-in-over-16s-diagnosis-and-management-pdf-66141600098245 [accessed 2019-12-02]
6. Seemungal TA, Donaldson GC, Paul EA, Bestall JC, Jeffries DJ, Wedzicha JA. Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 1998 May;157(5 Pt 1):1418-1422. [doi: 10.1164/ajrccm.157.5.9709032] [Medline: 9603117]
7. Guarascio AJ, Ray SM, Finch CK, Self TH. The clinical and economic burden of chronic obstructive pulmonary disease in the USA. Clinicoecon Outcomes Res 2013;5:235-245 [FREE Full text] [doi: 10.2147/CEOR.S34321] [Medline: 23818799]
8. Wilkinson TMA, Donaldson GC, Hurst JR, Seemungal TAR, Wedzicha JA. Early therapy improves outcomes of exacerbations of chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2004 Jun 15;169(12):1298-1303. [doi: 10.1164/rccm.200310-1443OC] [Medline: 14990395]
9. McLean S, Nurmatov U, Liu JL, Pagliari C, Car J, Sheikh A. Telehealthcare for chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2011(7):CD007718. [doi: 10.1002/14651858.CD007718.pub2] [Medline: 21735417]
10. McCabe C, McCann M, Brady AM. Computer and mobile technology interventions for self-management in chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2017 Dec 23;5:CD011425. [doi: 10.1002/14651858.CD011425.pub2] [Medline: 28535331]
11. van Buul AR, Wildschut TS, Bonten TN, Kasteleyn MJ, Slats AM, Chavannes NH, et al. A systematic diagnostic evaluation combined with an internet-based self-management support system for patients with asthma or COPD. Int J Chron Obstruct Pulmon Dis 2018;13:3297-3306 [FREE Full text] [doi: 10.2147/COPD.S175361] [Medline: 30349234]

12. Murphy LA, Harrington P, Taylor SJ, Teljeur C, Smith SM, Pinnock H, et al. Clinical-effectiveness of self-management interventions in chronic obstructive pulmonary disease: an overview of reviews. Chron Respir Dis 2017 Aug;14(3):276-288 [FREE Full text] [doi: 10.1177/1479972316687208] [Medline: 28774200]
13. Alcazar B, de Lucas P, Soriano JB, Fernández-Nistal A, Fuster A, González-Moro JMR, et al. The evaluation of a remote support program on quality of life and evolution of disease in COPD patients with frequent exacerbations. BMC Pulm Med 2016 Nov 08;16(1):140 [FREE Full text] [doi: 10.1186/s12890-016-0304-3] [Medline: 27821164]
14. Steventon A, Bardsley M, Billings J, Dixon J, Doll H, Hirani S, et al. Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. BMJ 2012;344:e3874 [FREE Full text] [Medline: 22723612]
15. Rixon L, Hirani SP, Cartwright M, Beynon M, Doll H, Steventon A, et al. A RCT of telehealth for COPD patient's quality of life: the whole system demonstrator evaluation. Clin Respir J 2017 Jul;11(4):459-469. [doi: 10.1111/crj.12359] [Medline: 26260325]
16. Gaveikaite V, Grundstrom C, Winter S, Chouvarda I, Maglaveras N, Priori R. A systematic map and in-depth review of European telehealth interventions efficacy for chronic obstructive pulmonary disease. Respir Med 2019;158:78-88. [doi: 10.1016/j.rmed.2019.09.005] [Medline: 31614305]
17. Kruse C, Pesek B, Anderson M, Brennan K, Comfort H. Telemonitoring to manage chronic obstructive pulmonary disease: systematic literature review. JMIR Med Inform 2019 Mar 20;7(1):e11496 [FREE Full text] [doi: 10.2196/11496] [Medline: 30892276]
18. Jakobsen AS, Laursen LC, Schou L, Emme C, Phanareth KV. [Varying effect of telemedicine in the treatment of chronic obstructive pulmonary disease—a systematic review]. Ugeskrift for Læger 2012 Apr 02;174(14):936-942. [Medline: 22469162]
19. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth interventions to support self-management of long-term conditions: a systematic metareview of diabetes, heart failure, asthma, chronic obstructive pulmonary disease, and cancer. J Med Internet Res 2017 May 17;19(5):e172 [FREE Full text] [doi: 10.2196/jmir.6688] [Medline: 28526671]
20. Farmer A, Williams V, Velardo C, Shah SA, Yu L, Rutter H, et al. Self-management support using a digital health system compared with usual care for chronic obstructive pulmonary disease: randomized controlled trial. J Med Internet Res 2017 May 03;19(5):e144 [FREE Full text] [doi: 10.2196/jmir.7116] [Medline: 28468749]
21. Mackay AJ, Kostikas K, Murray L, Martinez FJ, Miravitlles M, Donaldson G, et al. Patient-reported outcomes for the detection, quantification, and evaluation of chronic obstructive pulmonary disease exacerbations. Am J Respir Crit Care Med 2018 Sep 15;198(6):730-738. [doi: 10.1164/rccm.201712-2482CI] [Medline: 29671615]
22. Pombo-Juárez L, Könnölä T, Miles I, Saritas O, Schartinger D, Amanatidou E, et al. Wiring up multiple layers of innovation ecosystems: contemplations from Personal Health Systems Foresight. Technol Forecasting Soc Change 2017 Feb;115:278-288. [doi: 10.1016/j.techfore.2016.04.018]
23. WHO global strategy on people-centered and integrated health services: interim report. Geneva: World Health Organization; 2015. URL: https://apps.who.int/iris/bitstream/handle/10665/155002/WHO_HIS_SDS_2015.6_eng.pdf [accessed 2019-11-26]
24. Phanareth K, Vingtoft S, Christensen AS, Nielsen JS, Svenstrup J, Berntsen GKR, et al. The Epital Care Model: a new person-centered model of technology-enabled integrated care for people with long term conditions. JMIR Res Protoc 2017 Jan 16;6(1):e6 [FREE Full text] [doi: 10.2196/resprot.6506] [Medline: 28093379]
25. Lindskrog S, Christensen KB, Osborne RH, Vingtoft S, Phanareth K, Kayser L. Relationship between patient-reported outcome measures and the severity of chronic obstructive pulmonary disease in the context of an innovative digitally supported 24-hour service: longitudinal study. J Med Internet Res 2019 Jun 02;21(6):e10924 [FREE Full text] [doi: 10.2196/10924] [Medline: 31199320]
26. Vestbo J, Hurd SS, Agustí AG, Jones PW, Vogelmeier C, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2013 Feb 15;187(4):347-365. [doi: 10.1164/rccm.201204-0596pp]
27. Seemungal TAR, Hurst JR, Wedzicha JA. Exacerbation rate, health status and mortality in COPD—a review of potential interventions. Int J Chron Obstruct Pulmon Dis 2009;4:203-223 [FREE Full text] [doi: 10.2147/copd.s3385] [Medline: 19554195]
28. Vijayasaratha K, Stockley R. Reported and unreported exacerbations of COPD: analysis by diary cards. Chest 2008 Jan;133(1):34-41. [doi: 10.1378/chest.07-1692] [Medline: 17989153]
29. Langsetmo L, Platt RW, Ernst P, Bourbeau J. Underreporting exacerbation of chronic obstructive pulmonary disease in a longitudinal cohort. Am J Respir Crit Care Med 2008 Feb 15;177(4):396-401. [doi: 10.1164/rccm.200708-1290OC] [Medline: 18048806]
30. Kargiannakis M, Fitzsimmons DA, Bentley CL, Mountain GA. Does telehealth monitoring identify exacerbations of chronic obstructive pulmonary disease and reduce hospitalisations? An analysis of system data. JMIR Med Inform 2017 Mar 22;5(1):e8 [FREE Full text] [doi: 10.2196/medinform.6359] [Medline: 28330829]
31. Utens CMA, Goossens LMA, van Schayck OCP, Rutten–van Mölken MPMH, van Litsenburg W, Janssen A, et al. Patient preference and satisfaction in hospital-at-home and usual hospital care for COPD exacerbations: results of a randomised controlled trial. Int J Nurs Stud 2013 Nov;50(11):1537-1549. [doi: 10.1016/j.ijnurstu.2013.03.006] [Medline: 23582671]

32. Echevarria C, Gray J, Hartley T, Steer J, Miller J, Simpson AJ, et al. Home treatment of COPD exacerbation selected by DECAF score: a non-inferiority, randomised controlled trial and economic evaluation. Thorax 2018 Aug;73(8):713-722 [FREE Full text] [doi: 10.1136/thoraxjnl-2017-211197] [Medline: 29680821]

Abbreviations

COPD: chronic obstructive pulmonary disease
ECM: Epital Care Model
FEV_1: forced expiratory volume during the first second
GOLD: Global Initiative on Chronic Obstructive Lung Disease
IT: information technology
PRO: patient-reported outcome
RCC: response and coordination center
VCMRI: Virtual Clinical Management Resilience Index

Edited by R Kukafka, G Eysenbach; submitted 16.07.20; peer-reviewed by H Pinnock, C Kruse; comments to author 01.09.20; revised version received 16.10.20; accepted 08.08.21; published 19.10.21.

Please cite as:
Phanareth K, Dam AL, Hansen MABC, Lindskrog S, Vingtoft S, Kayser L
Revealing the Nature of Chronic Obstructive Pulmonary Disease Using Self-tracking and Analysis of Contact Patterns: Longitudinal Study
J Med Internet Res 2021;23(10):e22567
URL: https://www.jmir.org/2021/10/e22567
doi:10.2196/22567
PMID:34665151

Viewpoint

It Is Not Just the FEV1 That Matters, but the Personal Goals We Reach Along the Way: Qualitative, Multicenter, Prospective, Observational Study

Martinus C Oppelaar[1], BSc; Lara S van den Wijngaart[1], MD, PhD, MSc; Peter J F M Merkus[1], MD, PhD; Ellen A Croonen[2], MD, PhD; Cindy A C Hugen[1], MD; Marianne L Brouwer[2], MD; Annemie L M Boehmer[3], MD, PhD; Jolt Roukema[1], MD, PhD

[1]Department of Pediatric Pulmonology, Amalia Children's Hospital, Radboud University Medical Center, Nijmegen, Netherlands

[2]Department of Pediatrics, Canisius Wilhelmina Hospital, Nijmegen, Netherlands

[3]Department of Pediatrics, Spaarne Gasthuis, Haarlem, Netherlands

Corresponding Author:
Martinus C Oppelaar, BSc
Department of Pediatric Pulmonology
Amalia Children's Hospital
Radboud University Medical Center
Geert Grooteplein 10
Nijmegen, 6500 HB
Netherlands
Phone: 31 2414430
Fax: 31 2416428
Email: marc.oppelaar@radboudumc.nl

Abstract

Background: The COVID-19 pandemic has boosted the use of forced expiratory volume in 1 second (FEV_1) telemonitoring in pediatric asthma, but a consensus on its most efficient and effective implementation is still lacking. To find answers, it is important to study how such an intervention is perceived, experienced, and used by both patients and health care professionals (HCPs).

Objective: The aim of this study was to provide perspectives on how FEV_1 home monitoring should be used in pediatric asthma.

Methods: This is a qualitative, multicenter, prospective, observational study which included patients with asthma aged 6-16 and HCPs. Primary outcomes were results of 2 surveys that were sent to all participants at study start and after 3-4 months. Secondary outcomes consisted of FEV_1 device usage during 4 months after receiving the FEV_1 device.

Results: A total of 39 participants (26 patients and 13 HCPs) were included in this study. Survey response rates were 97% (38/39) at the start and 87% (34/39) at the end of the study. Both patients and HCPs were receptive toward online FEV_1 home monitoring and found it contributive to asthma control, self-management, and disease perception. The main concerns were about reliability of the FEV_1 device and validity of home-performed lung function maneuvers. FEV_1 devices were used with a median frequency of 7.5 (IQR 3.3-25.5) during the 4-month study period.

Conclusions: Patients and HCPs are receptive toward online FEV_1 home monitoring. Frequency of measurements varied largely among individuals, yet perceived benefits remained similar. This emphasizes that online FEV_1 home monitoring strategies should be used as a means to reach individual goals, rather than being a goal on their own.

(J Med Internet Res 2021;23(10):e29218) doi:10.2196/29218

KEYWORDS

eHealth; asthma; pediatrics; telemonitoring; lung function tests; lung function; spirometry; home monitoring; mHealth; app; smartphone; asthma control; child; outpatients; remote consultations; quality improvement; patient care management; telemetry; application; FEV1; pulmonary care

Introduction

The primary aim of asthma care is to reach and maintain asthma control by early recognition and treatment of pulmonary exacerbations (PEx) [1]. In children with asthma, objective measures to aid patients, caregivers, and health care professionals (HCPs) in detecting pulmonary deterioration are crucial [1-3]. Many objective measures have been studied and proposed, but the forced expiratory volume in 1 second (FEV_1) measured by spirometry remains undefeated as both an objective measure of pulmonary deterioration and a criterion for defining PEx [4-6]. The limitation of FEV_1 is that it is usually only measured during scheduled outpatient clinic visits and not preceding asthma exacerbations. As a response to the COVID-19 lockdowns, many hospitals have reduced their outpatient clinic capacities which further decreased our ability to timely recognize PEx. Additional unplanned outpatient visits during symptoms were also harder to schedule because of COVID-19 measures in hospitals. It is therefore no surprise that telemonitoring, including FEV_1 home measurements, has become more popular during the COVID-19 pandemic [7-9].

The value of FEV_1 telemonitoring has been a subject of debate since portable spirometers became available. Although generally accepted as a feasible intervention in children with limited disease perception, concerns regarding the reliability of the measurements persist, and studies have failed to convincingly show an added value of FEV_1 home monitoring in general asthma care [1,4,10-15]. Most of these studies used strict monitoring regimes in which patients measure their FEV_1 daily in order to reduce PEx and health care consumption. As a result, monitoring adherence declined, and HCPs were left with mounts of—mostly irrelevant—FEV_1 data and eventually the primary objectives were not reached [1,12-15]. This raises the question of what role FEV_1 telemonitoring should play in pediatric asthma. With the currently rapidly accelerating interests in FEV_1 telemonitoring, it is even more important to develop perspectives on this topic.

This study aimed to develop new perspectives on how to use FEV_1 telemonitoring in the future of pediatric asthma care. To achieve this we combined FEV_1 home monitoring with an online eHealth platform [16-18]. Our main research question was "How do patients, their parents, and HCPs want to make use of FEV_1 home monitoring?" Patients who already used the eHealth platform in regular pediatric asthma care to monitor asthma control with questionnaires received FEV_1 monitoring devices which were integrated in the platform. To realistically reflect a regular pediatric care setting, no fixed measurement schedules were used and patients themselves kept responsibility on how often they measured their lung function. Expectations and experiences of both patients and HCP were studied, as well as FEV_1 device usage.

Methods

This was a qualitative, multicenter, prospective, observational study on FEV_1 home monitoring combined with an online eHealth platform for 4 months. The eHealth platform is used in regular pediatric asthma care to monitor asthma control using the validated (childhood) Asthma Control Test ([C]-ACT), and to support self-management with personalized online asthma action plans [19,20]. Details of the eHealth platform have been published previously [16-18]. For this study the platform was expanded with a module for FEV_1 home measurements. FEV_1 measurements were performed with the Spirobank Smart and automatically uploaded to the online eHealth platform with a smartphone app [21]. The smartphone app was available on both Google Play and the Apple App Store. Participants could log in to the smartphone app with the same credentials as used for the online eHealth platform. After pairing with the FEV_1 device once, participants could use the app to perform measurements. Measured values were automatically sent to the app via Bluetooth and could be uploaded directly to the online eHealth platform (Figure 1). In addition to FEV_1, the Spirobank Smart devices measured the forced vital capacity (FVC), FEV_1-to-FVC ratio (FEV_1/FVC), peak expiratory flow (PEF), and forced expiratory flow at 25%-75% (FEF25-75). Automated feedback was given for each measurement. A green, orange, or red tag was provided based on individual thresholds depending on the personal best value of FEV_1 of that patient. A green tag prompted encouraging feedback and required no intervention, while an orange or red tag (ie, usually below 80% of the patient's best value) prompted an intervention based on the personal online asthma action plan of the patient. A red tag also sent an automatic email notification to the treating HCPs which they could follow-up on if deemed necessary. FEV_1 values were plotted over time in the online eHealth platform (Figure 2). The FEV_1 is the main lung function outcome on the eHealth platform, but additional values were visible when selecting specific measurements on the platform. During this study patients kept the responsibility on how often they used their devices, and no instructions were given on how often they needed to measure their FEV_1.

Figure 1. Screenshots of the instruction screen (left panel), measurement screen (middle panel) and feedback screen (right panel) of the smartphone app. The text in the instruction screen provides a short description on how to use the FEV_1 device appropriately. The measurement screen shows a successfully performed measurement with buttons to upload (upper) or repeat (lower) the measurement. The feedback in the feedback screen is based on individual thresholds. Language is in Dutch.

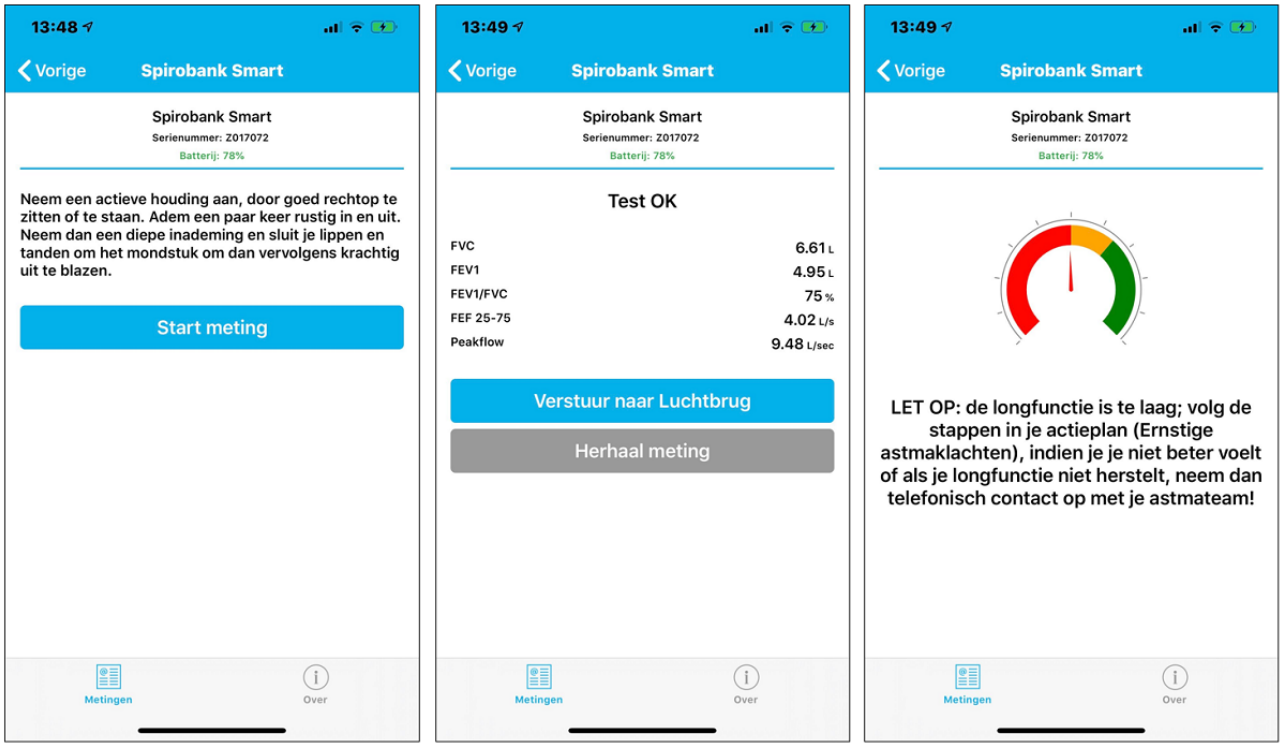
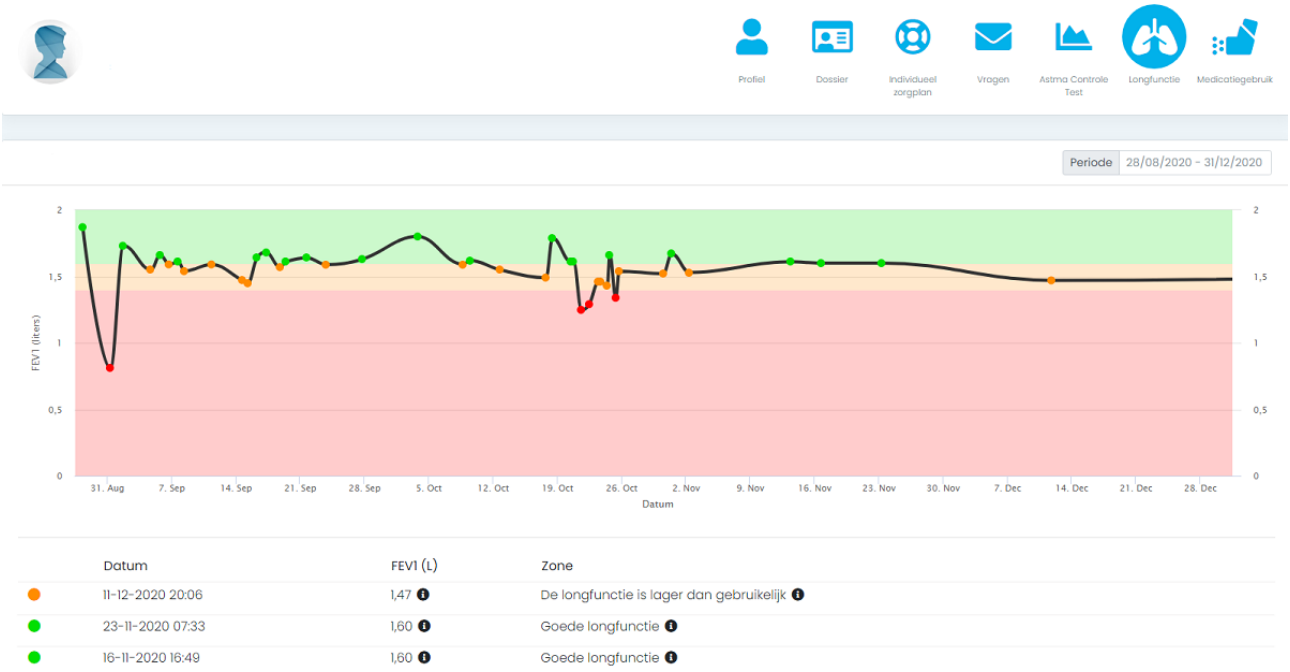


Figure 2. Screenshot of the graphical presentation of FEV_1 measurements on the online eHealth platform of one of the participants. Language is in Dutch. Colors represent individual color zones (green, orange, red).



Participants were recruited during outpatient visits in specialized asthma clinics from a university hospital (Radboudumc, Nijmegen, the Netherlands) and 2 general hospital (Canisius Wilhelmina Hospital, Nijmegen, the Netherlands and Spaarne Gasthuis, Haarlem, the Netherlands). Participants were eligible for inclusion if they had a doctor's diagnosis of asthma based on the Global Initiative for Asthma (GINA) criteria, were aged 6-18 years, and already used the online eHealth platform for regular pediatric asthma care [1]. All parents and participants aged ≥12 years had documented verbal consent for the anonymized use of their data. The local ethical committee waived approval of the Medical Research Involving Human Subjects Act (WMO) considering the negligible burden of this study and the absence of imposed risks.

The primary outcome included survey results of patients (or their parents in case of young patients) and HCPs before and 3-4 months after the introduction of FEV_1 devices. We used modified validated research questionnaires originally designed by Grol et al [22] to gather information from 5 perspectives: the receptiveness for using innovations in health care, the perceived contribution of online FEV_1 home monitoring to asthma care, the perceived contribution of online FEV_1 home monitoring to patient self-management, the user-friendliness of the service, and the possible undesired effects. The surveys consisted of 21-25 questions of which most were to be answered with a 5-item Likert scale (strongly disagree to strongly agree). The online survey also asked participants to comment on experienced benefits and disadvantages, and to provide suggestions to improve the service. Survey outcomes were described and analyzed on intergroup differences at both time points and intragroup differences over time.

The secondary outcome was FEV_1 device usage over the 4-month study period. All statistical analyses were performed using Statistical Package for the Social Sciences (SPSS version 25; IBM).

Results

A total of 39 participants (26 patients and 13 HCPs) were included in this study. Patient characteristics at baseline are summarized in Table 1. HCPs were employed as pediatric pulmonologists (5/13, 38%), general pediatricians (1/13, 8%), residents (2/13, 15%), specialist nurses (4/13, 31%), or doctor's assistants (1/13, 8%). The median (IQR) age of HCP was 54 years (38.5-58). Baseline lung function was defined as a measurement recorded within 31 days of FEV_1 device reception. Two patients did not record any lung function measurements and 1 patient did not record a lung function measurement within the 31-day baseline window. One patient quit the study due to loss of interest. During this study none of the patients experienced a PEx nor were hospitalized.

Table 1. Baseline[a] patient characteristics (N=26).

Characteristics	Value
Age (years), median (IQR)	13.4 (11.4-14.6)
Age group, n (%)	
6-11 years	8 (31)
12-18 years	18 (69)
Male, n (%)	12 (46)
Initial ICS[b] dose µg/day[c], mean (SD)	488.5 (353.6)
(C-)ACT[d] score (n=7), median (IQR)	20 (15-23)
ACT[d] score (n=17), median (IQR)	22 (18-27)
(C-)ACT <20 points (n=21), mean (SD)	8 (38.1)
6-11 years (n=6)	
z-FEV_1[e], mean (SD)	–0.30 (0.80)
z-FVC[f], mean (SD)	0.26 (0.50)
Tiff (%), median (IQR)	84.08 (73.19-92.00)
z-FEF25-75[g], mean (SD)	–0.22 (1.49)
Color zone not green, n (%)	0 (0)
12-18 years (n=17)	
z-FEV_1, mean (SD)	–1.53 (1.47)
z-FVC, mean (SD)	–0.65 (2.47)
Tiff (%), median (IQR)	81.87 (77.84-89.20)
z-FEF25-75, mean (SD)	–1.37 (1.07)
Color zone not green, n (%)	6 (35)

[a]Baseline lung function outcomes were defined as the first measurement within 31 days of receiving the FEV_1 measurement device. Three of the included participants did not perform a baseline lung function measurement.

[b]ICS: inhaled corticosteroids

[c]Beclomethasone or equivalent dose.

[d](C-)ACT: (Childhood) Asthma Control Test.

[e]FEV_1: forced expiratory volume in 1 second.

[f]FVC: forced vital capacity.

[g]FEF: forced expiratory flow.

Survey response rates at start were 96% (25/26) for patients and 100% (13/13) for HCPs, whereas those at the end were 85% (22/26) and 92% (12/13), respectively. Survey results and intragroup comparisons over time are summarized in Tables 2 and 3.

A total of 18 (69%) patients used their FEV_1 device more than 3 times during the study period. FEV_1 devices were used with a median frequency of 7.5 (IQR 3.3-25.5) distributed over 5.5 unique days (IQR 2.3-19.0). Two patients did not use their FEV_1 device at all, because one quit the study and another experienced technical difficulties. Most measurements were performed in the morning (154/421, 36.6%) or evening (156/421, 37.1%). 30.6% (129/421) of measurements were not in the personalized green zones, and 11.6% (49/421) of measurements were in the personalized red zone, leading to closer inspection by HCPs.

Seven patients reported that they did not want to decide themselves how often they measured their lung function (Table 2, Q12). Only 2 of these patients used their device with a frequency within the IQR of all participants. The others used their device as follows: not at all, once or twice, or very often (33 or 41 times).

Table 2. Patient survey outcomes.

Statement[a]	Start, median (IQR)	End, median (IQR)	Significance (*P* value)
Q1. I am experienced with using smartphones.	5 (4-5)	N/A[c]	—
Q2. Using innovations in health care is normal.	4 (3-5)	N/A	—
Q3. The FEV_1[b] measurement device looks nice.	N/A	4 (3-5)	—
Q4. The FEV_1 measurement device is easy to use.	N/A	4 (2-4)	—
Q5. The manual of the FEV_1 measurement device was clear.	N/A	4 (3-4)	—
Q6. Measuring my FEV_1 at home is new for me (innovative).	4 (4-5)	N/A	—
Q7. I should be able to measure my FEV_1 for a longer period of time.	4 (4-5)	N/A	—
Q8. Measuring my FEV_1 at home is a good addition to my daily asthma care.	4 (4-5)	4 (4-5)	.83
Q9. The FEV_1 measurements will provide me with more insights into my disease.	4 (4-5)	4 (3.75-4)	.20
Q10. Measuring my FEV_1 at home will cost a lot of time.	2 (2-2.5)	2 (2-3)	.23
Q11. I am glad that I am able to check my FEV_1 by myself.	4 (4-5)	4 (3-4.25)	.34
Q12. I would only measure my FEV_1 when I experience symptoms.	2 (2-3)	3 (2-4)	.33
Q13. Measuring my FEV_1 regularly will help me to better handle my disease.	4 (3-5)	4 (3-4)	.13
Q14. The home measurements of FEV_1 will make me insecure.	2 (1-2)	1.5 (1-2)	.49
Q15. The home measurements of FEV_1 will give me stress.	2 (1-2)	2 (1-2.25)	.42
Q16. I only want to know my FEV_1 when it's not going well.	2 (1.5-2.5)	2 (1-3)	.27
Q17. I want to decide myself how often I measure my lung function.	3 (3-4)	3 (2.75-4)	.86
Q18. I want to receive feedback on my home measurements.	4 (3-4)	4 (3-4.25)	>.99
Q19. I wouldn't mind to fill out a short symptom survey if my FEV_1 is lower than expected.	4 (4-5)	4 (4-4)	>.99
Q20. I would like to receive reminders in the online asthma clinic to measure my FEV_1.	4 (3.5-4.5)	4 (4-5)	.37
Q21. The graphical presentation of my FEV_1 measurements is useful.	N/A	4 (3-4)	—
Q22. If I don't succeed to measure my FEV_1 at home, I know whom to contact.	N/A	4 (4-5)	—
Q23. I don't worry as long as I feel good, even if my FEV_1 is lower than expected.	N/A	3 (3-4)	—
Q24. I wouldn't mind if my health care professionals can see my home measurements.	4 (4-5)	4 (4-5)	.48
Q25. I would feel controlled by my health care professionals if they can see my home measurements.	2 (1-3)	2 (1-2.25)	.21

[a]Responses were collected on a 5-point Likert scale, where 1=strongly disagree; 2=disagree; 3=neutral; 4=agree; and 5=strongly agree.

[b]FEV_1: forced expiratory volume in 1 second.

[c]N/A: not applicable.

Table 3. Health care professional survey outcomes.

Statement[a]	Start, median (IQR)	End, median (IQR)	Significance (*P* value)
Q1. I am experienced with using smartphones.	4 (3.5-4.5)	4 (4-4.75)	>.99
Q2. Using innovations in health care is normal.	4 (4-5)	4 (4-5)	>.99
Q3. The FEV_1[b] measurement device looks nice.	N/A[c]	4 (4-4.75)	—
Q4. The FEV_1 measurement device is easy to use.	N/A	4 (3-4)	—
Q5. The manual of the FEV_1 measurement device was clear.	N/A	4 (4-4)	—
Q6. Letting patients measure their FEV_1 at home is innovative.	4 (4-5)	N/A	—
Q7. Patients should be able to measure their FEV_1 at home for a long period of time.	4 (4-5)	N/A	—
Q8. Home monitoring of FEV_1 is a good addition to patients' daily asthma care.	4 (4-5)	4 (4-5)	.77
Q9. The graphical presentation of the FEV_1 measurements is useful.	N/A	4 (4-5)	—
Q10. The FEV_1 measurements will provide me with more insights into my patients' disease.	4 (4-4.5)	4 (4-4)	.48
Q11. I only want patients to measure their FEV_1 when they experience symptoms.	2 (2-3)	2 (2-2.75)	>.99
Q12. Possible deteriorations will be detected earlier thanks to the home measurements.	4 (4-4.5)	4 (4-4.75)	.74
Q13. If technical problems arise with the measurements, I know whom to contact.	4 (3.5-4)	4 (3.25-4.75)	.48
Q14. The FEV_1 measurements will pose an additional time burden for me.	3 (3-4)	3 (2-3)	.02[d]
Q15. I only want to know the FEV_1 measurements when they are lower than expected.	4 (2.5-4)	4 (2.25-4)	.41
Q16. Patients themselves should be responsible about how often they measure their FEV_1.	3 (2-3.5)	3.5 (3-4)	.84
Q17. If FEV_1 measurements are lower than expected, a short symptom survey will provide sufficient information for me.	3 (3-4)	4 (3.25-4)	.10
Q18. Patients should receive reminders in the online asthma clinic to measure their FEV_1.	4 (3-4)	4 (2.25-4)	.10
Q19. I know which patients are eligible for home monitoring of FEV_1.	4 (4-4)	4 (3.25-4.75)	.32

[a]Responses were collected on a 5-point Likert scale, where 1=strongly disagree; 2=disagree; 3=neutral; 4=agree; 5=strongly agree.

[b]FEV_1: forced expiratory volume in 1 second.

[c]N/A: not applicable.

[d]Statistically significant change between survey outcomes at end and start.

Patients were interested in knowing all of their measured FEV_1 values. By contrast, HCPs at both start and study end were more interested in only knowing their patients' FEV_1 outcomes when they were lower than expected (*P*=.006 and .001, respectively). At the start, HCPs were significantly more likely to expect an increased time burden compared with patients (*P*<.001). However their expectations were not matched by their experiences because at study end HCPs had significantly improved opinions on time burden compared with study start (*P*=.021; Table 3, Q14). Reported benefits, disadvantages, and suggestions by HCPs and patients are summarized in Textboxes 1 and 2.

Textbox 1. Patients reported benefits, disadvantages, and suggestions at the end of the study. FEV1: forced expiratory volume in 1 second.

Benefits

- More insights into my/my child's asthma
- Improved asthma control
- Less frequent hospital visits
- Always being able to know how I am/my child is doing
- It helps me with my/my child's disease perception
- More insights into my/my child's lung function
- Facilitates easy and quick adjustment of treatment
- Signaling of low lung function

Disadvantages

- Wrong values when the measurement is performed incorrectly
- The device does not always work
- It is easy to lose the device
- It is easy to forget to perform measurements
- The FEV_1 device is less reliable than the device in the hospital
- Performing the FEV_1 measurements is time consuming
- The mouthpiece is too large for small children

Suggestions

- Reminders should be sent to perform an FEV_1 measurement
- The FEV_1 device should be easier to use
- Instructions on how to clean the device
- Smaller mouthpiece for smaller children

Textbox 2. Health care professionals' reported benefits, disadvantages, and suggestions at the end of the study. FEV1: forced expiratory volume in 1 second.

Benefits

- Provision of objective measures of asthma control
- Aids patients with disease perception
- Continuity of monitoring
- Earlier recognition of pulmonary exacerbations
- Taking a simple and quick lung function test is possible when needed
- Facilitates easy and quick adjustment of treatment

Disadvantages

- More stress for both patients and caregivers
- More emphasis on patients' asthma
- FEV_1 measurements might be performed incorrectly which can falsely comfort or alarm patients
- Compulsive FEV_1 testing and less attention for perceived symptoms
- Decreased motivation when FEV_1 remains low while symptoms are barely present
- Time consuming for both patients and health care professionals
- Technological difficulties of FEV_1 devices

Suggestions

- Good instructions at baseline
- FEV_1 devices should be calibrated at every outpatient visit
- FEV_1 home measurements should be used as a means to aid patients in achieving their individual goals. FEV_1 home measurements should not be a goal on their own
- Better instructions for patients on what to do when their FEV_1 is lower than expected
- Less notifications for health care professionals during pulmonary exacerbations
- A notification when patients are back in their green zone
- Protective case for a fragile FEV_1 device

Discussion

Our findings show that both patients/their parents and HCPs are receptive toward online FEV_1 telemonitoring in pediatric asthma care. The participants in this study agreed that FEV_1 telemonitoring in pediatric asthma care is innovative and improves asthma care, self-management, and disease perception. Most of the expectations patients and HCPs had about FEV_1 telemonitoring matched their experiences. Initially, HCPs had concerns regarding the additional time burden and increased stress and insecurities of patients. However, HCPs' experienced time burden was eventually lower than expected and increased stress and insecurities of patients were not reflected in the patients' own experiences.

In our self-paced monitoring protocol, 69% (18/26) of patients used their FEV_1 device regularly, which affirms survey-reported receptiveness of patients. However, the large interindividual range in the frequency of use between patients emphasizes the different approaches of individuals. Only 11.6% (49/421) of measurements were in the "red zone." The vast majority of measurements required no intervention or could be resolved with automatic feedback prompts using patients' personalized online asthma action plans.

To our knowledge, this is the first study on patients' and HCPs' receptiveness of FEV_1 home monitoring combined with an online eHealth platform in pediatric asthma care. The existing body of literature often reports acceptance of FEV_1 home monitoring, but it is rarely quantified [12-15]. Because of the COVID-19 pandemic, there is an increased call for research into user satisfaction and receptiveness of eHealth interventions [8,9]. Before the COVID-19 pandemic, Simpson et al [23] surveyed 187 patients with asthma and 63 HCPs on eHealth interventions and found that both groups supported the use of eHealth and preferred eHealth over paper diaries and paper asthma action plans. They also found that in both groups lung function measurements were expected to be the most contributive additional measurement for asthma control in an eHealth program [23]. This highlights the increasing demand for home monitoring of lung function among patients. In our small population, expectations of both patients and HCPs were matched by their experiences, underlining that patients and their

parents can assess for themselves whether FEV_1 telemonitoring will aid them. Therefore, we argue that patients interested in the ability of online FEV_1 monitoring should receive this opportunity to assess personal benefits.

How patients achieve personal benefits should be a result of shared decision making between patients and their HCPs. We now know that standalone FEV_1 monitoring with strict monitoring regimes in most patients does not work out [12-15]. By contrast, we believe that in the future FEV_1 devices should be used to reach personalized goals, such as reassurance, as an aid to improve symptom perception or to quantify lung function during episodes with increased symptoms. When integrated into an eHealth platform, these devices can be used at the patient's convenience and provide immediate feedback. It is our opinion that these set ups can facilitate these goals more easily and quickly than standalone FEV_1 monitoring. The results of our study support this opinion. In this study most patients wanted to keep their own responsibility on how often they measured their lung function. Subsequently, we found a large variation in frequency of use between participants, but there was a consensus nonetheless on the perceived benefits of asthma control and disease perception by both patients and HCPs. Some patients did not want to keep the responsibility on how often they measured their lung function and as a result most of these patients either rarely used the devices or overused them. This also shows that for some patients a structured monitoring regime could be indicated. Self-management depends on several factors such as individual health skills, disease perception, beliefs, and the interaction between patients and their HCP, and thus one size does not fit all. To ensure longevity and efficiency of FEV_1 telemonitoring, a personalized approach should be used: patients who want to keep the responsibility to themselves should be offered this chance, whereas those who do not want to should be offered a schedule. To better understand this personalized approach, it is important that both the clinical and perceived benefits of the different approaches are studied experimentally in more detail. It is important to identify which patients benefit most from which intervention and how we define personal benefits.

Recurring concerns of FEV_1 telemonitoring are toward the reliability of the measurements. These were also present in our population and are generally shared among the scientific community [1,4,24,25]. Although perfect correlation of FEV_1 home measurements with hospital measurements is desirable, one can argue that FEV_1 monitoring will primarily be used to ascertain an FEV_1 above a certain threshold. In that case a somewhat lower FEV_1 may not be of any clinical relevance. By contrast, invalidly low FEV_1 values at home may lead to excessive health care consumption, medication use, and disease-related stress. Online FEV_1 home monitoring should function as an aid for symptom perception and a quick objective measure during symptoms. In our opinion these aims do not require an FEV_1 outcome as sensitive as that measured by a spirometer operated by a specialist nurse. Nevertheless, reliability should be studied in more detail to rule out counterproductive measurements.

This study was limited by its small study population and short study duration, which were chosen due to the nature of this viewpoint study. Only including patients and HCPs from specialized asthma clinics who also had experience using our eHealth platform would have introduced selection bias. In our opinion this does not invalidate our findings; by selecting patients and HCPs who already have experience with the eHealth platform we avoid confounding of our qualitative outcomes by the introduction of the eHealth platform. We also still observed a large interindividual variation in FEV_1 device usage. Finally we believe that home monitoring strategies like this will be reliant on some form of selection bias as they should primarily be applied in patients that are receptive. This might imply that our results are not extrapolatable to settings without previous eHealth experience, general practitioner services, or to low-resource settings. Furthermore, we performed no post measurement quality control of the FEV_1 measurements, which may have led to technically incorrect measurements being recorded. However, the software of the FEV_1 device performs a quality check based on the American Thoracic Society (ATS)/European Respiratory Society (ERS) consensus statement [26,27].

Viewpoint

Our findings show that patients and HCPs are receptive toward online FEV_1 home monitoring combined with an online eHealth platform. Frequency of measurements varies largely among individuals, yet perceived benefits remain similar. This emphasizes that online FEV_1 home monitoring strategies should be used as a means to reach individual goals, rather than being a goal on their own. During the COVID-19 pandemic we have seen eHealth being integrated more into daily medical practice and this will undoubtedly increase even more in the near future. With this increasing use of eHealth, we will see an equal surge of different ways to monitor our patients. In this data tangle it is easy to lose sight of what really matters: patient quality of life and needs. If we want to continuously monitor our patients, we should not only study the clinical relevance, but also the experienced benefits, invasiveness, and perceptions of our patients, while keeping in mind that one size does not fit all. Even after implementation we should repeatedly monitor patient satisfaction. eHealth is a dynamic entity and strongly bound to technologic advances. Patients should have a say in how these advances will be utilized and in which way they add value to their lives. HCPs have to remember that the goal should not be to gather outcomes, but to use these outcomes to reach personal goals in a personalized way.

Acknowledgments

This study was funded by the Dutch Healthcare Insurers Innovation Foundation (project number 3.712)

Conflicts of Interest

None declared.

References

1. Global Strategy for Asthma Management and Prevention, 2020. Global Initiative for Asthma. URL: http://www.ginasthma.org/ [accessed 2021-02-01]
2. Sonney JT, Gerald LB, Insel KC. Parent and child asthma illness representations: a systematic review. J Asthma 2016 Jun;53(5):510-516. [doi: 10.3109/02770903.2015.1116088] [Medline: 26785738]
3. Still L, Dolen WK. The Perception of Asthma Severity in Children. Curr Allergy Asthma Rep 2016 Jul;16(7):50. [doi: 10.1007/s11882-016-0629-2] [Medline: 27333779]
4. Pijnenburg MW, Baraldi E, Brand PLP, Carlsen K, Eber E, Frischer T, et al. Monitoring asthma in children. Eur Respir J 2015 Apr;45(4):906-925 [FREE Full text] [doi: 10.1183/09031936.00088814] [Medline: 25745042]
5. Kitch BT, Paltiel AD, Kuntz KM, Dockery DW, Schouten JP, Weiss ST, et al. A single measure of FEV1 is associated with risk of asthma attacks in long-term follow-up. Chest 2004 Dec;126(6):1875-1882. [doi: 10.1378/chest.126.6.1875] [Medline: 15596687]
6. Fuhlbrigge AL, Kitch BT, Paltiel AD, Kuntz KM, Neumann PJ, Dockery DW, et al. FEV(1) is associated with risk of asthma attacks in a pediatric population. J Allergy Clin Immunol 2001 Jan;107(1):61-67. [doi: 10.1067/mai.2001.111590] [Medline: 11149992]
7. Webster P. Virtual health care in the era of COVID-19. Lancet 2020 Apr 11;395(10231):1180-1181 [FREE Full text] [doi: 10.1016/S0140-6736(20)30818-7] [Medline: 32278374]
8. Taquechel K, Diwadkar AR, Sayed S, Dudley JW, Grundmeier RW, Kenyon CC, et al. Pediatric Asthma Health Care Utilization, Viral Testing, and Air Pollution Changes During the COVID-19 Pandemic. J Allergy Clin Immunol Pract 2020;8(10):3378-3387.e11 [FREE Full text] [doi: 10.1016/j.jaip.2020.07.057] [Medline: 32827728]
9. Hare N, Bansal P, Bajowala SS, Abramson SL, Chervinskiy S, Corriel R, et al. Work Group Report: COVID-19: Unmasking Telemedicine. J Allergy Clin Immunol Pract 2020 Sep;8(8):2461-2473.e3 [FREE Full text] [doi: 10.1016/j.jaip.2020.06.038] [Medline: 32603900]
10. Wensley DC, Silverman M. The quality of home spirometry in school children with asthma. Thorax 2001 Mar;56(3):183-185 [FREE Full text] [doi: 10.1136/thorax.56.3.183] [Medline: 11182009]
11. Pelkonen AS, Nikander K, Turpeinen M. Reproducibility of home spirometry in children with newly diagnosed asthma. Pediatr Pulmonol 2000 Jan;29(1):34-38. [doi: 10.1002/(sici)1099-0496(200001)29:1<34::aid-ppul6>3.0.co;2-o] [Medline: 10613784]
12. Ryan D, Price D, Musgrave SD, Malhotra S, Lee AJ, Ayansina D, et al. Clinical and cost effectiveness of mobile phone supported self monitoring of asthma: multicentre randomised controlled trial. BMJ 2012;344:e1756 [FREE Full text] [Medline: 22446569]
13. Rasmussen LM, Phanareth K, Nolte H, Backer V. Internet-based monitoring of asthma: a long-term, randomized clinical study of 300 asthmatic subjects. J Allergy Clin Immunol 2005 Jun;115(6):1137-1142. [doi: 10.1016/j.jaci.2005.03.030] [Medline: 15940125]
14. McLean S, Chandler D, Nurmatov U, Liu J, Pagliari C, Car J, et al. Telehealthcare for asthma. Cochrane Database Syst Rev 2010 Oct 06(10):CD007717. [doi: 10.1002/14651858.CD007717.pub2] [Medline: 20927763]
15. McLean G, Murray E, Band R, Moffat KR, Hanlon P, Bruton A, et al. Interactive digital interventions to promote self-management in adults with asthma: systematic review and meta-analysis. BMC Pulm Med 2016 May 23;16(1):83 [FREE Full text] [doi: 10.1186/s12890-016-0248-7] [Medline: 27215329]
16. van den Wijngaart LS, Roukema J, Boehmer ALM, Brouwer ML, Hugen CAC, Niers LEM, et al. A virtual asthma clinic for children: fewer routine outpatient visits, same asthma control. Eur Respir J 2017 Oct 5;50(4) [FREE Full text] [doi: 10.1183/13993003.00471-2017] [Medline: 28982775]
17. van den Wijngaart LS, Kievit W, Roukema J, Boehmer ALM, Brouwer ML, Hugen CAC, et al. Online asthma management for children is cost-effective. Eur Respir J 2017 Oct 5;50(4) [FREE Full text] [doi: 10.1183/13993003.01413-2017] [Medline: 28982768]
18. van den Wijngaart LS, Geense WW, Boehmer AL, Brouwer ML, Hugen CA, van Ewijk BE, et al. Barriers and Facilitators When Implementing Web-Based Disease Monitoring and Management as a Substitution for Regular Outpatient Care in Pediatric Asthma: Qualitative Survey Study. J Med Internet Res 2018 Oct 30;20(10):e284 [FREE Full text] [doi: 10.2196/jmir.9245] [Medline: 30377147]
19. Nathan RA, Sorkness CA, Kosinski M, Schatz M, Li JT, Marcus P, et al. Development of the asthma control test: a survey for assessing asthma control. J Allergy Clin Immunol 2004 Jan;113(1):59-65. [doi: 10.1016/j.jaci.2003.09.008] [Medline: 14713908]
20. Liu AH, Zeiger R, Sorkness C, Mahr T, Ostrom N, Burgess S, et al. Development and cross-sectional validation of the Childhood Asthma Control Test. J Allergy Clin Immunol 2007 Apr;119(4):817-825. [doi: 10.1016/j.jaci.2006.12.662] [Medline: 17353040]

21. MIR – Medical International Research – Spirometry Oximetry Telemedicine. URL: https://www.spirometry.com [accessed 2021-02-01]
22. Grol R, Wensing M. What drives change? Barriers to and incentives for achieving evidence-based practice. Med J Aust 2004 Mar 15;180(S6):S57-S60. [doi: 10.5694/j.1326-5377.2004.tb05948.x] [Medline: 15012583]
23. Simpson AJ, Honkoop PJ, Kennington E, Snoeck-Stroband JB, Smith I, East J, et al. Perspectives of patients and healthcare professionals on mHealth for asthma self-management. Eur Respir J 2017 May;49(5):1601966 [FREE Full text] [doi: 10.1183/13993003.01966-2016] [Medline: 28461291]
24. Gerzon FLGR, Jöbsis Q, Bannier MAGE, Winkens B, Dompeling E. Discrepancy between Lung Function Measurements at Home and in the Hospital in Children with Asthma and CF. J Clin Med 2020 May 26;9(6) [FREE Full text] [doi: 10.3390/jcm9061617] [Medline: 32466623]
25. Brouwer AFJ, Roorda RJ, Brand PLP. Comparison between peak expiratory flow and FEV(1) measurements on a home spirometer and on a pneumotachograph in children with asthma. Pediatr Pulmonol 2007 Sep;42(9):813-818. [doi: 10.1002/ppul.20660] [Medline: 17639585]
26. Beydon N, Davis SD, Lombardi E, Allen JL, Arets HGM, Aurora P, American Thoracic Society/European Respiratory Society Working Group on InfantYoung Children Pulmonary Function Testing. An official American Thoracic Society/European Respiratory Society statement: pulmonary function testing in preschool children. Am J Respir Crit Care Med 2007 Jun 15;175(12):1304-1345. [doi: 10.1164/rccm.200605-642ST] [Medline: 17545458]
27. Reddel HK, Taylor DR, Bateman ED, Boulet L, Boushey HA, Busse WW, et al. An official American Thoracic Society/European Respiratory Society statement: asthma control and exacerbations: standardizing endpoints for clinical asthma trials and clinical practice. Am J Respir Crit Care Med 2009 Jul 1;180(1):59-99. [doi: 10.1164/rccm.200801-060ST] [Medline: 19535666]

Abbreviations

ACT: Asthma Control Test
ATS: American Thoracic Society
(C-)ACT: (Childhood) Asthma Control Test
ERS: European Respiratory Society
FEF: forced expiratory flow
FEV_1: forced expiratory volume in 1 second
FVC: forced vital capacity
GINA: Global Initiative for Asthma
HCP: health care professional
PEx: pulmonary exacerbation

Edited by R Kukafka; submitted 30.03.21; peer-reviewed by S Stotland, KC Wong, B Nievas Soriano; comments to author 29.05.21; revised version received 08.07.21; accepted 27.07.21; published 20.10.21.

Please cite as:
Oppelaar MC, van den Wijngaart LS, Merkus PJFM, Croonen EA, Hugen CAC, Brouwer ML, Boehmer ALM, Roukema J
It Is Not Just the FEV1 That Matters, but the Personal Goals We Reach Along the Way: Qualitative, Multicenter, Prospective, Observational Study
J Med Internet Res 2021;23(10):e29218
URL: https://www.jmir.org/2021/10/e29218
doi:10.2196/29218
PMID:34668868

Original Paper

Multiparameter Continuous Physiological Monitoring Technologies in Neonates Among Health Care Providers and Caregivers at a Private Tertiary Hospital in Nairobi, Kenya: Feasibility, Usability, and Acceptability Study

Amy Sarah Ginsburg[1], MD; Mai-Lei Woo Kinshella[2,3], MA; Violet Naanyu[4], PhD; Jessica Rigg[3,5], MSc; Dorothy Chomba[6], MBA; Jesse Coleman[7], PhD; Bella Hwang[3], MBA; Roseline Ochieng[6], MMED; J Mark Ansermino[3,5], MBBCh; William M Macharia[6], MD

[1]UW Clinical Trial Center, University of Washington, Seattle, WA, United States

[2]Department of Obstetrics and Gynecology, British Columbia Children's and Women's Hospital and The University of British Columbia, Vancouver, BC, Canada

[3]Centre for International Child Health, BC Children's Hospital, Vancouver, BC, Canada

[4]School of Arts and Sciences, Moi University, Eldoret, Kenya

[5]Department of Anesthesiology, The University of British Columbia, Vancouver, BC, Canada

[6]Department of Pediatrics, Aga Khan University, Nairobi, Kenya

[7]Evaluation of Technologies for Neonates in Africa, Nairobi, Kenya

Corresponding Author:
Mai-Lei Woo Kinshella, MA
Department of Obstetrics and Gynecology
British Columbia Children's and Women's Hospital and The University of British Columbia
V344 - 950 W 28th Ave
Vancouver, BC, V5Z 4H4
Canada
Phone: 1 6048752253
Email: maggie.kinshella@cw.bc.ca

Abstract

Background: Continuous physiological monitoring technologies are important for strengthening hospital care for neonates, particularly in resource-constrained settings, and understanding user perspectives is critical for informing medical technology design, development, and optimization.

Objective: This study aims to assess the feasibility, usability, and acceptability of 2 noninvasive, multiparameter, continuous physiological monitoring technologies for use in neonates in an African health care setting.

Methods: We assessed 2 investigational technologies from EarlySense and Sibel, compared with the reference Masimo Rad-97 technology through in-depth interviews and direct observations. A purposive sample of health care administrators, health care providers, and caregivers at Aga Khan University Hospital, a tertiary, private hospital in Nairobi, Kenya, were included. Data were analyzed using a thematic approach in NVivo 12 software.

Results: Between July and August 2020, we interviewed 12 health care providers, 5 health care administrators, and 10 caregivers and observed the monitoring of 12 neonates. Staffing and maintenance of training in neonatal units are important feasibility considerations, and simple training requirements support the feasibility of the investigational technologies. Key usability characteristics included ease of use, wireless features, and reduced number of attachments connecting the neonate to the monitoring technology, which health care providers considered to increase the efficiency of care. The main factors supporting acceptability included caregiver-highlighted perceptions of neonate comfort and health care respondent technology familiarity. Concerns about the side effects of wireless connections, electromagnetic fields, and mistrust of unfamiliar technologies have emerged as possible acceptability barriers to investigational technologies.

Conclusions: Overall, respondents considered the investigational technologies feasible, usable, and acceptable for the care of neonates at this health care facility. Our findings highlight the potential of different multiparameter continuous physiological

monitoring technologies for use in different neonatal care settings. Simple and user-friendly technologies may help to bridge gaps in current care where there are many neonates; however, challenges in maintaining training and ensuring feasibility within resource-constrained health care settings warrant further research.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2019-035184

(J Med Internet Res 2021;23(10):e29755) doi:10.2196/29755

KEYWORDS

infants; Africa; medical technology design; user perspectives; in-depth interviews; direct observations

Introduction

Globally, neonatal mortality remains persistently high, with a disproportionate burden in Sub-Saharan Africa [1]. Technologies that allow for early detection of neonatal physiological instability and help guide appropriate interventions have the potential to reduce morbidity and mortality [2]. In resource-constrained health care settings where staffing shortages of trained health care providers (HCPs) may compromise capacities for adequate monitoring and management, such technologies may prove life-saving [2].

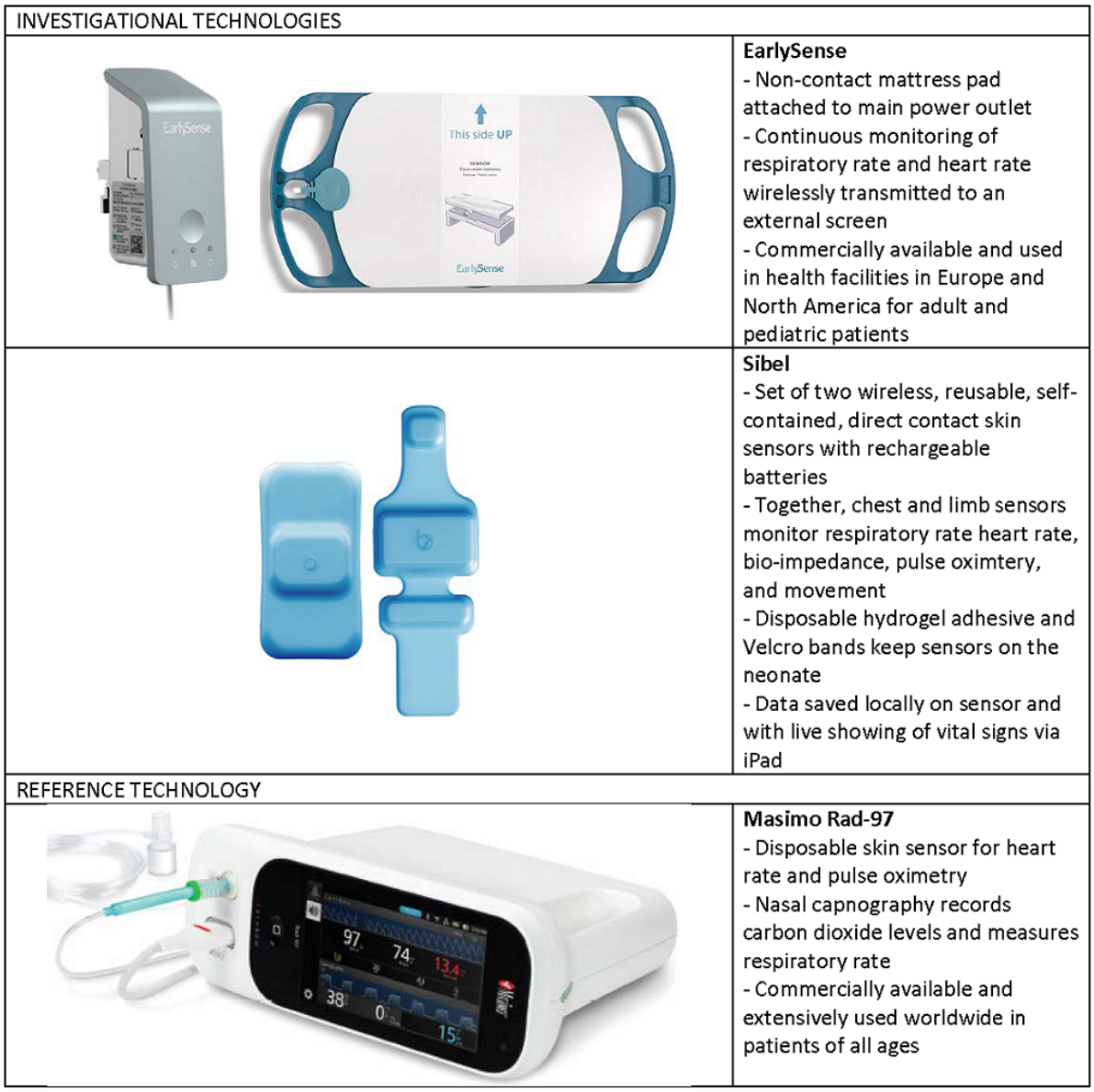
The Evaluation of Technologies for Neonates in Africa (ETNA) project was conceived with the goal of advancing and supporting development, as well as evaluation of technologies for use in neonates in resource-constrained settings. The project seeks to boost the development and optimization of promising neonatal diagnostic and care technologies that could be applied in resource-constrained settings by establishing an Africa-based evaluation platform. This is achieved through global collaboration with partners with expertise in medical technology development and evaluation, as well as neonatal and child health. Critical to medical technology design, development, deployment, and eventual uptake and acceptability is understanding user perspectives in the intended setting. Evidence of the feasibility, appropriateness, and acceptability of innovative approaches for improving maternal and neonatal health has not been adequately investigated, which has implications for scale-up [3]. We assessed the feasibility, usability, and acceptability of 2 existing noninvasive, multiparameter, continuous physiological monitoring (MCPM) technologies developed by technology developers EarlySense and Sibel for use in neonates in an African health care setting.

Methods

Study Design

We conducted a qualitative study comprising in-depth interviews and direct observations using a cross-sectional design. This substudy was part of the larger ETNA project to evaluate the accuracy, reliability, and performance of 2 investigational noninvasive MCPM technologies in neonates when compared with verified reference technologies (Figure 1) [4]. The qualitative component used a descriptive and interpretive approach to understand the meanings respondents ascribed to feasibility, usability, and acceptability [5]. Feasibility comprises systemic factors, including hospital infrastructure and operational capacities, as well as functional capacities of the HCP available [6]. Usability comprises design factors affecting user experience, including features that support or hinder the operation of the technology for its intended purpose, such as ease of and efficiency in use and frequency of errors, memorability to a casual user, and user satisfaction with the system [6,7]. Acceptability comprises 2 dimensions: the willingness of HCPs to use the technology during patient interactions and the willingness of caregivers to have the technology used with their neonates [6].

Figure 1. Overview of the 3 multiparameter continuous physiological monitoring technologies.



The study is reported based on the *Consolidated Criteria for Reporting Qualitative Research* (COREQ; Multimedia Appendix 1) [8]. Ethics approvals were obtained from Western Institutional Review Board 20 191 102 (Puyallup, Washington), and the Aga Khan University Nairobi Research Ethics Committee 2019/REC-02 (v2; Nairobi, Kenya).

Study Setting

The study was conducted at the Aga Khan University Hospital, Nairobi, a tertiary teaching and referral hospital with neonatal intensive care and high dependency units. The Aga Khan University Hospital is the busiest private hospital maternity unit in Nairobi, delivering approximately 4,500 neonates a year, and serves as a tertiary referral center for Kenya as well as the East Africa region. The neonatal high dependency unit has an 8-bed capacity and admits, on average, 5 to 10 neonates per week, with an approximate nurse-to-neonate ratio of 1:3. The ETNA project worked within the neonatal high-dependency unit and employed 2 research nurses to support the study.

Recruitment and Selection

A purposive study sample was drawn to include a wide variety of perspectives on the feasibility, usability, and acceptability of the 3 MCPM technologies. The sample consisted of health care administrators, including hospital leadership and administrative staff involved in the procurement of neonatal hospital equipment, HCPs who were direct users of the MCPM technologies (health care provider—direct [HCP-Ds]), indirect HCPs involved in neonatal care (HCP-Is), and caregivers. A sample size of 12 HCPs and 5 health care administrators was estimated to cover perspectives from the staffing positions available and selected from a predefined list of current hospital staff generated by the ETNA team. HCP-D were recruited from ETNA nursing staff. HCP-D was trained for the study and had experience working with the technologies, whereas HCP-I (facility-based neonatal consultants, pediatric residents, and nurses) were oriented to the technologies during the interviews. A sample size of 10 caregivers, including mothers and fathers

of neonates enrolled in the ETNA study, was estimated to reach data saturation because multiple technologies were used with each neonate during their hospital stay.

Study recruitment was publicized using flyers, and study participants were approached in person by a member of the qualitative study team, who were hired as part of the substudy and did not know the participants before the study. Interviewers first introduced themselves as members of the ETNA study team and explained the study in detail.

Data Collection

In-depth interviews with health care administrators, HCPs, and caregivers and direct observations were conducted between July and August 2020. A Kenyan research consultant (VN, PhD in sociology, female) and a research assistant (Diploma in health sciences, female) were hired by the ETNA substudy to collect data. The research assistant underwent a 3-day intensive training in qualitative research methods led by VN before conducting the interviews.

The semistructured interview guide and observation guide (Multimedia Appendices 2 and 3) were piloted within the Kenyan data collection team during training to refine the questions. Face-to-face interviews were conducted in a private place within the hospital after the study participants provided written informed consent. The 30- to 45-minute interview was conducted in English or Kiswahili, the major local language in Kenya, depending on participant preference. One participant opted for a mix of Kiswahili and English, whereas the rest of the participants opted for English. Observations were conducted after obtaining written informed consent from HCP-D and followed a structured guide covering preparation and initial technology application, ongoing monitoring or troubleshooting, and technology disconnect, removal, and cleaning. Interviews were audio-recorded with permission, and data collectors took field notes while conducting the interviews. No repeat interviews were conducted.

Data Analysis

Interviews were transcribed verbatim, translated into English as needed, and managed using NVivo 12 software (QSR International). We used a thematic approach to analyze the data following the methods described by Braun and Clarke to become familiar with the data, generating initial codes, collating identified codes into themes, and describing themes using illustrative quotes [9]. A coding framework was developed deductively from the study objectives to cover feasibility, usability, acceptability, and emergent themes from the transcripts. The coding framework was developed in consensus between the ETNA study team (ASG, MWK, VN, JR, DC, JC, and WMM), and VN conducted the primary coding with review by MWK (Multimedia Appendix 4). Confidentiality was maintained by limiting access of study materials to authorized personnel and ensuring that no identifying information was included in the analysis.

Results

Overview

Overall, the use of the relevant technologies was observed with 12 neonates, and observations took between 2 and 10 minutes per technology. In addition, 27 interviews were conducted, including 10 caregivers (9 mothers, 1 father), 2 HCP-D (study nurses), 10 HCP-I (4 medical doctors, 6 nurses), and 5 health care administrators (nurse managers, program administrators, and hospital unit supervisors). One HCP-I and no caregiver declined to participate. All health care administrators and HCPs had a postsecondary education. There were 2 HCPs with diplomas in nursing (1 HCP-D and 1 HCP-I), and all other health professionals had bachelor's or master's degrees. Health professionals had a median of 9 (<1-29) years of work experience in the medical field. In addition, all but 1 caregiver had a postsecondary education. The median age of caregivers was 33 (range 28-38) years, and they had a median of 2 (range 1-3) children. Caregivers were largely employed in professional occupations, including nursing, banking, human resource services, travel consultancy, business, sales, civil service, and farming.

Feasibility Factors for the Investigational Technologies

Health care administrators described challenges in staffing and maintenance of training in neonatal units as a key feasibility consideration for the development of neonatal MCPM technologies. A health care administrator described:

> *I am finding it difficult to get the expertise that we require because...we don't have many institutions who are training for critical care...neonatal nursing...[P]eople are learning on the job.*

Another health care administrator highlighted:

> *There has been a lot of turnover in the newborn unit. So...you need to now make sure that you are training...on a continuous basis. It is not just about the equipment; the staff also need to have a very good understanding of how that equipment function.*

Within the context of high staff turnover and on-the-job training, simple MCPM technologies were valued for the minimal training required and ease of application.

The minimal training required was a major facilitator reported for feasibility by the participants. Most HCPs and health care administrators (11/17, 65%) reported that the investigational technologies appeared to be easy to train for use and built on existing clinical skills. Referencing the EarlySense technology, an HCP-D nurse highlighted:

> *You only need very minimal training… just [place] it under the mattress and it monitors the baby, monitors the pressure. Very easy to use.*

An HCP-I nurse noted that the Sibel technology could be easily learned within a few hours mentored by a current user:

> *[it requires] like an on-job training, like maybe a few hours, because...it is not...totally new from what is being used.*

Feasibility challenges reported included the requirement of ancillary equipment, Wi-Fi requirements, and concerns about integration with existing facility equipment.

A minority of HCP and health care administrators expressed concerns about the requirements for external screens (2/17, 12%) and Wi-Fi (6/ 17, 35%). Equipment integration concerns were expressed particularly among health care administrators (3/5, 60%). A health care administrator said:

> *...we would want all the information in one place, and not this one here and the other on the other side...so that clinicians are not looking for information in two or three different places...*

Affordability has also been raised as an important issue for feasibility. An HCP-I nurse said of the EarlySense technology:

> *Number one thing will be the cost...If the cost is higher, then [the hospital administrators] will have to weigh which is a cheaper option that will give more or less the same results.*

Because the investigational MCPM technologies were perceived as simple, there was an expectation that costs would be affordable. A health care administrator said of the Sibel technology:

> *They should be about 10,000 shillings (approximately 90 USD) and not more than that. They are basic equipment.*

Usability Factors for the Investigational Technologies

Ease of use and efficiency of the technologies for neonatal care were major usability facilitators reported by most HCPs (9/12, 75%), health care administrators (4/5, 80%), and caregivers (7/10, 70%). An HCP-I nurse noted the EarlySense technology "is quite simple than our normal standard monitoring device here...it looks easier to use." Speaking about the Sibel technology, an HCP-D nurse said:

> *It even [has] more functions than our current cardiac monitor so that's a plus that we are having less manipulation to the baby in terms of attachment, but we are having much results. You can see more...heart rate, respirations, we can see movement, we can see temperature...*

Ease of use was also reflected in observations; trained HCP-D nurses were able to prepare and initiate the technologies, monitor, and disconnect smoothly without many errors or assistance from other HCPs. Caregivers shared that they found the investigational technologies easy to understand and memorable even as casual users while monitoring their neonate at the hospital. Regarding the EarlySense technology, a mother said:

> *At a glance, you're able to know all your readings...You're actually able to monitor at a glance; you don't need to worry.*

Another mother said of the Sibel technology:

> *.....on her iPad, I could see the oxygen [levels]. It was on the right levels.*

The potential for wireless features to improve work efficiency was another major usability factor reported by the participants. Approximately half of the caregivers, HCPs, and health care administrators (13/27, 48%) commented that the potential of the wireless or noncontact features of the investigational technologies to transmit information to an external screen and remotely monitor multiple neonates could increase the speed of HCP detection and reaction to changes in vital signs. An HCP-D nurse commented on the EarlySense technology:

> *The screen that you are using to display the results, you can put it at the nursing station...So you can monitor many babies at the same time. It will reduce workload...*

Speaking of the Sibel technology, a mother said:

> *You could be in a facility where babies are so many, so the service provider, they're overwhelmed. But if there are such devices that they're able to relay information faster, that means so many babies at least can be observed comfortably, so you save lives.*

Fewer attachments have also been reported to increase efficiency in care by reducing the need to disconnect and reconnect the neonate during HCP examinations and caregiving needs, such as breastfeeding. Speaking of the EarlySense technology, a health care administrator said:

> *...the coding is remote. It doesn't interfere with routine care...Like sometimes, I have to remove certain wires to be able to examine a baby properly. So, the fact that it leaves the baby unencumbered with all those things I think is a huge advantage*

However, with the EarlySense technology specifically, there were concerns from HCP and health care administrators (5/17, 29%) of monitoring during disruptions when the neonate was off the mattress during breastfeeding or otherwise being carried by caregivers.

Small size and portability were reported by some HCPs and health care administrators (8/17, 47%) as potential facilitators and challenges to usability. Although portability and convenience are linked, there was also concern about misplacement and theft because of their small size and portability. For example, a health care administrator said that it would be critical for the Sibel technology:

> *...to make sure that these things aren't lost by staff...We might buy them, but at the end of the year, they might all be lost. Because, you know something which is attached and something which is a little bit big might be better.*

Other usability challenges included concerns about infection control, shared by 41% (7/17) of the HCPs and health care administrators. In addition, 26% (7/27) of overall participants expressed concerns that equipment may be too large for preterm and low-birth-weight neonates. The plastic material of the Sibel technology was deemed easy to clean, but there was a preference for disposable items to streamline infection control processes at the hospital. Equipment that is too large for preterm and low-birth-weight neonates may lead to potentially poor application and inaccurate readings. For example, regarding

concerns that the sensors for the Sibel technology would not fit a preterm neonate, a health care administrator shared:

> *For the baby's chest, some of them like the preterms...I don't know whether it would be so big, and then if it is big, then it is not connecting well.*

Acceptability Factors for the Investigational Technologies

Perceptions of neonate comfort were a major acceptability facilitator reported by many of the HCPs (6/ 12, 50%), health care administrators (4/5, 80%), and caregivers (610, 60%). The investigational technologies were seen as more comfortable and did not interfere with the neonates' movements and sleep. Speaking about the EarlySense technology, one mother said:

> *It doesn't interfere in any way with the baby. The baby is sleeping; they can just sleep, you know?...It's painless while it's measuring.*

Another mother said, of the Sibel technology:

> *...the baby didn't seem uncomfortable...it didn't cause the baby any discomfort.*

Half of the caregivers interviewed also mentioned that the investigational technologies were simpler, less intimidating, and more acceptable than conventional monitoring technologies.

Concern about potential harm around electrical fields and wireless connectivity was a major acceptability challenge mentioned by caregivers (8/10, 80%) and HCPs (7/12, 58%). For example, as one mother said, about the EarlySense technology:

> *...okay, they're not using wires, so what are they using? Is there radiation, you know, that can harm my child?...of course, we asked about that, and we were told no, they're safe...The concerns...were put to rest.*

Speaking about the Sibel technology, an HCP-I nurse highlighted concerns shared by both caregivers and herself:

> *Now, you are not seeing any...wires moving from that device to...the screen next to the baby, so they [parents] want to know how that information is being passed...I am also wondering how it is working with that Bluetooth thing...so...personally I will go with the old version.*

Comparisons With the Reference Technology

A few HCPs and health care administrators (3/17, 18%) shared that in comparison with the investigational technologies, the complexity of the Masimo Rad-97 reference technology may require longer training. However, in contrast to the requirement of external screens or Wi-Fi with the investigational technologies, some HCPs and health care administrators shared that the reference technology is feasible within the Kenyan context because it is a stand-alone unit (3/17, 18%) and does not require Wi-Fi (2/17, 12%). For example, a health care administrator said:

> *...you don't necessarily need another device to monitor, unlike Sibel where you need a [tablet]...In terms of feasibility, I would go for the implementation of Rad-97 first...*

In addition, an HCP-I nurse said:

> *Aah, what I like about it is that...I don't think it requires those WiFi things...so it can be used anywhere, any part of the country.*

Of the 27 respondents, 4 (15%) mentioned that similar to the investigational technologies, the Masimo Rad-97 technology would be useful for care; 35% (6/17) of HCP and health care administrators shared that the Masimo Rad-97 technology seemed to have most of the features of the larger, wall-mounted conventional monitoring technology, but in a small and portable design.

In contrast to the investigational technologies, perceptions of neonate discomfort negatively impacted the acceptability of the reference technology, particularly with the nasal prong for capnography (12/ 27, 44%). Respondents (11/27, 41%) stated that wires and other attachments represent critical care and embody the seriousness of the health condition. A father shared:

> *Sometimes too many wires tend to shock...You might feel that the baby is in danger...Because the wires reflect the baby is, uh, is in dire need of help.*

The nasal prong was seen as part of oxygen delivery, which was especially stressful for caregivers with neonates previously in intensive care. A mother shared:

> *I freaked out because having been in the same situation for the past one week...the first question that came into my mind is the baby going to struggle again breathing using those tubing on the nose.*

An HCP-D nurse shared that caregivers were sometimes reluctant to accept the reference technology, as they perceived the nasal prong as invasive.

The facilitator of acceptability for the reference technology was brand familiarity, shared by HCPs (5/12, 42%) and health care administrators (3/ 5, 60%). An HCP-I physician described:

> *It is a device that has been used in the past, and its still being used all over the world. It's a no brainer. It's like going and asking someone "should you drive a Mercedes" it's a known brand.*

Feasibility, usability, and acceptability factors for the investigational and reference technologies included a myriad of facilitators and barriers (Table 1; Multimedia Appendices 5 and 6). Some factors have been reported to be potential facilitators and barriers.

Table 1. Feasibility, usability, and acceptability factors.

Factors	Investigational technologies		Reference technology	
	Facilitators	Barriers	Facilitators	Barriers
Feasibility	• Easy to train for use	• Requirement of ancillary equipment • Wireless connectivity requirements (Sibel Bluetooth or EarlySense Wi-Fi) • Integration capacities with existing equipment	• Stand-alone unit • Does not require Wi-Fi	• Longer training because of complexity
Usability	• Ease of use and useful for care • Wireless features improve work efficiency • Ability for infection control • Small size and portability	• Monitoring disruptions; such as when neonates are restless neonates or off mattress (EarlySense) • Need for appropriate preterm and low birth-weight sizing • Ability for infection control • Small size and portability	• Useful for care • Small size and portability	• Small size and portability
Acceptability	• Perceptions of neonate comfort • Simple and less intimidating	• Concerns about side effects from wireless connections	• Brand familiarity	• Perceptions of neonate discomfort because of nasal capnography • Wiring and tubing linked to critical care

Use Cases

HCPs and health care administrators shared that the investigational and reference technologies would be useful in different components of neonatal care at the hospital. As one health care administrator said, “I think in their own different capacity, they all have potential.”

Investigational technologies were especially recommended in the postnatal ward or nursery where healthy preterm and low-birth-weight neonates were being monitored. In these areas, there may be many neonates, largely under the care of their mothers and in stable health. Respondents suggested that the EarlySense technology would be useful “for babies who are being monitored [but] who are not so badly off” (HCP-I nurse), for “the postnatal babies...[where] there are many babies being continuously monitored” (HCP-D nurse), and for “a baby who we didn’t expect any sepsis or any challenges, and from this, we [would] be able to capture early signs of infection” (health care administrator). A health care administrator recommended the Sibel technology for the postnatal ward:

> *A mother who has delivered and has her baby needs to rest, and yet we have to monitor that baby. I would want that baby to be put on this. She can breastfeed and whatever she is doing, I can still be able to see the patterns and trends.*

The reference technology was described as more suitable for neonates requiring more critical care, where neonates may be largely under the care of an HCP rather than caregivers. “*Especially with capnography in place*,” an HCP-D nurse said that the reference technology is not appropriate “for the postnatal babies because it makes [parents] feel like their baby is very sick or maybe on oxygen.” She also said:

> *I don’t think it is appropriate for postnatal, the wiring and the limited space that is there...and...because the baby has to be unplugged from the wires for breastfeeding or bathing, so it is not exactly feasible where neonate is under almost complete care of the mother.*

However, she said that:

> *for babies in HDU or newborn ICU, it is very much feasible, as...it is familiar to what we are using.*

Discussion

Principal Findings

The purpose of this qualitative study was to assess the feasibility, usability, and acceptability of 2 noninvasive MCPM technologies for neonates in an African health care setting. Study participants reported that the investigational technologies were feasible and useful in the care of neonates at the Aga Khan University Hospital in Nairobi. Feasibility facilitators included simple training requirements, whereas infrastructural requirements such as Wi-Fi, external display screens, and limited integration with existing equipment were reported as potential barriers. Usability facilitators included ease of use and wireless features, and concerns of possible harm from wireless connections and mistrust of unfamiliar technologies emerged as potential barriers to acceptability. Appropriate sizing for preterm and low-birth-weight infants, portability, human resource requirements and training, and perceptions of wireless technologies were identified as key issues to consider during

the development and implementation of neonatal MCPM technologies.

Implications for Practice

Our experience with MCPM technologies highlights the potential of different technologies for different neonatal care settings. The EarlySense technology was recommended for the postnatal ward where neonates were largely in stable health conditions and there were more neonates than nurses who could regularly monitor. Because the EarlySense technology only monitored while the neonate was resting on the mattress, the Sibel technology may be more appropriate during kangaroo mother care when the neonate spends most of their time on the caregiver's chest. The investigational technologies' ease of use supported their function within less critical areas of neonatal care, where the neonate is largely under the care of family members. In contrast, the reference technology used in this study was valued as a more compact and portable version of the larger, wall-mounted existing monitoring systems in intensive care units. Contextually appropriate MCPM technologies are particularly needed for the management of clinically unstable neonates to support early and safe initiation of evidence-based interventions such as kangaroo mother care and to monitor emerging complications such as hypothermia during bubble continuous positive airway pressure for respiratory distress [10,11]. This is important because immediate kangaroo mother care of low-birth-weight infants in critical condition has been shown to reduce infant mortality rates compared with conventional kangaroo mother care initiated after stabilization [12].

Although wireless features of the investigational technologies supported usability and acceptability in certain dimensions, including the potential for remote monitoring, simple design, less interference with care, increased comfort, and concern for potential health risks with wireless connectivity emerged as an unexpected theme. A study evaluating the acceptability of a wireless fetal heart rate monitoring device among pregnant women in rural Uganda also reported concerns among mothers about possible negative effects of electromagnetic radiation [13]. An improved understanding of barriers to and enablers of innovative neonatal health technologies for resource-constrained settings is a recognized gap in the literature [3]. Two reviews of wearable continuous monitoring sensors for neonates compiled products and key features but did not investigate acceptability or implementation factors [14,15]. Concerns about potential side effects from wireless connections and electromagnetic fields emphasize the importance of caregiver engagement and the need to work with HCPs to address clients' mistrust of and fear of novel technologies. Caregivers expressed fear because of a lack of understanding of these technologies, but the fear appeared to be alleviated with HCP explanation for some.

Study findings such as identifying use case scenarios for different neonatal MCPM technologies and fears that wireless technologies may have adverse health effects highlight the importance of evaluating feasibility, usability, and acceptability during the development of medical technologies. Although medical technologies may demonstrate efficacy, their adoption, uptake, and use may be limited if implementation factors are not considered and incorporated during technology development. The potential impact of innovative neonatal MCPM technologies is substantial, particularly in resource-constrained settings. Frequently, there may be little to no neonatal continuous monitoring available in these settings, despite being routine in high-income settings for those who require it. This lack of monitoring may contribute to the higher rates of neonatal morbidity and mortality in resource-constrained settings [16-18]. An observational study at Kenyatta National Hospital in Nairobi reported that very few neonates had their vital signs recorded in the first hour of life, and more than half did not receive a temperature recording (54%), heart rate recording (56%), or respiratory rate recording (56%) on the first day of hospital admission [16]. Observations at 6 hospitals in Nairobi County found that missed vital sign monitoring and other nursing tasks were associated with nursing shortages and high patient workloads [17]. MCPM technologies are valuable for improving the quality of neonatal care by expanding nurses' capacities to monitor more neonates regularly and efficiently.

Strengths and Limitations

Of note, this qualitative study was conducted at a private, tertiary hospital where the study participants were highly educated, and almost all of the caregivers interviewed had university education and professional employment. In addition, limiting the generalizability of our study findings is that private and public hospitals in Kenya have dramatically different nursing workloads and infrastructure, with median ratios of 3 infants to 1 nurse at private hospitals around Nairobi and 19 infants to 1 nurse at public hospitals [17]. With a reliable back-up electrical system and maintenance team on staff at Aga Khan University Hospital, Nairobi, electrical outages, technology malfunction, and maintenance were not highlighted as feasibility concerns by our study participants. The feasibility of these investigational MCPM technologies for neonates has important implications for the sustainability and prevention of technology graveyards of nonfunctional or locally inappropriate technologies. Future research can explore whether feasibility, usability, and acceptability issues shift in a public hospital setting where resources may be more constrained. Another limitation of the study is that usability was not directly assessed among nontrained (HCP-I) users. The strengths of the study include the use of direct observations to support interview findings, as well as conducting in-depth interviews with caregivers, HCPs, and health care administrators to understand a diversity of perspectives.

Conclusions

MCPM for neonates is a critical component of comprehensive care that supports the effectiveness of other neonatal interventions. Our study examined the feasibility, usability, and acceptability of 2 investigational MCPM technologies for neonates compared with a reference MCPM technology and found that the different technologies fit different areas within the continuum of neonatal care at the hospital. Although each technology presented advantages suited for different neonatal care domains, challenges in maintaining training and ensuring

feasibility within resource-constrained health care settings warrant further research.

Acknowledgments

The authors are grateful to the ETNA (Evaluation of Technologies for Neonates in Africa) team for their support and to the health professionals and caregivers for sharing their thoughts and experiences with us. The authors would also like to thank Marianne Vidler for her insights on study design and the data collection tools as well as the research team who collected and managed the data without whom this study would not have been possible.

This trial was funded by the Bill & Melinda Gates Foundation (OPP12303136) and Save the Children Innovation Fund. Apart from the input provided for the trial design, funders had no role in data collection, analysis, interpretation, or writing of the report.

Authors' Contributions

ASG, JMA, and WMM conceptualized the study and components of the methodology with support from VN, JR, MC, BH, and RO. VN, JR, DC, and BH were responsible for data collection and project administration. MWK and VN performed the analyses. ASG and MWK wrote the first draft of the manuscript. All authors provided feedback and reviewed the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
COREQ (Consolidated Criteria for Reporting Qualitative Research) Checklist.
[DOCX File , 16 KB - jmir_v23i10e29755_app1.docx]

Multimedia Appendix 2
In-depth interview guides.
[PDF File (Adobe PDF File), 737 KB - jmir_v23i10e29755_app2.pdf]

Multimedia Appendix 3
Direct observation guide.
[PDF File (Adobe PDF File), 309 KB - jmir_v23i10e29755_app3.pdf]

Multimedia Appendix 4
Coding tree.
[DOCX File , 18 KB - jmir_v23i10e29755_app4.docx]

Multimedia Appendix 5
Feasibility, usability, and acceptability factors of investigational and reference technologies with illustrative quotes.
[DOCX File , 23 KB - jmir_v23i10e29755_app5.docx]

Multimedia Appendix 6
Themes by technology and participant group.
[DOCX File , 16 KB - jmir_v23i10e29755_app6.docx]

References

1. Lawn JE. The child survival revolution: what next? The Lancet 2014 Sep 13;384(9947):931-933. [doi: 10.1016/s0140-6736(14)61057-6]
2. Maynard KR, Causey L, Kawaza K, Dube Q, Lufesi N, Maria Oden Z, et al. New technologies for essential newborn care in under-resourced areas: what is needed and how to deliver it. Paediatr Int Child Health 2015 Aug;35(3):192-205. [doi: 10.1179/2046905515Y.0000000034] [Medline: 26053669]
3. Lunze K, Higgins-Steele A, Simen-Kapeu A, Vesel L, Kim J, Dickson K. Innovative approaches for improving maternal and newborn health--a landscape analysis. BMC Pregnancy Childbirth 2015 Dec 17;15:337 [FREE Full text] [doi: 10.1186/s12884-015-0784-9] [Medline: 26679709]
4. Ginsburg AS, Nkwopara E, Macharia W, Ochieng R, Waiyego M, Zhou G, et al. Evaluation of non-invasive continuous physiological monitoring devices for neonates in Nairobi, Kenya: a research protocol. BMJ Open 2020 Apr 12;10(4):e035184 [FREE Full text] [doi: 10.1136/bmjopen-2019-035184] [Medline: 32284391]

5. LeCompte M, Schensul J. Designing and Conducting Ethnographic Research 2nd Edition. Lanham, MD USA: AltaMira Press; 2010.
6. Ginsburg AS, Tawiah Agyemang C, Ambler G, Delarosa J, Brunette W, Levari S, et al. mPneumonia, an innovation for diagnosing and treating childhood pneumonia in low-resource settings: a feasibility, usability and acceptability study in Ghana. PLoS One 2016 Oct 27;11(10):e0165201 [FREE Full text] [doi: 10.1371/journal.pone.0165201] [Medline: 27788179]
7. Nielson J. Usability Engineering. Philadelphia: Elsevier Science; 1994.
8. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
9. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
10. Kinshella MW, Salimu S, Chiwaya B, Chikoti F, Chirambo L, Mwaungulu E, et al. "So sometimes, it looks like it's a neglected ward": health worker perspectives on implementing kangaroo mother care in southern Malawi. PLoS One 2020 Dec 17;15(12):e0243770 [FREE Full text] [doi: 10.1371/journal.pone.0243770] [Medline: 33332395]
11. Carns J, Kawaza K, Quinn M, Miao Y, Guerra R, Molyneux E, et al. Impact of hypothermia on implementation of CPAP for neonatal respiratory distress syndrome in a low-resource setting. PLoS One 2018 Mar 15;13(3):e0194144 [FREE Full text] [doi: 10.1371/journal.pone.0194144] [Medline: 29543861]
12. WHO Immediate KMC Study Group, Arya S, Naburi H, Kawaza K, Newton SA, Anyabolu CH, et al. Immediate "Kangaroo Mother Care" and survival of infants with low birth weight. N Engl J Med 2021 May 27;384(21):2028-2038 [FREE Full text] [doi: 10.1056/NEJMoa2026486] [Medline: 34038632]
13. Mugyenyi GR, Atukunda EC, Ngonzi J, Boatin A, Wylie BJ, Haberer JE. Functionality and acceptability of a wireless fetal heart rate monitoring device in term pregnant women in rural Southwestern Uganda. BMC Pregnancy Childbirth 2017 Jun 08;17(1):178 [FREE Full text] [doi: 10.1186/s12884-017-1361-1] [Medline: 28595604]
14. Zhu Z, Liu T, Li G, Li T, Inoue Y. Wearable sensor systems for infants. Sensors (Basel) 2015 Feb 05;15(2):3721-3749 [FREE Full text] [doi: 10.3390/s150203721] [Medline: 25664432]
15. Memon SF, Memon M, Bhatti S. Wearable technology for infant health monitoring: a survey. IET Circuits Devices Syst 2019 Oct;14(2):115-129. [doi: 10.1049/iet-cds.2018.5447]
16. Mitchell EJ, Qureshi ZP, Were F, Daniels J, Gwako G, Osoti A, et al. Feasibility of using an Early Warning Score for preterm or low birthweight infants in a low-resource setting: results of a mixed-methods study at a national referral hospital in Kenya. BMJ Open 2020 Oct 28;10(10):e039061 [FREE Full text] [doi: 10.1136/bmjopen-2020-039061] [Medline: 33115899]
17. Gathara D, Serem G, Murphy GA, Obengo A, Tallam E, Jackson D, et al. Missed nursing care in newborn units: a cross-sectional direct observational study. BMJ Qual Saf 2020 Jan;29(1):19-30 [FREE Full text] [doi: 10.1136/bmjqs-2019-009363] [Medline: 31171710]
18. Aluvaala J, Nyamai R, Were F, Wasunna A, Kosgei R, Karumbi J, SIRCLE/Ministry of Health Hospital Survey Group. Assessment of neonatal care in clinical training facilities in Kenya. Arch Dis Child 2015 Jan 19;100(1):42-47 [FREE Full text] [doi: 10.1136/archdischild-2014-306423] [Medline: 25138104]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
ETNA: Evaluation of Technologies for Neonates in Africa
HCP: health care provider
HCP-D: health care providers-direct users of the MCPM technologies
HCP-I: health care providers-indirect users of the MCPM technologies
MCPM: multiparameter continuous physiological monitoring

Edited by R Kukafka; submitted 19.04.21; peer-reviewed by E Green, S Lalla-Edward; comments to author 28.06.21; revised version received 01.07.21; accepted 27.07.21; published 28.10.21.

Please cite as:
Ginsburg AS, Kinshella MLW, Naanyu V, Rigg J, Chomba D, Coleman J, Hwang B, Ochieng R, Ansermino JM, Macharia WM Multiparameter Continuous Physiological Monitoring Technologies in Neonates Among Health Care Providers and Caregivers at a Private Tertiary Hospital in Nairobi, Kenya: Feasibility, Usability, and Acceptability Study
J Med Internet Res 2021;23(10):e29755
URL: https://www.jmir.org/2021/10/e29755
doi:10.2196/29755
PMID:34709194

Original Paper

How Food Marketing on Instagram Shapes Adolescents' Food Preferences: Online Randomized Trial

Marie Bragg[1], PhD; Samina Lutfeali[2], MPH; Tenay Greene[1], BA; Jessica Osterman[1], BA; Madeline Dalton[3], PhD

[1]Department of Population Health, New York University Grossman School of Medicine, New York City, NY, United States

[2]Stanford Graduate School of Business, Palo Alto, CA, United States

[3]Department of Pediatrics, Dartmouth Geisel School of Medicine, Hanover, NH, United States

Corresponding Author:
Marie Bragg, PhD
Department of Population Health
New York University Grossman School of Medicine
180 Madison Ave, 3rd Floor
New York City, NY, 10016
United States
Phone: 1 646 501 2717
Email: Marie.Bragg@nyulangone.org

Abstract

Background: Worldwide obesity rates have prompted 16 countries to enact policies to reduce children's exposure to unhealthy food marketing, but few policies address online advertising practices or protect adolescents from being targeted. Given adolescents spend so much time online, it is critical to understand how persuasive Instagram food advertisements (ads) are compared with traditional food ads. To strengthen online food marketing policies, more evidence is needed on whether social media ads are more persuasive than other types of ads in shaping adolescents' preferences.

Objective: This study examined whether adolescents could identify food companies' Instagram posts as ads, and the extent to which Instagram versus traditional food ads shape adolescents' preferences.

Methods: In Part 1, participants aged 13-17 years (N=832) viewed 8 pairs of ads and were asked to identify which ads originated from Instagram. One ad in each pair was selected from traditional sources (eg, print; online banner ad), and the other ad was selected from Instagram, but we removed the Instagram frame—which includes the logo, comments, and "likes." In Part 2, participants were randomized to rate food ads that ostensibly originated from (1) Instagram (ie, we photoshopped the Instagram frame onto ads); or (2) traditional sources. Unbeknownst to participants, half of the ads in their condition originated from Instagram and half originated from traditional sources.

Results: In Part 1, adolescents performed worse than chance when asked to identify Instagram ads (P<.001). In Part 2, there were no differences on 4 of 5 outcomes in the "labeled ad condition." In the "unlabeled ad condition," however, they preferred Instagram ads to traditional ads on 3 of 5 outcomes (ie, trendiness, P=.001; artistic appeal, P=.001; likeability, P=.001).

Conclusions: Adolescents incorrectly identified traditional ads as Instagram posts, suggesting the artistic appearance of social media ads may not be perceived as marketing. Further, the mere presence of Instagram features caused adolescents to rate food ads more positively than ads without Instagram features.

(J Med Internet Res 2021;23(10):e28689) doi:10.2196/28689

KEYWORDS

food marketing; traditional media; social media; adolescents; Instagram

Introduction

The high prevalence of childhood obesity persists in the United States, and the National Academy of Medicine identifies food marketing as a major contributing factor [1]. Despite both national and international efforts to restrict child-targeted food marketing, most policies do not protect adolescents aged 13-17 years, nor do they address social media advertising [2]. Yet adolescents spend an increasingly high number of hours online each day.

Recognizing social media's growing popularity, food companies have dramatically increased their advertising presence on social

media platforms in ways that appeal to young audiences. Coca-Cola, for example, dedicates 20% of their US $4 billion annual advertising budget to social media [3]. One descriptive study reported that 6.2 million adolescents followed a sample of 27 food and beverage brands on Instagram and Twitter [4]. This shift in marketing and exposure concerns public health experts [5] because studies have shown that food companies' official social media accounts promote mostly unhealthy products [6-8], and increase brand recognition among youth [9,10].

Adolescents' unique developmental stage may compound their vulnerability to social media food advertisements (ads). Social Norms Theory [11-13] suggests social media may capitalize on adolescents' exquisite sensitivity to peer behavior—"likes," for example, represent social norms that may signal to adolescents which social media accounts they should follow [14-16]. Experts have expressed concern that the ability to "like" posts also makes social media advertising uniquely interactive and may lead adolescents to perceive brands as friends more than companies [10]. Ads posted by companies may look similar to posts by friends because of context (ie, they appear interspersed with posts shared by friends) and content (ie, companies' posts mimic the aesthetic of everyday consumers' posts), and it is unclear whether these marketing strategies cloud adolescents' ability to identify company posts as advertising. A neuroimaging study showed heightened activity in the nucleus accumbens—a reward hub of the brain—among adolescents who viewed their own posts with high numbers of "likes" versus few "likes" [17]. Given the nucleus accumbens is more sensitive to reward among adolescents compared with adults and children [18], those findings reinforce adolescents' unique susceptibility to social media ads.

Despite the emerging research on adolescents' engagement with food marketing on social media, no studies have compared the power of unhealthy food ads on social media versus more traditional ads (eg, print ads; noninteractive website banner ads). It is not known whether adolescents are skilled at identifying social media ads. Besides, no studies have examined whether social media advertising is more powerful than traditional ads in its ability to affect adolescents' preferences. Finally, the combined effects of showing adolescents food ads and Instagram frames that include the Instagram logo, "likes," and comments have not been explored. Instagram frames that include the Instagram logo, "likes," and comments may have their own brand power and make ads highly appealing to adolescents.

The objective of this study is to address those gaps by examining whether adolescents could distinguish between food companies' Instagram ads and their more traditional ads, and the extent to which Instagram food ads generate more appeal compared with traditional ads. Given adolescents' frequent social media use [19], we hypothesized that participants would correctly identify Instagram ads, and that they would prefer Instagram ads more than traditional ads regardless of whether ads were photoshopped to appear as though they had originated from Instagram.

Methods

Study Population

We recruited 1044 adolescents aged 13-17 years who identified as either Black/African American or non-Latino White through Dynata, a firm that recruits research participants using online panels, digital networks, websites, SMS text messaging, and telephone alerts. The Institutional Review Board at New York University School of Medicine approved our study.

Of the 1044 adolescents who started the survey, 976 completed it, and 884 correctly answered our data integrity question (ie, "Type 'Facebook' in the box below."). A total of 52 adolescents identified as a race/ethnicity other than Black/African American or non-Latino White and were excluded from the analyses. We included only Black and non-Latino White adolescents in our sample because companies disproportionately target Black adolescents with their least healthy products [20,21], and White adolescents are featured in the majority of food ads [22]. Because of these differences in targeted marketing, we designed stimuli that featured either Black or non-Latino White individuals. Secondary analyses that examine a subset of racially targeted ads are under review elsewhere. Among the final sample (N=832), 387 (46.5%) adolescents identified as Black or African American and 445 (53.5%) adolescents identified as non-Latino White. Table 1 presents adolescents' self-reported demographic characteristics and social media usage.

Table 1. Demographic characteristics and social media usage of sample: January–June 2018.

Demographic characteristics	Total sample (N=832)	Part 2: Labeled advertisement condition (n=381)	Part 2: Unlabeled advertisement condition (n=451)
Age, years, mean, (SD)	14.73 (1.67)	14.75 (1.64)	14.71 (1.69)
Gender, n (%)			
Male	426 (51.2)	197 (51.7)	229 (50.8)
Female	406 (48.8)	184 (48.3)	222 (49.2)
Race, n (%)			
Non-Latino White	445 (53.5)	211 (55.4)	234 (51.9)
Black/African American	387 (46.5)	170 (44.6)	217 (48.1)
When do you use social media?, n (%)			
Right when you wake up	305 (36.7)	131 (34.4)	174 (38.6)
Before school	407 (48.9)	173 (45.4)	234 (51.9)
On the way to school	309 (37.1)	143 (37.5)	166 (36.8)
At school	258 (31.0)	115 (30.2)	143 (31.7)
During lunch	381 (45.8)	173 (45.4)	208 (46.1)
On the way home from school	323 (38.8)	151 (39.6)	172 (38.1)
After school	381 (45.8)	257 (67.5)	307 (68.1)
While doing homework	264 (31.7)	119 (31.2)	145 (32.2)
After doing homework	390 (46.9)	180 (47.2)	210 (46.6)
During dinner	158 (19.0)	74 (19.4)	84 (18.6)
Before bed	497 (59.7)	232 (60.9)	265 (58.8)
Right before going to sleep	234 (28.1)	102 (26.8)	132 (29.3)
Social media account history, n (%)			
Do you have...[check all that apply]			
Instagram?	582 (70.0)	279 (73.2)	303 (67.2)
Facebook?	710 (85.3)	321 (84.3)	389 (86.3)
Snapchat?	410 (49.3)	189 (49.6)	221 (49.0)
Tumblr?	71 (8.5)	40 (10.5)	31 (6.9)
Twitter?	394 (47.4)	194 (50.9)	200 (44.3)
Average number of social accounts per participant based on responses to question above, mean (SD)	2.60 (1.20)	2.69 (1.15)	2.54 (1.20)
Have you made a purchase through social media before?, n (%)	167 (20.1)	83 (21.8)	84 (18.6)

Advertisement Development

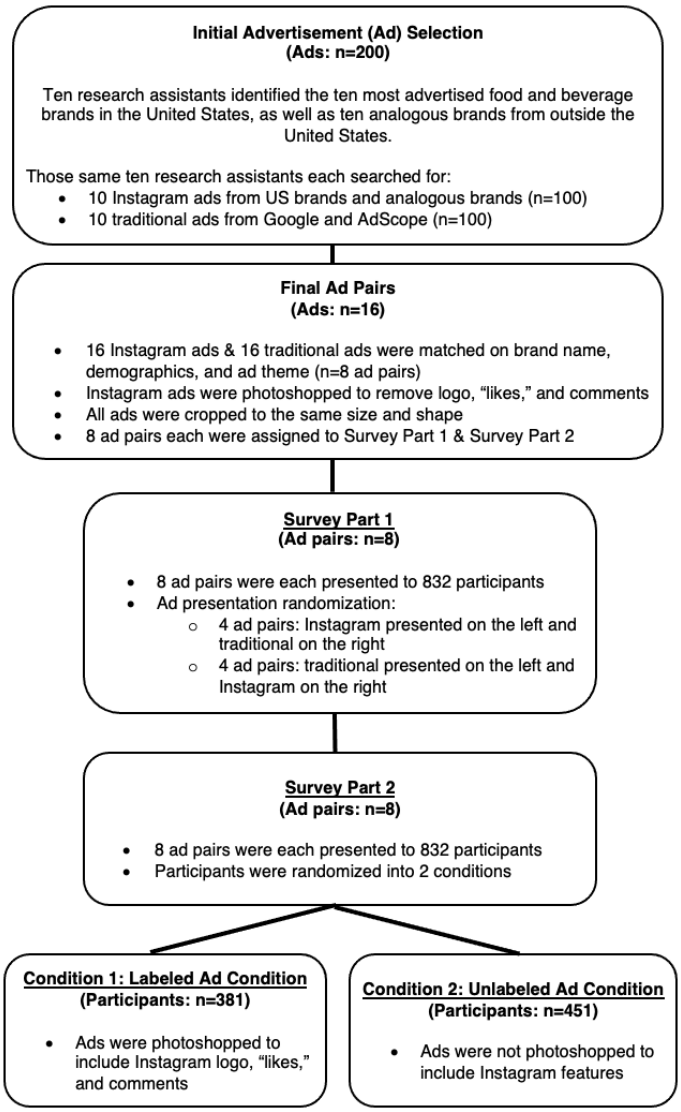
We identified the 10 most advertised food and beverage brands in the United States [23] and then identified 10 analogous brands originating outside of the United States (eg, Lay's Potato Chips from the United States vs. Walker's Crisps from the UK). We identified those analogous brands because consumers' familiarity with brands could potentially affect their survey responses [24]. We then aimed to select from those brands (1) 200 unhealthy food and beverage ads on Instagram; and (2) 200 traditional ads to serve as potential stimuli. To select the 200 Instagram ads for unhealthy food and beverages, we asked 10 research assistants to screen capture 10 Instagram ads from the official accounts of the most advertised brands and their analogous brands. Research assistants screen captured a random sample of 10 ads for unhealthy products from each brand (ie, 10 ads from 10 US brands + 10 ads from 10 analogous brands = 200 Instagram ads as potential stimuli). To select a random sample of 10 ads from each brand's Instagram account, research assistants were instructed to use a random number generator and then select the ad corresponding to the generated number by counting from the most recently posted ad. For example, if the number "10" was randomly generated, the research assistant would select the tenth most recent ad posted to the brand's Instagram feed. To identify traditional ads, we asked those same research assistants to capture 10 traditional ads for the same 10 most advertised and 10 analogous brands by searching on

Google and AdScope, a repository of all ads from television, print, radio, and the internet [25].

To assess the pairings of traditional and Instagram ads, we generated a codebook of ad themes using an iterative process. Researchers identified ad themes by assessing the ad image and text (eg, snowflakes or references to the cold for a "winter" theme). They then searched for traditional and Instagram ads with matching themes. In our first review of the matched ad pairs, we met 90% agreement. Discrepancies were resolved by consensus. The final ad themes included the following: features a human actor, features a nonhuman character, race of actor product visibility, product handling (eg, whether a person or character in the image holds the promoted product), seasonal (eg, fall, winter, spring, and summer themes), and number of actors or characters.

Researchers came to unanimous consensus on all matched pairs between traditional and Instagram ads based on demographic characteristics of the people in the ad and ad theme. We asked the researchers to identify 10 of the final 16 ad pairs from brands originating outside of the United States. By using ads from brands based in the United States we could evaluate reactions to ads that the participants would be most likely to see in their day-to-day life. By including some brands based outside of the United States, we could isolate reactions to the ad-specific effects of interest from potential confounding factors of preformed attitudes toward brands. As intended, this search process generated a final set of 32 unhealthy food and beverage ads, which included 16 ad pairs that were matched on brand name, demographics, and ad theme (Figure 1).

Figure 1. Advertisement development process and study procedure.



Survey Procedures

After parents provided informed consent, adolescents assented their participation and completed the online survey with a median completion time of 19 minutes. Qualtrics hosted the online survey. Data were collected in 2018 and analyzed in 2019.

In Part 1 of the survey, adolescents viewed 8 pairs of unhealthy food and beverage ads presented in random order (eg, a Starbucks magazine ad alongside a Starbucks Instagram ad). To remove cues indicating which ads originated from Instagram, we used Photoshop to compare advertising images with and without an "Instagram frame" which includes the Instagram

logo, "likes," and comments. We also cropped images to make ads the same shape because Instagram images are often square, which would have made it easier for adolescents to identify the Instagram ad. See Figure 2 for an example of our ad stimuli. We randomized the ad presentation so that half the pairs showed the Instagram ad on the left and the traditional ad on the right (Figure 1).

After viewing each pair, adolescents answered the following question, "Which of these photos do you think [brand name] would be most likely to post on their Instagram account?" See Figure 2 for an example of survey questions and images from Part 1.

Figure 2. Sample Question from Part 1: Identify the Instagram ad. In this question, the Instagram ad is on the right.

Adolescents were then randomly assigned to 1 of 2 conditions (Part 2) where ads ostensibly originated from (1) Instagram (ie, "labeled ad condition" in which advertising images have Instagram frames); or (2) traditional sources (ie, "unlabeled ad condition" in which advertising images do not have Instagram frames; Figure 1). Unbeknownst to participants, half of the ads in their condition originated from Instagram and half originated from traditional sources. Ads were presented in random order. The survey questions asked participants to rate the ad on how much they liked the image, trendiness, artistic appeal, how delicious they thought the featured product might be, and how likely they were to purchase the product (Table 2). We included the ratings of trendiness and artistic appeal because previous commentaries have noted that social media food ads are subtle and blend into the social media environment [10,26]. See Figures 3 and 4 for examples of survey questions and images from Part 2. Finally, participants responded to demographic questions. We included an attention check question (ie, "Type Facebook in the box") to ensure participants were carefully reading directions and questions. Those who did not type "Facebook" were excluded from analysis.

Table 2. Instagram and traditional advertisement ratings and differences by study condition, means, and SEs.

Participant advertisement ratings	Average rating of Instagram advertisement (SE)[a]	Average rating of traditional advertisement (SE)	Difference in ratings (SE)	*P* values[b]
Unlabeled advertisement condition, rating out of 0-100[c] (SE)				
How much do you like this image?	68.56 (0.93)	65.72 (0.93)	2.85 (0.49)	.001[d]
How artistic is this image?	68.83 (0.93)	66.86 (0.93)	1.97 (0.47)	.001[d]
How trendy is this image?	69.60 (0.90)	66.28 (0.90)	1.14 (0.54)	.001[d]
How delicious do you think this product is?	66.80 (0.94)	65.66 (0.94)	1.14 (0.54)	.07
How likely are you to purchase this product in the next 4 weeks?	56.25 (1.23)	55.27 (1.23)	0.97 (0.58)	.10
Labeled advertisement condition, rating out of 0-100[c] (SE)				
How much do you like this image?	67.43 (1.04)	67.11 (1.04)	0.32 (0.50)	>.99
How artistic is this image?	66.80 (1.07)	66.84 (1.07)	0.04 (0.50)	>.99
How trendy is this image?	68.16 (1.02)	66.92 (1.02)	1.24 (0.48)	.05[d]
How delicious do you think this product is?	66.62 (1.02)	67.51 (1.02)	0.88 (0.53)	.40
How likely are you to purchase this product in the next 4 weeks?	56.03 (1.38)	56.30 (1.38)	0.27 (0.60)	>.99

[a]SE: standard error.

[b]All *P* values are corrected for multiple comparisons using the Bonferroni–Holm procedure.

[c]Ratings were based on a scale from 0 (not at all) to 100 (very much).

[d]Statistical significance ($P<.05$).

Figure 3. Sample images from "Labeled Ad Condition." The traditional ad (top) and the Instagram ad (bottom) both feature the Instagram panel.

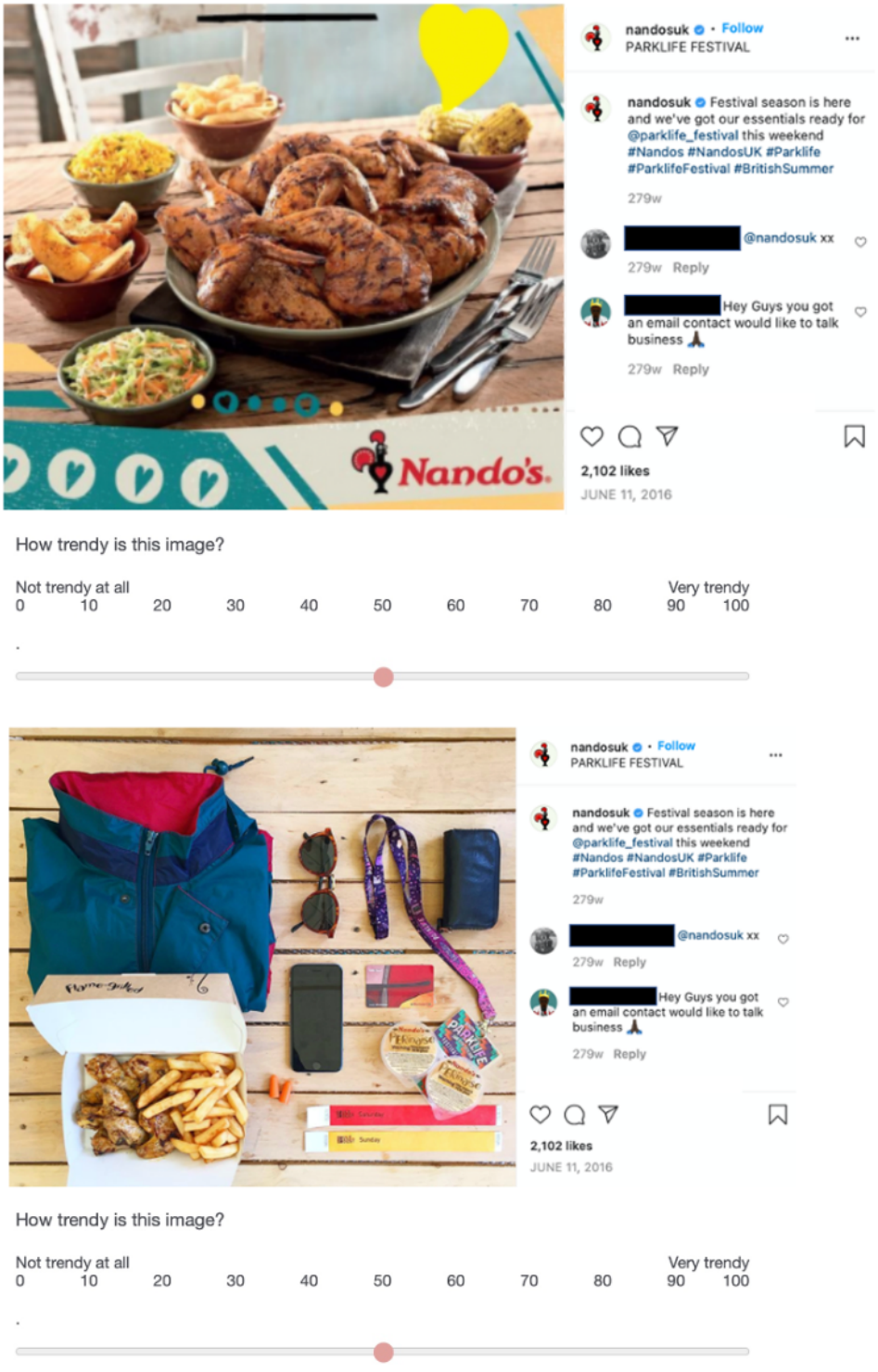
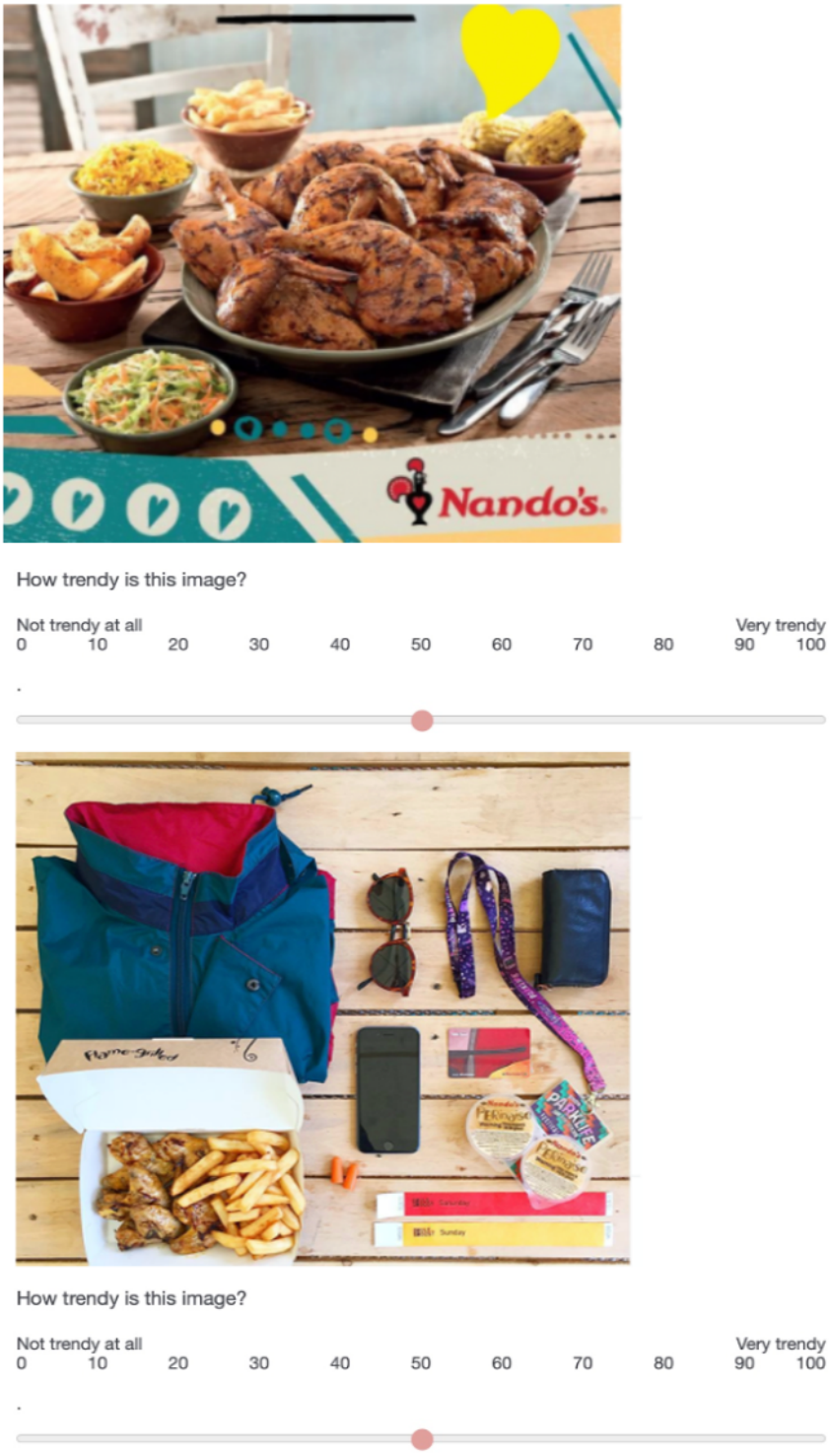


Figure 4. Sample images from "Unlabeled Ad Condition." The traditional ad (top) and the Instagram ad (bottom) do not feature the Instagram panel.



Statistical Analyses

Analyses were conducted in SAS 9.4 (SAS Institute). We used a score test to assess whether the proportion of correct responses was significantly different from 0.5, which is the percentage of correct responses we would expect if adolescents were guessing at random. We also used linear regression to compare adolescents' ad preferences between the unlabeled ad condition and the labeled ad condition. Ad preferences were scored from 0 to 100. Because each adolescent rated multiple ads, the model included a random effect for the participant to account for the repeated measures. Given adolescents spend approximately 7 hours per day using screens [27], it is critical to understand the extent to which heavy use of social media is associated with higher preferences for social media food ads compared with traditional food ads. To examine differences based on time spent on social media, we analyzed data according to time spent on social media. Specifically, analyses that adjusted for time spent

on social media used a median split at 3 hours, such that adolescents who reported spending more than 3 hours per day on social media were labeled "heavy social media users" and those who reported spending less than 3 hours per day on social media were labeled "light social media users." Analyses were then stratified by whether the ad had the "Instagram panel." We conducted chi-square tests and *t* tests to determine whether the randomization was successful and to verify that demographic characteristics did not differ between conditions. Because all tests were not significant, demographic characteristics were not included in the models. The Bonferroni–Holm procedure was used to correct for multiple comparisons.

Results

Part 1: Adolescents Were Asked to Identify Which Advertisements Originated from Instagram

Adolescents correctly identified the Instagram ad 39.1% of the time (z=17.293, P<.001). They correctly selected the Instagram ad significantly more often for international brands (42.8% of the time) than US brands (35.3% of the time; z=36.55, P<.001). We did not find a significant association between the likelihood of adolescents following the US brands on social media and their ability to accurately identify the Instagram ads for unhealthy food and beverages (z=.0022, P=.96), suggesting that the association was not confounded by brand familiarity.

In an analysis stratified by time spent using social media, we found that both light and heavy social media users performed worse than chance (z=15.79, P<.001 and z=8.4387, P<.001). However, heavy social media users correctly identified the Instagram ad more often (42.2%) than light users (36.3%; z=22.69, P<.001).

Part 2: Adolescents Rated Preferences for Instagram Advertisements and Traditional Advertisements

In the "unlabeled ad condition" where all ads for unhealthy products were presented without any Instagram features (Figure 4), analyses revealed adolescents reported higher preferences for Instagram ads than traditional ads on 3 of 5 measures after correcting for multiple comparisons (Table 2). Specifically, adolescents rated Instagram ads for unhealthy food and beverages as significantly trendier (P=.001) and more artistic (P=.001) than the traditional ads. They also reported liking the Instagram ads more than the traditional ads for unhealthy products (P=.001). There were no significant differences in how delicious they thought the featured products might be (P=.08) or how likely they would be to purchase them (P=.10).

Within the "labeled ad condition," we found significant differences on 1 of the 5 outcomes (Figure 3). Specifically, adolescents were more likely to rate the unhealthy food and beverage ads on Instagram as trendy compared with the traditional ads (P=.05), but they rated the ads similarly on the other 4 dimensions.

Heavy social media users responded more positively to social media ads for unhealthy food and beverages than light users across all 5 outcomes in both the "labeled ad condition" and the "unlabeled ad condition." In the " labeled ad condition," heavy social media users rated ads as significantly more artistic (9.46 points; standard error [SE] 2.02, P<.001), trendier (9.30 points; SE 1.93, P<.001), and delicious (7.69 points; SE 1.94, P<.001); they also reported liking ads more (9.37 points; SE 1.98, P<.001) and being more likely to purchase the featured products (15.84 points; SE 2.57, P<.001). In the "unlabeled ad condition," heavy social media users rated unhealthy food and beverage ads as more artistic (8.01 points; SE 1.76, P<.001), trendy (8.20 points; SE 1.70, P<.001), and delicious (7.92 points; SE 1.77, P<.001). They also reported liking them more (7.73 points; SE 1.76, P<.001) and being more likely to purchase the products pictured (13.68 points; SE 2.31, P<.001).

Discussion

Principal Findings

Our findings demonstrate that Instagram food ads were highly appealing to adolescents relative to traditional food ads. This is concerning given adolescents spend 7 hours online each day [27], and companies now spend US $41.5 billion on social media marketing each year [28]. Coca-Cola alone spends US $800 million—20% of their US $4 billion budget—on social media marketing [3], suggesting that companies may be increasingly able to place more ads in front of adolescents in digital spaces. Adolescents incorrectly identified traditional ads as Instagram ads, consistently, suggesting that the artistic quality of Instagram ads may not be perceived as marketing. The subtlety of these Instagram ads may create a public health challenge because adolescents might be more vulnerable to the persuasive influence of these visually artistic and entertaining ads. Additionally, when rating unlabeled, unhealthy food and beverage ads in Part 2, adolescents reported higher preferences for Instagram ads compared with traditional ads on 3 of 5 outcomes. This finding suggests that there is something uniquely appealing about the visual appearance of Instagram ads compared with traditional ads that piques adolescents' interests. But, when we photoshopped the Instagram frame onto those same ads, adolescents preferred the Instagram ads on just 1 of 5 outcomes. This suggests that the "Instagram frame" exerts a powerful influence on adolescents' perceptions by equalizing the appeal of traditional and Instagram ads for unhealthy products. In 2019, 2 bipartisan US Senators proposed expanding the Children's Online Privacy and Protection Act to reduce companies' ability to target youth in online advertising—their proposal would also extend the protected age range to include adolescents aged 13-16 years [29]. Providing policy protections for adolescents in digital spaces is critical given Instagram food ads are highly appealing to adolescents compared with traditional ads.

Contrary to our hypothesis that adolescents would be skilled at identifying which of 2 ads originated from Instagram in Part 1, adolescents performed significantly worse than chance by incorrectly choosing the traditional ad. One possibility is that adolescents could not discern between ads that originate on Instagram and traditional outlets—but adolescents' responses were worse than 50-50 chance, meaning they *consistently* thought the traditional ad was an Instagram ad. One explanation for the recurring misattribution is that companies mimic social media trends that are highly appealing to adolescents (eg, photo

filters or aerial photos of foods arranged in artistic ways), and such mimicry is not readily perceived as marketing [30-32]. These findings support that ads promoted on Instagram are highly appealing to adolescents compared with ads presented in a more traditional form.

The findings that "heavy users" reported higher ad preferences across all outcomes—and were more likely to "guess correctly" in Part 1 relative to "light users"—may be concerning because of the potential additive effect of social media: when users "like" posts, Instagram's algorithms place similar posts in their feed, thereby increasing ad exposure [33]. It is possible, however, that "heavy users" merely have more Instagram fluency, which may have accounted for their ability to correctly identify the Instagram ads. And it is possible that their heavy use of media makes them more inclined to like any type of image that stems from Instagram or other outlets.

This study contributes to the literature on food marketing in several ways. It is the first study, to our knowledge, to compare unhealthy food ads on social media with traditional ads. The power of the Instagram "halo" is concerning given adolescents' exposure to food ads on social media—one study found that 72% (n=101) of adolescents were exposed to social media food ads during a 5-minute data collection period [34], and another survey of 1564 adolescents in the United States found that 75% of adolescents followed, "liked," or shared food and beverage brands on social media [35]. This "halo" effect has also been shown to impact children and adolescents' eating behaviors. In a laboratory study, children who viewed influencers holding unhealthy snacks ate significantly more food than children who saw influencers holding nonfood products [36]. Another laboratory study randomized 132 adolescents aged 13-16 years to view a social media influencer who promoted unhealthy food, vegetables, or a nonfood item and found that adolescents exposed to the healthy food posts did not consume more vegetables [37]. Finally, one study of 72 adolescents (mean age 13 years) found that adolescents who viewed unhealthy food brand posts were more willing to share the post and report a positive attitude toward the product compared with adolescents who viewed healthy foods [38]. But one limitation of existing social media food ad studies is the absence of nonsocial media ads. Our study, therefore, builds upon previous research by demonstrating that the mere presence of the Instagram logo, "likes," and comments causes adolescents to rate ads more favorably than adolescents who see the same ads without those Instagram features.

Limitations

This study has several limitations. We did not ask participants about their rationale for choosing the traditional ad or the Instagram ad in Part 1, and additional research is needed to determine whether differences between heavy and light users' ratings translate into increased susceptibility to advertising. We did not collect data on self-reported height and weight, and it is possible that BMI could moderate the observed effects [39]. But studies suggest that BMI does not always predict higher preferences for ads [40,41]. Our study may not generalize to all adolescents given this was not a nationally representative sample. Despite these limitations, our study has several strengths. The randomized design and the within-subjects comparisons provide clear evidence of preferences for unhealthy food and beverage ads on Instagram relative to traditional ads. Further, "likes" are an ecologically valid signal of preferences, as most adolescents use social media regularly and are familiar with "liking" social media content [42]. Future research should examine how social media food ads affect food purchases.

Conclusion

This study provides the first evidence regarding adolescents' preferences for unhealthy food and beverage ads on social media relative to traditional food ads, sheds light on potential mechanisms that influence adolescents' behavior on social media, and can inform the extent to which self-regulatory food marketing pledges and food marketing policies should expand to include social media–based food marketing.

Acknowledgments

We thank the following NYU SeedProgram Research Assistants and staff who have no conflicts of interest to report: Joshua Arshonsky, Krystle Tsai, Ruchi Desai, Andrea Sharkey, Nasira Spells, Dana McIntyre, Robert Suss, Sana Husain, Rachael Biscocho, Ana Carmargo, Erica Finfer, Ingrid Wells, Chelsea Mangold, Shirley Valerio, Michelle Rosa, and Amaal Alruwaily. The study was funded by the National Institutes of Health (NIH) Early Independence Award (DP5OD021373-05; principal investigator MB) from the NIH Office of the Director and 1R01CA248441-01A1.

Conflicts of Interest

None declared.

References

1. Institute of Medicine, Board on Children, Youth, and Families, Food and Nutrition Board, Committee on Food Marketing and the Diets of Children and Youth. Food Marketing to Children and Youth: Threat or Opportunity?. Washington, D.C: National Academies Press; 2006.
2. Taillie L, Busey E, Stoltze F, Dillman Carpentier FR. Governmental policies to reduce unhealthy food marketing to children. Nutr Rev 2019 Nov 01;77(11):787-816 [FREE Full text] [doi: 10.1093/nutrit/nuz021] [Medline: 31329232]
3. Ignatius A. Shaking Things up at Coca-Cola. Harvard Business Review. 2011 Oct. URL: https://hbr.org/2011/10/shaking-things-up-at-coca-cola [accessed 2021-09-20]

4. Rummo PE, Cassidy O, Wells I, Coffino JA, Bragg MA. Examining the Relationship between Youth-Targeted Food Marketing Expenditures and the Demographics of Social Media Followers. Int J Environ Res Public Health 2020 Mar 03;17(5):1631 [FREE Full text] [doi: 10.3390/ijerph17051631] [Medline: 32138342]
5. Cheyne A, Phil C, Dorfman L, Gonzalez P, Mejia P. Food and Beverage Marketing to Adolescents: An Environment at Odds with Good Health. Healthy Eating Research. 2011. URL: http://healthyeatingresearch.org/wp-content/uploads/2013/12/HER_RS_FoodMarketing_FINAL_4-6-11.pdf [accessed 2021-09-20]
6. Freeman B, Kelly B, Baur L, Chapman K, Chapman S, Gill T, et al. Digital junk: food and beverage marketing on Facebook. Am J Public Health 2014 Dec;104(12):e56-e64. [doi: 10.2105/AJPH.2014.302167] [Medline: 25322294]
7. Laestadius LI, Wahl MM. Mobilizing social media users to become advertisers: Corporate hashtag campaigns as a public health concern. Digit Health 2017;3:2055207617710802 [FREE Full text] [doi: 10.1177/2055207617710802] [Medline: 29942600]
8. Dunlop S, Freeman B, Jones SC. Marketing to Youth in the Digital Age: The Promotion of Unhealthy Products and Health Promoting Behaviours on Social Media. MaC 2016 Jun 16;4(3):35-49 [FREE Full text] [doi: 10.17645/mac.v4i3.522]
9. Freeman B, Kelly B, Vandevijvere S, Baur L. Young adults: beloved by food and drink marketers and forgotten by public health? Health Promot Int 2016 Dec 13;31(4):954-961. [doi: 10.1093/heapro/dav081] [Medline: 26276799]
10. Montgomery KC, Chester J. Interactive food and beverage marketing: targeting adolescents in the digital age. J Adolesc Health 2009 Sep;45(3 Suppl):S18-S29. [doi: 10.1016/j.jadohealth.2009.04.006] [Medline: 19699433]
11. Lapinski MK, Rimal RN. An Explication of Social Norms. Commun Theory 2005 May;15(2):127-147. [doi: 10.1111/j.1468-2885.2005.tb00329.x]
12. Schultz PW, Nolan JM, Cialdini RB, Goldstein NJ, Griskevicius V. The constructive, destructive, and reconstructive power of social norms. Psychol Sci 2007 May 25;18(5):429-434. [doi: 10.1111/j.1467-9280.2007.01917.x] [Medline: 17576283]
13. Cialdini RB, Kallgren CA, Reno RR. A focus theory of normative conduct: a theoretical refinement and reevaluation of the role of norms in human behavior. In: Zanna MP, editor. Advances in Experimental Social Psychology. Waltham, MA: Academic Press; 1991:202-234.
14. Lakon CM, Hipp JR, Wang C, Butts CT, Jose R. Simulating Dynamic Network Models and Adolescent Smoking: The Impact of Varying Peer Influence and Peer Selection. Am J Public Health 2015 Dec;105(12):2438-2448. [doi: 10.2105/AJPH.2015.302789] [Medline: 26469641]
15. Lakon CM, Valente TW. Social integration in friendship networks: the synergy of network structure and peer influence in relation to cigarette smoking among high risk adolescents. Soc Sci Med 2012 May;74(9):1407-1417 [FREE Full text] [doi: 10.1016/j.socscimed.2012.01.011] [Medline: 22436575]
16. Brechwald W, Prinstein M. Beyond Homophily: A Decade of Advances in Understanding Peer Influence Processes. J Res Adolesc 2011 Mar 01;21(1):166-179 [FREE Full text] [doi: 10.1111/j.1532-7795.2010.00721.x] [Medline: 23730122]
17. Sherman LE, Greenfield PM, Hernandez LM, Dapretto M. Peer Influence Via Instagram: Effects on Brain and Behavior in Adolescence and Young Adulthood. Child Dev 2018 Jan 14;89(1):37-47 [FREE Full text] [doi: 10.1111/cdev.12838] [Medline: 28612930]
18. Casey B, Jones R, Hare T. The adolescent brain. Ann N Y Acad Sci 2008 Mar;1124:111-126 [FREE Full text] [doi: 10.1196/annals.1440.010] [Medline: 18400927]
19. Anderson M, Jiang J. Teens, Social Media & Technology 2018. Washington, DC: Pew Internet & American Life Project; 2018 May 31. URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ [accessed 2021-09-20]
20. Harris JL, Frazier WIII, Kumanyika S, Ramirez A. Rudd Center for Food Policy and Obesity. Choice Reviews Online 2011 Oct 01;49(02):1-167 [FREE Full text] [doi: 10.5860/choice.49-0620]
21. Harris JL. Targeted Food Marketing to Black and Hispanic Consumers: The Tobacco Playbook. Am J Public Health 2020 Mar;110(3):271-272. [doi: 10.2105/AJPH.2019.305518] [Medline: 32023097]
22. Frequency of witnessing people of same race and ethnicity portrayed in promotional materials according to adults in the United States as of as of June 2020, by ethnicity. Statista. 2020. URL: https://www.statista.com/statistics/1140059/perception-of-ethnic-representation-in-ads-usa-ethnicity/ [accessed 2021-08-05]
23. Harris JL, Schwartz MB, Shehan C, Hyary M, Appel J, Haraghey K. Snack FACTS 2015valuating Snack Food Nutrition and Marketing to Youth. Hartford, CT: UConn Rudd Center for Food Policy & Obesity; 2015. URL: http://www.uconnruddcenter.org/files/Pdfs/SnackFACTS_2015_Fulldraft03.pdf [accessed 2021-09-20]
24. Harris JL, Haraghey KS, Lodolce M, Semenza NL. Teaching children about good health? Halo effects in child-directed advertisements for unhealthy food. Pediatr Obes 2018 Apr 27;13(4):256-264. [doi: 10.1111/ijpo.12257] [Medline: 29076259]
25. AdScope Intelligence. Kantar Media. URL: http://products.adscope.com/TNSMI/CreativeServices/ClientUI/SecurityTS/Default.aspx?ReturnUrl=http%3a%2f%2fproducts.adscope.com%2fTNSMI%2fCreativeServices%2fClientUI%2fHomePage.aspx#/ [accessed 2017-06-23]
26. Montgomery KC, Chester J, Grier SA, Dorfman L. The new threat of digital marketing. Pediatr Clin North Am 2012 Jun;59(3):659-75, viii. [doi: 10.1016/j.pcl.2012.03.022] [Medline: 22643172]
27. Rideout V, Bobb M. The Common Sense Census: Media Use by Tweens and Teens, 2019. 2019. URL: https://www.commonsensemedia.org/research/the-common-sense-census-media-use-by-tweens-and-teens-2019 [accessed 2021-09-20]

28. IAB releases Internet Advertising Revenue Report for 2020. 2021 Jul 21. URL: https://www.iab.com/news/iab-internet-advertising-revenue/ [accessed 2021-08-05]
29. Senators Markey and Hawley Introduce Bipartisan Legislation to Update Children's Online Privacy Rules. 2019 Mar 12. URL: https://www.markey.senate.gov/news/press-releases/senators-markey-and-hawley-introduce-bipartisan-legislation-to-update-childrens-online-privacy-rules [accessed 2019-08-14]
30. Shearer C. Filters and photo manipulation on social media sites are creating a generation of deluded adolescents. The Independent. 2016 Feb 4. URL: https://www.independent.co.uk/student/istudents/filters-and-photo-manipulation-social-media-sites-are-creating-generation-deluded-adolescents-a6852736.html [accessed 2021-09-20]
31. Brucculieri J. Teenage girls in this photo series show The scary effects of editing apps. HuffPost. 2019 Mar 5. URL: https://www.huffpost.com/entry/rankin-selfie-harm-photo-series_l_5c7d35e0e4b0e5e313cd4a70 [accessed 2021-03-10]
32. Cavanagh E. "Snapchat dysmorphia" is leading teens to get plastic surgery based on unrealistic filters. Here's how parents can help. Insider. 2020 Jan 8. URL: https://www.insider.com/snapchat-dysmorphia-low-self-esteem-teenagers-2020-1 [accessed 2021-03-10]
33. Wagner K. Here's how Instagram's feed algorithm actually works. Vox. 2018 Jun 2. URL: https://www.vox.com/2018/6/2/17418476/instagram-feed-posts-algorithm-explained-reverse-chronological [accessed 2020-04-03]
34. Potvin Kent M, Pauzé E, Roy E, de Billy N, Czoli C. Children and adolescents' exposure to food and beverage marketing in social media apps. Pediatr Obes 2019 Jun 28;14(6):e12508 [FREE Full text] [doi: 10.1111/ijpo.12508] [Medline: 30690924]
35. Fleming-Milici F, Harris JL. Adolescents' engagement with unhealthy food and beverage brands on social media. Appetite 2020 Mar 01;146:104501. [doi: 10.1016/j.appet.2019.104501] [Medline: 31669579]
36. Coates AE, Hardman CA, Halford JCG, Christiansen P, Boyland EJ. Social Media Influencer Marketing and Children's Food Intake: A Randomized Trial. Pediatrics 2019 Apr;143(4):143 [FREE Full text] [doi: 10.1542/peds.2018-2554] [Medline: 30833297]
37. Folkvord F, de Bruijne M. The Effect of the Promotion of Vegetables by a Social Influencer on Adolescents' Subsequent Vegetable Intake: A Pilot Study. Int J Environ Res Public Health 2020 Mar 26;17(7):2243 [FREE Full text] [doi: 10.3390/ijerph17072243] [Medline: 32225032]
38. Murphy G, Corcoran C, Tatlow-Golden M, Boyland E, Rooney B. See, Like, Share, Remember: Adolescents' Responses to Unhealthy-, Healthy- and Non-Food Advertising in Social Media. Int J Environ Res Public Health 2020 Mar 25;17(7):2181 [FREE Full text] [doi: 10.3390/ijerph17072181] [Medline: 32218252]
39. Halford J, Boyland E, Hughes G, Stacey L, McKean S, Dovey T. Beyond-brand effect of television food advertisements on food choice in children: the effects of weight status. Public Health Nutr 2008 Sep;11(9):897-904. [doi: 10.1017/S1368980007001231] [Medline: 18005487]
40. Andreyeva T, Kelly I, Harris J. Exposure to food advertising on television: associations with children's fast food and soft drink consumption and obesity. Econ Hum Biol 2011;9:233. [doi: 10.3386/w16858]
41. Yokum S, Gearhardt AN, Harris JL, Brownell KD, Stice E. Individual differences in striatum activity to food commercials predict weight gain in adolescents. Obesity (Silver Spring) 2014 Dec 25;22(12):2544-2551 [FREE Full text] [doi: 10.1002/oby.20882] [Medline: 25155745]
42. Sherman LE, Payton AA, Hernandez LM, Greenfield PM, Dapretto M. The Power of the Like in Adolescence: Effects of Peer Influence on Neural and Behavioral Responses to Social Media. Psychol Sci 2016 Jul 31;27(7):1027-1035 [FREE Full text] [doi: 10.1177/0956797616645673] [Medline: 27247125]

Abbreviations

ads: advertisements
SE: standard error

Edited by R Kukafka; submitted 10.03.21; peer-reviewed by L Laestadius, JP Allem; comments to author 15.06.21; revised version received 16.08.21; accepted 12.09.21; published 22.10.21.

Please cite as:
Bragg M, Lutfeali S, Greene T, Osterman J, Dalton M
How Food Marketing on Instagram Shapes Adolescents' Food Preferences: Online Randomized Trial
J Med Internet Res 2021;23(10):e28689
URL: https://www.jmir.org/2021/10/e28689
doi:10.2196/28689
PMID:34677136

Original Paper

National-Level Disparities in Internet Access Among Low-Income and Black and Hispanic Youth: Current Population Survey

M Margaret Dolcini[1], PhD; Jesse A Canchola[2], MS, PStat; Joseph A Catania[1], PhD; Marissa M Song Mayeda[1], MPH; Erin L Dietz[1], MPH; Coral Cotto-Negrón[1], MS, PhD; Vasudha Narayanan[3], MA, MBA, MS

[1]Hallie Ford Center for Healthy Children and Families, College of Public Health and Human Sciences, Oregon State University, Corvallis, OR, United States

[2]StatCon Consulting, Hayward, CA, United States

[3]Acumen LLC, Burlingame, CA, United States

Corresponding Author:
M Margaret Dolcini, PhD
Hallie Ford Center for Healthy Children and Families
College of Public Health and Human Sciences
Oregon State University
2631 SW Campus Way
Corvallis, OR, 97331
United States
Phone: 1 541 737 3829
Email: Peggy.Dolcini@oregonstate.edu

Abstract

Background: Internet access is increasingly critical for adolescents with regard to obtaining health information and resources, participating in web-based health promotion, and communicating with health practitioners. However, past work demonstrates that access is not uniform among youth in the United States, with lower access found among groups with higher health-related needs. Population-level data yield important insights about access and internet use in the United States.

Objective: The aim of this study is to examine internet access and mode of access by social class and race and ethnicity among youth (aged 14-17 years) in the United States.

Methods: Using the Current Population Survey, we examined internet access, cell phone or smartphone access, and modes of connecting to the internet for adolescents in 2015 (unweighted N=6950; expanded weights N=17,103,547) and 2017 (unweighted N=6761; expanded weights N=17,379,728).

Results: Internet access increased from 2015 to 2017, but socioeconomic status (SES) and racial and ethnic disparities remained. In 2017, the greatest disparities were found for youth in low-income households (no home access=23%) and for Black youth (no home access=18%) and Hispanic youth (no home access=14%). Low-income Black and Hispanic youth were the most likely to lack home internet access (no home access, low SES Black youth=29%; low SES Hispanic youth=21%). The mode of access (eg, from home and smartphone) and smartphone-only analyses also revealed disparities.

Conclusions: Without internet access, web-based dissemination of information, health promotion, and health care will not reach a significant segment of youth. Currently, SES and racial and ethnic disparities in access prolong health inequalities. Moreover, the economic impact of COVID-19 on Black, Hispanic, and low-income communities may lead to losses in internet access for youth that will further exacerbate disparities.

(J Med Internet Res 2021;23(10):e27723) doi:10.2196/27723

KEYWORDS

internet access; smartphone use; Black youth; Hispanic youth; low-income youth; disparities; mobile phone

Introduction

Access to the internet is increasingly critical for adolescents with regard to health, employment, and education [1-7]. Furthermore, the COVID-19 pandemic has highlighted the fundamental importance of web-based access when in-person channels of communication are blocked [2,8]. Youth may use the internet to find health-related information, find resources,

or participate in web-based interventions addressing a variety of physical and mental health topics [3,9-11] and for communicating with health providers [12,13]. Given that health-related conditions (eg, obesity, diabetes, and sexually transmitted infections) disproportionately impact low-income and racial and ethnic minority youth [14-16], access may be particularly important for these groups. Those with intermittent, limited, or no access to the internet will be disadvantaged because their ability to benefit from web-based health-related resources and programs, job searches, and educational activities will be lower than among those who have stable internet access [17-21].

Rapid increases in internet access have been attained over the past decade; however, US national household data from the American Community Survey (ACS) and the Current Population Survey (CPS) have demonstrated that we have yet to achieve the goal of universal access to the internet [22]. Furthermore, they also document continued disparities in access for Black and Hispanic people, low-income households, and those in rural areas [23]. Although the CPS and ACS found that internet access is higher in households with members ≤18 years of age, published analyses of national probability household surveys infrequently focus on internet access for adolescents.

Valuable data on internet access among adolescents can be obtained from probability-based national surveys conducted by the US Census, including the CPS. Prior analyses of the CPS (2012 data) showed that youth aged 14-17 years living in low-income households had lower internet access than youth in moderate- and high-income households (household income <US $25,000=63%, US $25,000-$49,999=80%, and ≥US $50,000=95%) [4]. In addition, Black and Hispanic youth had lower in-home access than Asian and White youth (White=89%, Black=73%, Asian=93%, and Hispanic=76%). This research, which focused primarily on Black youth, also revealed disparities within racial and ethnic groups. That is, Black youth living in low-income households had lower access than Black youth living in households with middle or high incomes (household income <US $25,000=52%, US $25,000-$49,999=77%, and ≥US $50,000=93%) [4]. Cell phone access was lowest among both low- and high-income Black youth. Furthermore, Black youth have limited access to the internet in the community (schools=60%, libraries=27%, and community centers=5%), highlighting the importance of home access [4]. Given the broadscale uptick in internet access observed in the population at large, we anticipate that youth access will also have increased in recent years.

Other national studies on youth suggest substantial increases in access over time and a more even distribution of internet access across subgroups of youth [1,24,25]. A recent Pew Research Center report showed nearly universal internet access among youth in the United States (99%), with 95% reporting access to a smartphone (White=94%, Black=94%, and Hispanic=95%). Furthermore, relatively small differences in access by household income were reported (<US $30,000=93%, US $30,000-$74,999=93%, and ≥US $75,000=97%) [1]. However, these point estimates may conceal true differences. When CIs are taken into consideration, the Pew Research Center study shows that access for Black youth may range from 82.1% to 100%, and for Hispanic youth, the estimated range is 85.5%-100% (ie, margin of error by race and ethnicity: White=–7.2 to +7.2 points, Black=–11.9 to +11.9 points, and Hispanic=–9.5 to + 9.5 points). Similarly, CIs for income categories are wide (ie, margin of error by income: <US $30,000=–9.6 to +9.6 points, US $30,000-$74,999=–8.3 to +8.3 points, and ≥US $75,000=–8.1 to +8.1 points), suggesting that access by income may also be more variable. In light of this, it is critical that other sources, such as census data (eg, CPS and population studies with reduced error) are used to provide a sharper picture of access [22].

This research examines changes in access and use of the internet among those aged 14-17 years using the 2015 and 2017 CPS. We focus specifically on 2 racial and ethnic groups that have previously had lower access, Black and Hispanic youth, to determine whether disparities in access for racial and ethnic minorities and low-income youth have decreased over time. In light of recent research on adults, which suggests that the number of smartphone-only internet users in the United States is increasing [26], we also examined smartphone-only use among youth.

Methods

Data Source: The CPS

The CPS periodically collects data on computer and internet use, which provides national-level estimates of these topics. This household probability study conducted by the US Census Bureau is the primary source of labor force statistics [27,28]. Data for this study were obtained from the CPS Computer and Internet Use supplement conducted in July 2015 and November 2017. The supplement is sponsored by the National Telecommunications and Information Administration.

We used the 2 most recent surveys for which data are available to examine internet use inside and outside the home and additional modes (eg, devices) used to access the internet. Given our focus on youth, we limited our examination to household members aged 14-17 years in 2015 (total sample unweighted: N=6950; Black: n=917; Hispanic: n=1412) and 2017 (total sample unweighted: N=6761; Black: n=800; Hispanic: n=1377). This age group represents minors most likely to be in high school and living at home and corresponds to the age range examined in prior analyses of CPS 2012 data [4].

Measures and Analyses

All analyses were conducted for a subsample of the CPS representing youth aged 14-17 years. Data on youth were collected from an adolescent aged ≥15 years or from a proxy. We examined four primary questions (yes or no responses) that were assessed in 2015 and 2017. Three items were identical in both waves: “Do you/Does anyone in this household, including you, use the internet at home?” “Who uses the internet at home?” “Who uses a cellular phone or smartphone?” The item regarding smartphone use was asked slightly differently across the two waves (2015: “Do you/Does anyone in this household use a cellular phone or smartphone?” 2017: “What about a smartphone or a cell phone that connects to the internet? Do you/Does anyone in this household use a smartphone?”). In

addition, we constructed a variable to identify smartphone-only use, which was defined as accessing the internet from smartphone but not accessing the internet from home.

Analyses were conducted for all racial and ethnic groups, followed by analyses for Black and Hispanic youth separately. Cross tabulations are weighted using the final person weight [29]. Data analyses were generated using SAS software, version 9.3 of the SAS System (SAS Institute Inc; SAS and all other SAS Institute Inc product or service names are registered trademarks or trademarks of SAS Institute Inc). Standard survey weights adjusted with expansion weights provide estimates of the size of the population to which data are generalized (expanded weights for total sample: N=17,103,547 [2015] and 17,379,728 [2017]; Black: N=2,626,139 [2015] and 2,560,101 [2017]; Hispanic: N=3,867,713 [2015] and 3,981,899 [2017]). Owing to the population sizes, the CIs around the point estimates are very tight (ie, for proportions, –1% to +1% on either side) and can be interpreted as population parameters. For illustration purposes, we calculated the 95% CI for access by income for Black households with those aged 14-17 years and found that the CIs for all 3 categories of income were less than 1% (eg, <US $25,000=27.37% (95% CI 27.32-27.42), US $25,000-$50,000=27.7% (95% CI 27.65-27.67), and >US $50,000=44.93% (95% CI 44.87-44.99). Finally, we do not conduct statistical hypothesis testing for differences between groups because very large population sizes will invariably produce statistically significant differences [30-32].

Results

Overview

The results examine 2 major elements: internet access and the mode of internet access. To report on internet access, we examined the proportion of youth residing in households that have internet connection and the proportion of youth residing in households in which an adolescent, aged 14-17 years, has a cell phone or smartphone. The mode of internet access is reflected in data on (1) the proportion of adolescents accessing the internet through home internet connections, (2) cell phone or smartphone internet connections, and (3) those who have accessed the internet only through smartphones (ie, smartphone only). We examined internet access and mode of internet access for 2015 and 2017 with a focus on racial and ethnic minority and social class differences for the population as a whole and social class differences within Black and Hispanic populations.

Table 1 presents the percentage of those aged 14-17 years (1) living in a household with internet, (2) who have a cell phone or smartphone, and (3) with modes of accessing the internet (eg, home and smartphone) for 2015 and 2017 by racial and ethnic group and household income for the total population of households with an adolescent resident. Table 2 presents data separately by household income for Black and Hispanic youth, the 2 racial and ethnic groups that evidence the lowest access. In the text, we highlight the disparities in access and use by discussing the proportion of youth who lack access overall and through specific modalities (Figures 1-3 present data on lack of access; parallel data on modalities are not included in figures). Our disparity analysis assumes the ideal goal of universal internet access (ie, 100%).

Table 1. Reported internet use in the United States: youth aged 14-17 years by selected characteristics (Current Population Survey 2015 and 2017)[a].

Characteristic	Individual lives in household with internet use[b], n (%)		Individual uses a cell phone or smartphone, n (%)		Mode of internet access, n (%)			
	Year, 2015	Year, 2017	Year, 2015	Year, 2017	Home		Smartphone	
					Year, 2015	Year, 2017	Year, 2015	Year, 2017
Race and Hispanic origin[c]								
White non-Hispanic	7.9 (86)	8.2 (88)	8.2 (89)	8.1 (87)	7.3 (79)	7.3 (78)	6.5 (70)	6.5 (70)
Black	2.0 (77)	2.1 (82)	2.2 (85)	2.1 (82)	1.7 (67)	1.8 (70)	1.6 (61)	1.5 (59)
Asian	0.7 (85)	0.8 (89)	0.8 (88)	0.8 (88)	0.6 (75)	0.7 (78)	0.7 (68)	0.6 (71)
Hispanic (of any race)	2.9 (74)	3.4 (86)	3.2 (84)	3.4 (85)	2.5 (66)	3.0 (75)	2.3 (59)	2.5 (63)
Household income[c] (US $)								
<25,000	1.9 (63)	2.1 (77)	2.4 (78)	2.2 (79)	1.6 (52)	1.8 (66)	1.5 (49)	1.6 (56)
25,000-49,999	3.0 (78)	3.0 (81)	3.3 (85)	3.1 (82)	2.7 (70)	2.6 (71)	2.4 (62)	2.3 (61)
≥50,000	9.0 (88)	9.9 (91)	9.2 (90)	9.7 (89)	8.3 (81)	8.8 (80)	7.4 (73)	7.8 (72)

[a]Data derived from the CPS 2015 [33] and 2017 [34]. Percentages are rounded to whole numbers. Counts are presented in the millions and rounded to the nearest tenth of a million or nearest one-hundredth of a million if less than 1 million. Thus, 7.9 represents 7.9 million.

[b]At least 1 member of the individual's household reported using the internet from home, even if that individual did not report use themselves.

[c]Data collected from respondents aged 14-17 years. Standard survey weights were adjusted with expansion weights to estimate the size of the population to which the data are generalized. Expanded weights N=17,103,547 (2015) and N=17,379,728 (2017). Unweighted N=6950 (2015) and N=6761 (2017).

Table 2. Reported internet usage in the United States by income: Black and Hispanic youth aged 14-17 years by selected characteristics (Current Population Survey 2015 and 2017)[a].

Race and household income	Individual lives in household with internet use[b], n (%)		Individual uses a cell phone or smart phone, n (%)		Mode of internet access, n (%)			
	Year, 2015	Year, 2017	Year, 2015	Year, 2017	Home		Smartphone	
					Year, 2015	Year, 2017	Year, 2015	Year, 2017
Black[c]								
Household income[c] (US $)								
<25,000	0.4 (61)	0.5 (71)	0.6 (79)	0.5 (78)	0.3 (48)	0.4 (61)	0.3 (47)	0.4 (57)
25,000-49,999	0.5 (78)	0.6 (82)	0.5 (83)	0.6 (79)	0.5 (70)	0.5 (66)	0.4 (66)	0.4 (54)
≥50,000	1.1 (86)	1.0 (88)	1.1 (89)	1.0 (86)	0.9 (76)	0.9 (78)	0.8 (67)	0.7 (65)
Hispanic origin[c]								
Household income[c] (US $)								
≤25,000	0.7 (60)	0.8 (79)	0.9 (77)	0.8 (79)	0.6 (52)	0.7 (68)	0.6 (48)	0.5 (54)
25,000-49,999	0.9 (74)	1.0 (81)	1.0 (81)	1.0 (84)	0.8 (66)	0.9 (72)	0.7 (57)	0.7 (61)
≥50,000	1.2 (86)	1.7 (93)	1.3 (92)	1.6 (89)	1.1 (77)	1.4 (80)	1.0 (70)	1.2 (69)

[a]Data derived from the CPS 2015 [33] and 2017 [34]. Percentages are rounded to whole numbers. Counts are presented in the millions and rounded to the nearest tenth of a million or nearest one-hundredth of a million if less than 1 million. Thus, 2.0 represents 2.0 million.

[b]At least one member of the individual's household reported using the internet from home, even if that individual did not report use themselves.

[c]Data collected from respondents aged 14-17 years. Standard survey weights were adjusted with expansion weights to provide an estimate of the size of the population to which the data are generalized. Black: expanded weights, N=2,626,139 (2015) and N=2,560,101 (2017). Unweighted N=917 (2015) and N=800 (2017). Hispanic: expanded weights N=3,867,713 (2015) and N=3,981,899 (2017). Unweighted N=1412 (2015) and N=1377 (2017).

Disparities in Internet and Cell Phone or Smartphone Access 2015-2017

Youth Access: Population Profile

In 2015, no group had 100% access, and racial, ethnic, and income disparities were evident (Figures 1 and 2). Up to one-fourth of youth lived in households without internet access; Black and Hispanic youth evidenced the greatest disparities. Overall, low-income youth were most likely to have no home internet access. Similarly, more Black and Hispanic youth and youth in middle-income households reported not using cell phones or smartphones. In 2017, there continued to be disparities despite gains in household internet access from 2015 to 2017 (Table 1). Although disparities by income and across racial and ethnic groups decreased in 2017 (Figures 1 and 2), low-income and Black and Hispanic youth continued to lag behind other groups. In some groups, fewer youth reported having cell phones or smartphones in 2017, revealing small increases in disparities overall from 2015 to 2017. Racial and ethnic disparities in cell phone or smartphone use in 2017 remained (Figure 1).

Figure 1. Disparities in access by race and ethnicity 2015-2017. The figure shows the percentage of youth (aged 14-17 years) who do not have access; household internet: internet access in the home; cell phone/smartphone: has cell phone or smartphone.

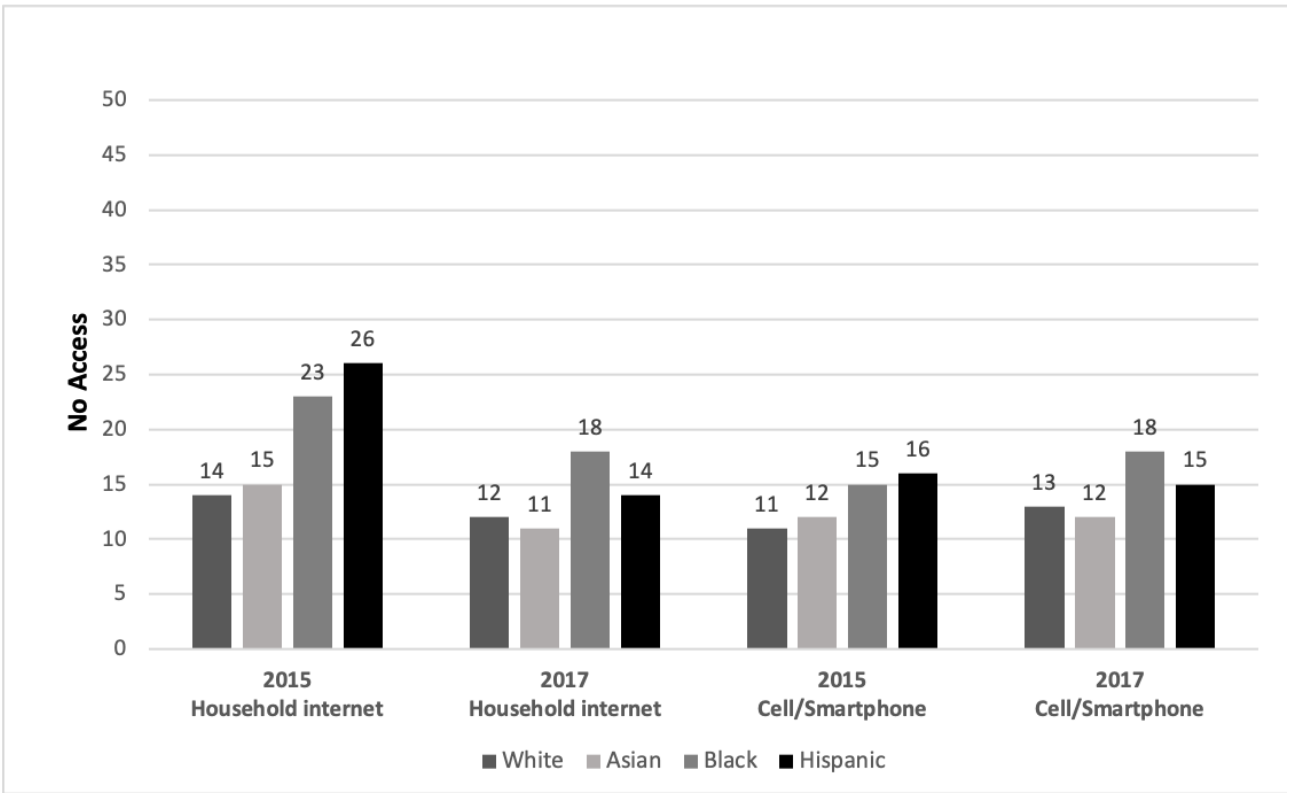
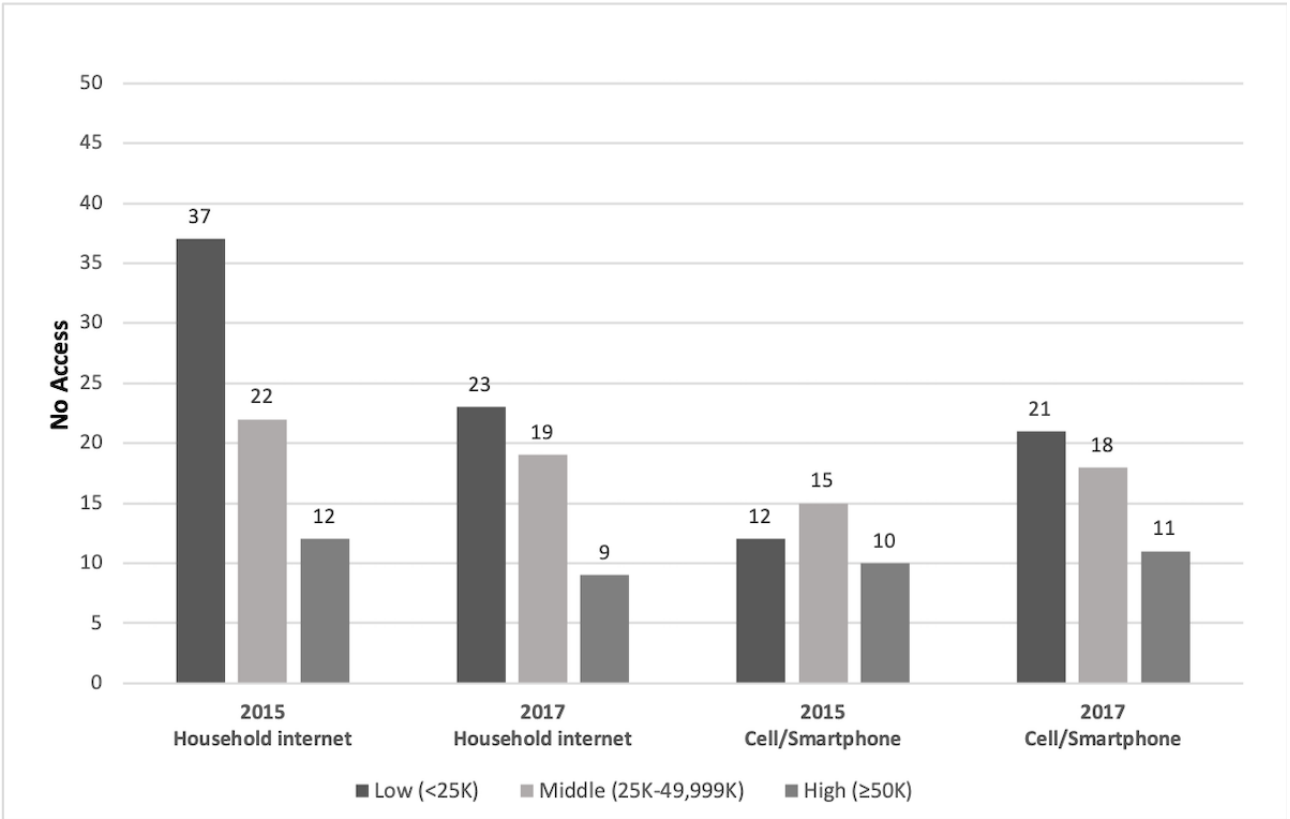


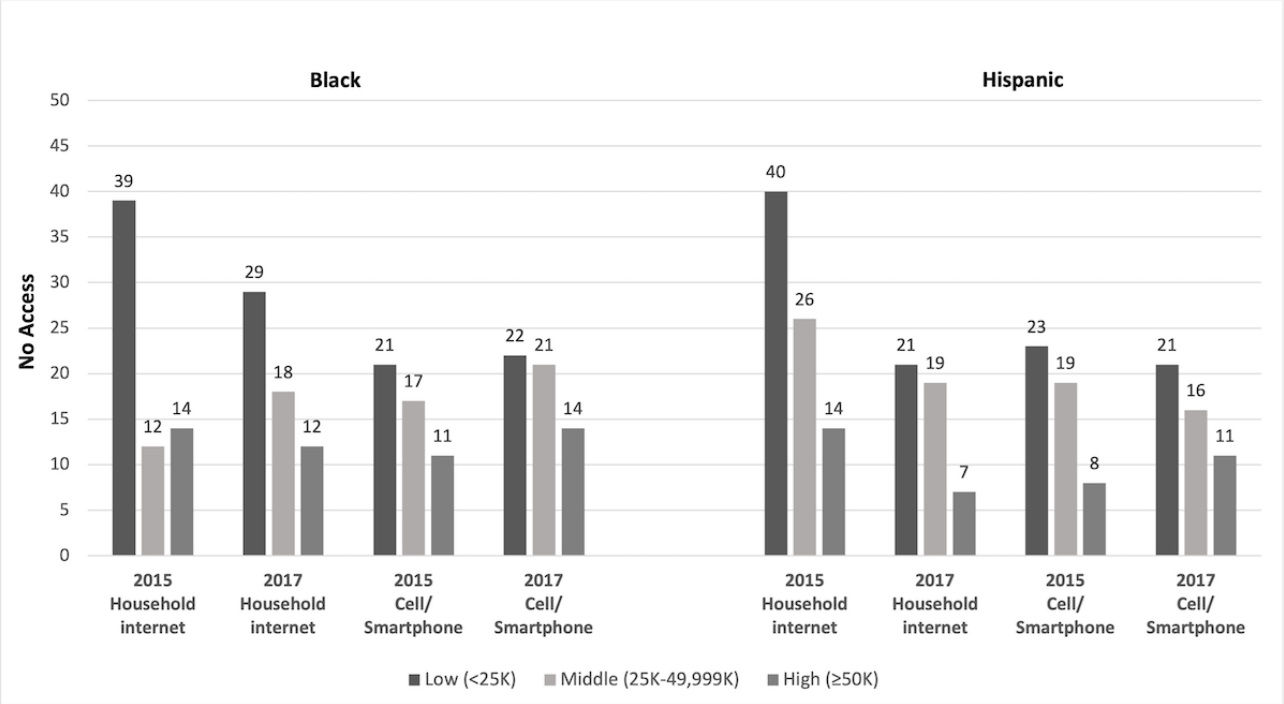
Figure 2. Disparities in access by income 2015-2017. The figure shows the percentage of youth (aged 14-17 years) who do not have access; household internet: internet access in the home; cell phone/smartphone: has cell phone or smartphone.



Youth Access: Disparities Among Black and Hispanic Youth

Socioeconomic status differences were evident among Black and Hispanic youth in 2015 (Figure 3). Nearly 3 times as many low-income compared with high-income Black and Hispanic youth lacked home access to the internet in 2015. Approximately 40% of low-income Black and Hispanic youth did not have home internet access. Cell phone and smartphone use followed a similar pattern, with greater disparities for low-income youth. In 2017, Black and Hispanic youth in the lowest-income households continued to evidence the greatest disparities. A total of 29% and 21% of low-income Black and Hispanic youth, respectively, lacked home access. In contrast, 12% and 7% of high-income Black and Hispanic youth, respectively, lacked access. Cell phone and smartphone use followed similar patterns (Figure 3).

Figure 3. Disparities in access by income among Black and Hispanic youth. The figure shows the percentage of youth (aged 14-17 years) who do not have access; household internet: internet access in the home; cell phone/smartphone: has cell phone or smartphone.



Mode of Internet Access 2015-2017

Youth's Mode of Access: Population Profile

We examined modes of youth use for connecting to the internet, including accessing the internet from home or from a smartphone (Table 1). In 2015, more than 20% of youth did not access the internet from home, and at least 30% did not access the internet from a smartphone; Black and Hispanic youth were less likely than other groups to access from either mode. Overall, youth in low-income households were least likely to access the internet from home or from a smartphone in 2015; gaps between the lowest and highest income households were substantial. In 2017, a large proportion of youth still did not access the internet from home. Some racial and ethnic disparities observed in 2015 were reduced because more Hispanic youth reported accessing the internet from home. Disparities in the use of smartphones to access the internet also remained in 2017. Overall, nearly one-third of the youth did not access the internet from a smartphone (population average=33%). An income gap in accessing the internet from home continued to exist in 2017, although a substantial increase in access among low-income households narrowed the gap. Surprisingly, there were only modest changes in accessing the internet from smartphones between 2015 and 2017, with lower use reported overall. Youth in low-income households continue to be the most likely to not access the internet from a smartphone.

Mode of Access: Disparities Among Black and Hispanic Youth

Table 2 includes the mode of access data for Black and Hispanic youth. In 2015, Black and Hispanic youth in the lowest-income households were substantially more likely than those in households with higher income to lack access to the internet from home. Strikingly, about half of low-income Black and Hispanic youth did not access the internet from home or from a smartphone. Income disparities in access modality decreased from 2015 to 2017, with more low-income Black and Hispanic youth reporting access to the internet from home or smartphones in 2017. Despite this change, the disparity between low- and upper-income brackets continues to be evident in 2017. More specifically, 39% and 32% of low-income Black and Hispanic youth, respectively, did not access the internet from home in 2017. Similarly, 43% and 46% of low-income Black and Hispanic youth, respectively, did not access the internet from a smartphone in 2017.

Mode of Access: Smartphone Only

We examined whether the youth accessed the internet *only* from a smartphone for both time points (eg, smartphone-only users; data not presented). In 2015, nearly 10% of the youth were

smartphone-only users, with Hispanic youth having the largest percentage with smartphone-only access (White=5%, Asian=7%, Black=8%, and Hispanic=9%). In addition, youth in low-income households were the most likely to have only smartphone access (12%); this was also true among Black and Hispanic youth. In 2017, there was a small reduction in reports of smartphone-only access; Hispanic youth were again the most likely to report only smartphone access (White=3%, Asian=3%, Black=3%, and Hispanic=5%). Youth in low-income households continued to be the most likely to have smartphone-only access in 2017 (6%). Among Blacks, low-income youth were more likely to report smartphone-only access in 2017, whereas among Hispanics, youth in middle-income households were the most likely to report smartphone-only use.

Discussion

Principal Findings

Using the CPS, we found that internet access has not reached universal levels, despite the critical role that the internet plays in connecting youth to health information, interventions, and providers. On a positive note, we observed a reduction in disparities from 2015 to 2017, with the largest gains typically found among the racial, ethnic, and income groups that lagged behind. Despite these gains, Hispanic and Black youth continue to have lower levels of household access. Regardless of racial or ethnic group membership, youth in low-income households remain at a disadvantage. Even with recent gains, nearly 30% of low-income Black youth and over 20% of low-income Hispanic youth aged 14-17 years did not have home internet access in 2017. Thus, disparities in access found at earlier time points [4,35,36] have not been eliminated, and these will continue to impact youths' ability to benefit from web-based health-related resources.

With increasing reliance on the internet for information and for a host of health and educational services, youth without home internet access are at a disadvantage [37]. Although smartphones can also be used to access the internet, and indeed many youth use smartphones, there are recognized limits to having *only* smartphone internet access [24,38]. Our finding that in 2017, 3%-5% of youth accessed the internet only from smartphones, suggests that some youth will experience challenges (eg, with connectivity and poor audio and video) [17,24,38]. The disadvantages of smartphone-only access will have a greater impact on Black and Hispanic youth living in low- and middle-income households. Similarly, research by the National Center for Educational Statistics found that 6% of children aged 3-18 years had smartphone-only access to the internet, a situation that was more common among Black, Hispanic, and low-income children [39]. In addition to challenges with connectivity, the demands of video connection (eg, telehealth visit and educational interactions) create disadvantages for those who rely only on smartphones for accessing the internet. Furthermore, even when the internet is available on a computer or smartphone, other limitations exist, such as digital literacy (eg, ability to download and install), limited internet speed that can impede using video, and incompatibility of the recipient's devices with the platform on which the health care is being delivered [8,21,40]. Thus, although this study's findings provide information on access and internet use in the home and via smartphone, there are additional factors that influence whether one can effectively participate in health programs or health care via the internet.

Our findings reinforce the importance of income. At both the population level and within racial and ethnic groups (ie, Black and Hispanic youth), we consistently found disparities in access for youth in low-income households. As the effects of unemployment because of the COVID-19 pandemic continue to unfold, an increasing number of households will make decisions about what is affordable. Without public and private efforts to boost internet access [41,42], more low-income households may lose this important service. Given the scope of the impact of COVID-19 on Black and Hispanic communities [43,44], gains in access between 2015 and 2017, especially among low-income households, are at risk.

Home access may have advantages over community access (eg, school and library) for youth seeking health-related information or assistance on sensitive topics (eg, sexual health and mental health; see the study by Smith-East [17]). For example, with rare exceptions, sexual content is blocked on school computers (see the study by Dolcini et al [4]), and community sites may not offer adequate privacy for health consultations [17]. Regardless of these limitations, community internet access follows similar patterns as home access, such that low-income youth have lower access than high-income youth [4]. Community closures because of COVID-19 will have further reduced access because schools, libraries, and community centers were closed for extended periods in many communities.

Although some prior research is viewed as providing evidence of nearly universal access to the internet for youth in the United States (eg, the study by Anderson and Jiang [1]), this study suggests a different reality. Our findings show continued disparities in internet access that are in alignment with other published work using population-level data (eg, ACS and the CPS) [22,23]. The perceived discrepancies in findings across studies may be explained by large margins of error for data on Black and Hispanic youth in studies that include relatively small numbers of these subpopulations. In fact, our findings with respect to household internet access for Black and Hispanic youth roughly correspond to the lower ranges of access identified by Anderson and Jiang [1] (ie, when the margin of error is taken into account). This study underscores the importance of large population-based surveys, which produce estimates with narrow CIs that avoid camouflaging population disparities. We recommend that surveys provide CIs along with point estimates to allow for cross-survey comparisons.

Strengths and Limitations

This study has several limitations. First, the rates of internet access and use may have changed since 2017. On the basis of past trends, we anticipate that disparities by racial and ethnic groups and income would continue [4,35,36]. In addition, the 2020 recession may reverse prior gains in access. Second, we were unable to examine regional differences because the CPS was not designed to examine patterns at that level; other studies provide evidence of regional disparities in access [22]. Third,

the CPS questions on internet use were updated in 2015 and one item used in this study differed between 2015 and 2017. Wording changes in items can influence responses, especially those related to smartphone use [45]. Wording changes in the item on smartphone use could have contributed to the finding showing lower use in 2017; responses to the 2015 item could have included cell phones without capacity to connect to the internet. Despite item changes, patterns observed in 2012 and in later waves of CPS were similar and broadly aligned with patterns observed in the ACS. Finally, CPS data rely on interviews with adolescents and proxies. The survey literature provides a good case for accepting proxy reports in large surveys [46-50]. Agreement between a proxy and the primary respondent is high when questions are general, observable, and use a yes/no format [49]. Fulton et al [46] demonstrated strong agreement for items in the CPS that assess computer and internet use, and laboratory studies show reliability above 90% for general items [49,50]. As noted earlier, the limitations of this study are balanced by the advantages of using population-level data sets, especially when examining racial and ethnic minority populations.

Conclusions

Despite continued calls for universal access to the internet and the recent proposition that broadband internet access is a social determinant of health [51], this study, along with other sources [52], reveals continued disparities in access and use among youth in the United States. A variety of approaches, including private-public collaborations, increasing community *hot spots*, government sponsored expansion of broadband to rural communities, and public access broadband have potential to increase internet access of youth in the United States [37,53-55]. Along with efforts to make the internet more available to all youths to ensure access to important health resources, continued monitoring of internet access among youth using population-level data is warranted.

Acknowledgments

This research was supported by the National Institutes of Mental Health (MH120512, principal investigator: MMD). We thank the Bureau of the US Census for making the Current Population Survey data available. We thank Marissa Lovell (Master of Public Health) and Ashley Schuyler (Master of Public Health) for assistance in manuscript preparation.

Conflicts of Interest

None declared.

References

1. Anderson M, Jiang J. Teens, social media and technology 2018. Pew Research Center. 2018. URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ [accessed 2020-09-24]
2. Auxier B, Anderson M. As schools close due to the coronavirus, some U.S. students face a digital 'homework gap'. Pew Research Center. 2020. URL: https://www.pewresearch.org/fact-tank/2020/03/16/as-schools-close-due-to-the-coronavirus-some-u-s-students-face-a-digital-homework-gap/ [accessed 2020-09-24]
3. Dolcini MM, Warren J, Towner SL, Catania JA, Harper GW. Information age: do urban african american youth find sexual health information online? Sex Res Social Policy 2015 Jun 01;12(2):110-114 [FREE Full text] [doi: 10.1007/s13178-014-0174-5] [Medline: 26052264]
4. Dolcini MM, Catania JA, Cotto-Negron C, Canchola JA, Warren J, Ashworth C, et al. Challenges in reach with online sexual health information among African American youth: assessing access and engagement. Sex Res Soc Policy 2018 May 8;16(1):12-21. [doi: 10.1007/s13178-018-0332-2]
5. Guilamo-Ramos V, Lee JJ, Kantor LM, Levine DS, Baum S, Johnsen J. Potential for using online and mobile education with parents and adolescents to impact sexual and reproductive health. Prev Sci 2015 Jan;16(1):53-60. [doi: 10.1007/s11121-014-0469-z] [Medline: 24522898]
6. Kang C. Bridging a digital divide that leaves schoolchildren behind. The New York Times. 2016 Feb 22. URL: https://www.nytimes.com/2016/02/23/technology/fcc-internet-access-school.html [accessed 2021-08-10]
7. Park E, Kwon M. Health-related internet use by children and adolescents: systematic review. J Med Internet Res 2018 Apr 03;20(4):e120 [FREE Full text] [doi: 10.2196/jmir.7731] [Medline: 29615385]
8. Katzow MW, Steinway C, Jan S. Telemedicine and health disparities during COVID-19. Pediatrics 2020 Aug;146(2):e20201586 [FREE Full text] [doi: 10.1542/peds.2020-1586] [Medline: 32747592]
9. Dolcini MM, Catania JA, Harper GW, Boyer CB, Richards KA. Sexual health information networks: what are urban African American youth learning? Res Hum Dev 2012;9(1):54-77 [FREE Full text] [doi: 10.1080/15427609.2012.654432] [Medline: 22505842]
10. Van Meter AR, Birnbaum ML, Rizvi A, Kane JM. Online help-seeking prior to diagnosis: can web-based resources reduce the duration of untreated mood disorders in young people? J Affect Disord 2019 Jun 01;252:130-134 [FREE Full text] [doi: 10.1016/j.jad.2019.04.019] [Medline: 30981056]
11. Patten CA, Rock E, Meis TM, Decker PA, Colligan RC, Pingree S, et al. Frequency and type of use of a home-based, internet intervention for adolescent smoking cessation. J Adolesc Health 2007 Nov;41(5):437-443. [doi: 10.1016/j.jadohealth.2007.05.016] [Medline: 17950163]

12. Davis SA, Coyne I, Carpenter DM, Thomas KC, Lee C, Garcia N, et al. Adolescent preferences regarding a web site to empower adolescents to talk with their healthcare providers. J Adolesc Health 2021 Mar;68(3):629-631. [doi: 10.1016/j.jadohealth.2020.06.032] [Medline: 32713739]
13. Miklin DJ, Vangara SS, Delamater AM, Goodman KW. Understanding of and barriers to electronic health record patient portal access in a culturally diverse pediatric population. JMIR Med Inform 2019 Apr 26;7(2):e11570 [FREE Full text] [doi: 10.2196/11570] [Medline: 31066681]
14. Frederick CB, Snellman K, Putnam RD. Increasing socioeconomic disparities in adolescent obesity. Proc Natl Acad Sci U S A 2014 Jan 28;111(4):1338-1342 [FREE Full text] [doi: 10.1073/pnas.1321355110] [Medline: 24474757]
15. Borschuk AP, Everhart RS. Health disparities among youth with type 1 diabetes: a systematic review of the current literature. Fam Syst Health 2015 Sep;33(3):297-313. [doi: 10.1037/fsh0000134] [Medline: 25984737]
16. Harling G, Subramanian S, Bärnighausen T, Kawachi I. Socioeconomic disparities in sexually transmitted infections among young adults in the United States: examining the interaction between income and race/ethnicity. Sex Transm Dis 2013 Jul;40(7):575-581 [FREE Full text] [doi: 10.1097/OLQ.0b013e31829529cf] [Medline: 23965773]
17. Smith-East M, Starks S. COVID-19 and mental health care delivery: a digital divide exists for youth with inadequate access to the internet. J Am Acad Child Adolesc Psychiatry 2021 Jul;60(7):798-800. [doi: 10.1016/j.jaac.2021.04.006] [Medline: 33910040]
18. Chou WS, Prestin A, Lyons C, Wen K. Web 2.0 for health promotion: reviewing the current evidence. Am J Public Health 2013 Jan;103(1):9-18. [doi: 10.2105/AJPH.2012.301071] [Medline: 23153164]
19. The digital divide and economic benefits of broadband access. Council of Economic Advisers Issue Brief. 2016 Mar. URL: https://obamawhitehouse.archives.gov/sites/default/files/page/files/20160308_broadband_cea_issue_brief.pdf [accessed 2021-08-10]
20. Din HN, McDaniels-Davidson C, Nodora J, Madanat H. Profiles of a health information-seeking population and the current digital divide: cross-sectional analysis of the 2015-2016 California health interview survey. J Med Internet Res 2019 May 14;21(5):e11931 [FREE Full text] [doi: 10.2196/11931] [Medline: 31094350]
21. Veinot T, Mitchell H, Ancker J. Good intentions are not enough: how informatics interventions can worsen inequality. J Am Med Inform Assoc 2018 Aug 01;25(8):1080-1088 [FREE Full text] [doi: 10.1093/jamia/ocy052] [Medline: 29788380]
22. Ryan C. Computer and internet use in the United States (ACS-39). United States Census Bureau. 2018. URL: https://www.census.gov/content/dam/Census/library/publications/2018/acs/ACS-39.pdf [accessed 2020-09-25]
23. Ryan C, Lewis J. Computer and internet use in the United States (ACS-37). United States Census Bureau. 2017. URL: https://www.census.gov/content/dam/Census/library/publications/2017/acs/acs-37.pdf [accessed 2020-09-25]
24. Horrigan J, Duggan M. Home broadband 2015. Pew Research Center. 2015. URL: https://www.pewresearch.org/internet/2015/12/21/home-broadband-2015/ [accessed 2020-09-24]
25. Lenhart A. Teens, social media and technology overview 2015. Pew Research Center. 2015. URL: https://www.pewresearch.org/internet/2015/04/09/teens-social-media-technology-2015/ [accessed 2020-09-24]
26. Anderson M. Mobile technology and home broadband 2019. Pew Research Center. 2019. URL: https://www.pewresearch.org/internet/2019/06/13/mobile-technology-and-home-broadband-2019/ [accessed 2020-09-24]
27. Supplemental surveys. United States Census Bureau. 2018. URL: https://www.census.gov/programs-surveys/cps/about/supplemental-surveys.html [accessed 2020-09-25]
28. Labor force statistics from the current population survey. U.S. Bureau of Labor Statistics. 2021. URL: https://www.bls.gov/cps/ [accessed 2021-06-09]
29. Current Population Survey, August 2016: Veterans Supplement Technical Documentation. United States Census Bureau. URL: https://cps.ipums.org/cps/resources/codebooks/cpsaug16.pdf [accessed 2021-09-13]
30. Filho DB, Paranhos R, Rocha EC, Batista M, Silva Jr JA, Santos ML, et al. When is statistical significance not significant? Braz Polit Sci Rev 2013;7(1):31-55. [doi: 10.1590/s1981-38212013000100002]
31. Kunte S, Gore A. The paradox of large samples. Curr Sci 1992 Mar 10;62(5):393-395 [FREE Full text]
32. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol 2016 Apr;31(4):337-350 [FREE Full text] [doi: 10.1007/s10654-016-0149-3] [Medline: 27209009]
33. Current Population Survey (CPS). US Census Bureau. 2016. URL: https://www.census.gov/programs-surveys/cps.html [accessed 2021-09-13]
34. Current Population Survey, November 2017. US Census Bureau. 2018. URL: https://www.census.gov/programs-surveys/cps.html [accessed 2021-09-13]
35. Horrigan J. The numbers behind the broadband 'homework gap'. Pew Research Center. 2015. URL: https://www.pewresearch.org/fact-tank/2015/04/20/the-numbers-behind-the-broadband-homework-gap/ [accessed 2020-09-18]
36. Goldberg R. Digital divide among school-age children narrows, but millions still lack internet connections. National Telecommunications and Information Administration. 2018. URL: https://www.ntia.gov/blog/2018/digital-divide-among-school-age-children-narrows-millions-still-lack-internet-connections [accessed 2020-10-18]

37. Maher K. Remote schooling out of reach for many students in West Virginia without internet. The Wall Street Journal. 2020. URL: https://www.wsj.com/articles/remote-schooling-out-of-reach-for-many-students-in-west-virginia-without-internet-11599989401 [accessed 2020-09-25]
38. Rideout V, Robb M. The common sense census: media use by tweens and teens, 2019. Common Sense Media. 2019. URL: https://www.commonsensemedia.org/research/the-common-sense-census-media-use-by-tweens-and-teens-2019 [accessed 2021-08-10]
39. Children's internet access at home. National Center for Education Stataistics. 2020. URL: https://nces.ed.gov/programs/coe/indicator/cch [accessed 2021-08-10]
40. Kelly SL, Steinberg EA, Suplee A, Upshaw NC, Campbell KR, Thomas JF, et al. Implementing a home-based telehealth group adherence intervention with adolescent transplant recipients. Telemed J E Health 2019 Nov;25(11):1040-1048. [doi: 10.1089/tmj.2018.0164] [Medline: 30681398]
41. Public-private partnerships provide internet connection for students. American Federation of Teachers. 2020. URL: https://www.aft.org/news/public-private-partnerships-provide-internet-connection-students [accessed 2020-09-24]
42. 2019 broadband deployment report FCC 19-44. Federal Communications Commission. 2019 May 29. URL: https://docs.fcc.gov/public/attachments/FCC-19-44A1.pdf [accessed 2020-09-25]
43. Bibbins-Domingo K. This time must be different: disparities during the COVID-19 pandemic. Ann Intern Med 2020 Aug 04;173(3):233-234 [FREE Full text] [doi: 10.7326/M20-2247] [Medline: 32343767]
44. Double jeopardy: COVID-19 and behavioral health disparities for Black and Latino Communities in the U.S. (submitted by OBHE)VID-19 and behavioral health disparities for Black and Latino communities in the U.S. Substance Abuse and Mental Health Services Administration. Substance Abuse and Mental Health Services Administration. 2020. URL: https://www.samhsa.gov/sites/default/files/covid19-behavioral-health-disparities-black-latino-communities.pdf [accessed 2020-09-25]
45. Lewis J, Barth D. 2016 American community survey content test evaluation report: computer and internet use. United States Census Bureau. 2017. URL: https://www.census.gov/content/dam/Census/library/working-papers/2017/acs/2017_Lewis_01.pdf [accessed 2020-09-25]
46. Fulton J, Katz J, Luck J, Henderson A, Davis M. Is a proxy response good enough? Using paired cognitive interviews to assess the accuracy of proxy responses. Research and Methodology Directorate, Center for Behavioral Science Methods Research Report Series. 2020. URL: https://www.census.gov/content/dam/Census/library/working-papers/2020/adrm/rsm2020-06.pdf [accessed 2021-08-10]
47. Davin B, Joutard X, Paraponaris A. "If you were me": proxy respondents' biases in population health surveys. HAL. 2019. URL: https://halshs.archives-ouvertes.fr/halshs-02036434/document [accessed 2021-08-10]
48. Shields M. Proxy reporting in the National Population Health Survey. Health Rep 2000 Oct;12(1):21-39. [Medline: 11565112]
49. Krosnick J, Presser S, Husbands FK, Ruggles S. The future of survey research: challenges and opportunities. The National Science Foundation Advisory Committee for the Social, Behavioral and Economic Sciences Subcommittee on Advancing SBE Survey. 2015. URL: https://www.nsf.gov/sbe/AC_Materials/The_Future_of_Survey_Research.pdf [accessed 2021-08-10]
50. Cobb C. Answering for someone else: proxy reports in survey research. In: Vannette D, Krosnick J, editors. The Palgrave Handbook of Survey Research. Cham: Palgrave Macmillan; Oct 2017:87-93.
51. Benda NC, Veinot TC, Sieck CJ, Ancker JS. Broadband internet access is a social determinant of health!. Am J Public Health 2020 Aug;110(8):1123-1125. [doi: 10.2105/AJPH.2020.305784] [Medline: 32639914]
52. Busby J, Tanberk J. FCC reports broadband unavailable to 21.3 million Americans, BroadbandNow study indicates 42 million do not have access. Broadbandnow Research. 2020. URL: https://broadbandnow.com/research/fcc-underestimates-unserved-by-50-percent [accessed 2020-09-25]
53. Bauer JM, Hampton KN, Fernandez L, Robertson C. Overcoming Michigan's Homework Gap: The role of broadband internet connectivity for student success and career outlooks. SSRN Journal 2020 Oct 19:3714752 (forthcoming). [doi: 10.2139/ssrn.3714752]
54. Stelitano L, Doan S, Woo A, Diliberti M, Kaufman J, Henry D. The digital divide and COVID-19 teachers' perceptions of inequities in students' internet access and participation in remote learning. Rand Corporation. 2020. URL: https://www.rand.org/pubs/research_reports/RRA134-3.html [accessed 2021-08-10]
55. Porter E. A rural-urban broadband divide, but not the one you think of. The New York Times. 2021. URL: https://www.nytimes.com/2021/06/01/business/rural-urban-broadband-biden.html [accessed 2021-06-03]

Abbreviations

ACS: American Community Survey
CPS: Current Population Survey

Edited by R Kukafka; submitted 04.02.21; peer-reviewed by L Marc, H Lam; comments to author 02.05.21; revised version received 18.06.21; accepted 05.07.21; published 12.10.21.

Please cite as:
Dolcini MM, Canchola JA, Catania JA, Song Mayeda MM, Dietz EL, Cotto-Negrón C, Narayanan V
National-Level Disparities in Internet Access Among Low-Income and Black and Hispanic Youth: Current Population Survey
J Med Internet Res 2021;23(10):e27723
URL: https://www.jmir.org/2021/10/e27723
doi:10.2196/27723
PMID:34636728

Original Paper

Sources of Health Information, Technology Access, and Use Among Non–English-Speaking Immigrant Women: Descriptive Correlational Study

Steve Chae[1*], MSN, RN; Yoon-Jae Lee[1*], BSN, RN; Hae-Ra Han[1,2*], PhD, RN, FAAN

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

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

[*]all authors contributed equally

Corresponding Author:
Hae-Ra Han, PhD, RN, FAAN
Johns Hopkins University School of Nursing
525 N Wolfe St, Room 533
Baltimore, MD, 21205
United States
Phone: 1 310 614 2669
Email: hhan3@jhu.edu

Abstract

Background: As the world is becoming increasingly connected by the World Wide Web, the internet is becoming the main source of health information. With the novel COVID-19 pandemic, ubiquitous use of the internet has changed the daily lives of individuals, from working from home to seeking and meeting with health care providers through web-based sites. Such heavy reliance on internet-based technologies raises concerns regarding the accessibility of the internet for minority populations who are likely to already face barriers when seeking health information.

Objective: This study aims to examine the level of technology access and common modes of technology used by Korean American women and to investigate how key psychosocial determinants of health such as age, education, English proficiency, and health literacy are correlated with sources of health information used by Korean American women and by their use of the internet.

Methods: We used data from a subsample of Korean American women (N=157) who participated in a community-based randomized trial designed to test a health literacy–focused cancer screening intervention. In addition to descriptive statistics to summarize Korean American women's internet access and common modes of technology use, we conducted backward stepwise logistic regression analyses to substantiate the association between the psychosocial determinants of health and internet use.

Results: Approximately two-thirds (103/157, 65.6%) of the sample had access to the internet, and nearly all had access to a mobile phone. The internet was the most commonly used channel to obtain health information 63% (99/157), and 70% (110/157) of the sample used text messaging. Nevertheless, only approximately 38.8% (40/103) of the sample were very confident in using the internet, and only 29.9% (47/157) were very confident in using text messaging. Multivariate analyses revealed that older age (>50 years) was associated with 79% lower odds of using the internet to seek health information (adjusted odds ratio [AOR] 0.21, 95% CI 0.10-0.46). The higher health literacy group (19+ on Rapid Estimate of Adult Literacy in Medicine) had 56% lower odds of using the internet to acquire health information (AOR 0.44, 95% CI 1.13-11.18). Higher education (college+) was associated with both internet use (AOR 4.42, 95% CI 1.88-9.21) and text messaging (AOR 3.42, 95% CI 1.55-7.54). Finally, English proficiency was associated with text messaging (AOR 4.20, 95% CI 1.44-12.24).

Conclusions: The differences in modes of technology access, use, and confidence by some of the key psychosocial determinants, as observed in our study sample, have important implications when health care teams develop dissemination plans.

(J Med Internet Res 2021;23(10):e29155) doi:10.2196/29155

KEYWORDS

technology use; internet; text messaging; health literacy; English proficiency; immigrant; health disparities; Korean American; women; mobile phone

Introduction

Background

The internet is a unique technology that is constantly developing and encompasses various forms of technologies, such as social media, e-communications, e-commerce, telecommunications, and telehealth. The internet is currently used as a platform that connects people, stores and shares information, and virtually provides health services [1]. Especially during the COVID-19 pandemic, with the suspension of many face-to-face medical and health care services, telehealth and telecommunication have been widely used to offset the loss of manpower within face-to-face medical services [2]. Furthermore, many forms of internet-based health technologies ease the daily lives of Americans by providing ways to self-manage their illnesses and conditions such as diabetes, physical activity, diet, weight loss, heart failure, and more [3]. These daily health technologies use combinations of already existing technologies that are updated with the current technology. For example, sphygmomanometer, weight scale, electrocardiogram recorder, and smartphone apps are combined to monitor heart failure with cost-saving benefits compared with telephone-based monitoring [4].

Despite the proven benefits and rapidly increasing, ubiquitous use of the internet and internet-based health technologies for all Americans, this rapid incline is exposing the health disparities and digital divide in underserved populations: racial and ethnic minorities, older adults, rural populations, lower-income groups, and populations with a lower educational background [5,6]. These underserved populations are more likely to have lower health literacy levels, lack health care coverage, lack provider availability, experience lower quality of care, and are more likely to face discrimination in general and in health care [7]. They are also less likely to have adequate internet connectivity and are less likely to use the internet to seek health information [5,8].

Asian racial or ethnic minorities are rapidly growing in the United States: 72% growth between 2000 and 2015 compared with 60% in Hispanic racial or ethnic minority populations and are currently projected to be the largest immigrant racial or ethnic group by 2055, accounting for 38% of the immigrant population [9]. Overall, 97% of English-speaking Asian Americans are estimated to use the internet compared with 85% of non-Hispanic Whites, 81% of Hispanics, and 78% of non-Hispanic Blacks [10]. However, the healthier and more affluent Asian American subgroups tend to be overrepresented in these reports. This tends to overcast the underserved Asian American subgroups: non–English-speaking, lower socioeconomic status, elderly, disabled, and unauthorized immigrant subgroups comprise more than 50 ethnic and 100 language groups [11]. Some of these smaller Asian ethnic and language groups live in poverty rates that are three times higher than the US average of 10.5%, such as Hmong (28.3%), Bhutanese (33.3%), and Burmese (35%) [9,12]. In addition, some subgroups, such as Koreans, culturally do not express their hardships or deficiencies to the public because they do not want to show their weakness—financially or socially [9].

According to the 2019 American Community Survey, Korean Americans are the fifth largest Asian population and represent one of the fastest growing ethnic groups in the United States [13]; 59% of Korean Americans are first-generation immigrants, significantly higher than the 14% immigrant share among general Americans [14]. Research has revealed health disparities in the Korean American population. For example, Korean American women are 1.5 times more likely to die from cervical cancer than non-Hispanic White women, yet are nearly 20% less likely to receive triennial Papanicolaou tests [15,16]. Korean Americans are also more susceptible to acquiring hepatitis B virus than White Americans by 10-fold because of significant language barriers, limited health knowledge, financial issues, and poor access to care [17].

With their well-documented benefits, internet-based health technologies provide an opportunity to eliminate the health disparities experienced by traditionally underserved populations with limited resources and limited English proficiency, such as Korean Americans [5]. However, information addressing technology access and the use of the internet among Korean Americans is currently scarce. As health care professionals are rapidly adapting and expanding the use of the internet and technology, a better understanding of how Korean American women use the internet may help researchers and clinicians to implement and disseminate health programs to address health disparities in Korean American populations.

Objective

The purpose of this study was to understand the internet use among Korean American women. Specifically, we examined the level of internet access and the common modes of technology used by the target group of Korean American women. We also investigated how key social determinants of health, such as age, education, English proficiency, and health literacy, are associated with Korean American women's use of internet technology.

Methods

Study Design and Sample

This study used data obtained from a sample of Korean American women who participated in a community-based, cluster randomized controlled trial that was designed to test the efficacy of a community health worker–led health literacy–focused intervention program, including mammograms and Papanicolaou test screening, among Korean American women. Details regarding the study design and outcomes have been published elsewhere (NCT00857636) [18]. Briefly, trained community health workers from 23 ethnic churches (intervention=11 and waitlist control=12) in the Maryland-Washington metropolitan area recruited the study sample. Sample inclusion criteria included Korean American women aged 21-65 years who were overdue for either a mammogram (for women aged ≥40 years) or Papanicolaou tests. Those with an acute or terminal condition (eg, cancer diagnosis or life expectancy of <6 months), psychiatric diagnosis, or other conditions that precluded participation in the study activities were excluded. A total of 560 eligible women were recruited and enrolled in the study (intervention, n=278; control, n=282)

and completed the baseline assessment. Follow-up data were collected at 3 and 6 months from the start of the intervention. At 6-month follow-up, 527 women participated in the final data collection (intervention, n=261; control, n=266). For the women participating only in the intervention, a brief phone survey was conducted 1 year after the final data collection between 2013 and 2017. The main goal of this 1-year postsurvey was to examine the long-term effects of the study intervention on cancer screening behaviors and the intervention of women's dissemination of cancer screening knowledge to neighbors in the community (Multimedia Appendix 1). The postsurvey also included questions about access to and use of the internet. A total of 157 intervention women participated in the 1-year postsurvey (response rate=60%). For this analysis, we combined baseline data for key sociodemographic and 1-year postsurvey data for internet use.

Procedures and Measures

All study procedures were approved by the Johns Hopkins Medicine Institutional Review Board. The baseline questionnaire was administered face-to-face by trained bilingual research staff to assess participant characteristics. The baseline questionnaire included sociodemographic questions such as age in years, educational level, income comfortability, marital status, health insurance, and English proficiency. In addition, health literacy was measured using the Rapid Estimate of Adult Literacy in Medicine (REALM), a well-validated 66-item screening instrument that assesses an adult's ability to read common medical words and lay terms for body parts and illnesses [19]. Each correctly pronounced word was coded as *1*, with a total score ranging from 0 to 66. The REALM has been significantly correlated with other standardized reading tests. The REALM was validated in a Korean sample [20].

The 1-year postsurvey was completed through phone interviews by trained bilingual research assistants. Trained bilingual research assistants called the study participants in the intervention group who agreed to participate in a follow-up survey. Part of the postsurvey asked questions about the main sources of health information (eg, internet, television [TV], radio, newspaper, book, or magazine) and access to the internet (with access location) and mobile phones. For those who reported the internet as the main source of health information, additional questions were asked about the frequency of internet use and types of health information sought. Similarly, for those who reported having access to mobile phones, the frequency of use of text messaging was asked. Finally, the level of confidence in using the internet and text messaging was assessed for those who indicated access to these technologies. Each phone interview lasted, on average, approximately 10 minutes.

Data Analysis

We used descriptive statistics such as means, SDs, frequencies, and percentages to summarize the sample characteristics. We conducted a series of backward stepwise logistic regression analyses to examine the association between the key social determinants of health (age, education, English proficiency, and health literacy), sources of health information, use of internet technology, and confidence in using technologies. Age, education, and English proficiency were categorized as young and old (<50 years vs ≥50), low and high education (less than college vs college or more), and limited and proficient English (none, little, and well vs fluent). More than 80% of the analysis sample scored zero on the REALM; hence, instead of using the suggested cutoff of 60 to represent adequate health literacy (ie, high school reading level), we used a cutoff of 18 to categorize the sample as low versus high health literacy groups. The cutoff of 18 on the REALM indicates a reading level of third grade and below, where people will not be able to read most low-literacy materials and may require repetitive oral instructions or written materials composed of illustrations or audiotapes or videotapes [19]. All statistical significance was set at $P<.05$.

Results

Sample Characteristics

Table 1 summarizes the sample characteristics. The analysis sample was mostly middle-aged (mean age 46.3, SD 8.2 years) and married (133/157, 84.7%) women. More than half (100/157, 63.7%) received college or more education, but only 1 in 5 (33/157, 21%) said they felt comfortable with their income level. Slightly more than one-third (55/157, 35%) of study participants indicated at baseline that they had health insurance and a primary care provider (53/157, 33.7%). The majority of the sample had limited English proficiency; only 12.1% (19/157) reported being fluent in English. Similarly, more than three-fourths (134/157, 85.4%) scored ≤18 on the REALM, indicating a reading level of third grade or lower.

Table 1. Sample characteristics (N=157).

Characteristics	Values
Age (years; range 24-64), mean (SD)	46.3 (8.2)
<50, n (%)	96 (61.1)
≥50, n (%)	61 (38.9)
Married, n (%)	133 (84.7)
Education (years; range 8-22), mean (SD)	14.7 (2.5)
Less than college (12 years), n (%)	57 (36.4)
College and more (13 years), n (%)	100 (63.7)
Income comfortability, n (%)	
Difficult and very difficult	69 (43.9)
Okay	55 (35.0)
Comfortable and very comfortable	33 (21.0)
Have health insurance, n (%)	55 (35.0)
Have primary care provider, n (%)	53 (33.8)
English proficiency, n (%)	
None, little and well	138 (87.9)
Fluent	19 (12.1)
Health literacy (range 0-66), mean (SD)	7.0 (17.0)
Third grade reading level or less, n (%)	134 (85.4)
More than third grade reading level, n (%)	23 (14.6)

Technology Accessibility, Usability, and Confidence

Table 2 shows the technology accessibility, usability, and confidence of Korean American women. The most commonly used method for obtaining health information was the internet (99/157, 63%). Of those who indicated the internet as the main source of health information, the main types of health information they searched for were disease (64/99, 65%), followed by alternative medicine (26/99, 26%) and health supplements (21/99, 21%); the majority (65/99, 66%) indicated that they searched on average 1-2 times or less per month. Approximately two-thirds (103/157, 65.6%) of the study sample indicated that they had access to the internet; most of them (101/103, 98.1%) had internet access at home. In addition, nearly all (153/157, 97.5%) participants reported that they had access to a mobile phone; 70.1% (110/157) used text messaging. Nevertheless, only 39% (40/103) were very confident in using the internet; 30% (47/157) were very confident in using text messaging.

Table 2. Accessibility and use of technology (N=157).

Characteristics	Values, n (%)
Sources of health information[a]	
Internet	99 (63.1)
Newspaper, magazine, and book	58 (36.9)
Television and radio	42 (26.8)
Friend, family, and acquaintances	26 (16.6)
Physician	3 (1.9)
Community	3 (1.9)
Types of health information sought on the internet[a,b]	
Disease	64 (64.6)
Alternative medicine	26 (26.3)
Health supplement	21 (21.2)
Medication	13 (13.1)
Hospital and physician	8 (8.1)
General health	2 (2.0)
Missing	21 (21.2)
Frequency of internet use for health information[b]	
1-2 times or less per month	65 (65.7)
1-2 times per week	11 (11.1)
≥3 times per week	8 (8.1)
Missing	15 (15.1)
Have internet access	103 (65.6)
Location of access[c]	
Home	101 (98.1)
Work	15 (14.6)
Public library	5 (4.9)
School	1 (1)
Have mobile access	153 (97.5)
Use text messaging	110 (70.1)
Confidence with internet use[c]	
Not at all, fairly, and somewhat	63 (61.1)
Very confident	40 (38.9)
Confidence with text messaging	
Not at all, fairly, and somewhat	87 (55.4)
Very confident	47 (29.9)
Missing	23 (14.6)

[a]Multiple choice.

[b]Women who used the internet to obtain health information only (n=99).

[c]Women who had access to the internet only (n=103).

Relationship Among Social Determinants of Health, Sources of Health Information, Technology Use, and Confidence

Multivariate analyses revealed that older age was associated with approximately 79% lower odds of using the internet as the main source of health information (adjusted odds ratio [AOR] 0.213, 95% CI 0.100-0.455) but had at least 2.4 times higher odds of using print media (eg, books and magazines) to obtain health information (AOR 2.403, 95% CI 1.166-4.950). The high health literacy group had 56% lower odds (AOR 0.440, 95% CI 1.134-11.182) of using the internet to acquire health information but more than 1.5 times higher odds (AOR 1.535, 95% CI 1.004-6.400) of using TV and radio. When these variables were examined compared with technology use and confidence, educational level was significantly associated with both internet use (AOR 4.419, 95% CI 1.870-9.205) and text messaging (AOR 3.417, 95% CI 1.549-7.540), with high education favoring the use of both technologies. Although no variables were associated with confidence in using the internet, English proficiency was associated with more than four times higher odds of confidence in using text messaging (AOR 4.198, 95% CI 1.439-12.244; Table 3).

Table 3. Multiple logistic regression analysis to explain sources of health information, technology use, and confidence[a].

Outcome	Social determinants of health, adjusted odds ratio (95% CI)			
	Age	English proficiency	Education	Health literacy
Main source of health information				
Internet	0.213 (0.100-0.455)	N/A[b]	N/A	0.440 (1.134-11.182)
Print	2.403 (1.166-4.950)	N/A	N/A	N/A
Television and radio	N/A	N/A	N/A	1.535 (1.004-6.400)
Use of technology				
Internet	N/A	N/A	4.419 (1.870-9.205)	N/A
Text messaging	N/A	N/A	3.417 (1.549-7.540)	N/A
Confidence in using technology				
Internet	N/A	4.198 (1.439-12.244)	N/A	N/A
Text messaging	N/A	N/A	N/A	N/A

[a]Variables were added if the *P* value was less than .05. The adjusted odds ratio (95% CI) is displayed. The different groups were as follows: younger age (<50 years), limited English proficiency (less than fluent English), low education (less than college), and low health literacy (≤18 on the Rapid Estimate of Adult Literacy in Medicine).

[b]N/A: not applicable.

Discussion

Principal Findings

We found that only approximately two-thirds of our study sample had internet access. However, the internet was the most popular source of health information used by Korean American women in this study. We also found that although nearly all study participants (153/157, 97.4%) had access to a mobile phone, slightly more than two-thirds (110/157, 70%) of them used text messaging. The Korean American women in the study had overall low levels of confidence in using the internet and text messaging. Older women, lower education, and limited English proficiency were associated with less use of these technologies. This is one of the first studies to comprehensively examine the common modes of technology and to investigate how key social determinants of health are associated with technology use in one of the fastest growing yet understudied ethnic groups in the nation (ie, Korean American) [21].

The level of home-based internet access among our study sample of Korean American women was comparable with that in Black populations (66%) but lower than that in the general US population (73%) or White populations (79%) and higher than that in Hispanic (61%) populations [5]. Relying on the internet to seek health information may have been due in part to the finding that Korean Americans represent one of the most uninsured ethnic groups in the nation. For example, in a survey of 498 patients from primary care clinics mainly serving Hispanic and Black patients, Gutierrez et al [21] revealed that the primary source of health information most commonly reported was their health care professionals, followed by various forms of media, such as news, radio, or books. The internet was used less frequently (<25%). In our study sample, only approximately one-third of women had health insurance (35%) and reported having a primary care provider (34%). For individuals with limited access to care, the internet may serve as a key resource for health information. As the internet plays an increasingly important role in health information access, it is important to identify strategies to promote and control the quality of health information posted on the internet. For example, a recent systematic review yielded three criteria that are most important in evaluating web-based health information: trustworthiness, expertise, and objectivity [22]. These indicators can be used to guide those who develop web-based health information.

Age was significantly associated with the sources of health information used by Korean American women in this study. Specifically, older age (≥50 years) was associated with an

approximately 80% lower likelihood of using the internet but more than two times higher likelihood of using print media as the main source of health information. Using data from the 2011-2014 Health Information National Trends Survey (58% female and 70% White with mean age 54 years), Jacobs et al [23] reported that younger US populations were more likely to use web-based sources to seek health information compared with older populations. In this study, the older population with less internet proficiency was more likely to use health care professionals as a source of health information [23]. Taken together, our findings suggest the importance of satisfying health information-seeking behavior with appropriate dissemination approaches based on the characteristics of the target population.

Korean American women with higher education (college or more) were four times more likely to use the internet and three times more likely to use text messaging than women with less than a college-level education. The ability to use the internet and text messaging may be attributable to multiple factors such as generation gaps (eg, younger generations being exposed to digital technologies early on in their life), regularized use, and required higher reading levels for many publicly available websites. In particular, people with higher education are more likely to be easily adapted to use new technologies [24,25]. Korean American women with higher education might have learned to use the internet to seek information during years of education and used text messaging for social interactions and group projects. The national guideline states that health information on public websites should be written at a lower than eighth grade reading level [26,27]. Nevertheless, evidence suggests that the health information found on the web is above a grade 10 level or even 12th [28,29]. Most Korean American women in our study had an average reading level of third grade and lower on REALM, indicating a significant barrier to finding adequate health information on the web. Future efforts should focus on the development of plain language health information that is accessible and understandable to wider audiences, including recent immigrants with limited health literacy, such as Korean Americans.

We found that high health literacy was associated with a lower likelihood of using the internet but a higher likelihood of using TV and radio as the main source of health information. This finding conflicts with a study (N=498, 62% Hispanic) in which patients with limited health literacy had lower odds of seeking health information on the web after controlling for age, sex, and race and ethnicity [21]. It may be that Korean American women might have lacked the necessary skills to use health information on the internet. Overall, our study sample had very low health literacy, with 85% of participants being lower than the third-grade level. In addition, nearly 9 of 10 women (88%) had limited English proficiency, a risk factor for low health literacy [30]. In fact, confidence in using the internet was more than four times higher among women with higher English proficiency in the study. The US Department of Health and Human Services' Healthy People 2020 published a report that states an upward trend of access to the internet in the general population, from 69% in 2007 to 81% in 2017 [31]. However, the proportion of web-based health information seekers who report that they can easily access health information decreased from 41% in 2008 to 38% in 2017 [31]. The findings suggest that having access to the internet does not necessarily mean that health information seekers have the skills and resources to adequately use the information obtained from the internet [32]. To collectively provide a wide range of health information to the general public, health information published on the web not only needs to be accurate and reliable in English and Spanish but also lower-grade English reading levels and diverse languages. As the US is projected to be a *minority-majority* country as early as 2045, it is particularly important to provide lower English reading levels and accurately translate health information [33].

Despite nearly every woman in our sample indicating access to a mobile phone, only 70% of those reported ever using text messaging, and only 30% were very confident. More than 9 of 10 (97%) US adults use text messaging as the most frequently and widely used feature of smartphones [34]. Text messaging is an increasingly popular way of getting connected to the health care system for appointment alerts, information sharing, and methods for quick communication to question and answer between health care providers and patients. As studies show higher efficiency and result through text messaging for health care [35,36], it is important to understand the reasons why Korean American women face difficulties in using text messaging features. Ladley et al [37] found that the use of text messaging to disseminate health information to infant caregivers with low health literacy improved information retention and resulted in fewer visits to the emergency department during the first year. The use of text messaging also benefits health care providers, as it offers advantages such as time management [38]. Active smartphone use may be a gateway for telehealth communications and web-based patient portals such as MyChart, which provide reliable sources of health information. Currently, digital technology health resources are primarily used by the highly educated White population and are often out of reach for those with less than high school education and lower English proficiency [39,40]. These underserved populations are likely to face barriers in accessing health care similar to MyChart because of external factors such as low health literacy and low technology proficiency or lack of access to the internet or devices to access web-based patient portals [40,41]. According to a recent review of technology use among underserved populations, having a proxy person to provide support helped to increase health technology uptake [32]. Future research should explore whether and how the availability of social resources to impart knowledge and skills can serve as a facilitator of digital technology use for health information in underserved populations.

Limitations

Our study limitations include a lack of generalizability of the study findings beyond the targeted sample. Women in the study were recruited from ethnic churches in one geographical area (Maryland-Washington metropolitan area). People in urban areas have more access to the internet than those in rural areas (97% vs 65%) [42]. Nevertheless, most Korean Americans reside in metropolitan areas (eg, Los Angeles, New York, and Washington, DC) [43]. In addition, we did not specify the use of social media as part of the postintervention survey, mainly focusing on some of the most common sources of information

used in our study sample. However, the literature indicates that certain cultural and age groups (eg, Hispanic women aged between 45 and 55 years) use social media frequently, especially when a message is centered on familism [44]. This was a secondary analysis of data that were already collected for its parent study and were pulled only from the intervention group of its parent study sample. For the purposes of the parent study, participants completed a sociodemographic survey about 1.5 years before the internet use survey. However, we do not believe that any significant change might have occurred in the psychosocial determinants of health we examined (ie, age, education, English proficiency, and health literacy) in relation to technology use.

Conclusions

As the use of digital technology is ubiquitous in our daily lives with better health outcomes, it is important to allow all populations to use these technological advances to narrow the gap between health inequality. Although working with ethnic minority populations, particularly recent immigrants, it is important to understand the role of social determinants of health in relation to the types of technologies being used, how they are used, and their level of confidence in using the technologies. Such specific information may help provide effective strategies for optimal health information dissemination to engage underserved populations, such as Korean American women. Future research should address barriers to and facilitators of using digital technologies, health information, and health care. In addition, given that people with limited access to health care, such as Korean Americans, are not likely to obtain health information from a health provider, diverse health information dissemination strategies beyond web-based information (eg, using friends and family as main sources of health information) should be considered and tested for its efficacy [45]. The Plain Writing Act of 2010 stipulates that the government-issued documents to the public be written clearly with the audiences' recommended reading level [46,47]. This act can be expanded to not only federally supported agencies but also all agencies subsidized by federal and state governments and include set guidelines of all documents disseminated be written with a readability score of eighth grade and lower in accordance with the National Institute of Health and American Medical Association's readability recommendations [48,49]. The rapid evolution of technology from word of mouth to telephones, radio, and TV and then to the internet and smartphones are differentiating the methods of obtaining health information and health care seeking behavior. The differences in the modes of obtaining health information in younger and older persons, as observed in our study sample, have important implications when health care teams develop dissemination plans. Having multiple modes such as paper, text message, and web-based along with dissemination with respect to age groups may lead to more successful health information delivery.

Acknowledgments

This study was supported, in part, by a grant from the National Cancer Institute (R01CA129060). Additional funding was received from the National Center for Advancing Translational Sciences (UL1TR003098), the National Institute of Nursing Research (P30NR018093), and the National Institute on Aging (R01AG062649). The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and the decision to submit the manuscript for publication. The authors are grateful to all Korean American women and community health workers who participated in this study and to the research staff for their assistance with data collection.

Authors' Contributions

All authors approved the final version of the manuscript. HRH originated the study and led to writing. SC and YJL contributed to the development of the study concept and design. SC, YJL, and HRH contributed to the analysis and interpretation of the data. SC drafted the manuscript, and all authors contributed to the critical revision of the manuscript. HRH also supervised the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
"Obtaining Health Information" section of the postsurvey.
[DOCX File , 15 KB - jmir_v23i10e29155_app1.docx]

References

1. Gordon NP, Crouch E. Digital information technology use and patient preferences for internet-based health education modalities: cross-sectional survey study of middle-aged and older adults with chronic health conditions. JMIR Aging 2019 Apr 04;2(1):e12243 [FREE Full text] [doi: 10.2196/12243] [Medline: 31518291]
2. Calton B, Abedini N, Fratkin M. Telemedicine in the time of coronavirus. J Pain Symptom Manage 2020 Jul;60(1):12-14 [FREE Full text] [doi: 10.1016/j.jpainsymman.2020.03.019] [Medline: 32240756]

3. Delva S, Mendez KJ, Cajita M, Koirala B, Shan R, Wongvibulsin S, et al. Efficacy of mobile health for self-management of cardiometabolic risk factors: a theory-guided systematic review. J Cardiovasc Nurs 2021;36(1):34-55 [FREE Full text] [doi: 10.1097/JCN.0000000000000659] [Medline: 32040072]
4. Cajita MI, Gleason KT, Han H. A systematic review of mHealth-based heart failure interventions. J Cardiovasc Nurs 2016;31(3):10-22 [FREE Full text] [doi: 10.1097/JCN.0000000000000305] [Medline: 26544175]
5. Internet/broadband fact sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/internet-broadband [accessed 2021-03-10]
6. Lewis J. Internet access and racial/ethnic disparities in using internet health resources. U.S. Census Bureau. URL: https://www.census.gov/content/dam/Census/library/working-papers/2017/demo/SEHSD-WP2017-31paper.pdf [accessed 2020-03-11]
7. Ndugga N, Artiga S. Disparities in health and health care: 5 key questions and answers. Kaiser Family Foundation. URL: https://www.kff.org/disparities-policy/issue-brief/disparities-in-health-and-health-care-five-key-questions-and-answers/ [accessed 2020-03-11]
8. Anderson M, Kumar M. Digital divide persists even as lower-income Americans make gains in tech adoption. Urbanism Next. URL: https://www.urbanismnext.org/resources/digital-divide-persists-even-as-lower-income-americans-make-gains-in-tech-adoption [accessed 2020-03-11]
9. Key facts about Asian Americans, a diverse and growing population. Pew Research Center. URL: https://www.pewresearch.org/fact-tank/2021/04/29/key-facts-about-asian-americans/ [accessed 2021-03-10]
10. Perrin A, Duggan M. Americans' internet access: 2000-2015. Pew Research Center. URL: https://www.pewresearch.org/internet/2015/06/26/americans-internet-access-2000-2015/ [accessed 2021-03-10]
11. The rise of Asian Americans. Pew Research Center. URL: https://www.pewsocialtrends.org/2012/06/19/the-rise-of-asian-americans [accessed 2021-03-10]
12. Income and poverty in the United States: 2019. U.S. Census Bureau. URL: https://www.census.gov/content/dam/Census/library/publications/2020/demo/p60-270.pdf [accessed 2021-03-10]
13. 2019: ACS 5-Year estimates data profiles. U.S. Census Bureau. URL: https://data.census.gov/cedsci/table?q=United%20States&g=0100000US&tid=ACSDP5Y2019.DP05 [accessed 2021-03-10]
14. Budiman A, Ruiz N. Key facts about Asian origin groups in the U.S. Pew Research Center. URL: https://www.pewresearch.org/fact-tank/2021/04/29/key-facts-about-asian-origin-groups-in-the-u-s/ [accessed 2021-06-17]
15. Chawla N, Breen N, Liu B, Lee R, Kagawa-Singer M. Asian American women in California: a pooled analysis of predictors for breast and cervical cancer screening. Am J Public Health 2015 Feb;105(2):98-109. [doi: 10.2105/AJPH.2014.302250] [Medline: 25521898]
16. Miller JW, King JB, Joseph DA, Richardson LC, Centers for Disease Control and Prevention (CDC). Breast cancer screening among adult women--Behavioral Risk Factor Surveillance System, United States, 2010. MMWR Suppl 2012 Jun 15;61(2):46-50. [Medline: 22695463]
17. Hyun CS, Kim S, Kang SY, Jung S, Lee S. Chronic hepatitis B in Korean Americans: decreased prevalence and poor linkage to care. BMC Infect Dis 2016 Aug 15;16(1):415 [FREE Full text] [doi: 10.1186/s12879-016-1732-7] [Medline: 27526685]
18. Han H, Song Y, Kim M, Hedlin HK, Kim K, Ben Lee H, et al. Breast and cervical cancer screening literacy among korean american women: a community health worker-led intervention. Am J Public Health 2017 Jan;107(1):159-165. [doi: 10.2105/AJPH.2016.303522] [Medline: 27854539]
19. Davis TC, Long SW, Jackson RH, Mayeaux EJ, George RB, Murphy PW, et al. Rapid estimate of adult literacy in medicine: a shortened screening instrument. Fam Med 1993 Jun;25(6):391-395. [Medline: 8349060]
20. Han H, Kim J, Kim MT, Kim KB. Measuring health literacy among immigrants with a phonetic primary language: a case of Korean American women. J Immigr Minor Health 2011 Apr;13(2):253-259 [FREE Full text] [doi: 10.1007/s10903-010-9366-0] [Medline: 20585985]
21. Gutierrez N, Kindratt TB, Pagels P, Foster B, Gimpel NE. Health literacy, health information seeking behaviors and internet use among patients attending a private and public clinic in the same geographic area. J Community Health 2014 Feb;39(1):83-89. [doi: 10.1007/s10900-013-9742-5] [Medline: 23900880]
22. Sun Y, Zhang Y, Gwizdka J, Trace CB. Consumer evaluation of the quality of online health information: systematic literature review of relevant criteria and indicators. J Med Internet Res 2019 May 02;21(5):e12522 [FREE Full text] [doi: 10.2196/12522] [Medline: 31045507]
23. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: An analysis of health information seeking behavior among US adults. Cog Soc Sci 2017 Mar 13;3(1):1302785. [doi: 10.1080/23311886.2017.1302785]
24. Alexander B. Higher education, digital divides, and a Balkanized internet. EDUCAUSE Review. URL: https://er.educause.edu/-/media/files/articles/2017/10/erm17612.pdf [accessed 2021-03-10]
25. Berner J, Rennemark M, Jogréus C, Anderberg P, Sköldunger A, Wahlberg M, et al. Factors influencing Internet usage in older adults (65 years and above) living in rural and urban Sweden. Health Informatics J 2015 Sep;21(3):237-249 [FREE Full text] [doi: 10.1177/1460458214521226] [Medline: 24567416]

26. Eltorai AE, Ghanian S, Adams CA, Born CT, Daniels AH. Readability of patient education materials on the American association for surgery of trauma website. Arch Trauma Res 2014 Jun 30;3(2):e18161 [FREE Full text] [doi: 10.5812/atr.18161] [Medline: 25147778]
27. Hutchinson N, Baird GL, Garg M. Examining the reading level of internet medical information for common internal medicine diagnoses. Am J Med 2016 Jun;129(6):637-639. [doi: 10.1016/j.amjmed.2016.01.008] [Medline: 26829438]
28. Cajita MI, Rodney T, Xu J, Hladek M, Han H. Quality and health literacy demand of online heart failure information. J Cardiovasc Nurs 2017;32(2):156-164 [FREE Full text] [doi: 10.1097/JCN.0000000000000324] [Medline: 26938508]
29. Mishra V, Dexter JP. Comparison of readability of official public health information about COVID-19 on websites of international agencies and the governments of 15 countries. JAMA Netw Open 2020 Aug 03;3(8):e2018033 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.18033] [Medline: 32809028]
30. Sentell T, Braun KL. Low health literacy, limited English proficiency, and health status in Asians, Latinos, and other racial/ethnic groups in California. J Health Commun 2012;17 Suppl 3:82-99 [FREE Full text] [doi: 10.1080/10810730.2012.712621] [Medline: 23030563]
31. Health communication and health information technology data details. Office of Disease Prevention and Health Promotion. URL: https://www.healthypeople.gov/node/3508/data-details [accessed 2021-03-10]
32. Huh J, Koola J, Contreras A, Castillo A, Ruiz M, Tedone K, et al. Consumer health informatics adoption among underserved populations: thinking beyond the digital divide. Yearb Med Inform 2018 Aug;27(1):146-155 [FREE Full text] [doi: 10.1055/s-0038-1641217] [Medline: 30157518]
33. The US will become “minority white” in 2045, census projects. Brookings. URL: https://www.brookings.edu/blog/the-avenue/2018/03/14/the-us-will-become-minority-white-in-2045-census-projects/ [accessed 2021-03-10]
34. U.S. smartphone use in 2015. Pew Research Center. URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2015/03/PI_Smartphones_0401151.pdf [accessed 2021-03-10]
35. Azulay R, Valinsky L, Hershkowitz F, Magnezi R. Repeated automated mobile text messaging reminders for follow-up of positive fecal occult blood tests: randomized controlled trial. JMIR Mhealth Uhealth 2019 Feb 05;7(2):e11114 [FREE Full text] [doi: 10.2196/11114] [Medline: 30720439]
36. Ershad Sarabi R, Sadoughi F, Jamshidi Orak R, Bahaadinbeigy K. The effectiveness of mobile phone text messaging in improving medication adherence for patients with chronic diseases: a systematic review. Iran Red Crescent Med J 2016 May;18(5):e25183 [FREE Full text] [doi: 10.5812/ircmj.25183] [Medline: 27437126]
37. Ladley A, Broom M, Arthur J. Update on the evaluation of text messaging as an educational method to improve health care utilization. Acad Pediatr 2020;20(7):889-890. [doi: 10.1016/j.acap.2020.04.004] [Medline: 32302755]
38. Leahy D, Lyons A, Dahm M, Quinlan D, Bradley C. Use of text messaging in general practice: a mixed methods investigation on GPs' and patients' views. Br J Gen Pract 2017 Nov;67(664):744-750 [FREE Full text] [doi: 10.3399/bjgp17X693065] [Medline: 28947620]
39. Genes N, Violante S, Cetrangol C, Rogers L, Schadt EE, Chan YY. From smartphone to EHR: a case report on integrating patient-generated health data. NPJ Digit Med 2018;1:23 [FREE Full text] [doi: 10.1038/s41746-018-0030-8] [Medline: 31304305]
40. Han H, Gleason KT, Sun C, Miller HN, Kang SJ, Chow S, et al. Using patient portals to improve patient outcomes: systematic review. JMIR Hum Factors 2019 Dec 19;6(4):e15038 [FREE Full text] [doi: 10.2196/15038] [Medline: 31855187]
41. Bloom PP, Marx M, Wang TJ, Green B, Ha J, Bay C, et al. Attitudes towards digital health tools for outpatient cirrhosis management in patients with decompensated cirrhosis. BMJ Innov 2020 Jan 14;6(1):18-25. [doi: 10.1136/bmjinnov-2019-000369]
42. Homework gap and connectivity divide. Federal Communications Commission. URL: https://www.fcc.gov/about-fcc/fcc-initiatives/bridging-digital-divide-all-americans [accessed 2021-03-10]
43. Top 10 U.S. metropolitan areas by Korean population, 2015. Pew Research Center. 2017. URL: https://www.pewresearch.org/social-trends/chart/top-10-u-s-metropolitan-areas-by-korean-population/ [accessed 2021-03-10]
44. Blanco LR, Rodriguez LM. Delivering information about retirement saving among Hispanic women: two Facebook experiments. Behav Public Policy 2020 Nov;4(3):343-369 [FREE Full text] [doi: 10.1017/bpp.2018.33] [Medline: 34084943]
45. Cudjoe J, Budhathoki C, Roter D, Gallo JJ, Sharps P, Han H. Exploring health literacy and the correlates of pap testing among African immigrant women: findings from the Afropap study. J Cancer Educ 2021 Jun;36(3):441-451. [doi: 10.1007/s13187-020-01755-9] [Medline: 32410109]
46. Plain Writing Act of 2010. Digital.gov. URL: https://digital.gov/resources/plain-writing-act-of-2010/ [accessed 2021-08-12]
47. Federal plain language guidelines. plainlanguage.gov. URL: https://www.plainlanguage.gov/guidelines/ [accessed 2021-06-15]
48. Weiss B. Health literacy and patient safety: help patients understand. American Medical Association Foundation. 2007. URL: http://www.partnershiphp.org/Providers/HealthServices/Documents/Health%20Education/CandLToolKit/2%20Manual%20for%20Clinicians.pdf [accessed 2021-08-12]
49. Clear & Simple: Clear Communication. National Institutes of Health. 2018. URL: https://www.nih.gov/institutes-nih/nih-office-director/office-communications-public-liaison/clear-communication/clear-simple [accessed 2021-05-26]

Abbreviations

AOR: adjusted odds ratio
REALM: Rapid Estimate of Adult Literacy in Medicine
TV: television

Edited by R Kukafka; submitted 28.03.21; peer-reviewed by MA Bahrami, H Moon; comments to author 13.05.21; revised version received 23.06.21; accepted 05.07.21; published 29.10.21.

Please cite as:
Chae S, Lee YJ, Han HR
Sources of Health Information, Technology Access, and Use Among Non–English-Speaking Immigrant Women: Descriptive Correlational Study
J Med Internet Res 2021;23(10):e29155
URL: https://www.jmir.org/2021/10/e29155
doi:10.2196/29155
PMID:

Original Paper

Circadian Rhythm Analysis Using Wearable Device Data: Novel Penalized Machine Learning Approach

Xinyue Li[1,2], PhD; Michael Kane[3], PhD; Yunting Zhang[2,4], PhD; Wanqi Sun[5], MD; Yuanjin Song[5], MD; Shumei Dong[5], MSc; Qingmin Lin[5], PhD; Qi Zhu[5], MSc; Fan Jiang[4,5], MD, PhD; Hongyu Zhao[3,6], PhD

[1]School of Data Science, City University of Hong Kong, Hong Kong, China (Hong Kong)

[2]Child Health Advocacy Institute, Shanghai Children's Medical Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[3]Department of Biostatistics, Yale School of Public Health, New Haven, CT, United States

[4]School of Public Health, Shanghai Jiao Tong University, Shanghai, China

[5]Department of Developmental and Behavioral Pediatrics, Shanghai Children's Medical Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[6]Yale Joint Center for Biostatistics, Shanghai Jiao Tong University, Shanghai, China

Corresponding Author:
Hongyu Zhao, PhD
Department of Biostatistics
Yale School of Public Health
300 George Street
Suite 503
New Haven, CT, 06511
United States
Phone: 1 203 785 3613
Fax: 1 203 785 6912
Email: hongyu.zhao@yale.edu

Abstract

Background: Wearable devices have been widely used in clinical studies to study daily activity patterns, but the analysis remains a major obstacle for researchers.

Objective: This study proposes a novel method to characterize sleep-activity rhythms using actigraphy and further use it to describe early childhood daily rhythm formation and examine its association with physical development.

Methods: We developed a machine learning–based Penalized Multiband Learning (PML) algorithm to sequentially infer dominant periodicities based on the Fast Fourier Transform (FFT) algorithm and further characterize daily rhythms. We implemented and applied the algorithm to Actiwatch data collected from a cohort of 262 healthy infants at ages 6, 12, 18, and 24 months, with 159, 101, 111, and 141 participants at each time point, respectively. Autocorrelation analysis and Fisher test in harmonic analysis with Bonferroni correction were applied for comparison with the PML. The association between activity rhythm features and early childhood motor development, assessed using the Peabody Developmental Motor Scales-Second Edition (PDMS-2), was studied through linear regression analysis.

Results: The PML results showed that 1-day periodicity was most dominant at 6 and 12 months, whereas one-day, one-third–day, and half-day periodicities were most dominant at 18 and 24 months. These periodicities were all significant in the Fisher test, with one-fourth–day periodicity also significant at 12 months. Autocorrelation effectively detected 1-day periodicity but not the other periodicities. At 6 months, PDMS-2 was associated with the assessment seasons. At 12 months, PDMS-2 was associated with the assessment seasons and FFT signals at one-third–day periodicity (P<.001) and half-day periodicity (P=.04), respectively. In particular, the subcategories of stationary, locomotion, and gross motor were associated with the FFT signals at one-third–day periodicity (P<.001).

Conclusions: The proposed PML algorithm can effectively conduct circadian rhythm analysis using time-series wearable device data. The application of the method effectively characterized sleep-wake rhythm development and identified the association between daily rhythm formation and motor development during early childhood.

(J Med Internet Res 2021;23(10):e18403) doi:10.2196/18403

KEYWORDS

wearable device; actigraphy; circadian rhythm; physical activity; early childhood development

Introduction

Background

Wearable devices have been increasingly used in research recently because they can provide continuous objective monitoring of activities and vital signs data such as body temperature and pulse rates [1-3]. In sleep and activity studies, researchers have focused on the actigraphy data generated from wearable devices to study sleep and activity patterns as an alternative to sleep diaries and polysomnography [1,4]. The device usually uses an accelerometer that monitors acceleration in one or more directions, and this wristwatch-like device is often worn on the wrist to record activity continuously for several days. Either the raw data or the transformed activity count data can be used to study sleep-wake patterns and screen sleep disorders [4,5]. Actigraphy not only avoids the subjectivity and bias issues with sleep diaries but also overcomes the drawbacks of polysomnography, such as high costs, in-laboratory setting, intrusive measures, and difficulty in long-term monitoring.

Continuous objective monitoring using wearable devices provides researchers with the opportunity to conduct circadian rhythm studies. Circadian rhythms are endogenous and entrainable biological processes that follow a period of approximately 24 hours, and many physiological phenomena such as sleep-wake patterns, body temperature, and hormone levels exhibit circadian rhythms. For humans, most circadian rhythms are under the control of the pacemaker located in the suprachiasmatic nuclei in the anterior hypothalamus of the central nervous system, and suprachiasmatic nuclei accept environmental information such as the light and dark cycle to adjust the 24-hour cycle [6]. However, 24-hour human circadian rhythms are not mature at birth, when the predominant rhythm is ultradian, and the circadian rhythms of sleep-wake cycles and body temperature gradually develop during the first year after birth [6-9]. Many studies have investigated how circadian rhythms develop through childhood into adolescence and adulthood and how they are related to health issues such as sleep problems, mental problems, and disease risks, to name a few [8,10-14]. It is noteworthy that the development of circadian rhythms during early childhood is associated with disease risk factors and can affect both childhood and adult life [8]. Therefore, it is important to conduct circadian rhythm studies to gain a thorough understanding of the formation and consolidation of daily activity rhythms during early development as well as the association between the changes in daily rhythms and health conditions.

Actigraphy data generated from wearable devices have been validated to provide reliable information on sleep and circadian rhythms [15]. However, the analysis of time-series data from actigraphy remains a major obstacle for researchers. Current major statistical methods are either parametric, based on cosinor analysis, or nonparametric [16-21]. These methods do not specifically focus on periodic information and are not specifically suitable for populations whose sleep-wake rhythms are not sinusoidal, such as patients with Circadian Rhythm Sleep Disorder, or not mature, such as young infants and toddlers [6-10,12]. Therefore, traditional approaches targeting normal daily rhythms might not work because detailed activity rhythms cannot be captured. There is a need to develop an appropriate methodology to extract periodic information and study detailed circadian patterns of all populations to better characterize daily rhythms.

Objective

In this paper, we propose a Penalized Multiband Learning (PML) approach that can complement current methods to characterize daily rhythms based on periodic information in time-series wearable device data. PML extracts periodic information using the Fast Fourier Transform (FFT) algorithm and then performs penalized selection based on regularization, a classic approach used in machine learning, to identify dominant periodicities and further characterize daily rhythms [22,23]. In this paper, we first present the proposed PML approach in detail and discuss its usefulness and advantages compared with other methods. Subsequently, we present an application of the method to early childhood wearable device activity data, in which we characterize the formation and consolidation of sleep-activity rhythms and further study its association with physical development during early childhood.

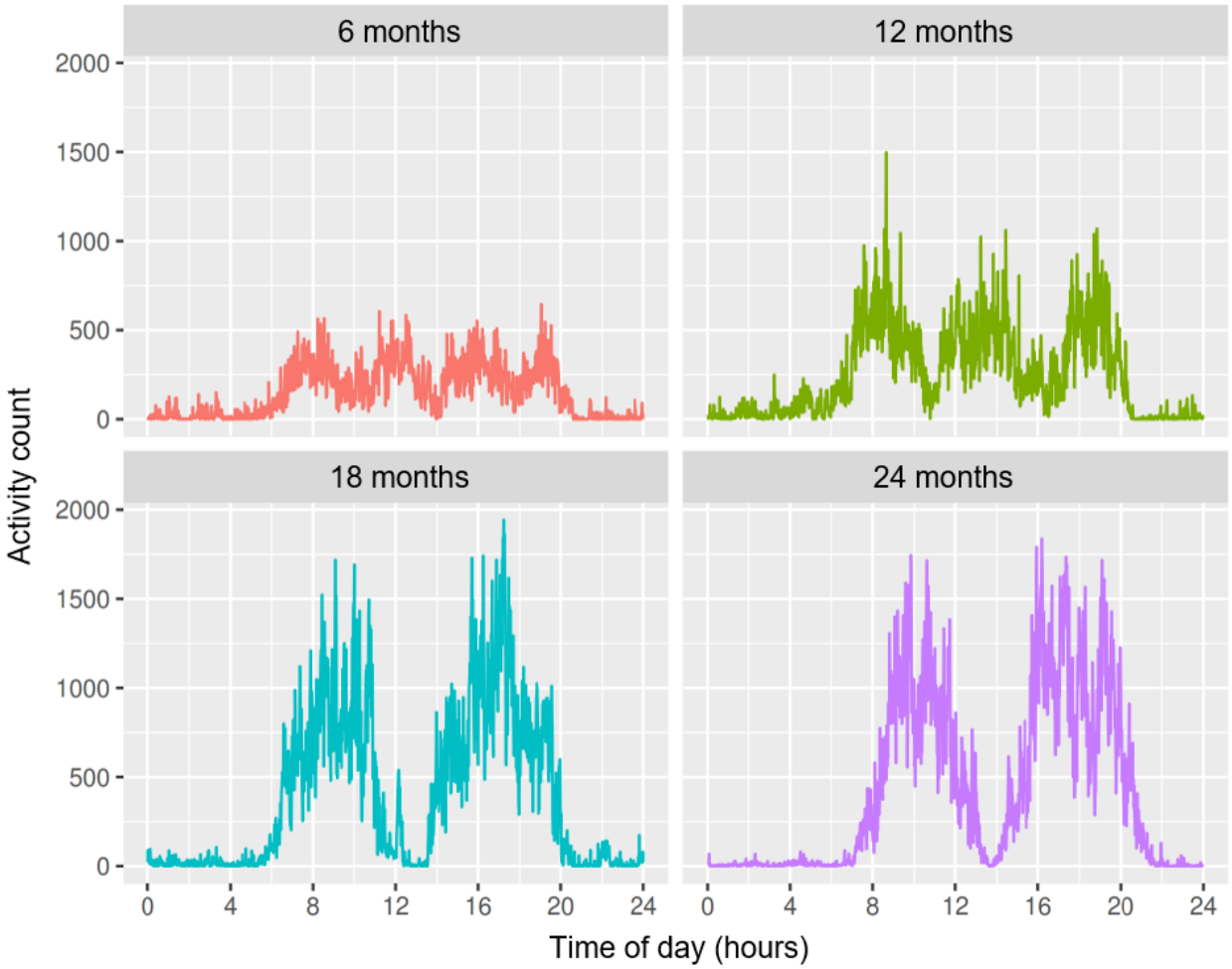
Methods

Data

The study participants were 262 healthy newborns recruited in 2012-2013 by the Shanghai Children's Medical Center, Shanghai, China. Actiwatch data were collected at ages 6, 12, 18, and 24 months, with 159, 101, 111, and 141 participants at each time point, respectively; not all participants from the cohort participated each time. The infants and toddlers were required to wear Actiwatch 2 (Philips Respironics Mini-Mitter) on the ankle for 7 consecutive days. Wearing such devices on the ankle is commonly recommended for young infants or toddlers [24]. The Actiwatch 2 uses a piezoelectric sensor to detect accelerations between 0.5 and 2.0 g with a frequency response range between 0.35 and 7.5 Hz, and activity counts summarize the accelerations over each epoch. The data output format for Actiwatch 2 was configured to be the activity count per 1-minute epoch. On the basis of sleep diaries and activity plots for each individual, the days showing nonwear periods with straight lines of zero activity counts were removed. Nonwear periods can be differentiated from sedentary behaviors or sleep because the former gives almost all zero activity counts, whereas the latter gives nonzero activity counts every now and then. Figure 1 shows the activity plots for participant ID 17, and it can be seen that at 6 months, low and high activities were intermittent during the day, suggesting multiple daytime naps, whereas near-zero activity levels at night suggest long nighttime sleep. At 12 months, 3 activity peaks, one morning nap, and one afternoon nap can be identified. At 18 and 24 months, 2 activity peaks

formed and stabilized, showing only one afternoon nap. The daily activity rhythm developed and stabilized as the infant grew. In addition to Actiwatch data, demographic information and family information were collected at baseline, such as child's sex, child's date of birth, parents' age, child's birth weight and body length, parents' height and weight, parents' educational levels and working status, and family income. The Peabody Developmental Motor Scales-Second Edition (PDMS-2) was used to assess early childhood physical development at 6, 12, 18, and 24 months [25]. The institutional review board of the Shanghai Children's Medical Center, Shanghai Jiao Tong University, approved the study (approval number: SCMCIRB-2012033). The parents of the children who participated in the study provided written informed consent.

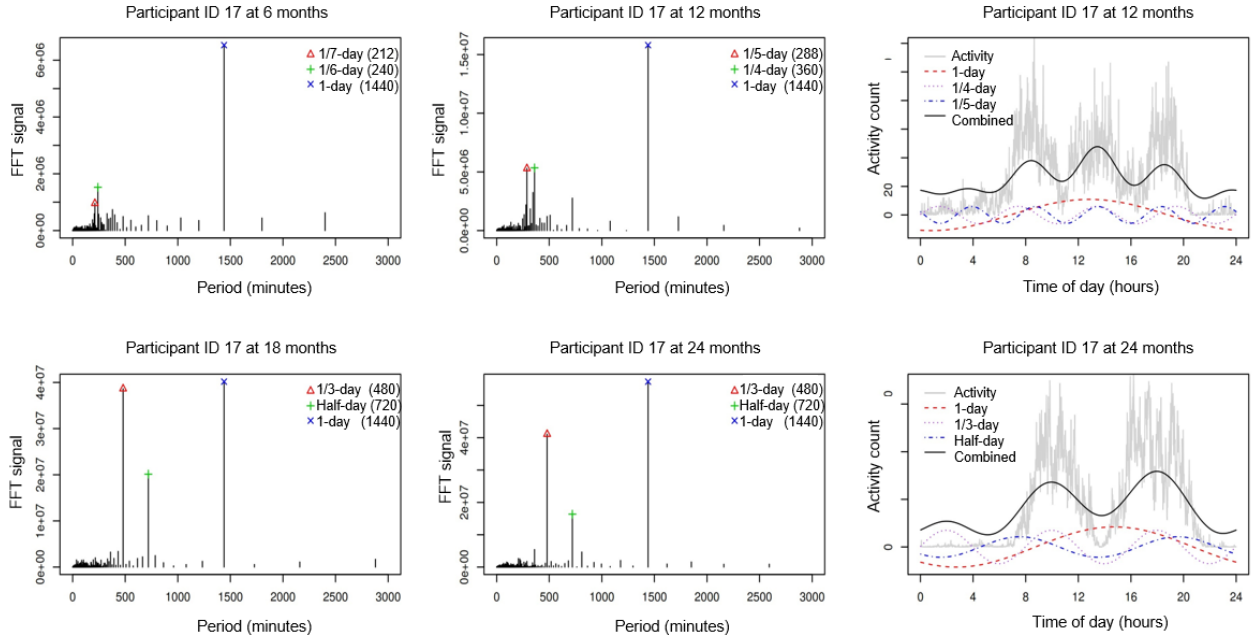
Figure 1. The activity plots for participant ID 17 at 6, 12, 18, and 24 months, with activity counts averaged across 7 days at each time point.



FFT Algorithm

To describe the consolidation of sleep-activity rhythms during early childhood, we used periodic information to characterize the daily rhythms. Specifically, we used the FFT algorithm to convert time domain signals into a frequency domain spectrum to extract periodic information. We analyzed the original data to allow for non–24-hour sleep-wake rhythm detection. Figure 2 shows the FFT results for participant ID 17 at each age. The 1-day periodicity was the most dominant at all time points. The one-fifth–day and one-fourth–day periodicities can be identified at 12 months. The half-day and one-third–day periodicities did not become dominant until 18 and 24 months. It is noteworthy that each periodicity is not interpreted alone; rather, the periodicities are combined to understand the overall pattern. As suggested in the 2 right panels of Figure 2, the combined one-fifth–day and one-fourth–day periodicities form a three-peak, two-nap pattern at 12 months. Similarly, the combined half-day and one-third–day periodicities exhibited a two-peak, one-nap pattern at 18 and 24 months. Therefore, the combination of dominant periodicities can be used to capture the main sleep-activity patterns at each age and describe the gradual consolidation of daily rhythms in early childhood development.

Figure 2. Four panels on the left: Fast Fourier Transform results for participant ID 17 at 6, 12, 18, and 24 months. Two panels on the right: top three periodicities and the combined periodicities plotted on 1-day observation for participant ID 17 at 12 and 24 months. FFT: Fast Fourier Transform.



Identification of Dominant Periodicities

The PML algorithm is as follows: let matrix $X \in R^{n \times p}$ denote the FFT results, where denotes the number of individual observations and denotes the number of periodicities from FFT. Specifically, $X = (x_1,x_2,...,x_p)$, where x_j is the vector of length n for the jth periodicity.

Let Θ be the diagonal matrix selecting columns from X such that  and $0 \leq \theta_{j,j} \leq 1$, $j=1,...,p$:

Θ identifies columns of dominant periodicities from X in such a way that dominant periodicities corresponding to nonzero $\theta_{j,j}$'s are selected. We minimized the squared Frobenius norm  , which is the sum of the squared elements of the matrix. Using the properties of the Frobenius norm, we obtained the following:

As X^TX is fixed, it is equivalent to minimize as follows:

To estimate Θ and identify dominant periodicities, we used a penalized selection method similar to Lasso, a widely used method in shrinkage and selection of a subset of features in regression models and machine learning approaches [23]. In regression analysis, the Lasso penalty is most effective in selecting a few important features while suppressing the regression coefficients of other nonselected features to 0 [23]. In our case, the Lasso penalty served to select a few dominant periodicities through diagonal elements of Θ instead of regression coefficients. Furthermore, we added an elastic net–like penalty term to the Frobenius norm, namely a combination of L1 and L2 norms [22]:

where λ is the turning parameter and α controls the balance between the L1 and L2 norms. Note that $\theta_{j,j}$'s is nonnegative; thus, we do not need to take the absolute value for the L1 norm. By setting λ large enough, all diagonal elements of Θ, namely all $\theta_{j,j}$'s are suppressed to zero, and no periodicities are selected. As λ decreases, some $\theta_{j,j}$'s become nonzero, and they correspond to the most dominant periodicities that are selected sequentially according to their dominance.

To minimize $g(\theta)$, we took the partial derivative of $g(\theta)$ with respect to each  $= -2||x_k||^2 + 2\theta_{k,k}||x_k||^2 + (1\text{-}a)\lambda\theta_{k,k} + a\lambda$,which is convex and subject to the constraint $0 \leq \theta_{k,k} \leq 1$. Thus, we have the following:

If we only have the L1 penalty, then $\alpha=1$ and  . In our case, we used the Lasso L1 penalty alone and trained λ because we wanted to select the most important periodicities while suppressing other periodicities to 0. However, we still kept the L2 norm in the original model as an option because it might be helpful in future tasks such as prediction, classification, and reconstruction of curves.

We used the mean squared error (MSE), which is equivalent to the squared Frobenius norm  , as the criterion for choosing λ and the number of nonzero $\theta_{j,j}$'s (the number of dominant periodicities selected), as well as to evaluate the variability that was not explained by the selected periodicities. We did not

choose cross-validation because the results showed that the test data set error curve was monotonous. We trained λ from  to 0 because λ =  suppresses all $\theta_{j,j}$'s to and λ=0 gives no penalty. By decreasing λ, we identified dominant periodicities sequentially and characterized the daily sleep-activity rhythm at each age. An R package named PML was developed [26] for the implementation of the PML algorithm [27].

Comparison With Other Methods

To rigorously conduct statistical tests and select significant periodicities, we applied the Fisher test in harmonic analysis [28]. It is a sequential test for ordered statistics, and periodicities are first ordered and then tested for significance. If one periodicity is statistically significant, the next one will be tested further. Otherwise, the sequential test will be stopped. At each step, the critical value at which to declare statistical significance was different. In some studies, the method may not be implemented correctly; therefore, we included the sequential test in the R package for easy implementation. As we performed multiple testing, Bonferroni correction was used to adjust the *P* values. If we conduct the tests at significance level α, we reject the null hypothesis if *P* value≤α/*p*, where *p* is the number of periodicities, and conclude that the periodicity is significant.

To evaluate the effectiveness of the PML algorithm, we compared it with the autocorrelation of the standard approach. The autocorrelation r_k is calculated between activity measurements with a time lag *k*, and the coefficient r_{24} denoting a 24-hour time lag is of primary interest in circadian studies [29]. r_k ranges between –1 and 1, and a *k*-hour periodic pattern can show a higher value of the correlation coefficient r_k. In the plot of r_k against the time lag *k*, a peak around *k*=24 can be observed when there is a dominant circadian pattern of 24-hour periodicity. We plotted autocorrelation against the time lag to compare the autocorrelation method with our algorithm.

Association Between Daily Rhythms and Motor Development

Next, we conducted linear regression analysis to study the association between the consolidation of daily activity rhythms and early childhood physical development. PDMS-2 is considered an early childhood developmental assessment tool, and the score is used as the outcome. If the PDMS-2 total motor standard scores are found to be associated with daily rhythm features, the standard scores for the subtests, including stationary, locomotor, object manipulation, grasping, and visual-motor integration as well as gross motor and fine motor, are used as the outcome to examine which specific subcategory is associated. Gross motor represents the overall performance on stationary, locomotion, and reflexes (6 months) or object manipulation (12, 18, and 24 months) for infants, and fine motor represents the overall performance on grasping and visual-motor integration.

The FFT signals at dominant periodicities identified by the PML were used as daily rhythm features and considered covariates in the model. In addition to periodic features, demographic information and family information as potential confounders were also considered in the model. Backward selection was used in the model-fitting process. Although some variable (denoted as variable A here) may seem to be statistically significant in the complete model, after removal of insignificant variables in the variable selection process, variable A may become insignificant. In such cases, variable A is removed in the final model to achieve parsimony.

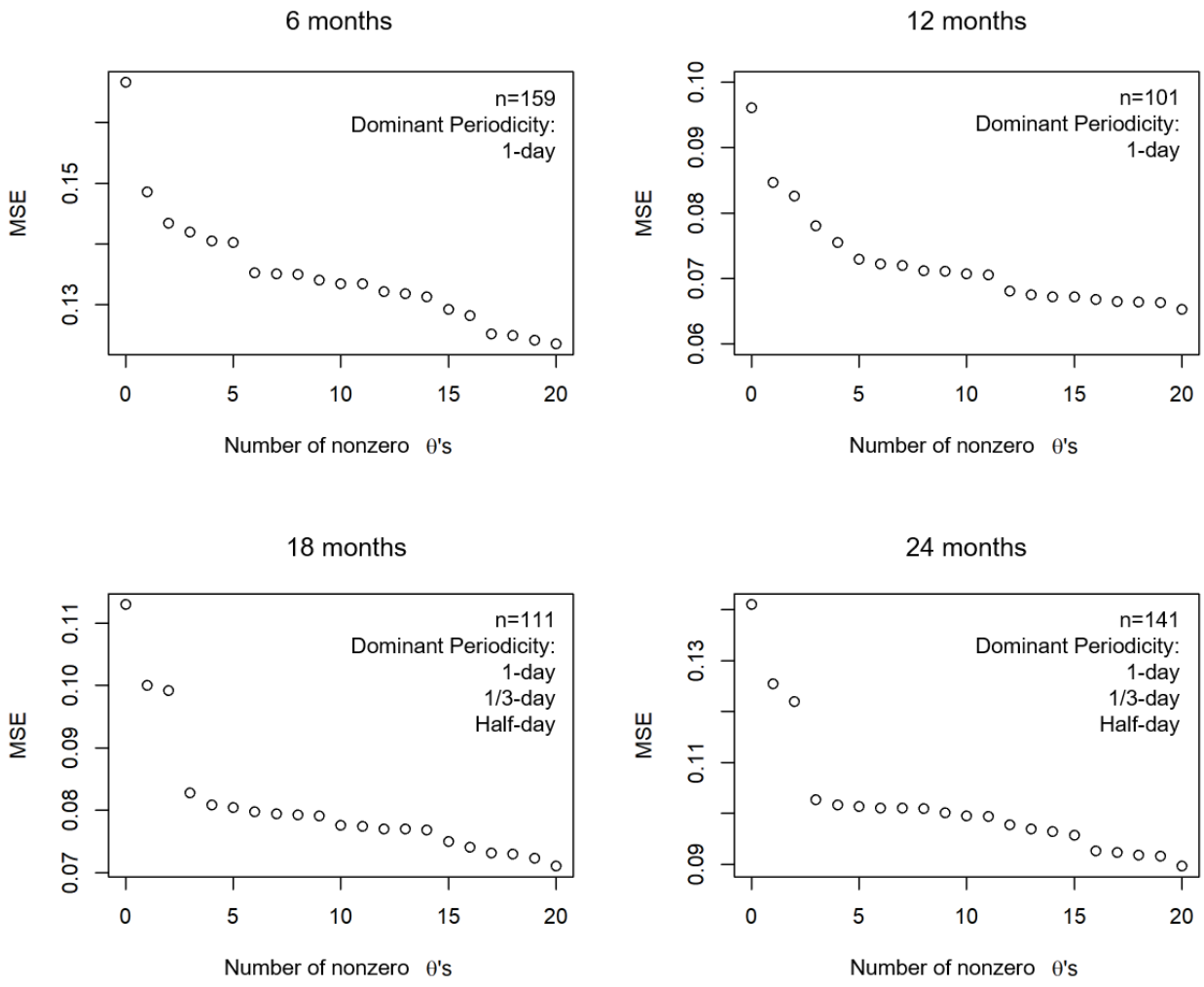
Linear regression analysis was conducted at 6, 12, 18, and 24 months to study the association between daily rhythms and motor development. For the final model comparison, r^2, which measures the proportion of variance in the outcome explained by the model, and the adjusted r^2, which modifies r^2 based on the number of predictor variables, were also calculated. All statistical analyses were conducted using R version 3.3.2 (R Foundation for Statistical Computing).

Results

Identification of Dominant Periodicities

As shown in Figure 3, at each age, we plotted the MSE against the number of nonzero θ's. Specifically, we plotted only the points where the number of nonzero θ's (periodicities selected) increased as the penalty term λ decreased. For 6 months and 12 months, the first harmonic at 1-day periodicity is the most dominant because we can observe a large dip in the MSE when the first periodicity is selected, whereas the periodicities that are further selected do not cause the same level of decrease. For 18 and 24 months, the first three periodicities of 1 day, one-third day, and half day are the most dominant, and selecting the first three can lead to a relatively large decrease in the MSE. The results indicate that during the first year, only one-day periodicity was formed and stabilized in the infant population. The sleep-activity rhythm did not stabilize until 18 months, showing the pattern of 1 nighttime sleep, 1 daytime nap, and 2 daytime activity peaks.

Figure 3. Mean squared error plotted against the number of nonzero θ's as the penalty term λ decreased at 6, 12, 18, and 24 months. MSE: mean squared error.



The Fisher tests yielded similar results. As shown in Table 1, only 1-day periodicity was significant at 6 months ($P<10^{-5}$) because for infants at this stage, sleep-activity patterns have already adjusted to a 24-hour cycle. However, daytime activities have not yet been stabilized and variations exist across days. At 12 months, the 4 periodicities were significant ($P<10^{-5}$). This is because infants' sleep-activity patterns start to stabilize, but there are variations across individuals: some take one nap in the afternoon, whereas others take two naps, one in the morning and one in the afternoon. The one-nap pattern can be captured by the one-third–day periodicity, whereas the two-nap pattern can be captured by the one-fourth–day periodicity, as shown in Figure 2. Furthermore, at 18 and 24 months, three periodicities were significant ($P<10^{-5}$): 1 day, half day, and one-third day, indicating the final consolidation of daily sleep-activity rhythms with only one daytime nap in the afternoon. In addition, the proportions of variance explained by the half day and one-third day periodicities at 18 months and 24 months were approximately the same, both higher than those at 12 months.

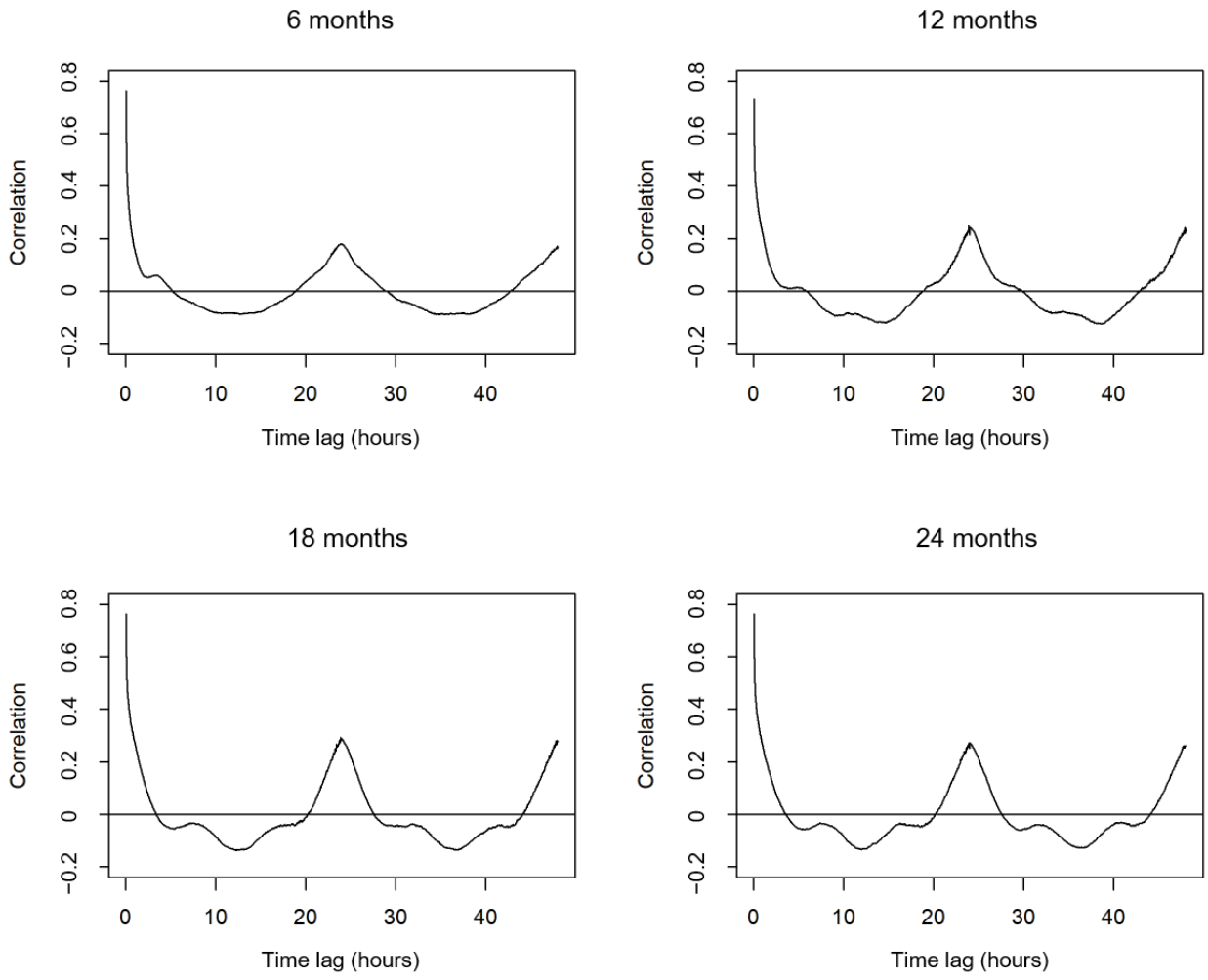
Table 1. Significant periodicities at 6, 12, 18, and 24 months, with the corresponding proportions of variances among all Fast Fourier Transform signals and *P* values.

Age (months)	Proportions of variance	Period		*P* value
		Minutes	Days	
6	0.0110	1440	One	6.59×10^{-8}
12	0.0120	1440	One	1.75×10^{-8}
12	0.0068	720	Half	3.75×10^{-7}
12	0.0056	360	One-fourth	2.82×10^{-7}
12	0.0045	480	One-third	6.49×10^{-6}
18	0.0130	1440	One	1.46×10^{-9}
18	0.0085	480	One-third	1.87×10^{-10}
18	0.0083	720	Half	5.02×10^{-15}
24	0.0130	1440	One	1.97×10^{-9}
24	0.0086	480	One-third	1.17×10^{-10}
24	0.0080	720	Half	3.78×10^{-14}

Comparison With Autocorrelation

To compare the PML algorithm with autocorrelation, the plot of correlation estimates against time lags is shown in Figure 4. The circadian rhythm at 24 hours can be observed at all time points because the peaks of estimated correlation are at time lags between 23.8 hours and 24.3 hours. We can also observe some local maximal correlation estimates at other time lags: 3.3 hours at 6 months, 4.7 and 10.7 hours at 12 months, 7.5 hours at 18 months, and 7.5 and 16.3 hours at 24 months. Although 3.3 hours at 6 months may seem reasonable because of the infant feeding schedule, other cycles are difficult to explain [30,31]. Autocorrelation estimates can be biased because of the presence of multiple periodicities, and researchers generally use these estimates to verify the most dominant periodicity such as 24 hours. Thus, from the autocorrelation plots, the most dominant 24-hour rhythm that yields the global maximal correlation estimate can be identified at each age, and this dominant periodicity was also identified by PML.

Figure 4. Estimated autocorrelation against time lags at 6, 12, 18, and 24 months.



Association Between Daily Rhythms and Motor Development

The summary of the PDMS-2 standard scores for each category is presented in Table 2.

As shown in Table 3, at 6 months, the PDMS-2 total motor scores were found to be associated with the assessment seasons (P<.001). Infants receiving the PDMS-2 assessment in winter and spring tended to have lower PDMS-2 total motor scores than those assessed in summer and autumn. At 12 months, the PDMS-2 total motor scores were associated with both seasons and FFT signals: infants assessed in summer tended to have higher PDMS-2 total motor scores, and infants with higher FFT signals detected at one-third–day and half-day periodicities also tended to have higher PDMS-2 total motor scores (P<.001 and P=.04, respectively). r^2 was 0.25, and the adjusted r^2 was 0.21. At 18 and 24 months, no association was identified between the PDMS-2 total motor scores and any other variables.

Table 2. The PDMS-2[a] standard scores for the subtests, gross motor, fine motor, and total motor.

PDMS-2 category	Standard scores, mean (SD)			
	6 months (n=246)	12 months (n=225)	18 months (n=192)	24 months (n=170)
Reflexes	10.56 (1.02)	—[b]	—	—
Stationary	9.67 (1.44)	9.48 (1.00)	9.94 (0.36)	9.03 (1.38)
Locomotor	10.09 (1.08)	8.82 (1.76)	9.26 (1.24)	8.61 (1.70)
Object manipulation	—	9.92 (0.78)	9.49 (1.16)	8.74 (1.33)
Grasping	10.69 (1.02)	11.08 (1.44)	9.55 (0.78)	9.99 (0.98)
Visual-motor integration	11.19 (1.20)	10.69 (1.13)	11.04 (1.74)	9.79 (1.74)
Gross motor	30.34 (2.72)	28.25 (2.65)	28.63 (2.21)	26.16 (3.17)
Fine motor	21.86 (2.02)	21.83 (2.17)	20.67 (1.93)	19.80 (2.22)
Total	52.24 (4.18)	50.23 (5.05)	49.43 (3.19)	45.98 (4.69)

[a]PDMS-2: Peabody Developmental Motor Scales-Second Edition.

[b]Reflexes is only for 6-month infants, and object manipulation is only for 12-, 18-, and 24-month infants.

Table 3. Linear regression analysis of Peabody Developmental Motor Scales-Second Edition total motor standard scores by season and Fast Fourier Transform signals at 6 months and 12 months, respectively.

Time and variable	Estimate (SE)	*t* test (*df*)	*P* value
6 months			
(intercept)	50.10 (0.56)	89.52 (155)	<.001
spring	0.48 (0.85)	0.56 (155)	.58
summer	4.48 (0.82)	5.43 (155)	<.001
autumn	4.06 (0.89)	4.54 (155)	<.001
12 months			
(intercept)	45.52 (1.29)	35.25 (96)	<.001
summer	1.66 (1.58)	1.05 (96)	.30
half day[a]	0.19 (0.09)	2.10 (96)	.04
one-third day[a]	0.31 (0.09)	3.55 (96)	<.001
summer: one-third day[a]	-0.48 (0.16)	−3.00 (96)	.004

[a]The Fast Fourier Transform signals were multiplied by 10,000 in regression models, so that the estimated effect sizes were for every 10,000-unit increase in the Fast Fourier Transform signals.

As PDMS-2 total motor scores are associated with FFT signals at 12 months, further linear regression between each subtest score and FFT signal was also examined. As shown in Table 4, subtests for stationary and locomotion as well as gross motor and fine motor were found to be associated with the one-third–day periodicity. The gross motor represents the overall performance on the three subtests of stationary, locomotion, and object manipulation for infants at 12 months, and because the association of FFT signals at one-third–day periodicity with stationary and locomotion subtests is strong, it is expected that the association of FFT signals at one-third–day periodicity with gross motor is also strong. r^2 and the adjusted r^2 were 0.05 and 0.04 for the stationary model, 0.23 and 0.20 for the locomotion model, 0.21 and 0.17 for the gross motor model, and 0.15 and 0.11 for the fine motor model, respectively.

Table 4. Linear regression analysis of Peabody Developmental Motor Scales-Second Edition standard scores by season and Fast Fourier Transform signals at 12 months: stationary and locomotion subtests and gross motor and fine motor as the outcome, respectively.

Category and variable[a]	Estimate (SE)	*t* test (*df*)	*P* value
Stationary			
(intercept)	9.23 (0.19)	47.9 (99)	<.001
one-third day	0.04 (0.02)	2.17 (99)	.03[b]
Locomotion			
(intercept)	7.16 (0.43)	16.48 (97)	<.001
summer	0.71 (0.75)	0.94 (97)	.35
one-third day	0.19 (0.04)	4.61 (97)	<.001[b]
summer: one-third day	–0.13 (0.07)	–1.93 (97)	.06[c]
Gross motor			
(intercept)	26.41 (0.63)	42.03 (97)	<.001
summer	0.57 (1.09)	0.52 (97)	.60
one-third day	0.24 (0.06)	3.97 (97)	<.001[b]
summer: one-third day	–0.17 (0.10)	–1.73 (97)	.09[c]
Fine motor			
(intercept)	21.08 (0.5)	42.15 (97)	<.001
summer	0.34 (0.86)	0.39 (97)	.70
one-third day	0.11 (0.05)	2.36 (97)	.02[b]
summer: one-third day	–0.15 (0.08)	–1.89 (97)	.06[c]

[a]The Fast Fourier Transform signals were multiplied by 10,000 in regression models so that the estimated effect sizes were for every 10,000-unit increase in the Fast Fourier Transform signals.

[b]Statistical significance level at α=.05.

[c]Statistical significance level at α=.10.

Discussion

Method Evaluation

The PML approach is very effective in studying daily activity rhythms among infants and toddlers. At 6 and 12 months, the dominant 1-day periodicity suggests the formation of the 24-hour cycle. At 18 and 24 months, the combination of the dominant 1-day, one-third–day, and half-day periodicities forms a consolidated daily activity pattern with 2 activity peaks during the day and 1 afternoon nap. PML not only effectively identified population-level dominant periodicities, but also characterized sleep-activity patterns without complex functional analysis. PML can complement current methods for circadian rhythm analysis and is especially useful for populations whose daily rhythm patterns are nonsinusoidal and irregular. On the other hand, because PML is applicable to time-series data with a nature similar to that of actigraphy, the application of PML can be extended to other types of circadian rhythm studies using information such as body temperature and hormone data to study and characterize daily rhythms effectively.

Comparison With Other Methods

In comparison, the Fisher test in harmonic analysis tends to identify many significant periodicities unless a stringent threshold is used for statistical significance. In this study, we used the Bonferroni correction to adjust for multiple testing and used a significance level of 10^{-5} to select the periodicities, even though we did not conduct as many statistical tests simultaneously. In sequential testing procedures, as was the case in our study, people often use less-conservative multiple testing correction methods such as the Benjamini-Hochberg procedure [32]. We chose the most stringent threshold to avoid selecting too many periodicities that are not helpful in characterizing daily activity patterns at each age.

We also compared our PML algorithm with the standard approach autocorrelation. Plots of correlation estimates against time lags are useful for identifying the correlation peak at 24 hours but not for shorter periods of rhythmicity. This is because the estimation of correlation can be biased because of the presence of multiple periodic rhythms, and the identification of multiple periodicities by simple calculation of autocorrelation may not be accurate. Therefore, the standard approach using autocorrelation is effective in confirming the most dominant 24-hour periodicity but is not as effective in identifying other periodicities, which the PML algorithm can achieve.

Other standard approaches such as periodograms and cosinor analysis were not used in this study because there are, in fact,

connections between PML and the two methods. It is important to point out their connections as well as differences. Among periodograms, the Fourier, Enright, and Lomb-Scargle periodograms are commonly used [18]. Both our PML algorithm and the Fourier periodogram use Fourier analysis to identify dominant periodicities, but the PML algorithm uses a shrinkage method in machine learning to identify dominant patterns in the population, whereas Fourier periodograms are focused on individuals to manually identify dominant periodicities based on individual plots. The Enright periodogram, although suitable for equidistant activity measurements in our scenario, may not be applicable here because it requires ≥10 days of data [18]. In addition, the estimation method only holds when there is one periodic component, but in our case, the presence of multiple periodic components may have attenuated the results [18]. The Lomb-Scargle periodogram is a modification of Fourier analysis to accommodate unevenly spaced data or missing data. As our data do not have this issue, the Lomb-Scargle periodogram is equivalent to the Fourier periodogram in our case. Compared with the PML algorithm, the Fourier periodogram involves more manual work to generate periodograms for each individual and visually identify dominant periodicities, whereas PML is more automated and more effective in studying the population as a whole and further identifying the periodicities that are characteristic of the population. In addition, researchers often use periodograms to validate the most dominant periodicity such as 24 hours but do not specifically examine information on secondary dominant periodicities or use periodic curves to reconstruct or approximate activity patterns, even though the connection between dominant frequencies or periodicities and functional curves can be made and the periodic information can be fully used. Therefore, the PML algorithm makes full use of the information from more than one dominant periodicity and links the FFT results in the frequency domain with their corresponding cosine curves to effectively characterize activity patterns.

For cosinor analysis, one may recall that the FFT results consist of real parts and imaginary parts that correspond to cosine curves and sine curves, respectively; thus, FFT is equivalent to fitting the cosine model. We fitted cosine models to the activity data with 1 to 3 cosine curves at dominant periodicities identified by the PML algorithm. Even though the estimated amplitudes for the cosine curves are different from the amplitudes in the FFT results, the Pearson correlation between the cosine coefficients and the FFT signals of the same periodicity is 1, indicating equivalence. Although the final results are equivalent, the procedures are different. For cosinor analysis, if prior knowledge is available, simple least squares methods can be used to fit the model [33]. However, if there is no prior knowledge of periodic information, the least squares method cannot be used because the dominant periodicity needs to be estimated first, and the cosinor model can no longer be linearized in its parameters. One has to either start from an initial guess and use iterative procedures to minimize the residual sum of squares or use simulated annealing alternatively to fit the model, the process of which can be exhaustive [34-36]. In comparison, without prior knowledge of the dominant periods, the PML algorithm based on shrinkage in machine learning is still easy to implement without computational burden in extracting periodic information, and the results are as effective as the cosinor model to characterize daily activity patterns using cosine curves.

In summary, the proposed PML algorithm is effective in extracting periodic information, identifying dominant periodicities, and characterizing activity patterns. In the presence of multiple periodicities, PML does not have the estimation problem that autocorrelation encounters. To identify dominant periodicities, PML uses shrinkage in machine learning methods that can help researchers avoid manual work in periodograms, which require individual plots and visual identification. PML can also characterize activity patterns by making full use of the cosine curves represented by FFT signals and avoiding the computationally intensive process of choosing and fitting cosinor models when prior knowledge of the dominant periodicities is not available.

Sleep-Activity Rhythm Characterization

Our study confirms previous findings that infants already form 24-hour sleep-wake cycles at 6 months due to entrainment by cyclical changes in the environment, whether it is due to light-dark cycles or maternal rest-activity cycles [6,9,10,37,38]. It is noteworthy that although 24-hour cycles are formed, sleep-activity patterns over the 24-hour period are not stabilized: infants often take multiple naps at different times of the day, and their daily activity patterns vary across days and across individuals.

Our study indicates that from 6 months to 12 months, the infant sleep-activity pattern gradually develops: some infants take only 1 afternoon nap, whereas others take two naps: one in the morning and one in the afternoon. Strong FFT signals at one-third–day periodicity capture two-peak, one-nap activity patterns, whereas strong FFT signals at one-fourth–day periodicity capture three-peak, two-nap patterns. The results are in line with those of previous sleep studies which indicated that the duration of nighttime sleep gradually lengthens and sleep patterns become increasingly consolidated during the first year after birth [12,39].

Although the timing for the stabilization of infant sleep-activity patterns varies across individuals, by the time infants reach the age of 18 months, their daily activity patterns have consolidated into a predominant nighttime pattern with 1 afternoon nap only, which can be obtained by combining the 3 dominant periodicities at 1 day, half day, and one-third day. The consolidation of sleep-activity patterns is confirmed by increased FFT signals at half-day and one-third–day periodicities and decreased FFT signals at other periodicities compared with previous ages. The results for 24 months remained the same as those at 18 months, showing no changes and confirming further that sleep-activity patterns are consolidated by 18 months and are stable from that age onward.

In our study, the 3-hour periodicity, normally for feeding behaviors, was not detected, and there might be 2 reasons for this. First, in the feeding guidelines for infants, the feeding of infants aged 6-8 months is advised to be 5-6 times in 24 hours, less frequent than the 3-hour (8 times) schedule, and it is advised that infants aged 12-24 months should have 3 meals with family

and have additional snacks 2-3 times [30,31]. Infants aged below 6 months may have a more frequent feeding schedule, but our activity data were collected at 6 months or later. Second, there might be desynchronization between the feeding schedules and activity patterns. Although infants aged 6 months might be fed 5-6 times per day, they do not nap or sleep 5-6 times within the same time frame. We referred to the sleep diaries recorded by the parents as a reference for napping information. Most infants aged 6 months have 1 to 2 naps in the morning and 1 nap in the afternoon. Infants aged 12 months generally have 1 or no naps in the morning and 1 nap in the afternoon. Most infants aged 18 months and 24 months have 1 nap in the afternoon. Therefore, sleep-activity patterns are desynchronized with feeding schedules because feeding behaviors might not dominate infant sleep-activity patterns at this age. For the aforementioned reasons, feeding cycles such as 3-hour periodicity were not detected in our data.

Association Between Daily Rhythms and Early Childhood Development

Using FFT signals at dominant periodicities identified by PML, we were able to find an association between the consolidation of sleep-activity rhythms and early childhood motor development. At 6 months, the association between PDMS-2 total motor scores and assessment seasons may be explained by the different number of layers of clothing worn by infants in different seasons. In winter, infants are likely to wear many layers of clothing, which may restrict their behaviors and result in suboptimal performance compared with infants wearing light clothes and taking the PDMS-2 assessment in summer. As a result, infants aged 6 months who were assessed in summer and autumn obtained relatively higher PDMS-2 scores than infants assessed in winter and spring.

At 12 months, after controlling for the assessment seasons, stronger FFT signals at half-day and one-third–day periodicities were generally associated with higher PDMS-2 scores. The period of 12 months is critical for sleep-activity rhythm consolidation, which was captured by the growing FFT signals at half-day and one-third–day periodicities. It is noteworthy that all the infants at this age had strong FFT signals at 1-day periodicity, indicating that they exhibited 1-day periodicity in their sleep-activity patterns and that their 24-hour periodic activity patterns tended to be stabilized. As a result, there was not much variation in the strength of 1-day periodicity, which might not explain much of the variability in the PDMS-2 scores among individuals. In comparison, the activity pattern over the 24-hour period was not consolidated, and the activity pattern could be characterized by the one-third–day and half-day periodicities. The larger variability in the strength of one-third–day and half-day periodicities can describe the degree to which the daily sleep-activity pattern is consolidated, which is further associated with child development as evaluated by the PDMS-2 scores. The infants with a more consolidated activity pattern tended to have better motor development evaluations. In addition, activity rhythm consolidation is strongly associated with the subcategories of locomotion and stationary, which belong to the gross motor and measure how the large muscle system is used to move from place to place or assume a stable posture when not moving. Therefore, we obtained new insights into early childhood development that the degree to which the sleep-activity pattern is consolidated at 12 months is associated with infant motor development and with large muscle system development in particular.

At 18 and 24 months, the PDMS-2 scores were not associated with the season, FFT signals, or any other variables in our data set. Most of the toddlers had stabilized daily activity patterns with strong periodic rhythmicity at this age. FFT signals as characteristics of sleep-activity rhythms were no longer associated with the PDMS-2 scores, and this is likely because the critical age at which the daily activity rhythm stabilizes had passed.

Limitations and Future Work

One limitation of our study is that we collected Actiwatch data every 6 months, and thus we were not able to capture more detailed monthly changes over the period. Future work may collect Actiwatch data in a more frequent manner, such as every 3 months or every month, to capture gradual changes in the sleep-activity rhythm during early childhood. For infants aged ≤6 months, more frequent observations can also allow us to observe how the predominant rhythm of infants changes from ultradian to circadian by adjusting to the 24-hour cycle in the environment. Another limitation of this study is that although we identified the association of sleep-activity daily rhythm consolidation with early childhood motor development and with large muscle system development in particular, the mechanism behind this association is not clear. Future work should investigate how daily rhythm consolidation and motor development interrelate and contribute to early childhood development.

Conclusions

In summary, the proposed PML algorithm provides a new method for circadian rhythm analysis and is particularly useful for studying populations whose daily patterns are not regular. In addition, the PML algorithm is applicable to other types of wearable device data in the format of a time series with a nature similar to that of actigraphy; therefore, it can be extended to other types of circadian studies using information such as body temperature, heart rate, and hormone data. Therefore, the PML algorithm can be widely applied to other wearable device studies to help characterize periodic information. Using the proposed method, our study provides novel insights into sleep-activity rhythm development in early childhood. First, in our study, the critical period for the consolidation of sleep-activity rhythms was between 6 and 18 months. This is because at 6 months, 24-hour sleep-wake cycles are formed, but their daily activity patterns are not stabilized, and by the time toddlers reach the age of 18 months, their sleep-activity patterns have consolidated into a fixed pattern with 2 activity peaks and 1 afternoon nap. The period between 6 and 18 months is critical for early childhood sleep-activity rhythm development and consolidation. Second, we identified the association between the consolidation of daily rhythms and early childhood motor development and large muscle system development in particular. This association has not been identified in previous studies. Infants with more consolidated circadian rhythms tend to have better motor development assessments. Although the mechanism is not clear,

maintaining a regular and stable sleep-activity pattern and maintaining a healthy circadian system are important for healthy physical development in early childhood.

Acknowledgments

The authors would like to thank all the parents and children who participated in this study. The authors also thank Shanghai Jiao Tong University for its support.

This study was supported by the Chinese National Natural Science Foundation (81773443), Ministry of Science and Technology of China (2016YFC1305203); Ministry of Health of the People's Republic of China (201002006); Shanghai Science and Technology Commission (17XD1402800 and 18695840200); Construction Project of Key and Weak Discipline of Shanghai Municipal Commission of Health and Family Planning (2016ZB0104); and Shanghai Jiao Tong University (YG2016ZD04).

Conflicts of Interest

None declared.

References

1. Henriksen A, Haugen MM, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, et al. Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables. J Med Internet Res 2018 Mar 22;20(3):e110 [FREE Full text] [doi: 10.2196/jmir.9157] [Medline: 29567635]
2. Sano A, Taylor S, McHill AW, Phillips AJ, Barger LK, Klerman E, et al. Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: observational study. J Med Internet Res 2018 Jun 08;20(6):e210 [FREE Full text] [doi: 10.2196/jmir.9410] [Medline: 29884610]
3. Kim J, Campbell AS, de Ávila BE, Wang J. Wearable biosensors for healthcare monitoring. Nat Biotechnol 2019 Apr;37(4):389-406. [doi: 10.1038/s41587-019-0045-y] [Medline: 30804534]
4. Guillodo E, Lemey C, Simonnet M, Walter M, Baca-García E, Masetti V, HUGOPSY Network, et al. Clinical applications of mobile health wearable-based sleep monitoring: systematic review. JMIR Mhealth Uhealth 2020 Apr 01;8(4):e10733 [FREE Full text] [doi: 10.2196/10733] [Medline: 32234707]
5. Zhang Z, Cajochen C, Khatami R. Social jetlag and chronotypes in the Chinese population: analysis of data recorded by wearable devices. J Med Internet Res 2019 May 11;21(6):e13482 [FREE Full text] [doi: 10.2196/13482] [Medline: 31199292]
6. Zornoza-Moreno M, Fuentes-Hernández S, Sánchez-Solis M, Rol MA, Larqué E, Madrid JA. Assessment of circadian rhythms of both skin temperature and motor activity in infants during the first 6 months of life. Chronobiol Int 2011 May;28(4):330-337. [doi: 10.3109/07420528.2011.565895] [Medline: 21539424]
7. Peirano P, Algarín C, Uauy R. Sleep-wake states and their regulatory mechanisms throughout early human development. J Pediatr 2003 Oct;143(4 Suppl):S70-S79. [doi: 10.1067/s0022-3476(03)00404-9] [Medline: 14597916]
8. Kennaway DJ. Programming of the fetal suprachiasmatic nucleus and subsequent adult rhythmicity. Trends Endocrinol Metab 2002 Nov;13(9):398-402. [doi: 10.1016/s1043-2760(02)00692-6] [Medline: 12367822]
9. Heraghty JL, Hilliard TN, Henderson AJ, Fleming PJ. The physiology of sleep in infants. Arch Dis Child 2008 Nov;93(11):982-985. [doi: 10.1136/adc.2006.113290] [Medline: 18653626]
10. Rivkees SA. Developing circadian rhythmicity in infants. Pediatrics 2003 Aug;112(2):373-381. [doi: 10.1542/peds.112.2.373] [Medline: 12897290]
11. Merikanto I, Lahti J, Kuula L, Heinonen K, Räikkönen K, Andersson S, et al. Circadian preference and sleep timing from childhood to adolescence in relation to genetic variants from a genome-wide association study. Sleep Med 2018 Oct;50:36-41. [doi: 10.1016/j.sleep.2018.04.015] [Medline: 29982088]
12. Guilleminault C, Leger D, Pelayo R, Gould S, Hayes B, Miles L. Development of circadian rhythmicity of temperature in full-term normal infants. Neurophysiol Clin 1996;26(1):21-29. [doi: 10.1016/0987-7053(96)81531-0] [Medline: 8657095]
13. Farhud D, Aryan Z. Circadian rhythm, lifestyle and health: a narrative review. Iran J Public Health 2018 Aug;47(8):1068-1076 [FREE Full text] [Medline: 30186777]
14. Jagannath A, Taylor L, Wakaf Z, Vasudevan SR, Foster RG. The genetics of circadian rhythms, sleep and health. Hum Mol Genet 2017 Oct 01;26(R2):R128-R138 [FREE Full text] [doi: 10.1093/hmg/ddx240] [Medline: 28977444]
15. Smith MT, McCrae CS, Cheung J, Martin JL, Harrod CG, Heald JL, et al. Use of actigraphy for the evaluation of sleep disorders and circadian rhythm sleep-wake disorders: an American academy of sleep medicine systematic review, meta-analysis, and GRADE assessment. J Clin Sleep Med 2018 Jul 15;14(7):1209-1230 [FREE Full text] [doi: 10.5664/jcsm.7228] [Medline: 29991438]
16. Bingham C, Arbogast B, Guillaume GC, Lee JK, Halberg F. Inferential statistical methods for estimating and comparing cosinor parameters. Chronobiologia 1982;9(4):397-439. [Medline: 7168995]
17. Nelson W, Tong YL, Lee JK, Halberg F. Methods for cosinor-rhythmometry. Chronobiologia 1979;6(4):305-323. [Medline: 548245]

18. Refinetti R, Lissen GC, Halberg F. Procedures for numerical analysis of circadian rhythms. Biol Rhythm Res 2007;38(4):275-325 [FREE Full text] [doi: 10.1080/09291010600903692] [Medline: 23710111]
19. Gonçalves B, Adamowicz T, Louzada FM, Moreno CR, Araujo JF. A fresh look at the use of nonparametric analysis in actimetry. Sleep Medicine Reviews 2015 Apr;20:84-91. [doi: 10.1016/j.smrv.2014.06.002]
20. Huang Y, Liu R, Wang Q, Van Someren EJ, Xu H, Zhou J. Age-associated difference in circadian sleep-wake and rest-activity rhythms. Physiol Behav 2002 Aug;76(4-5):597-603. [doi: 10.1016/s0031-9384(02)00733-3] [Medline: 12126998]
21. Ortiz-Tudela E, Martinez-Nicolas A, Campos M, Rol MA, Madrid JA. A new integrated variable based on thermometry, actimetry and body position (TAP) to evaluate circadian system status in humans. PLoS Comput Biol 2010 Nov 11;6(11):e1000996 [FREE Full text] [doi: 10.1371/journal.pcbi.1000996] [Medline: 21085644]
22. Zou H, Hastie T. Regularization and variable selection via the elastic net. J Royal Stat Soc B 2005 Apr;67(2):301-320. [doi: 10.1111/j.1467-9868.2005.00503.x]
23. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Series B Stat Methodol 2018 Dec 05;58(1):267-288. [doi: 10.1111/j.2517-6161.1996.tb02080.x]
24. Meltzer LJ, Montgomery-Downs HE, Insana SP, Walsh CM. Use of actigraphy for assessment in pediatric sleep research. Sleep Med Rev 2012 Oct;16(5):463-475 [FREE Full text] [doi: 10.1016/j.smrv.2011.10.002] [Medline: 22424706]
25. Folio MR, Fewell RR. Peabody Developmental Motor Scales: Examiner's Manual. Second Edition. Austin: Pro-ed; 2000.
26. xinyue L/PML. Github. URL: https://github.com/xinyue-L/PML [accessed 2021-09-16]
27. Li X, Kane M. PML: Penalized Multi-Band Learning for Circadian Rhythm Analysis Using Actigraphy. 2019. URL: https://CRAN.R-project.org/package=PML [accessed 2019-05-08]
28. Fisher RA. Tests of significance in harmonic analysis. In: Proceedings of the Royal Society of London Series. 1997 Jan Presented at: Proceedings of the Royal Society of London Series; 1929; London p. 54-59. [doi: 10.1098/rspa.1929.0151]
29. Dowse HB, Ringo JM. The search for hidden periodicities in biological time series revisited. J Theoretical Biol 1989 Aug;139(4):487-515. [doi: 10.1016/s0022-5193(89)80067-0]
30. Editorial Board, Chinese Journal of Pediatrics, Subspecialty Group of Child Health Care, the Society of Pediarics, Chinese Medical Association. Infant and young child feeding guideline (0 - 3 years): information for primary health workers. Zhonghua Er Ke Za Zhi 2016 Dec 02;54(12):883-890. [doi: 10.3760/cma.j.issn.0578-1310.2016.12.002] [Medline: 27938586]
31. American Academy of Pediatrics. Caring for Your Baby and Young Child: Birth to Age 5. 6th ed. New York: Bantam; Nov 4, 2014.
32. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 2018 Dec 05;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
33. Tong YL. Parameter estimation in studying circadian rhythms. Biometrics 1976 Mar;32(1):85-94. [Medline: 1276369]
34. Marquardt DW. An algorithm for least-squares estimation of nonlinear parameters. J Soc Ind Appl Math 1963 Jun;11(2):431-441. [doi: 10.1137/0111030]
35. Czaplicki J, Cornélissen G, Halberg F. GOSA, a simulated annealing-based program for global optimization of nonlinear problems, also reveals transyears. J Appl Biomed 2006 Jun;4(2):87-94 [FREE Full text] [Medline: 20414480]
36. Cornelissen G. Cosinor-based rhythmometry. Theor Biol Med Model 2014 Apr 11;11:16 [FREE Full text] [doi: 10.1186/1742-4682-11-16] [Medline: 24725531]
37. Thomas KA, Burr RL, Spieker S, Lee J, Chen J. Mother-infant circadian rhythm: development of individual patterns and dyadic synchrony. Early Hum Dev 2014 Dec;90(12):885-890 [FREE Full text] [doi: 10.1016/j.earlhumdev.2014.09.005] [Medline: 25463836]
38. Thomas KA, Burr RL, Spieker S. Maternal and infant activity: analytic approaches for the study of circadian rhythm. Infant Behav Dev 2015 Nov;41:80-87 [FREE Full text] [doi: 10.1016/j.infbeh.2015.08.002] [Medline: 26360916]
39. Ficca G, Fagioli I, Salzarulo P. Sleep organization in the first year of life: developmental trends in the quiet sleep-paradoxical sleep cycle. J Sleep Res 2000 Mar;9(1):1-4 [FREE Full text] [doi: 10.1046/j.1365-2869.2000.00172.x] [Medline: 10733682]

Abbreviations

FFT: Fast Fourier Transform
MSE: mean squared error
PDMS-2: Peabody Developmental Motor Scales-Second Edition
PML: Penalized Multiband Learning

Edited by G Eysenbach; submitted 26.02.20; peer-reviewed by J Liu, E Moon; comments to author 20.04.20; revised version received 23.04.20; accepted 13.05.20; published 14.10.21.

Please cite as:
Li X, Kane M, Zhang Y, Sun W, Song Y, Dong S, Lin Q, Zhu Q, Jiang F, Zhao H
Circadian Rhythm Analysis Using Wearable Device Data: Novel Penalized Machine Learning Approach
J Med Internet Res 2021;23(10):e18403
URL: https://www.jmir.org/2021/10/e18403
doi:10.2196/18403
PMID:34647895

Original Paper

Clustering of Hypoglycemia Events in Patients With Hyperinsulinism: Extension of the Digital Phenotype Through Retrospective Data Analysis

Chris Worth[1,2], BMedSci, MBChB; Simon Harper[2], MEng (Hons), PhD; Maria Salomon-Estebanez[1], MBBS, MPhil; Elaine O'Shea[1], Bsc (Hons); Paul W Nutter[2], BEng (Hons), MSc, PhD; Mark J Dunne[3], BSc, PhD; Indraneel Banerjee[1,3], MBBS, MD

[1]Department of Paediatric Endocrinology, Royal Manchester Children's Hospital, Manchester, United Kingdom

[2]Department of Computer Science, University of Manchester, Manchester, United Kingdom

[3]Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom

Corresponding Author:
Chris Worth, BMedSci, MBChB
Department of Paediatric Endocrinology
Royal Manchester Children's Hospital
Oxford Road
Manchester, M13 9WL
United Kingdom
Phone: 44 07837740913
Email: christophersimon.worth@mft.nhs.uk

Abstract

Background: Hyperinsulinism (HI) due to excess and dysregulated insulin secretion is the most common cause of severe and recurrent hypoglycemia in childhood. High cerebral glucose use in the early hours results in a high risk of hypoglycemia in people with diabetes and carries a significant risk of brain injury. Prevention of hypoglycemia is the cornerstone of the management of HI, but the risk of hypoglycemia at night or the timing of hypoglycemia in children with HI has not been studied; thus, the digital phenotype remains incomplete and management suboptimal.

Objective: This study aims to quantify the timing of hypoglycemia in patients with HI to describe glycemic variability and to extend the digital phenotype. This will facilitate future work using computational modeling to enable behavior change and reduce exposure of patients with HI to injurious hypoglycemic events.

Methods: Patients underwent continuous glucose monitoring (CGM) with a Dexcom G4 or G6 CGM device as part of their clinical assessment for either HI (N=23) or idiopathic ketotic hypoglycemia (IKH; N=24). The CGM data were analyzed for temporal trends. Hypoglycemia was defined as glucose levels <3.5 mmol/L.

Results: A total of 449 hypoglycemic events totaling 15,610 minutes were captured over 237 days from 47 patients (29 males; mean age 70 months, SD 53). The mean length of hypoglycemic events was 35 minutes. There was a clear tendency for hypoglycemia in the early hours (3-7 AM), particularly for patients with HI older than 10 months who experienced hypoglycemia 7.6% (1480/19,370 minutes) of time in this period compared with 2.6% (2405/92,840 minutes) of time outside this period (P<.001). This tendency was less pronounced in patients with HI who were younger than 10 months, patients with a negative genetic test result, and patients with IKH. Despite real-time CGM, there were 42 hypoglycemic events from 13 separate patients with HI lasting >30 minutes.

Conclusions: This is the first study to have taken the first step in extending the digital phenotype of HI by describing the glycemic trends and identifying the timing of hypoglycemia measured by CGM. We have identified the early hours as a time of high hypoglycemia risk for patients with HI and demonstrated that simple provision of CGM data to patients is not sufficient to eliminate hypoglycemia. Future work in HI should concentrate on the early hours as a period of high risk for hypoglycemia and must target personalized hypoglycemia predictions. Focus must move to the human-computer interaction as an aspect of the digital phenotype that is susceptible to change rather than simple mathematical modeling to produce small improvements in hypoglycemia prediction accuracy.

(J Med Internet Res 2021;23(10):e26957) doi:10.2196/26957

KEYWORDS

hyperinsulinism; continuous glucose monitoring; digital phenotype; hypoglycemia; nocturnal hypoglycemia

Introduction

Background

Hyperinsulinism (HI) is a diverse collection of disorders united by the pathology of inappropriate insulin secretion causing hyperinsulinemic hypoglycemia with simultaneous suppression of alternative fuel sources. It has an estimated incidence of 1:28,000 in the United Kingdom [1] and is the most common cause of severe and recurrent hypoglycemia in childhood. This recurrent hypoglycemia, with the corresponding suppression of ketones as an alternative fuel source, results in brain damage in an unacceptably high proportion of cases, up to 48% [2]. The risk of brain damage is independent of the chronicity of the disease [3], as damage often occurs early in life when the neonatal brain is highly susceptible to such insults [4,5].

Significant progress in the understanding of the underlying pathophysiology of HI has been made since its first detailed description in 1953 [6]. Increased knowledge of changes at the organic [7], cellular [8], and genetic [9,10] levels has led to improvements in the care of these patients, and recent studies even suggest a lowering of the subsequent rate of brain injury [11]. As first suggested by Richard Dawkins in 1976, there is an “extended phenotype” of all conditions, not just limited to observable physical traits or cellular changes [12]. The most recent extension of this is that of a digital phenotype, encompassing aspects of patients’ behaviors related to and measured by technology [13]. The digital phenotype includes everything from interactions with others on social media to digitally collected location data and continuously measured physiological parameters, such as glucose levels and heart rate. These measures sit alongside the traditional characterization of diseases to form a more comprehensive picture and facilitate a more nuanced approach to management. Current management of HI is complex, balancing the risks and benefits of a limited repertoire of medications, all of which have a small therapeutic window and significant side effect profiles [14]. Because of the dysregulated secretion of insulin in these patients, hypoglycemic events are often very difficult to predict. The standard of care for home monitoring of hypoglycemia is intermittent fingerpick testing for blood glucose. However, patients rarely achieve more than three to four such measurements per day and this infrequent testing strategy risks missing hypoglycemia between tests, particularly overnight [15]. This practice also offers little in the way of disease characterization and does very little to extend the phenotype or the scientific understanding of HI.

In recent years, continuous glucose monitoring (CGM) has emerged as an alternative, offering insight into glucose trends. CGM measures subcutaneous glucose at frequent intervals over extended periods (7-10 days) to provide glycemic phenotypes in patients with hypoglycemia and diabetes and contributes to the digital phenotype [16]. The application of CGM may not yet be readily applicable in patients with HI as a reliable means of hypoglycemia detection or prediction [17]. However, as CGM is a passive form of monitoring, it can record data at a high granularity with very minimal response burden on users [18]; therefore, it has the potential to collect detailed glycemic data while being acceptable to patients on a long-term basis. There have been a limited number of studies describing the utility of CGM in patients with HI [19-22], and none have described the timing of hypoglycemia events or glycemic trends.

There is good empirical evidence [23] to suggest that hypoglycemic events may not be evenly distributed throughout the day and that the risk of hypoglycemia may be disproportionately higher during periods of reduced food intake. In all but the youngest children, each day is divided into two distinct phases, one of activity and eating and the other of fasting and rest [23]. It is well established that hormones such as cortisol vary throughout the day, with peak levels varying in relation to the time of sunrise [24]. Glucose homeostasis also varies with the time of day [25], but this variability is not likely to be directly related to cortisol [26,27]. Rather, there is likely a direct circadian control involving the suprachiasmatic nucleus in the hypothalamus [28,29] and peripheral clock-gene–regulated components in the pancreas [23].

Before waking, early morning is a high-risk period for hypoglycemia, as glucose use is at its highest [23]. Normally, this is counteracted by a high rate of gluconeogenesis [23]; however, in patients with HI, this is suppressed owing to inappropriate secretion of insulin, resulting in an imbalance of glucose homeostasis weighted toward hypoglycemia [30]. The high risk in early morning is exacerbated by the time-independent, high-glucose requirement of the brain during late sleep [31].

Despite over 30 years of investigation into nocturnal hypoglycemia in children with diabetes [32], the estimated incidence remains as high as 68% [33]. More than half of all severe hypoglycemic events occur overnight [34], and up to 18% of deaths in younger patients with diabetes are attributed to nocturnal hypoglycemia [35].

The potential for CGM data to contribute to the digital phenotype of HI has not yet been investigated. Despite the physiological and empirical evidence from healthy participants and participants with diabetes regarding the risks of nocturnal hypoglycemia, no study has investigated the timing of hypoglycemic events in children with HI. It is important to identify periods of greater risk of hypoglycemia to design targeted detection, prediction, and prevention strategies, as traditional medical management techniques do not allow for this. Hall et al [36], Colas et al [37], and Lunt et al [38] used CGM as a short-term phenotyping tool to better understand patient profiles and categorize risks. Larkin et al [39] detailed their intention to use CGM as part of a long-term phenotyping tool on a large scale to provide personalized insights into disease.

Digital phenotyping allows for two important changes in disease management. First, the detailed analysis of CGM data provides an extension of the digital phenotype for both the disease and

the individual and allows for the targeting of interventions to times when they will achieve the optimum effect. Second, patients and parents are not passive bystanders in the management of HI, and an analysis of how they interact with and respond to the technology further extends the digital phenotype [40], as well as enabling and enhancing behavior change [41]. Knowledge of an extended digital phenotype will not, in and of itself, improve outcomes but does improve understanding of how future interventions can be adapted to achieve the most significant and lasting behavior change [42].

Objective

In this study, we take the first step in extending the digital phenotype of HI by describing glycemic trends and identifying the timing of hypoglycemia measured by CGM. The following findings provide a basis for future work concentrating on using the newly extended digital phenotype with human-computer interactions and ultimately altering care behaviors of parents to reduce the incidence of damaging hypoglycemia. Our original code is provided as an appendix (Multimedia Appendix 1) and is freely available on GitHub [43].

Methods

Recruitment

Patients were recruited between July 2017 and October 2020. Three distinct groups of patients were enrolled in the study: (1) those with a diagnosis of HI who were in hospital for acute management (n=6), (2) those in an outpatient setting with a diagnosis of HI and for whom glycemic control was suboptimal (n=17), and (3) those with a diagnosis of idiopathic ketotic hypoglycemia (IKH) for whom glycemic control was unstable (n=24). The IKH group was selected to investigate CGM profiles in an alternative clinical model of hypoglycemia that does not involve excess insulin secretion. All patients underwent CGM with the primary intention of better understanding their glucose control for clinical purposes. All patients were approached by the research team to seek consent to use anonymized CGM data for research purposes, as per a local research ethics protocol (REC/H1010/88).

Within the HI group, further subcategorization was done as either diffuse or focal disease [44]. Focal HI is characterized by the formation of a focal lesion within the pancreas comprising hyperfunctioning islets and is potentially curative by focal lesionectomy. Diffuse HI implies some specific histopathological features but practically implies nonfocal HI in which medical therapy should be prioritized over surgery when possible.

Patient recruitment to the study was based on a pragmatic design in the absence of previous studies to determine sample size in a rare disease. Previous studies investigating CGM in patients with HI recruited 11 to 15 patients [19,21]. We recruited all patients undergoing CGM for clinical reasons over a 3-year period.

Data collection

Patients undergoing inpatient monitoring had their CGM device (Dexcom G4 or G6 depending on the date) attached between 1 and 5 days before pancreatic surgery (lesionectomy or subtotal pancreatectomy) and removed after plasma glucose levels stabilized, reducing the need for frequent monitoring. All other patients were brought to the Royal Manchester Children's hospital to have their CGM device attached by a specialist nurse, after which they returned home for the remainder of the monitoring period. The patients returned CGM devices to the department at the end of the monitoring period. CGM devices were always inserted in the daytime to ensure that full calibration had occurred before the evening and data collected overnight were reliable. Data were collected for between 4 and 10 days from each patient.

For patients who underwent a controlled fast in hospital (n=0 for HI, n=2 for IKH) during CGM, data during the fasting period were deleted from the analysis and the only data used were those acquired from home monitoring after discharge from hospital. Dexcom G4 devices were used from the beginning of the study period until March 2019, from which point all patients were monitored using a Dexcom G6 device. All devices were unblinded so that parents and staff could see glucose values in real time and alarms would sound if glucose levels dropped (or was predicted to drop) below 3.5 mmol/L for those with HI and 3.3 mmol/L for those with IKH (as per lowest allowable device settings). The fall rate and urgent low soon alarms were set to *on* when the device was given to the patient. Patient modifications of the alarms were not routinely investigated.

Data Analysis

Hypoglycemia was defined as any glucose value <3.5 mmol/L (63 mg/dL) as a safe cutoff used by most specialized centers for the everyday management of HI [14]. Dexcom CGM devices report a glucose value every 5 minutes, and therefore, hypoglycemia events were measured at 5-minute intervals with a minimum duration of 5 minutes. The term *early hours* is used throughout the manuscript to refer to the period 3 AM to 7 AM.

Data were downloaded from CGM devices to Dexcom CLARITY database and raw data were downloaded for analysis using Python 3.8. We analyzed the total number of hypoglycemia events by start time irrespective of length and separately the total amount of time spent in hypoglycemia within each hour period.

As data were not normally distributed, nonparametric tests (Mann–Whitney *U*) were used to assess for differences between continuous variables. Chi-square tests were used to assess differences in proportions or percentages between the groups. The results provided are raw test statistics and associated *P* values.

Glucose testing using alternative methods (plasma glucose measured by point-of-care testing or handheld home glucose monitoring) was not routinely performed alongside CGM because assessment of CGM accuracy was not the intention or focus of this study. Correlations between CGM-derived subcutaneous glucose and plasma glucose have been previously reported in patients with HI [19,21].

Results

Overview

Baseline data showed a male predominance (29 males, 18 females) and mean age in months at the time of CGM was higher in those with IKH than in those with HI (82, SD 43 vs 57, SD 61; *P*=.03), as expected, given that HI is prevalent at a younger age than IKH. At the time of CGM testing, the mean time in months from diagnosis did not differ significantly between those with IKH and HI (33, SD 34 vs 49, SD 53; *P*=.43) (Table 1). A total of 28 patients underwent monitoring with a Dexcom G4 device and 19 with a Dexcom G6.

Table 1. Demographics of hyperinsulinism and idiopathic ketotic hypoglycemia groups.

Demographics	Hyperinsulinism (N=23)	IKH[a] (N=24)	*P* value[b]
Male, n (%)	17 (74)	12 (50)	.09
Female, n (%)	6 (26)	12 (50)	N/A[c]
Age (months), mean (SD)	57 (60)	82 (43)	.03
Time since diagnosis (months), mean (SD)	49 (52)	33 (44)	.43

[a]IKH: idiopathic ketotic hypoglycemia.

[b]*P* value for difference between groups calculated via chi square test for sex and Mann–Whitney *U* test for continuous values.

[c]N/A: not applicable.

Demographic data demonstrate that the only difference between groups at baseline was that patients with IKH had a higher mean age at time of CGM.

A total of 449 hypoglycemic events (189 in HI and 260 in IKH) were captured over 237 days. The time spent in hypoglycemia was 15,610 out of a total of 342,355 minutes (4.6%). The mean duration of hypoglycemic events was 35 (SD 57) minutes and was longer in those with IKH than in those with HI (40 vs 28). Mean lowest glucose per hypoglycemia event was 3.1 (SD 0.37) mmol/L. The mean number of hypoglycemic events per patient was 9.5 (SD 9.6) with a positive skew to the distribution (Figure 1), illustrating the small number of patients with a very large number of hypoglycemic events. In patients with HI, there were 42 hypoglycemia events lasting more than 30 minutes from 13 separate patients. The mean duration of such prolonged hypoglycemia events was 79 (SD 72) minutes, with a mean lowest glucose of 2.8 mmol/L. The characteristics of patients with HI are shown in Table 2.

Figure 1. Frequency plot of number of hypoglycemia events (hypos) per patient. This plot demonstrates the positive skew to the distribution of hypoglycemic events. Mean number of hypoglycemia events was 9.5 per patient, and median was 6.0 with 6 patients having no episodes of hypoglycemia and 1 patient having more than 30 separate hypoglycemic events. The majority of patients had 5 to 20 hypoglycemic events.

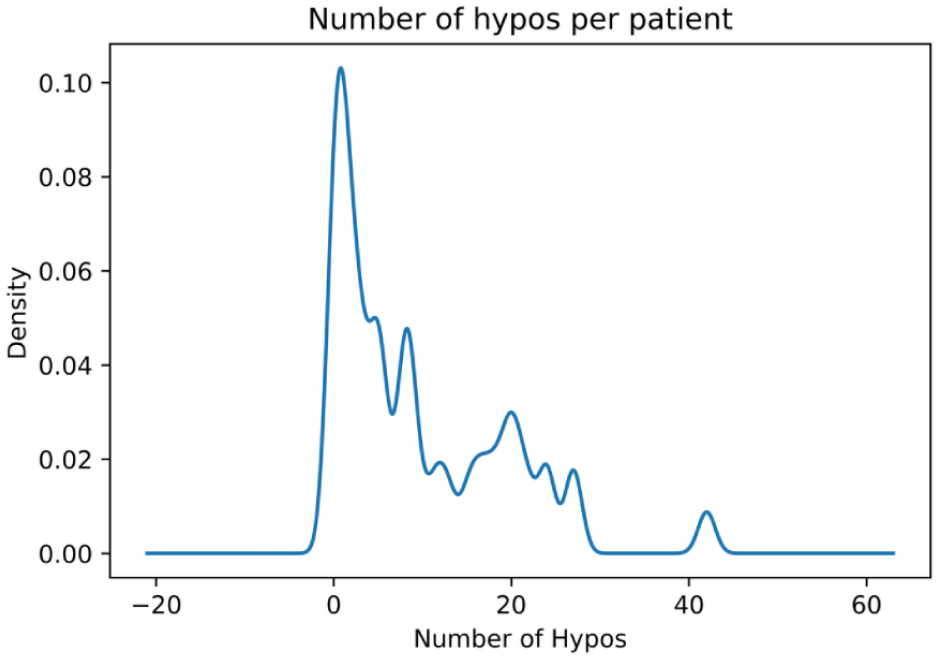


Table 2. Characteristics of patients with hyperinsulinism included in the study.[a]

	Sex	Mutation	Type	Surgery	Age (months)	Location	Rx[b]
1	Male	Not done	Diffuse	None	88	Outpatient	None
2	Male	Not done	Diffuse	None	8	Outpatient	None
3	Female	Negative	Diffuse	None	92	Outpatient	None
4	Male	*SLC16A1*	Diffuse	None	34	Outpatient	CHO[c]
5	Male	*SLC16A1*	Diffuse	None	37	Outpatient	CHO[c]
6	Female	Negative	Diffuse	None	119	Outpatient	None
7	Male	*ABCC8*	Diffuse	None	58	Outpatient	Diazoxide
8	Male	*GCK*	Diffuse	None	190	Outpatient	Diazoxide
9	Male	*ABCC8*	Diffuse	Subtotal[d]	141	Outpatient	None
10	Male	*ABCC8*	Diffuse	Subtotal[d]	132	Outpatient	None
11	Male	*ABCC8*	Focal	Lesionectomy	2	Inpatient	Octreotide
12	Male	*ABCC8*	Diffuse	None	36	Outpatient	None
13	Female	*ABCC8*	Diffuse	Subtotal[d]	51	Outpatient	Octreotide
14	Male	*ABCC8*	Focal	Lesionectomy	36	Outpatient	Octreotide
15	Female	Negative	Diffuse	None	17	Outpatient	Diazoxide
16	Female	*ABCC8*	Focal	Lesionectomy	3	Inpatient	Octreotide
17	Male	Negative	Diffuse	None	63	Outpatient	Diazoxide
18	Male	Negative	Diffuse	None	10	Outpatient	Diazoxide
19	Male	*ABCC8*	Diffuse	Subtotal[d]	193	Outpatient	None
20	Male	*ABCC8*	Focal	Lesionectomy	1	Inpatient	None
21	Male	*ABCC8*	Focal	Lesionectomy	3	Inpatient	Octreotide
22	Female	*ABCC8*	Focal	Lesionectomy	1	Inpatient	Octreotide
23	Male	*ABCC8*	Focal	Lesionectomy	3	Inpatient	Octreotide

[a]Important characteristics relating to hyperinsulinism and continuous glucose monitoring are listed for all hyperinsulinism patients individually.

[b]Rx: medical treatment at the time of continuous glucose monitoring.

[c]CHO: carbohydrate supplementation in feeds.

[d]Subtotal: subtotal pancreatectomy.

Timings of Hypoglycemia in Patients With HI

Figure 2 presents the number of hypoglycemic events by start time in patients with HI (n=23). This does not account for the duration of hypoglycemia episodes and is only representative of the hour in which the event started. There was a higher risk of a hypoglycemia event beginning in the later part of the night/early morning compared with the rest of the day. Figure 3 illustrates the percentage of time spent in hypoglycemia by patients with HI within each hour of the day. There was an increase in the prevalence of hypoglycemia in the early hours (3 AM-7AM), 6.4% (1665/25,875 minutes) of this time hypoglycemic compared with only 2.9% (3585/123,490 minutes) of time outside this period (χ^2_1=98.4, P<.001). This represents a doubling of risk during this period. Although the frequency of hypoglycemia was greater in the early hours, the mean duration (minutes) of individual hypoglycemic episodes in this period was the same as the rest of the day (28.2 minutes vs 28.1 minutes, P=.99).

Figure 2. Number of hypoglycemic events (hypos) plotted by start time in patients with HI. The X-axis represents hours of each 24 hour period. Bars represent the number of hypoglycemic events starting at any particular point in the day but do not indicate the duration of each episode. What is demonstrated is the increased number of hypoglycemic episodes starting in the later hours of the night and early morning (black) compared with the rest of the day (grey). HI: hyperinsulinism.

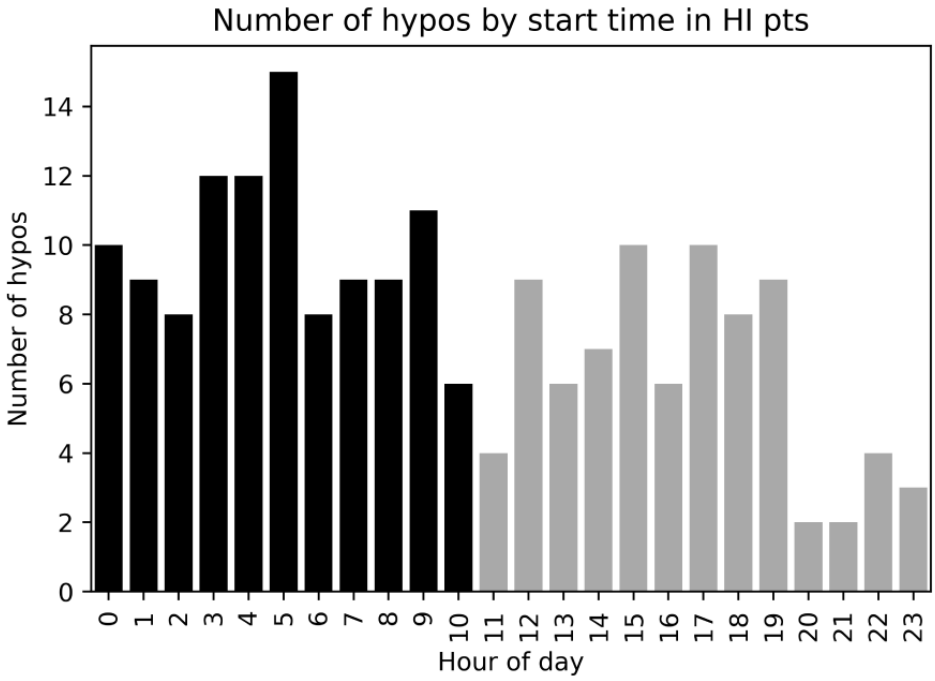
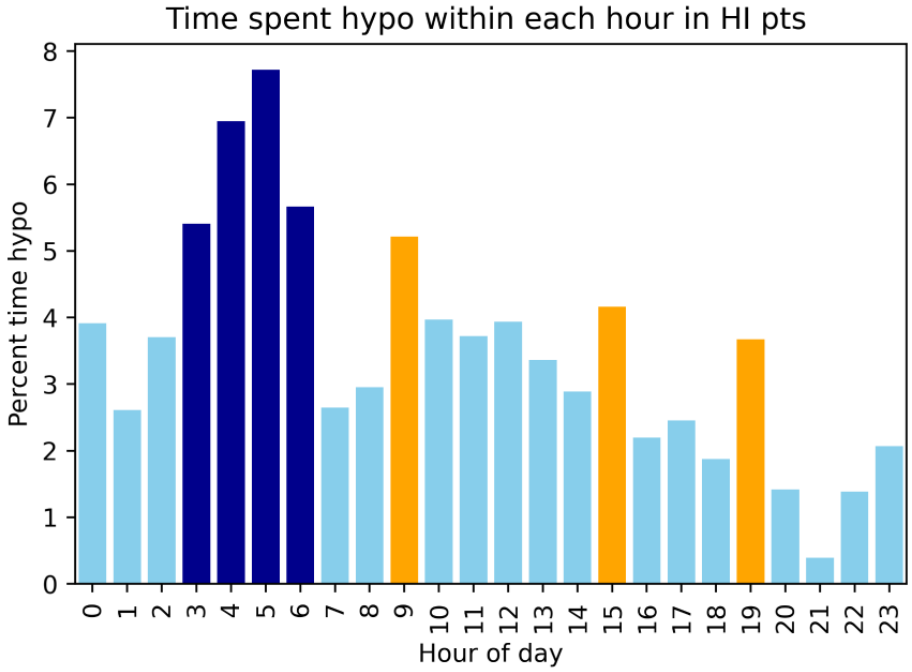


Figure 3. Percentage time spent in hypoglycemia by hour of the day in patients with HI.There is a clear period of high risk for hypoglycemia between 3 AM and 7 AM (dark blue) which represents the early hours. There are also three distinct spikes of increased hypoglycemia prevalence at 9 AM, 3 PM and 7 PM (orange), which may represent postprandial hypoglycemia. HI: hyperinsulinism.



The other periods of increased risk were the separate hours of 9 AM, 3 PM, and 7 PM, within which there was a higher proportion of minutes spent hypoglycemic than the rest of the daytime/evening (7 AM to midnight), that is 5.2% (320/6140 minutes), 4.2% (250/6015 minutes), and 3.7% (230/6265 minutes), respectively, versus 2.8% (2915/103,910 minutes; P<.001).

Subgroup Analysis of Patients With HI

Analysis of time spent in hypoglycemia by patients with HI above (n=16) and below (n=7) the age of 10 months (ie, the age at which a weaning diet with solid food is well established) demonstrated that the risk of early hours of hypoglycemia was even more pronounced in the group above the age of 10 months (Figure 4). Within the early hours, 7.6% (1480/19,370 minutes)

of the time was hypoglycemic compared with 2.6% (2405/92,840 minutes) of time outside of this period (χ^2_1=146.4, *P*<.001), indicating an almost trebling risk of hypoglycemia. In the group below the age of 10 months, no obvious patterns of hypoglycemia were visible with risk of hypoglycemia distributed randomly throughout the day, except for an unexplained peak of risk at 7 PM (Figure 5).

Figure 4. Percentage time spent hypoglycemic by hour of the day in patients with HI > 10 months of age. Analysis of timings of hypoglycemia in this subgroup show a greater tendency to early hours hypoglycemia between the hours of 3 AM and 7 PM (dark blue) with a persistence of spikes in hypoglycemia risk at 9 AM and 3 PM (orange). No spike is observed at 7 PM in contrast to the analysis for all ages. HI: hyperinsulinism.

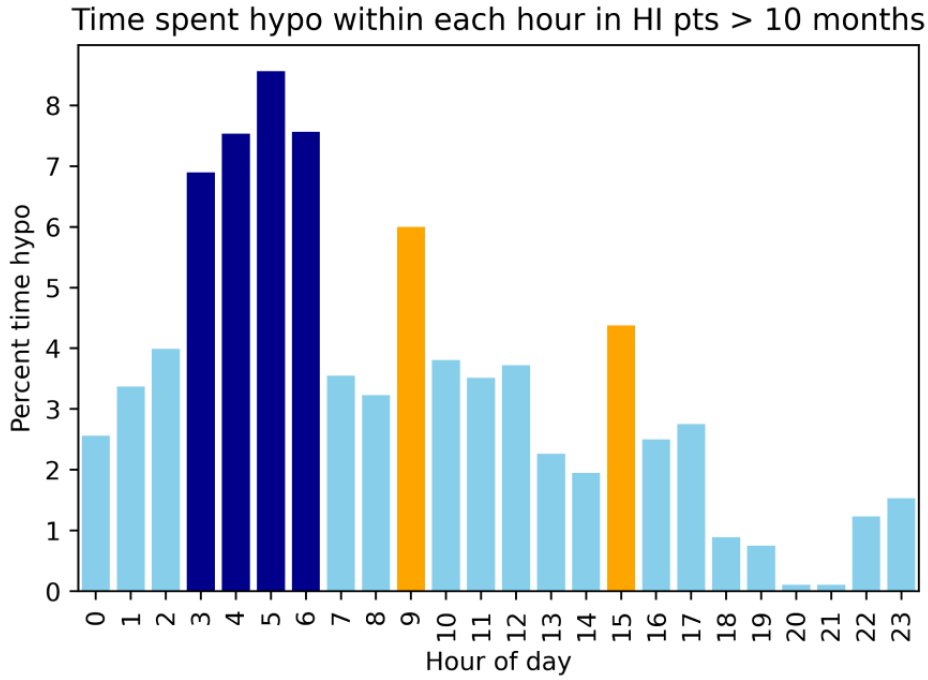
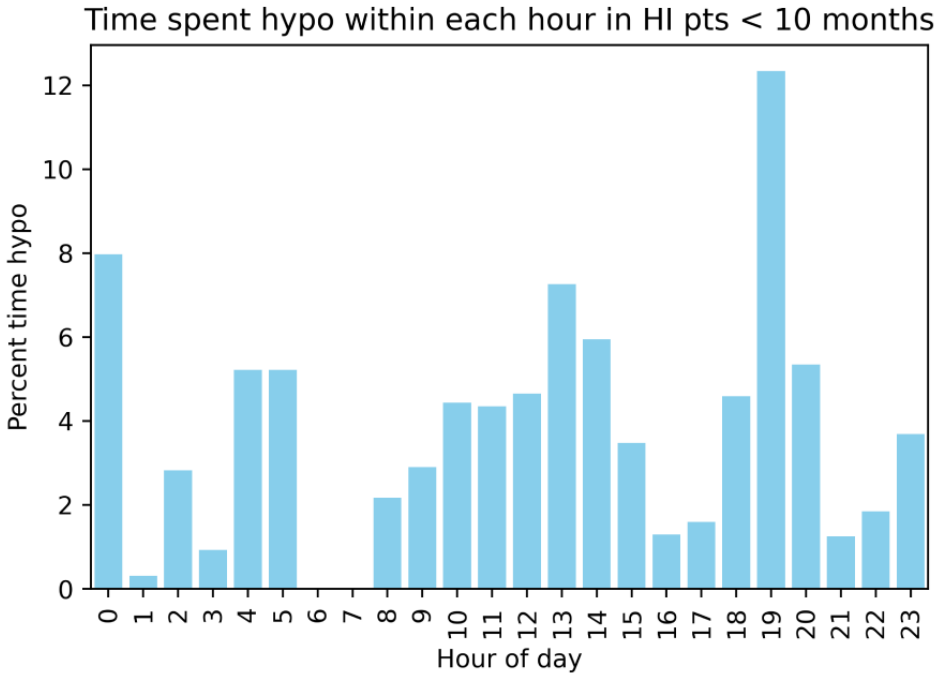


Figure 5. Percentage time spent hypoglycemic by hour of the day in patients with HI < 10 months of age. Analysis of timings of hypoglycemia in this subgroup are limited by numbers but clearly show a very different pattern of hypoglycemia compared with the group > 10 months of age. There is no obvious pattern of hypoglycemia and no obvious periods of higher risk. HI: hyperinsulinism.



Further comparisons were performed between the following HI subgroups: mutation-positive versus mutation-negative and medication versus off-medication. The results are summarized in Table 3 and reported in more detail (along with Figures S1-S4) in Multimedia Appendix 2: further HI subgroup comparisons.

Table 3. Comparisons of timings of hypoglycemia between hyperinsulinism (HI) subgroups.[a]

	Time hypoglycemic in early hours (minutes)		Time hypoglycemic outside early hours (minutes)		Minutes hypo contained in early hours (expected 16.7%; minutes)		Period of risk (compared with all HI)
	Total, N	Value, n (%)	Total, N	Value, n (%)	Total, N	Value, n (%)	
All HI	25,875	1665 (6.43)	123,490	3585 (2.9)	5250	1665 (31.71)	Early hours
Age (months), mean (SD)							
>10	19,370	1480 (7.64)	92,840	2405 (2.59)	3885	1480 (38.09)	Early hours ++[b]
<10	6505	185 (2.84)	37,150	1365 (3.85)	1365	185 (13.55)	Evenly distributed
HI causing mutation							
Positive	17,005	1135 (6.67)	81,810	2290 (2.79)	3425	1135 (33.14)	Early hours +[c]
Negative	8870	530 (5.98)	41,680	1295 (3.11)	1825	530 (29.04)	Early hours –[d]
Medication							
On	14,415	790 (5.48)	69,485	2395 (3.45)	3185	790 (24.8)	Early Hours ––[e]
Off	11,460	875 (7.64)	54,005	1190 (2.2)	2065	875 (42.37)	Early Hours ++

[a]Detailed is the difference in percentage time hypoglycemic in the early hours (3 AM-7 AM). Also reported is the percentage of all hypoglycemia minutes spent in the early hours as a comparison to the expected 16.7% that would be seen if hypoglycemia was distributed evenly. The table demonstrates an exaggerated tendency to hypoglycemia at early hours in those above the age of 10 months, off medication, and with a known hyperinsulinism causing mutation.

[b]++ is used to denote a very strong tendency to early hours hypoglycemia.

[c]+ is used to denote a strong tendency to early hours hypoglycemia.

[d]– is used to denote a relatively weak tendency to early hours hypoglycemia.

[e]–– is used to denote a weak tendency to early hours hypoglycemia.

Timings of Hypoglycemia in Those With IKH

The timing of hypoglycemia in patients with IKH showed a more evenly distributed pattern than in those with HI. The number of hypoglycemic episodes starting overnight was slightly higher than that during the day (Figure 6), but not as markedly as in the patients with HI. Total minutes spent hypoglycemic were also more evenly distributed throughout the day in patients with IKH than in those with HI (Figure 7). There was no clear period of higher than average risk of hypoglycemia; rather, a period of relatively low risk was observed in the evening and early night (6 PM-1 AM). Within this period, the risk of hypoglycemia was 2.8% (1575/57,195 minutes) compared with 6.5% (8785/135,805 minutes) outside of this period (X^2_1=132.2, P<.001).

Subgroup analysis of patients with IKH was not performed, as no patients within this group were below the age of 10 months and this group was primarily analyzed as a comparison group for those with HI.

Figure 6. Number of hypoglycemic events (hypos) plotted by hour of the day in patients with IKH. This demonstrates the risk of a hypoglycemic event starting at any particular point in the day but does not account for the length of this episode. An increased number of hypoglycemic episodes starting in the later hours of the night and early morning (dark blue) is observed, compared with the rest of the day (light blue). This, however, is less pronounced than in those patients with HI. HI: hyperinsulinism; IKH: idiopathic ketotic hypoglycemia.

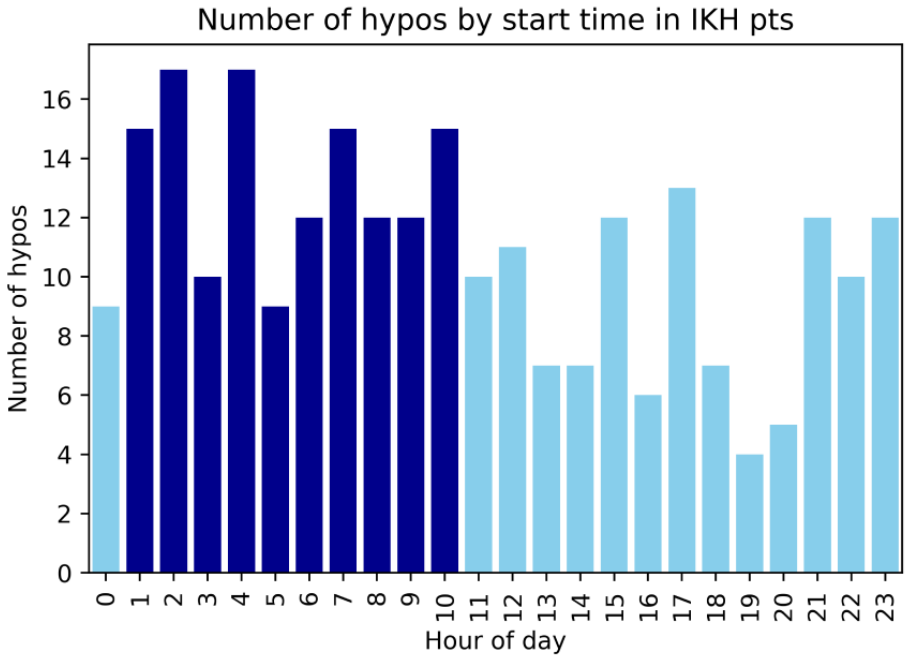
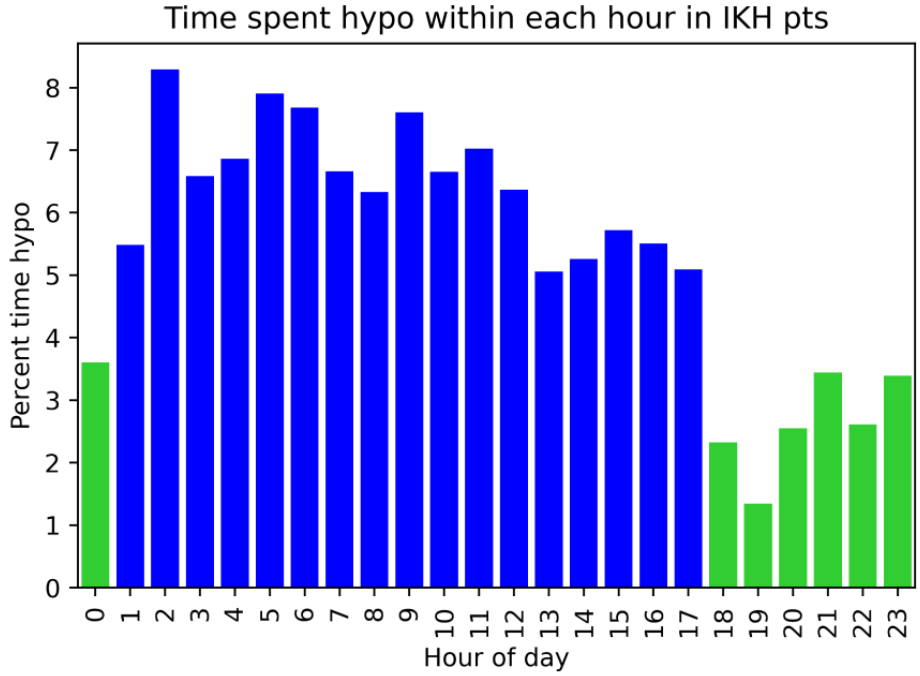


Figure 7. Percentage time hypoglycemic by hour of the day in patients with IKH. There is no period of particularly high risk as seen in patients with HI. In contrast, a short period of lower than average risk in the evening and early night (green) can be observed. HI: hyperinsulinism; IKH: idiopathic ketotic hypoglycemia.



Altering Thresholds for Hypoglycemia

As the definition of hypoglycemia differs between countries, regions, and even hospital departments, we performed further analysis by altering the threshold for hypoglycemia. The cutoffs of 3.9 mmol/L, 3.5 mmol/L, and 3.0 mmol/L were chosen as commonly used definitions for hypoglycemia in children [44,45]. Further analysis of values below this were deemed unlikely to be meaningful as they are rare events, and the Dexcom CGM devices only measure glucose down to 2.2 mmol/L, below which values are reported as *Low*.

Data for HI patients above the age of 10 months were compared with those of patients with IKH, as these 2 groups were more comparable in terms of age. As hypoglycemia threshold was reduced from 3.9 mmol/L to 3.5 mmol/L and then 3.0 mmol/L,

the tendency to hypoglycemia during early hours was emphasized in patients with HI. In patients with HI, the early hours contained 41% of all minutes (63/152 minutes) spent with a glucose level <3.0 mmol/L despite this period representing only 16.7% of the 24-hour period. A similar trend began to emerge in patients with IKH when the hypoglycemia threshold was reduced, and the proportion of all minutes spent hypoglycemic that lay within the early hours increased from 21% (842/3998 minutes) to 27% (157/577 minutes). Figures S5 and S6 provide a visualisation of this change and are provided in Multimedia Appendix 3.

Discussion

Principal Findings

We have provided a novel analysis of the timing of hypoglycemic events in patients with HI using CGM data. Our data provide new and clinically useful information to extend the digital phenotype of HI. This aspect of the newly described digital phenotype demonstrates a tendency for hypoglycemia during the early hours (3 AM-7 AM) where the risk is 2 to 3 times higher than that at other times of the day. The relative risk of nocturnal hypoglycemia compared with daytime hypoglycemia was greater in patients with HI than in those with IKH. This risk is greater still in patients with HI above the age of 10 months, those with genetic mutations, and those off medication.

A strength of this study is the novel examination of a glycemic phenotype using CGM in patients with HI while using an alternative model of hypoglycemia in children with IKH as a contrasting paradigm, demonstrating the specificity of glycemic profiles within each condition's digital phenotype. Although home blood glucose monitoring remains the standard of care for monitoring in patients with HI, we have used high-granularity glucose data to expand the phenotype of HI and highlight an important role for CGM, that is, describing nocturnal glycemic status in real time with high ecological validity. Alternative strategies to identify periods of hypoglycemia on the basis of parent interviews would be open to recall bias and unable to identify unexpected hypoglycemia, as self-monitoring of blood glucose is rarely performed overnight. With increasing refinements in CGM technology and the increasing popularity of the use of CGM in children with hypoglycemia, our study highlights the need for a targeted application of CGM. Although previous studies have used simplistic correlation methods to test the accuracy of CGM in the detection of hypoglycemia vis-à-vis home blood glucose monitoring, our study has investigated a deeper phenotype with significant clinical impact.

Our study is exploratory, as a similar analysis of CGM data has never been attempted. Therefore, it is not possible to validate the strength of our observations, except that the glycemic phenotype is replicable across the whole group with HI, in contrast to an alternative model of hypoglycemia in IKH. The number of patients with HI was large for a rare disease, and the total number of measurements in the data set adds strength to the rigor of the study.

It is not possible to investigate the cause of early hours hypoglycemia from the design of our study, nor was this the intended purpose. We speculate that the unavailability of carbohydrates due to the nocturnal fasting period, high glucose demand in the brain at this time [23,31], and the suppression of counter-regulatory gluconeogenesis in patients with HI [30] could be the probable causes. The latter view is supported by the observation of a reduced tendency to early hours hypoglycemia in patients with IKH and an increased tendency in those positive for genetic mutations known to cause HI. Patients with IKH do not have underlying metabolic disturbances and are therefore capable of mounting adequate counter-regulatory responses. In contrast, those positive for HI mutations tend to have more severe disease [11] and less ability to mount counter-regulatory responses. Further investigation of the metabolic and counter-regulatory hormonal milieu in patients with HI may be required to refine specific causation. It is not clear why patients negative for mutations had a relatively high incidence of hypoglycemia between 9 PM and 1 AM, and further work will be required to investigate this apparent trend.

We observed a tendency for early hours hypoglycemia in patients with HI above the age of 10 months (Figure 4); this may be explained by older children sleeping longer and requiring less frequent feeds through the night. We did not record sleep-wake cycles in our study to confirm or refute this possibility. However, children below the age of 10 months were also more likely to be inpatients during the period of monitoring, and the lack of early hours hypoglycemia tendency may reflect 24-hour nursing care and a reduction in ecological validity. The reason for the significant increase in time spent in hypoglycemia between 7 PM and 8 PM is unclear and may reflect the small number of patients from whom this sample was taken. Further data are required to validate or refute this finding.

Patients with HI who were no longer receiving HI medications showed a clear tendency toward hypoglycemia during early hours seen in other groups. This response was not observed in patients receiving medications for HI (diazoxide or octreotide). This may reflect the efficacy of HI medications in preventing nocturnal hypoglycemia but may also be simply reflective of the much higher median age in the off-treatment group and the subsequent effects described above. Because this study was completed, it was not possible to assess the relative impacts of age and medication on the timing of hypoglycemia in patients with HI, and further work is required.

Tendency to hypoglycemia was shown to reduce throughout the day in patients with HI, but with small increase in risk at 9 AM, 3 PM, and 7 PM (Figure 3). These periods of increased risk are likely to correlate with postmeal times in those having three meals a day, suggesting possible postprandial hypoglycemia secondary to the hyperinsulinemic response to food sometimes observed in patients with HI [46]. This trend was not observed in patients with IKH who had fasting rather than postprandial hypoglycemia. This was not observed in patients with HI below the age of 10 months who would have been receiving a high proportion of caloric intake as milk feeds distributed more frequently throughout the day and night.

As hypoglycemia thresholds were reduced from 3.9 mmol/L to 3.5 mmol/L and 3.0 mmol/L, the number of hypoglycemic events and total minutes spent hypoglycemic also reduced significantly, as reflected in the differing Y axes in Figures S5 and S6 (Multimedia Appendix 3). This is unsurprising, as patients with HI are told to maintain glucose levels >3.5 mmol/L and those with IKH >3.0 mmol/L. Reducing the hypoglycemia threshold emphasized the tendency for early hours hypoglycemia in those with HI and allowed it to become apparent in those with IKH. These likely better reflect the true hypoglycemic events that would be acted on by parents.

There were 42 hypoglycemia events lasting more than 30 minutes in patients with HI despite them wearing an unblinded CGM device programmed to alarm at glucose levels <3.5 mmol/L and parents being aware of the importance of keeping glucose levels ≥3.5 mmol/L at all times. This high prevalence of prolonged and potentially dangerous hypoglycemia events may reflect CGM inaccuracy, and these events may represent false positives not acted on by parents. However, this is unlikely to explain all the prolonged hypoglycemic events.

Although CGM provides vital information, the volume of data provided by continuous monitoring such as CGM can be overwhelming for both health care professionals and patients [47], and simple detection and reporting of glucose values will not be sufficient to eliminate all hypoglycemia. It is well recognized that alarm fatigue is a significant problem [48] and that only 37% of parents will wake to hypoglycemia alarms that sound overnight [49]. Other behavioral explanations are possible, including voluntary alarm switching off and leaving the receiver out of earshot. The engagement of parents with the CGM device and their behaviors in response to data provided comprise a vital extension of the digital phenotype of HI [40,42].

Future work must further evaluate this aspect of the digital phenotype of HI to better understand not only the underlying pathophysiology but also the human in the loop. Simple mathematical modeling to generate ever better glycemic predictions is unable to eradicate hypoglycemic events in the real world if it does not factor in how human behaviors change in response to the data.

Conclusion

We provide the first analysis of the timing of hypoglycemia in patients with hypoglycemia due to HI using CGM data and, in doing so, expand the digital phenotype. In contrast to the phenotype of hypoglycemia in children with IKH, a clear period of high risk for hypoglycemia was observed in patients with HI in the early hours (3 AM-7 AM). Such early hours hypoglycemia was particularly frequent in those with genetic mutations known to cause HI and in children older than 10 months. Despite the unblinded state-of-the-art technology, prolonged and potentially harmful hypoglycemic events were detected in patients with HI. Hypoglycemia during early hours poses a high risk for neuroglycopenic brain injury, and behavioral aspects of the digital phenotype in HI must be evaluated so that interventions can be designed to maximize effect.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Notepad file with original Python code used for data analysis. This can also be found freely available on GitHub [<xref ref-type="bibr" rid="35ref49">49</xref>].
[TXT File , 10 KB - jmir_v23i10e26957_app1.txt]

Multimedia Appendix 2
Further hyperinsulinism subgroup comparisons.
[DOC File , 302 KB - jmir_v23i10e26957_app2.doc]

Multimedia Appendix 3
Comparison of reducing hypoglycemia thresholds.
[DOC File , 130 KB - jmir_v23i10e26957_app3.doc]

References

1. Yau D, Laver T, Dastamani A, Senniappan S, Houghton JA, Shaikh G, et al. Using referral rates for genetic testing to determine the incidence of a rare disease: the minimal incidence of congenital hyperinsulinism in the UK is 1 in 28,389. PLoS One 2020 Feb 06;15(2):e0228417 [FREE Full text] [doi: 10.1371/journal.pone.0228417] [Medline: 32027664]
2. Lord K, Radcliffe J, Gallagher PR, Adzick NS, Stanley CA, De León DD. High risk of diabetes and neurobehavioral deficits in individuals with surgically treated hyperinsulinism. J Clin Endocrinol Metab 2015 Nov;100(11):4133-4139 [FREE Full text] [doi: 10.1210/jc.2015-2539] [Medline: 26327482]
3. Avatapalle HB, Banerjee I, Shah S, Pryce M, Nicholson J, Rigby L, et al. Abnormal neurodevelopmental outcomes are common in children with transient congenital hyperinsulinism. Front Endocrinol (Lausanne) 2013 May 20;4:60 [FREE Full text] [doi: 10.3389/fendo.2013.00060] [Medline: 23730298]

4. Tam EW, Haeusslein LA, Bonifacio SL, Glass HC, Rogers EE, Jeremy RJ, et al. Hypoglycemia is associated with increased risk for brain injury and adverse neurodevelopmental outcome in neonates at risk for encephalopathy. J Pediatr 2012 Jul;161(1):88-93 [FREE Full text] [doi: 10.1016/j.jpeds.2011.12.047] [Medline: 22306045]
5. Lucas A, Morley R, Cole TJ. Adverse neurodevelopmental outcome of moderate neonatal hypoglycaemia. BMJ 1988 Nov 19;297(6659):1304-1308 [FREE Full text] [doi: 10.1136/bmj.297.6659.1304] [Medline: 2462455]
6. McQuarriei I. Idiopathic spontaneously occurring hypoglycemia in infants; clinical significance of problem and treatment. AMA Am J Dis Child 1954 Apr;87(4):399-428. [Medline: 13147532]
7. Craigie RJ, Salomon-Estebanez M, Yau D, Han B, Mal W, Newbould M, et al. Clinical diversity in focal congenital hyperinsulinism in infancy correlates with histological heterogeneity of islet cell lesions. Front Endocrinol (Lausanne) 2018 Oct 17;9:619 [FREE Full text] [doi: 10.3389/fendo.2018.00619] [Medline: 30386300]
8. Kane C, Shepherd RM, Squires PE, Johnson PR, James RF, Milla PJ, et al. Loss of functional KATP channels in pancreatic beta-cells causes persistent hyperinsulinemic hypoglycemia of infancy. Nat Med 1996 Dec;2(12):1344-1347. [doi: 10.1038/nm1296-1344] [Medline: 8946833]
9. Dunne MJ, Kane C, Shepherd RM, Sanchez JA, James RF, Johnson PR, et al. Familial persistent hyperinsulinemic hypoglycemia of infancy and mutations in the sulfonylurea receptor. N Engl J Med 1997 Mar 06;336(10):703-706. [doi: 10.1056/NEJM199703063361005] [Medline: 9041101]
10. Stanley CA. Perspective on the genetics and diagnosis of congenital hyperinsulinism disorders. J Clin Endocrinol Metab 2016 Mar;101(3):815-826 [FREE Full text] [doi: 10.1210/jc.2015-3651] [Medline: 26908106]
11. Worth C, Hashmi LA, Yau D, Salomon-Estebanez M, Ruiz DP, Hall C, et al. Longitudinal Auxological recovery in a cohort of children with Hyperinsulinaemic Hypoglycaemia. Orphanet J Rare Dis 2020 Jun 24;15(1):162 [FREE Full text] [doi: 10.1186/s13023-020-01438-0] [Medline: 32580746]
12. Dawkins R. The Selfish Gene, 1st Edition. London: Oxford University Press; 1976.
13. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. Nat Biotechnol 2015 May;33(5):462-463. [doi: 10.1038/nbt.3223] [Medline: 25965751]
14. Worth C, Yau D, Estebanez MS, O'Shea E, Cosgrove K, Dunne M, et al. Complexities in the medical management of hypoglycaemia due to congenital hyperinsulinism. Clin Endocrinol (Oxf) 2020 May;92(5):387-395. [doi: 10.1111/cen.14152] [Medline: 31917867]
15. Jauch-Chara K, Schultes B. Sleep and the response to hypoglycaemia. Best Pract Res Clin Endocrinol Metab 2010 Oct;24(5):801-815. [doi: 10.1016/j.beem.2010.07.006] [Medline: 21112027]
16. Torous J, Onnela J, Keshavan M. New dimensions and new tools to realize the potential of RDoC: digital phenotyping via smartphones and connected devices. Transl Psychiatry 2017 Mar 07;7(3):e1053 [FREE Full text] [doi: 10.1038/tp.2017.25] [Medline: 28267146]
17. Worth C, Dunne M, Ghosh A, Harper S, Banerjee I. Continuous glucose monitoring for hypoglycaemia in children: perspectives in 2020. Pediatr Diabetes 2020 Aug;21(5):697-706. [doi: 10.1111/pedi.13029] [Medline: 32315515]
18. Skinner AL, Attwood AS, Baddeley R, Evans-Reeves K, Bauld L, Munafò MR. Digital phenotyping and the development and delivery of health guidelines and behaviour change interventions. Addiction 2017 Jul;112(7):1281-1285. [doi: 10.1111/add.13746] [Medline: 28472848]
19. Rayannavar A, Elci O, Mitteer L, De León DD. Continuous glucose monitoring systems: are they useful for evaluating glycemic control in children with hyperinsulinism? Horm Res Paediatr 2019;92(5):319-327 [FREE Full text] [doi: 10.1159/000506230] [Medline: 32208390]
20. Saif M, Kapoor A, Kochar I, Jindal R. Continuous glucose monitoring system for congenital hyperinsulinemia. Indian Pediatr 2013 Apr 25;50(4):421-422. [doi: 10.1007/s13312-013-0103-3] [Medline: 23665604]
21. Alsaffar H, Turner L, Yung Z, Didi M, Senniappan S. Continuous flash glucose monitoring in children with congenital hyperinsulinism; first report on accuracy and patient experience. Int J Pediatr Endocrinol 2018;2018:3 [FREE Full text] [doi: 10.1186/s13633-018-0057-2] [Medline: 29599801]
22. Braune K, Wäldchen M, Raile K, Hahn S, Ubben T, Römer S, et al. Open-source technology for real-time continuous glucose monitoring in the neonatal intensive care unit: case study in a neonate with transient congenital hyperinsulinism. J Med Internet Res 2020 Dec 04;22(12):e21770 [FREE Full text] [doi: 10.2196/21770] [Medline: 33275114]
23. Kalsbeek A, la Fleur S, Fliers E. Circadian control of glucose metabolism. Mol Metab 2014 Mar 19;3(4):372-383 [FREE Full text] [doi: 10.1016/j.molmet.2014.03.002] [Medline: 24944897]
24. Hadlow NC, Brown S, Wardrop R, Henley D. The effects of season, daylight saving and time of sunrise on serum cortisol in a large population. Chronobiol Int 2014 Mar;31(2):243-251. [doi: 10.3109/07420528.2013.844162] [Medline: 24156521]
25. Mason I, Qian J, Adler G, Scheer FA. Impact of circadian disruption on glucose metabolism: implications for type 2 diabetes. Diabetologia 2020 Mar;63(3):462-472 [FREE Full text] [doi: 10.1007/s00125-019-05059-6] [Medline: 31915891]
26. Kalsbeek A, Yi CX, La Fleur SE, Fliers E. The hypothalamic clock and its control of glucose homeostasis. Trends Endocrinol Metab 2010 Jul;21(7):402-410. [doi: 10.1016/j.tem.2010.02.005] [Medline: 20303779]
27. Bright GM, Melton TW, Rogol AD, Clarke WL. Failure of cortisol blockade to inhibit early morning increases in basal insulin requirements in fasting insulin-dependent diabetics. Diabetes 1980 Aug;29(8):662-664. [doi: 10.2337/diab.29.8.662] [Medline: 7002678]

28. Huang W, Ramsey KM, Marcheva B, Bass J. Circadian rhythms, sleep, and metabolism. J Clin Invest 2011 Jun;121(6):2133-2141 [FREE Full text] [doi: 10.1172/JCI46043] [Medline: 21633182]
29. Cailotto C, La Fleur SE, Van Heijningen C, Wortel J, Kalsbeek A, Feenstra M, et al. The suprachiasmatic nucleus controls the daily variation of plasma glucose via the autonomic output to the liver: are the clock genes involved? Eur J Neurosci 2005 Nov;22(10):2531-2540. [doi: 10.1111/j.1460-9568.2005.04439.x] [Medline: 16307595]
30. Christesen H, Brusgaard K, Hussain K. Recurrent spontaneous hypoglycaemia causes loss of neurogenic and neuroglycopaenic signs in infants with congenital hyperinsulinism. Clin Endocrinol (Oxf) 2012 Apr;76(4):548-554. [doi: 10.1111/j.1365-2265.2011.04250.x] [Medline: 21981106]
31. Jauch-Chara K, Hallschmid M, Gais S, Oltmanns KM, Peters A, Born J, et al. Awakening and counterregulatory response to hypoglycemia during early and late sleep. Diabetes 2007 Jul;56(7):1938-1942 [FREE Full text] [doi: 10.2337/db07-0044] [Medline: 17400929]
32. Whincup G, Milner RD. Prediction and management of nocturnal hypoglycaemia in diabetes. Arch Dis Child 1987 Apr;62(4):333-337 [FREE Full text] [doi: 10.1136/adc.62.4.333] [Medline: 3296962]
33. van Beers CA, DeVries JH. Continuous glucose monitoring: impact on hypoglycemia. J Diabetes Sci Technol 2016 Nov 01;10(6):1251-1258 [FREE Full text] [doi: 10.1177/1932296816653411] [Medline: 27257169]
34. Graveling AJ, Frier BM. The risks of nocturnal hypoglycaemia in insulin-treated diabetes. Diabetes Res Clin Pract 2017 Nov;133:30-39. [doi: 10.1016/j.diabres.2017.08.012] [Medline: 28888993]
35. Mukherjee E, Carroll R, Matfin G. Endocrine and metabolic emergencies: hypoglycaemia. Ther Adv Endocrinol Metab 2011 Apr;2(2):81-93 [FREE Full text] [doi: 10.1177/2042018811401644] [Medline: 23148173]
36. Hall H, Perelman D, Breschi A, Limcaoco P, Kellogg R, McLaughlin T, et al. Glucotypes reveal new patterns of glucose dysregulation. PLoS Biol 2018 Jul 24;16(7):e2005143 [FREE Full text] [doi: 10.1371/journal.pbio.2005143] [Medline: 30040822]
37. Colas A, Vigil L, de Castro CR, Vargas B, Varela M. New insights from continuous glucose monitoring into the route to diabetes. Diabetes Metab Res Rev 2018 Jul;34(5):e3002. [doi: 10.1002/dmrr.3002] [Medline: 29516622]
38. Lunt H, Heenan HF. Diagnosing the underlying cause of mild hyperglycemia using interstitial glucose data. J Diabetes Sci Technol 2020 Nov;14(6):1139-1140 [FREE Full text] [doi: 10.1177/1932296820929369] [Medline: 32462915]
39. Larkin ME, Nathan DM, Bebu I, Krause-Steinrauf H, Herman WH, Higgins JM, GRADE Research Group. Rationale and design for a GRADE substudy of continuous glucose monitoring. Diabetes Technol Ther 2019 Dec;21(12):682-690 [FREE Full text] [doi: 10.1089/dia.2019.0202] [Medline: 31393176]
40. Radhakrishnan K, Kim MT, Burgermaster M, Brown RA, Xie B, Bray MS, et al. The potential of digital phenotyping to advance the contributions of mobile health to self-management science. Nurs Outlook 2020;68(5):548-559. [doi: 10.1016/j.outlook.2020.03.007] [Medline: 32402392]
41. Ziegler R, Heinemann L, Freckmann G, Schnell O, Hinzmann R, Kulzer B. Intermittent use of continuous glucose monitoring: expanding the clinical value of CGM. J Diabetes Sci Technol 2021 May;15(3):684-694 [FREE Full text] [doi: 10.1177/1932296820905577] [Medline: 32064909]
42. Yang Q, Hatch D, Crowley M, Lewinski AA, Vaughn J, Steinberg D, et al. Digital phenotyping self-monitoring behaviors for individuals with type 2 diabetes mellitus: observational study using latent class growth analysis. JMIR Mhealth Uhealth 2020 Jun 11;8(6):e17730 [FREE Full text] [doi: 10.2196/17730] [Medline: 32525492]
43. Chrisworth88 - Basic-CGM-analysis-for-CHI. GitHub. URL: https://github.com/chrisworth88/Basic-CGM-analysis-for-CHI [accessed 2021-09-28]
44. Banerjee I, Salomon-Estebanez M, Shah P, Nicholson J, Cosgrove KE, Dunne MJ. Therapies and outcomes of congenital hyperinsulinism-induced hypoglycaemia. Diabet Med 2019 Jan;36(1):9-21 [FREE Full text] [doi: 10.1111/dme.13823] [Medline: 30246418]
45. Ghosh A, Banerjee I, Morris AA. Recognition, assessment and management of hypoglycaemia in childhood. Arch Dis Child 2016 Jun;101(6):575-580. [doi: 10.1136/archdischild-2015-308337] [Medline: 26718813]
46. Demirbilek H, Hussain K. Congenital hyperinsulinism: diagnosis and treatment update. J Clin Res Pediatr Endocrinol 2017 Dec 30;9(Suppl 2):69-87 [FREE Full text] [doi: 10.4274/jcrpe.2017.S007] [Medline: 29280746]
47. Liang Y, Zheng X, Zeng DD. A survey on big data-driven digital phenotyping of mental health. Inf Fusion 2019;52:290-307. [doi: 10.1016/j.inffus.2019.04.001]
48. Pickup JC, Ford HM, Samsi K. Real-time continuous glucose monitoring in type 1 diabetes: a qualitative framework analysis of patient narratives. Diabetes Care 2015 Apr;38(4):544-550. [doi: 10.2337/dc14-1855] [Medline: 25552422]
49. Buckingham B, Block J, Burdick J, Kalajian A, Kollman C, Choy M, Diabetes Research in Children Network. Response to nocturnal alarms using a real-time glucose sensor. Diabetes Technol Ther 2005 Jun;7(3):440-447 [FREE Full text] [doi: 10.1089/dia.2005.7.440] [Medline: 15929675]

Abbreviations

CGM: continuous glucose monitoring
HI: hyperinsulinism

IKH: idiopathic ketotic hypoglycemia

Edited by R Kukafka, G Eysenbach; submitted 05.01.21; peer-reviewed by D Yau, P Conti; comments to author 25.06.21; revised version received 30.06.21; accepted 23.08.21; published 29.10.21.

Please cite as:
Worth C, Harper S, Salomon-Estebanez M, O'Shea E, Nutter PW, Dunne MJ, Banerjee I
Clustering of Hypoglycemia Events in Patients With Hyperinsulinism: Extension of the Digital Phenotype Through Retrospective Data Analysis
J Med Internet Res 2021;23(10):e26957
URL: https://www.jmir.org/2021/10/e26957
doi:10.2196/26957
PMID:34435596

Original Paper

Clinician-Created Educational Video Resources for Shared Decision-making in the Outpatient Management of Chronic Disease: Development and Evaluation Study

Joshua G Kovoor[1], BHlthMedSc (Hons); Daniel McIntyre[2], BSc; William W B Chik[3], MBBS, PhD; Clara K Chow[2,3], MBBS, PhD; Aravinda Thiagalingam[2,3], MB ChB, PhD

[1]University of Adelaide, The Queen Elizabeth Hospital, Adelaide, Australia

[2]Westmead Applied Research Centre, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[3]Department of Cardiology, Westmead Hospital, Sydney, Australia

Corresponding Author:
Aravinda Thiagalingam, MB ChB, PhD
Department of Cardiology
Westmead Hospital
Cnr Darcy and Hawkesbury Rds
Sydney, 2145
Australia
Phone: 61 298456795
Email: aravinda.thiagalingam@sydney.edu.au

Abstract

Background: The provision of reliable patient education is essential for shared decision-making. However, many clinicians are reluctant to use commonly available resources, as they are generic and may contain information of insufficient quality. Clinician-created educational materials, accessed during the waiting time prior to consultation, can potentially benefit clinical practice if developed in a time- and resource-efficient manner.

Objective: The aim of this study is to evaluate the utility of educational videos in improving patient decision-making, as well as consultation satisfaction and anxiety, within the outpatient management of chronic disease (represented by atrial fibrillation). The approach involves clinicians creating audiovisual patient education in a time- and resource-efficient manner for opportunistic delivery, using mobile smart devices with internet access, during waiting time before consultation.

Methods: We implemented this educational approach in outpatient clinics and collected patient responses through an electronic survey. The educational module was a web-based combination of 4 short videos viewed sequentially, followed by a patient experience survey using 5-point Likert scales and 0-100 visual analogue scales. The clinician developed the audiovisual module over a 2-day span while performing usual clinical tasks, using existing hardware and software resources (laptop and tablet). Patients presenting for the outpatient management of atrial fibrillation accessed the module during waiting time before their consultation using either a URL or Quick Response (QR) code on a provided tablet or their own mobile smart devices. The primary outcome of the study was the module's utility in improving patient decision-making ability, as measured on a 0-100 visual analogue scale. Secondary outcomes were the level of patient satisfaction with the videos, measured with 5-point Likert scales, in addition to the patient's value for clinician narration and the module's utility in improving anxiety and long-term treatment adherence, as represented on 0-100 visual analogue scales.

Results: This study enrolled 116 patients presenting for the outpatient management of atrial fibrillation. The proportion of responses that were "very satisfied" with the educational video content across the 4 videos ranged from 93% (86/92) to 96.3% (104/108) and this was between 98% (90/92) and 99.1% (107/108) for "satisfied" or "very satisfied." There were no reports of dissatisfaction for the first 3 videos, and only 1% (1/92) of responders reported dissatisfaction for the fourth video. The median reported scores (on 0-100 visual analogue scales) were 90 (IQR 82.5-97) for improving patient decision-making, 89 (IQR 81-95) for reducing consultation anxiety, 90 (IQR 81-97) for improving treatment adherence, and 82 (IQR 70-90) for the clinician's narration adding benefit to the patient experience.

Conclusions: Clinician-created educational videos for chronic disease management resulted in improvements in patient-reported informed decision-making ability and expected long-term treatment adherence, as well as anxiety reduction. This form of patient

education was also time efficient as it used the sunk time cost of waiting time to provide education without requiring additional clinician input.

(J Med Internet Res 2021;23(10):e26732) doi:10.2196/26732

KEYWORDS

Shared decision-making; chronic disease; outpatients; audiovisual aids; atrial fibrillation; educational technology; teaching materials; referral and consultation; physician-patient relations; physicians

Introduction

Chronic disease is the leading cause of disease burden and mortality worldwide, with increasing prevalence due to an aging global population [1]. Cardiovascular disease is one of the major categories of chronic disease, and atrial fibrillation (AF) is widely recognized as one of the most common chronic conditions [2].

Ongoing outpatient consultations are an essential component of chronic disease management [3], and one strategy that could have considerable utility in this setting is that of "shared decision-making" [4]. Shared decision-making involves bidirectional information exchange within the clinician-patient relationship before making management decisions [5], and can optimize the practice of evidence-based medicine [6]. Patients are informed and care is patient-centered as the patient is empowered to communicate their personal values and management preferences for the clinician to individualize suggested management options [7]. A patient's level of involvement in shared decision-making is influenced by their level of health literacy [8], and accordingly educational decision aids are facilitative [9]. Decision aids have shown potential benefit [10] when implemented while patients are waiting for the consultation [11].

Audiovisual education aids, including videos, can be an effective method of improving patient health literacy [10,12,13]. However, clinicians may be reluctant to use those that are commonly available (eg, those publicly available on health care websites or YouTube) if they are generic or contain information of insufficient accuracy, quality, or currency [14]. Patient care may benefit from clinicians creating their own audiovisual content and delivering it to their patients within the clinical setting.

Based on findings from the prior literature [15], we hypothesized that clinician-created audiovisual content could be created with limited resources and would be acceptable and improve the ability of patients to contribute to the decision-making process. Integral to this is understanding effects of the intervention on anxiety around consultation ("white coat" effects) [16] and potential long-term compliance to formulated management plans [17]. We aimed to evaluate this approach within outpatient management of chronic disease (represented by AF in this instance [2]), whereby clinicians create audiovisual patient education in a time- and resource-efficient manner for opportunistic delivery, using mobile smart devices with internet access, during waiting time before consultation. Specifically, we aimed to assess the following: patient satisfaction with the individual videos and the approach overall, patient response to clinician narration within the content, and the patient-perceived effect of the approach on patient decision-making, patient anxiety around consultation, and potential long-term treatment adherence.

Methods

Study Design

We conducted a prospective, nonrandomized, observational study of 116 patients presenting specifically for the outpatient management of AF. Patients provided informed consent through a validated electronic form before commencing the audiovisual module on a smart device. Ethical approval for the study was obtained from the Western Sydney Local Health Network Human Research Ethics Committee (item number 2011–18).

Setting and Participants

Consecutive patients presenting for outpatient management of AF were prospectively recruited at a specialist outpatient clinic within a large university teaching hospital in Sydney, Australia. The site was within a public metropolitan hospital that serves patients of the Western Sydney Local Health District. As of 2018, the district provides AUD $1.8 billion (US $1.3 billion) in public health care to over 120 suburbs consisting of a demographically diverse population; approximately 50% of the residents have chronic conditions [18]. Potential participants were identified by study personnel via screening of the specialist outpatient clinic lists.

Patients were eligible for inclusion if they had presented to the specialist outpatient clinic specifically for consultation for the management of AF, and were English speaking, willing and able to use smart devices, and could provide informed consent. Patients were excluded if they were unable to speak English, refused consent, or had visual impairment or any other factor that prevented them from using the provided tablet (eg, rheumatoid arthritis). All prospective patients who met the inclusion criteria were enrolled in the study.

Development of the Audiovisual Module

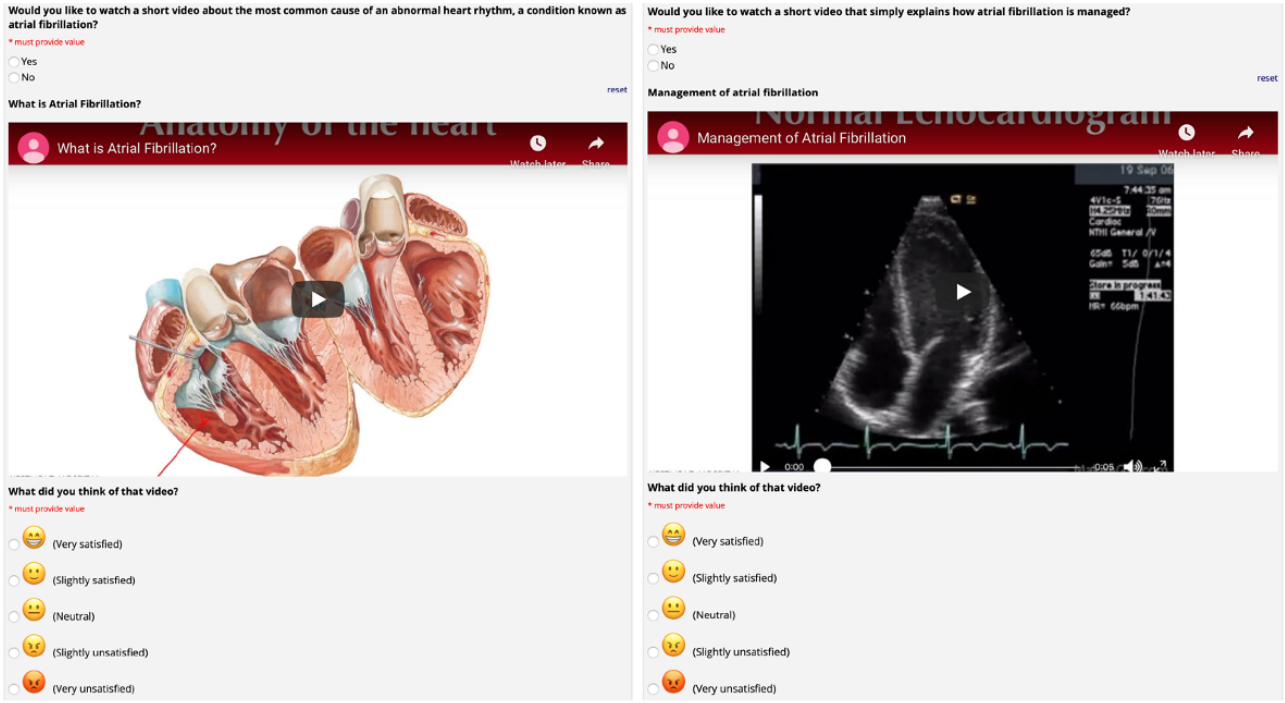
The educational audiovisual module was designed to be a web-based combination of 4 short videos (privately hosted on YouTube) to be viewed sequentially, accompanied by questions gauging patient experience. The module was consolidated using Research Electronic Data Capture (REDCap) [19]. Data from the module were collected and coded on REDCap via a secure database.

The audiovisual module sought to provide a baseline level of disease-specific health literacy to patients that supplemented information provision and improved shared decision-making

in the subsequent outpatient consultation. For this study, two of the authors (AT and JGK) consulted the latest guidelines and peer-reviewed literature, and decided on an up-to-date, reliable, yet simple syllabus covering the fundamental concepts of the pathophysiology, clinical presentation, diagnosis, and management of AF. For this, we searched for studies of any design, in any setting, within Scopus using the search term "atrial fibrillation" in all fields. To ensure that the most prominent recent AF literature was reviewed, the search was date restricted from January 2016 to December 2018, and the 46,393 resultant records were sorted in order of total number of citations. From the prominent recent AF literature that was reviewed, the two most cited guidance statements specifically relevant to the management of AF [20,21] were discussed by two authors (AT and JGK) and concepts and the syllabus outline were subsequently formulated by consensus. The resultant module contained four succinct videos encompassing the following topics: (1) "What is AF?" (2) "AF Management," (3) "Stroke risk and anticoagulation," and (4) "Lifestyle modification."

For the measurement of outcomes, we used Likert scales and visual analogue scales, which have both demonstrated reliability and validity as health measurement tools [22]. The level of patient satisfaction for each individual video and satisfaction with the videos overall were recorded on 5-point Likert scales [23]. Further, 0-100 visual analogue scales were used to gauge the patient's response to clinician narration, as well as the module's patient-perceived utility in improving patient decision-making, anxiety around consultation, and potential long-term treatment adherence. The visual analogue scales and Likert scales were created and combined with the 4 videos for the audiovisual module using REDCap (Figure 1).

Figure 1. A sample of the web-based audiovisual educational module consolidated using REDCap.

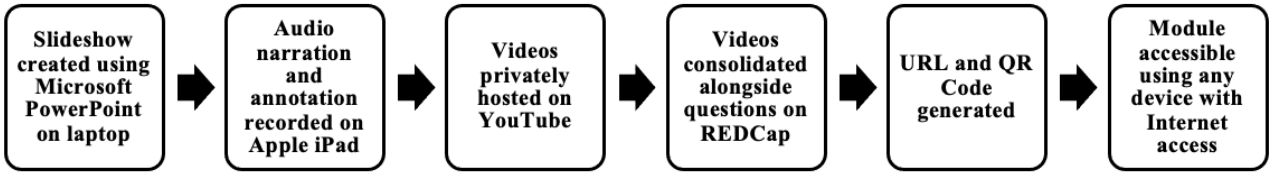


After the simple syllabus was formulated, the clinician (AT) developed the audiovisual module over a 2-day span between clinical commitments and within normal working hours (approximately 3 hours total). No disruption to regular clinical duties was experienced. The only material resources used in the process were that of a laptop containing Microsoft PowerPoint (Microsoft Corp) and an Apple iPad (Apple Inc), both of which were already owned by the clinician, thus not incurring any additional financial costs.

To make the 4 educational videos within the module, the clinician created 4 brief slideshows using PowerPoint on a laptop, then recorded audio narration simultaneously to annotate the slideshows using the Apple iPad using the Screen Recording function available by default on iOS 13 or later (Figure 2). The latter was done in a single take; on average, it took the clinician 3 minutes and 42 seconds to record the audio narration and annotation that complemented the slideshows within each individual video. The videos were subsequently privately hosted on YouTube in order to be consolidated alongside the patient experience survey on REDCap. This ensured that no advertising material was delivered with the videos by the YouTube website. REDCap then generated a URL, also known as a web address, and a Quick Response (QR) code, both of which enabled access to the educational module using any device with internet access.

All information provided within the module was only included following brief inspection of the latest peer-reviewed literature on the associated topics by the clinician to update existing knowledge. Further, the module was recorded in English with language and readability aimed below an eighth grade level to enhance accessibility for the general adult population [24]. The audiovisual module, along with all included images and visual media, was consolidated solely for private use limited to the duration of the study. It was not marketed or sold for commercial purposes.

Figure 2. Technique for audiovisual module creation by a clinician.



Delivery During Waiting Time

To minimize potential bias attributable to the halo effect [25], patients were approached by members of staff in the specialist clinic other than the clinician who created the educational videos. Following informed consent, patients completed the audiovisual educational module during the otherwise nonclinically utilized waiting time before their consultation for AF. Patients accessed the web-based module through either the URL or QR code generated by REDCap. This was done using either a tablet that was already owned by the treating specialist in the clinic or mobile smart devices that the patients already owned.

The entire audiovisual module was designed to take a maximum of 20 minutes for the patient to complete, so as to comfortably be completed during the expected waiting time before outpatient consultation [26]. Further, the cumulative duration of all four videos totaled 14 minutes and 46 seconds, so as to maximize concentration and minimize the chances of attention decline [27].

Data Elements

The primary outcome of the study was the audiovisual educational module's utility in improving patient decision-making ability during the subsequent outpatient consultation [17], which was measured in an anonymous manner using a 0-100 visual analogue scale. The secondary outcomes were the level of patient satisfaction for each video and satisfaction with the videos overall (both measured using 5-point Likert scales), as well as the patient's value for clinician narration, and the module's patient-perceived utility in improving anxiety around consultation and potential long-term treatment adherence (all measured using 0-100 visual analogue scales in an anonymous manner). Data were extracted by two authors (JGK and AT) from the secure REDCap database as a CSV file.

Statistical Analysis

Data were assessed in Python (open source, Python Software Foundation) using the Pandas library (version 1.0.4, open source, PyData), with results reported using medians and interquartile ranges. Figures were prepared using the Plotly (version 4.8.1) library. Friedman test and post hoc analysis were performed using Statsmodels (version 0.10.2, open source) as the data were not normally distributed, with repeated observations on the same individuals.

Results

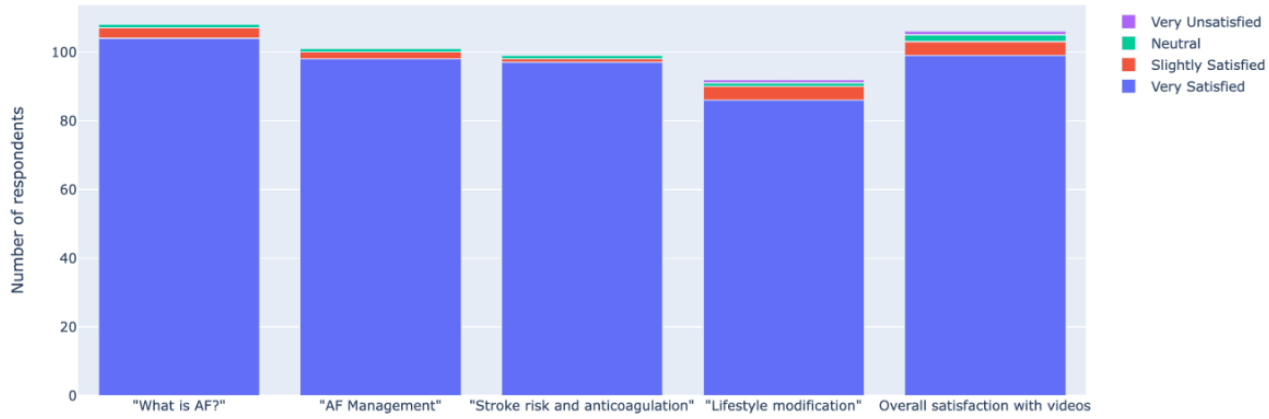
Overview

Between January 2019 and August 2019, all 116 prospective patients who met the aforementioned inclusion criteria were enrolled in the study. Response rates for the 5-point Likert scales progressively decreased with each successive video, with 93.1% (108/116) of the total cohort conveying their level of satisfaction with the first video ("What is AF?"), 87.1% (101/116) for the second ("AF Management"), 85.3% (99/116) for the third ("Stroke risk and anticoagulation"), and 79.3% (92/116) for the fourth ("Lifestyle modification"). In addition, 85.3% (99/116) of the total cohort reported the effect of the audiovisual module on consultation anxiety, decision-making ability, and likelihood of potential treatment adherence through the 0-100 visual analogue scales, with 84.5% (98/116) reporting the importance of clinician narration.

Likert Scales

The proportion of participants that were "very satisfied" with the educational video content ranged from 93% (86/92) to 96.3% (104/108) for the individual videos. In the final 5-point Likert scale assessing overall satisfaction with all 4 videos, 93.4% (99/106) of responders were "very satisfied," with 0.9% (1/106) reporting dissatisfaction. For the first video ("What is AF?"), 99.1% (107/108) of responders reported satisfaction with the content, compared to 99.0% (100/101) for the second ("AF Management"), 99.0% (98/99) for the third ("Stroke risk and anticoagulation"), and 98% (90/92) for the fourth ("Lifestyle modification"). There were no reports of dissatisfaction for the first 3 videos, and 1% (1/92) of responders reported dissatisfaction for the fourth video. Patient satisfaction with the clinician-created educational videos, as obtained through 5-point Likert scales, is represented in Figure 3.

Figure 3. Patient satisfaction with clinician-created videos. AF: atrial fibrillation.



Visual Analogue Scales

Median scores on the four 0-100 visual analogue scales ranged from 82 to 90. The audiovisual module resulted in median reported patient-perceived scores of 90 (IQR 82.5-97) for improving patient decision-making, 89 (IQR 81-95) for improving consultation anxiety, 90 (IQR 81-97) for improving potential treatment adherence, and 82 (IQR 70-90) for the clinician's narration adding benefit to the patient experience. Patient responses to the 0-100 visual analogue scales are presented in Figure 4.

Friedman test results yielded a Friedman statistic value of 33.3 (P<.001). The Nemenyi post hoc analysis showed that only "Importance of clinician narration" was significantly different (P=.001) from all the other evaluations. The other evaluations did not differ significantly from each other. These results are presented in Table 1.

Figure 4. Patient responses to the audiovisual module.

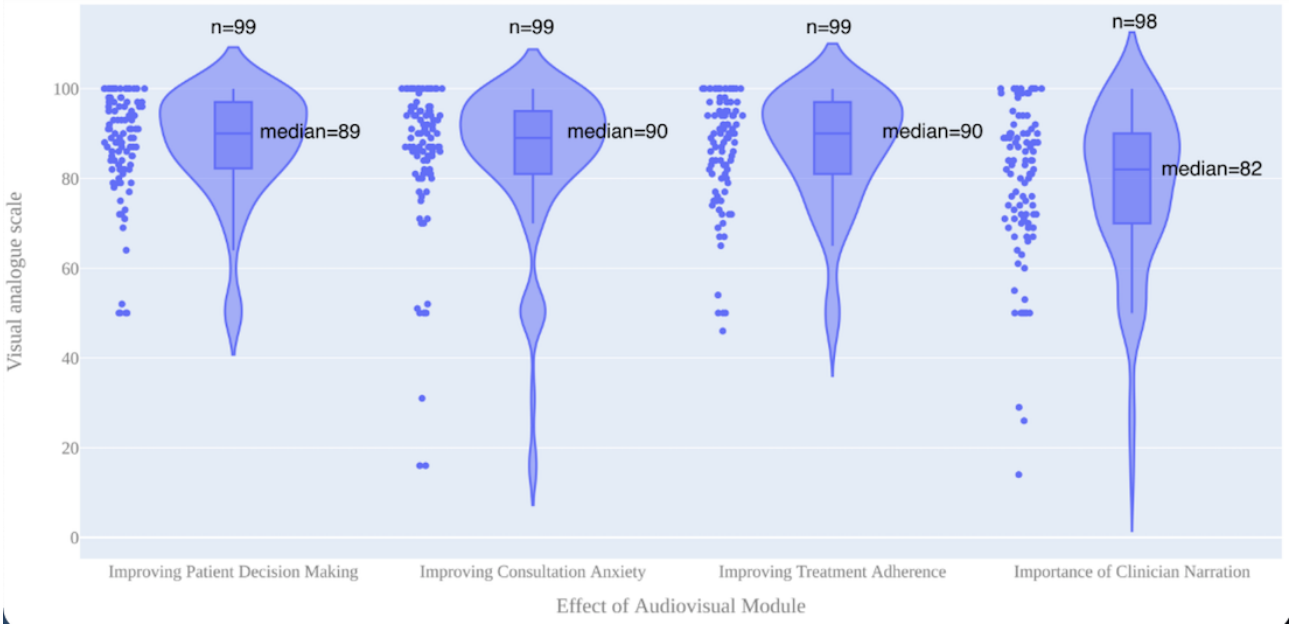


Table 1. Analysis of patient responses using the Friedman test.

Evaluation	Improving consultation anxiety	Improving patient decision making	Improving potential treatment adherence	Importance of clinician narration
Improving consultation anxiety	1	.70	.90	.001
Improving patient decision making	.70	1	.49	.001
Improving potential treatment adherence	.90	.49	1	.001
Importance of clinician narration	.001	.001	.001	1

Discussion

Principal Findings

A clinician-created audiovisual patient education module with content focusing on improving consultation efficiency and shared decision-making received scores indicating a high level of acceptability and patient-perceived utility. The video content was developed by a clinician using electronic resources that are relatively common in the developed world, during interstitial time between clinical duties, and was delivered using mobile smart devices while patients were waiting for their appointments. The majority of participants perceived the clinician's creation and narration of the module to be positive and effective in delivering the educational content of the disease-specific videos.

Shared decision-making provides a potential solution for achieving adequate efficiency while improving patient autonomy [28]; however, this may not translate to clinical practice if the appropriate approach to implementation is not taken [29]. An element of usefulness in our approach is that it not only improves shared decision-making, but also uses waiting time (sunk time cost) for an improved patient experience that does not add to the overall length of the consultation. Further, our approach demonstrates that clinicians can use off-the-shelf tools to simply create their own highly customized educational content tailored specifically for their patients. Further, it is implementable with minimal additional resources.

Patient education has been shown to be more effective when delivered in an audiovisual format than when presented solely through visual pamphlets or verbal one-to-one dialogue [30]. However, clinical utility for the management of chronic disease is limited when audiovisual patient education has required additional outpatient appointments [31]. This intervention used shorter duration, highly focused videos, allowing delivery in the waiting room so that patients would not have to make additional visits. Increasing access to unreliable, inaccurate, and outdated sources via the internet can increase confusion and cause difficulties during consultations for the management of patients' conditions [32]. This can potentially be reduced through the provision of up-to-date and reliable information directly from a patient's own treating clinician, as a supplement to scheduled consultations. Our approach allows for the provision of reliable health education in a manner that is streamlined for integration within both the clinician (through time and resource efficiency) and patient (through tailored, accessible education) experience of outpatient consultation.

The audiovisual educational module demonstrated utility in improving patient-reported decision-making ability for the subsequent outpatient consultation for AF management. However, due to a lack of any of the validated outcome measures or objective endpoints that are found within the existing literature of studies investigating shared decision-making in AF management [33], comparison between previous studies and our study is compromised. The high levels of patient satisfaction with both the videos and clinician narration suggest that there may be potential usefulness in clinician-created audiovisual educational content for the management of chronic disease. Further, anxiety [34], particularly that associated with the "white coat" effect [16], and long-term treatment adherence [17,35] have a significant impact on a patient's decision-making ability regarding the management of any chronic disease. Accordingly, our positive findings imply potential value in these domains. Our finding of patient-reported improvements to potential treatment adherence is in line with the existing literature, which has demonstrated the efficacy of video-assisted patient education in positively modifying the behaviors of patients with chronic disease if designed and delivered correctly [36].

There were limitations to this study that require future evaluation. Our study did not have a control group that enabled comparison; the study was observational and baseline characteristics of the participants were not collected. However, we prospectively included a consecutive series of eligible patients to limit patient bias, and all patients answered the same questionnaire. Our study was limited by subjective responses from patients as outcome measures, instead of outcome measures such as recurrence of AF or occurrence of its complications. The data collected referred specifically to the patients' own perceptions, which may carry inherent bias. We did not collect data using validated measures of patient-reported experience or health literacy. Further, we delivered the educational module only in English, and excluded non–English-speaking patients. Within the literature search that informed the curriculum for our educational module, sorting search results in Scopus by total numbers of citations identified the most prominent articles within the search time frame, but added a source of bias regarding the data informing our intervention. As our study was conducted in 2019, any evidence published since then has not been integrated. Further, only one database was searched prior to the development of our curriculum, so some literature may have been missed. The study was limited to a single center, so this approach's multicenter applicability has not been tested. Additionally, although the clinician in our study was able to carry out our approach in a time- and resource-efficient manner, this approach may not translate to certain doctors with different levels of resources, time within schedules, and technology skills.

Findings from this pilot study may be useful for future research in this area. Although audio narration was used in this study, future studies may benefit from the inclusion of video of the treating clinician and investigation of subsequent effects on patient trust, anxiety reduction, and potential adherence. Clinical interactions within settings outside the management of chronic disease at outpatient clinics may also benefit from this approach and should be investigated. Usefulness of the approach relative to the socioeconomic status of the health care provider may be important to delineate for the maximization of global scalability. Further, exploration of utility for health professionals in other areas of medicine, surgery, nursing, and allied health ought to be explored. Study design in this future research can be improved by incorporating a comparison (ideally with randomization) against other forms of patient education as well as more concrete endpoints, such as objective health metrics or validated scoring schemas. The measures of patient perceptions that were recorded in this pilot study may be better assessed via a longitudinal approach that facilitates the evaluation of temporal trends. Explanations for changes in patient perceptions

throughout the course of the videos may have greater clarity if more qualitative data points are also collected in addition to measures of satisfaction.

Conclusions

This approach to outpatient consultation for the management of chronic disease may provide benefit for shared decision-making between clinicians and patients, overall leading to the improvement of care while maintaining a patient-centered focus. It benefits clinicians by combining their individual content knowledge with technology to create highly customized disease-specific audiovisual educational material for their patients. It benefits patients by allowing them to learn about their condition in a nonconfrontational situation without additional investment of time or effort. Patient satisfaction may be improved by converting a potentially negative situation (the sunk time cost of waiting for a medical appointment) to a positive experience. This pilot study demonstrated the potential utility of this approach in a specific setting; however, it may have widespread applicability across a large number of clinical scenarios, including those outside the management of chronic disease and outside the outpatient setting. Future research should explore this potential widespread applicability of our approach through studies of larger size and strong design.

Authors' Contributions

JGK was involved in the design of the study, the development of the associated intervention, and the preparation of this manuscript. DM was involved in the development of the associated intervention, data collection, and the preparation of this manuscript. WWBC was involved in data collection and the preparation of this manuscript. CKC was involved in the preparation of this manuscript. AT was involved in the design of the study, the development of the associated intervention, and the preparation of this manuscript.

Conflicts of Interest

CKC is supported by an NHMRC (National Health and Medical Research Council) Investigator grant. The other authors have no conflicts to declare.

References

1. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
2. January CT, Wann LS, Calkins H, Chen LY, Cigarroa JE, Cleveland JC, et al. 2019 AHA/ACC/HRS Focused Update of the 2014 AHA/ACC/HRS Guideline for the Management of Patients With Atrial Fibrillation: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society in Collaboration With the Society of Thoracic Surgeons. Circulation 2019 Jul 09;140(2):e125-e151. [doi: 10.1161/CIR.0000000000000665] [Medline: 30686041]
3. Grover A, Joshi A. An overview of chronic disease models: a systematic literature review. Glob J Health Sci 2014 Oct 29;7(2):210-227 [FREE Full text] [doi: 10.5539/gjhs.v7n2p210] [Medline: 25716407]
4. Murray E, Charles C, Gafni A. Shared decision-making in primary care: tailoring the Charles et al. model to fit the context of general practice. Patient Educ Couns 2006 Aug;62(2):205-211. [doi: 10.1016/j.pec.2005.07.003] [Medline: 16139467]
5. Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: What does it mean? (or it takes at least two to tango). Social Science & Medicine 1997 Mar;44(5):681-692. [doi: 10.1016/s0277-9536(96)00221-3]
6. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996 Jan 13;312(7023):71-72 [FREE Full text] [doi: 10.1136/bmj.312.7023.71] [Medline: 8555924]
7. Barry MJ, Edgman-Levitan S. Shared decision making--pinnacle of patient-centered care. N Engl J Med 2012 Mar 1;366(9):780-781. [doi: 10.1056/NEJMp1109283] [Medline: 22375967]
8. Yin HS, Dreyer BP, Vivar KL, MacFarland S, van SL, Mendelsohn AL. Perceived barriers to care and attitudes towards shared decision-making among low socioeconomic status parents: role of health literacy. Acad Pediatr 2012;12(2):117-124 [FREE Full text] [doi: 10.1016/j.acap.2012.01.001] [Medline: 22321814]
9. O'Connor AM, Llewellyn-Thomas HA, Flood AB. Modifying unwarranted variations in health care: shared decision making using patient decision aids. Health Aff (Millwood) 2004;Suppl Variation:VAR63-VAR72. [doi: 10.1377/hlthaff.var.63] [Medline: 15471770]
10. Clarkesmith DE, Pattison HM, Khaing PH, Lane DA. Educational and behavioural interventions for anticoagulant therapy in patients with atrial fibrillation. Cochrane Database Syst Rev 2017 Dec 05;4:CD008600. [doi: 10.1002/14651858.CD008600.pub3] [Medline: 28378924]
11. Lipkus IM, Peters E, Kimmick G, Liotcheva V, Marcom P. Breast cancer patients' treatment expectations after exposure to the decision aid program adjuvant online: the influence of numeracy. Med Decis Making 2010;30(4):464-473 [FREE Full text] [doi: 10.1177/0272989X09360371] [Medline: 20160070]

12. Denny MC, Vahidy F, Vu KYT, Sharrief AZ, Savitz SI. Video-based educational intervention associated with improved stroke literacy, self-efficacy, and patient satisfaction. PLoS ONE 2017 Mar 23;12(3):e0171952. [doi: 10.1371/journal.pone.0171952] [Medline: 28333925]
13. Stacey D, Légaré F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2017 Dec 12;4:CD001431. [doi: 10.1002/14651858.CD001431.pub5] [Medline: 28402085]
14. Légaré F, Ratté S, Gravel K, Graham ID. Barriers and facilitators to implementing shared decision-making in clinical practice: update of a systematic review of health professionals' perceptions. Patient Educ Couns 2008 Dec;73(3):526-535. [doi: 10.1016/j.pec.2008.07.018] [Medline: 18752915]
15. Joosten EAG, DeFuentes-Merillas L, de Weert GH, Sensky T, van der Staak CPF, de Jong CAJ. Systematic review of the effects of shared decision-making on patient satisfaction, treatment adherence and health status. Psychother Psychosom 2008;77(4):219-226. [doi: 10.1159/000126073] [Medline: 18418028]
16. Verdecchia P, Schillaci G, Borgioni C, Ciucci A, Zampi I, Gattobigio R, et al. White coat hypertension and white coat effect. Similarities and differences. Am J Hypertens 1995 Aug;8(8):790-798. [doi: 10.1016/0895-7061(95)00151-E] [Medline: 7576395]
17. Donovan JL. Patient decision making. The missing ingredient in compliance research. Int J Technol Assess Health Care 1995;11(3):443-455. [doi: 10.1017/s0266462300008667] [Medline: 7591546]
18. Western Sydney Local Health District. Western Sydney Local Health District 2017-18 Year in Review. 2018. URL: https://www.wslhd.health.nsw.gov.au/ArticleDocuments/1232/WSLHD_year-in-review_2017-18.pdf.aspx [accessed 2021-09-20]
19. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
20. Kirchhof P, Benussi S, Kotecha D, Ahlsson A, Atar D, Casadei B, et al. 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. Eur J Cardiothorac Surg 2016 Nov;50(5):e1-e88. [doi: 10.1093/ejcts/ezw313] [Medline: 27663299]
21. Calkins H, Hindricks G, Cappato R, Kim Y, Saad EB, Aguinaga L, et al. 2017 HRS/EHRA/ECAS/APHRS/SOLAECE expert consensus statement on catheter and surgical ablation of atrial fibrillation. Europace 2018 Jan 01;20(1):e1-e160 [FREE Full text] [doi: 10.1093/europace/eux274] [Medline: 29016840]
22. Streiner D, Norman G, Cairney J. Health measurement scales: a practical guide to their development and use. Oxford: Oxford University Press; 2015.
23. Dawes J. Do Data Characteristics Change According to the Number of Scale Points Used? An Experiment Using 5-Point, 7-Point and 10-Point Scales. International Journal of Market Research 2008 Jan 01;50(1):61-104. [doi: 10.1177/147078530805000106]
24. Safeer RS, Keenan J. Health literacy: the gap between physicians and patients. Am Fam Physician 2005 Aug 01;72(3):463-468 [FREE Full text] [Medline: 16100861]
25. Nisbett RE, Wilson TD. The halo effect: Evidence for unconscious alteration of judgments. Journal of Personality and Social Psychology 1977;35(4):250-256. [doi: 10.1037/0022-3514.35.4.250]
26. Hart M. Improving out-patient clinic waiting times: methodological and substantive issues. Int J Health Care Qual Assur 1995;8(6):14-22. [doi: 10.1108/09526869510098813] [Medline: 10165401]
27. Stuart J, Rutherford RJ. Medical student concentration during lectures. Lancet 1978 Sep 02;2(8088):514-516. [doi: 10.1016/s0140-6736(78)92233-x] [Medline: 79879]
28. O'Connor AM, Wennberg JE, Legare F, Llewellyn-Thomas HA, Moulton BW, Sepucha KR, et al. Toward the 'tipping point': decision aids and informed patient choice. Health Aff (Millwood) 2007;26(3):716-725 [FREE Full text] [doi: 10.1377/hlthaff.26.3.716] [Medline: 17485749]
29. Tak HJ, Ruhnke GW, Meltzer DO. Association of patient preferences for participation in decision making with length of stay and costs among hospitalized patients. JAMA Intern Med 2013 Jul 08;173(13):1195-1205 [FREE Full text] [doi: 10.1001/jamainternmed.2013.6048] [Medline: 23712712]
30. Marshall WR, Rothenberger LA, Bunnell SL. The efficacy of personalized audiovisual patient-education materials. J Fam Pract 1984 Nov;19(5):659-663. [Medline: 6208312]
31. Mulrow C, Bailey S, Sönksen PH, Slavin B. Evaluation of an Audiovisual Diabetes Education Program: negative results of a randomized trial of patients with non-insulin-dependent diabetes mellitus. J Gen Intern Med 1987;2(4):215-219. [doi: 10.1007/BF02596442] [Medline: 2441013]
32. Hartzband P, Groopman J. Untangling the Web--patients, doctors, and the Internet. N Engl J Med 2010 Mar 25;362(12):1063-1066. [doi: 10.1056/NEJMp0911938] [Medline: 20335581]
33. Seaburg L, Hess EP, Coylewright M, Ting HH, McLeod CJ, Montori VM. Shared decision making in atrial fibrillation: where we are and where we should be going. Circulation 2014 Feb 11;129(6):704-710. [doi: 10.1161/CIRCULATIONAHA.113.004498] [Medline: 24515956]
34. Turner J, Kelly B. Emotional dimensions of chronic disease. West J Med 2000 Feb;172(2):124-128 [FREE Full text] [doi: 10.1136/ewjm.172.2.124] [Medline: 10693376]

35. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005 Aug 4;353(5):487-497. [doi: 10.1056/NEJMra050100] [Medline: 16079372]
36. Abu AM, Himmel W, Vormfelde S, Koschack J. Video-assisted patient education to modify behavior: a systematic review. Patient Educ Couns 2014 Oct;97(1):16-22. [doi: 10.1016/j.pec.2014.06.015] [Medline: 25043785]

Abbreviations

AF: atrial fibrillation
QR: Quick Response
REDCap: Research Electronic Data Capture

Edited by R Kukafka; submitted 23.12.20; peer-reviewed by P van der Keylen, HYC Wong; comments to author 15.03.21; revised version received 11.04.21; accepted 05.07.21; published 11.10.21.

Please cite as:
Kovoor JG, McIntyre D, Chik WWB, Chow CK, Thiagalingam A
Clinician-Created Educational Video Resources for Shared Decision-making in the Outpatient Management of Chronic Disease: Development and Evaluation Study
J Med Internet Res 2021;23(10):e26732
URL: https://www.jmir.org/2021/10/e26732
doi:10.2196/26732
PMID:34633292

Original Paper

Improvement and Evaluation of the TOPCOP Taxonomy of Patient Portals: Taxonomy-Evaluation-Delphi (TED) Approach

Michael Glöggler[1], MSc, MEng; Elske Ammenwerth[1], MSc, MET, PhD

Institute of Medical Informatics, UMIT – Private University for Health Sciences, Medical Informatics and Technology, Hall in Tirol, Austria

Corresponding Author:
Michael Glöggler, MSc, MEng
Institute of Medical Informatics
UMIT – Private University for Health Sciences, Medical Informatics and Technology
Eduard-Wallnöfer-Zentrum 1
Hall in Tirol, 6060
Austria
Phone: 43 50 8648 ext 3809
Fax: 43 50 8648 673809
Email: michael.gloeggler@edu.umit-tirol.at

Abstract

Background: Patient portals have been introduced in many countries over the last 10 years, but many health information managers still feel they have too little knowledge of patient portals. A taxonomy can help them to better compare and select portals. This has led us to develop the TOPCOP taxonomy for classifying and comparing patient portals. However, the taxonomy has not been evaluated by users.

Objective: This study aimed to evaluate the taxonomy's usefulness to support health information managers in comparing, classifying, defining a requirement profile for, and selecting patient portals and to improve the taxonomy where needed.

Methods: We used a modified Delphi approach. We sampled a heterogeneous panel of 13 health information managers from 3 countries using the criterion sampling strategy. We conducted 4 anonymous survey rounds with qualitative and quantitative questions. In round 1, the panelists assessed the appropriateness of each dimension, and we collected new ideas to improve the dimensions. In rounds 2 and 3, the panelists iteratively evaluated the taxonomy that was revised based on round 1. In round 4, the panelists assessed the need for a taxonomy and the appropriateness of patient engagement as a distinguishing concept. Then, they compared 2 real portals with the final taxonomy and evaluated its usefulness for comparing portals, creating an initial requirement profile, and selecting patient portals. To determine group consensus, we applied the RAND/UCLA Appropriateness Method.

Results: The final taxonomy consists of 25 dimensions with 65 characteristics. Five new dimensions were added to the original taxonomy, with 8 characteristics added to already existing dimensions. Group consensus was achieved on the need for such a taxonomy to compare portals, on patient engagement as an appropriate distinguishing concept, and on the comprehensibility of the taxonomy's form. Further, consensus was achieved on the taxonomy's usefulness for classifying and comparing portals, assisting users in better understanding portals, creating a requirement profile, and selecting portals. This allowed us to test the usefulness of the final taxonomy with the intended users.

Conclusions: The TOPCOP taxonomy aims to support health information managers in comparing and selecting patient portals. By providing a standardized terminology to describe various aspects of patient portals independent of clinical setting or country, the taxonomy will also be useful for advancing research and evaluation of patient portals.

(J Med Internet Res 2021;23(10):e30701) doi:10.2196/30701

KEYWORDS

taxonomy; classification system; patient portal; EHR portal; online EHR access; evaluation; Delphi study; electronic health records; digital health; health information; information management; user perspectives

Introduction

Background

The delivery of knowledge-based care depends on patient engagement, where patients take an active role in their care [1-3]. Patient portals are considered a health information technology that promotes patient engagement [4-6] by providing patients with online tools to take an active role in their care [7-9]. Since patient portals are more than just static repositories for patient data [10,11], they support the new vision of health services that enable patient-provider information sharing [12,13], thus contributing to empowering patients [14,15], supporting shared decision making [16], and engaging patients actively in their care [6,17]. A patient portal is an internet-based application combining knowledge and software tools [18,19] that allow patients to have autonomous access to their electronic health record (EHR) anywhere at any time [20,21]. Besides its core function of providing EHR access [22], the features of a patient portal range from viewing visit notes, requesting medication refills, appointment scheduling, access to test and lab results, secure messaging with the health provider, e-visits, or reporting patient-generated health data [17,23-26]. Patient portals are used in different organizational settings such as independent physician practices and hospitals, group practices, or large, integrated health care delivery organizations [27-29].

The Need for a Taxonomy of Patient Portals

The widespread use of the internet, rise of mobile computing, and progress in patients' technical aptness have led to an increase in the use of patient portals in various countries such as the United States, Denmark, and Australia [12,28,30]. However, there are countries where patient portals are still not widely used [31,32]. A benchmarking study presented by Ammenwerth et al [33] in 2020 analyzed the eHealth progress of 14 countries worldwide with different health systems and different levels of economic development. The study showed that the use of patient portals and the provided patient portal functionalities vary significantly between the countries. While Finland and South Korea, for example, allow patients the best access to their health record data, 6 of the 14 analyzed countries do not offer their population any access to their health data online [33].

The low use of patient portals in both developing and high-income countries [31,34] creates a problematic situation for the health informatics professionals who are responsible for strategic and tactical information technology management in their health care institution or department; we will call these professionals "health information managers." On the one hand, there already exists a very heterogeneous landscape and a broad diversity of patient portals [12,27,35,36] regarding their intended deployment and functionalities [25,30,37]. In contrast, many health information managers still feel they have too little knowledge of patient portals [32]. They admit having difficulties understanding the portals' various application areas and scopes, defining their general requirements, and selecting a patient portal for their specific context or problem [32]. Health information managers are responsible for planning, organizing, and following up on all activities related to health information technology [38,39]. This also involves selecting, introducing, and managing patient portals for their health care institution [32].

To support health information managers with a tool for comparing patient portals and defining which general type and functionalities of patient portals they need, we developed the TOPCOP taxonomy (Taxonomy of Patient Portals based on Characteristics of Patient Engagement) [32]. The need for such a patient portal taxonomy had already been stressed in a recently published Cochrane Review on the impact of patient portals [40]. The TOPCOP taxonomy is shown in Figure 1. A comprehensive description of the dimensions is published elsewhere [32].

Figure 1. The TOPCOP taxonomy of patient portals [32]: A patient portal is described regarding 7 aspects that cover 20 dimensions. A patient portal can be described by selecting 1 characteristic per dimension.

ASPECTS	DIMENSIONS	CHARACTERISTICS		
Portal Design	D1: Care Sector Target	primary care	secondary care	generic
	D2: Portal Specialization	universal	disease-specific	
	D3: Activity Monitoring	no insight	with insight	
	D4: Patient Target	outpatient	in & outpatient	
Management	D5: Appointment Booking	no booking	request	schedule
	D6: Prescription Renewal	no renewal	basic renewal	advanced renewal
	D7: Portal Customizability	not customizable	customizable	
Communication	D8: E-Consult	no e-consult	asynchronous	synchronous
	D9: System Notifications	no notifications	reminder	alerts
Instruction	D10: Patient Education	no education	non-personalized	personalized
	D11: Therapy Instructions	non-protocol-based	protocol-based	
Self-Management	D12: Health Monitoring	no monitoring	self-reported	self-tracked
	D13: Visit Preparation	no preparation	with preparation	
Self-Determination	D14: Declaration of Will	no registration	with registration	
	D15: Second Opinion	no inquiry	with inquiry	
	D16: Study Sign-Up	no sign-up	with sign-up	
Data Management	D17: Record Access	no control	shared control	full control
	D18: Records Management	no management	with management	
	D19: Health Data Amend	review	correct	delete
	D20: Health Data Upload	no upload	with upload	

The Development of the TOPCOP Taxonomy

A taxonomy is a classification system to classify similar objects of a domain into groups based on distinct characteristics and offers a set of decision rules [41-44]. The reduction of complexity, identifying similarities and differences among objects [45,46], and the understanding of interrelationships are major advantages of taxonomies [44,47,48]. Taxonomies thus support researchers and practitioners in better understanding a domain and distinguishing among its objects [49,50]. The TOPCOP taxonomy was built by applying the formal taxonomy-building method proposed by Nickerson et al [51]. This method specifies the necessary steps and integrates 2 optional, iterative development approaches to conceptionally and empirically build and empirically evaluate a taxonomy [51,52]. The TOPCOP taxonomy was created conceptionally based on a literature review to assess the characteristics and functionalities of patient portals [32]. It was evaluated empirically by classifying patient portals offered on the market and health providers' portals online [32].

A taxonomy is determined by the user's intended purpose, which guides the taxonomy's development by focusing on the specific phenomenon of interest [50,53]. Different users or different purposes may therefore lead to a different taxonomy [51,54]. We created the TOPCOP taxonomy for health information managers to classify and compare patient portals [32]. Further, the taxonomy should serve the health information managers in defining the general type and functionalities of patient portals and should help them select the most suitable solution offered on the market. The dimensions were built to distinguish among patient portals with the scope on patient engagement.

Since there is no objective metric to define the usefulness or quality of a taxonomy [44,51,55], the method by Nickerson et al [51] provides a set of conditions to determine usefulness. Applying these conditions during the building process, the taxonomy's usefulness was empirically validated by classifying all patient portals of interest with the taxonomy [32].

The Requirement for a Taxonomy's Evaluation

According to design science research, taxonomies are fundamental design artifacts to provide knowledge and understanding of a problem domain [49]. In design science research, the design process for an artifact is divided first into building and then into evaluating [56,57]. Following the design science research paradigm, one suitable criterion to evaluate a taxonomy is by having users assess its usefulness in achieving its intended purpose [58,59]. Therefore, we now wanted to evaluate and further improve the TOPCOP taxonomy together with the projected users, guided by the evaluation criteria of the taxonomy's usefulness related to its intended purpose.

Methods

The Delphi Technique

Overview

We applied a modified Delphi approach to evaluate the TOPCOP taxonomy. The Delphi technique is a qualitative method, first described by Dalkey and Helmer [60]. It is used

in many research areas such as business, policy science, education, health sciences, information science, and health informatics [61-65]. Since there is no consistency in the methods used for evaluating a taxonomy [49], we opted for the Delphi technique because it is commonly agreed that Delphi research elicits sound scientific evidence [66]. Further, various researchers argue that qualitative methods may be particularly appropriate for evaluating design artifacts [67] including taxonomies.

The Delphi technique was particularly suited for our study as it aims to obtain a highly reliable consensus of group opinions on the research items [68] and has been used by researchers to evaluate taxonomies in the past [69-73]. The method is adequate to explore a domain [63], elicit new evidence, and generate new ideas [68,74]. Since the aim of this study was to collect new ideas on dimensions, improve the existing TOPCOP taxonomy, and achieve consensus on the appropriateness of the dimensions by the taxonomy's users, we considered Delphi to be the best approach as it goes beyond collecting simple intuitive expert opinions [75]. Further, the method applies relatively rigorous control over the interviewing methods, the controlled opinion feedback, and the summary of the results [76].

Using a series of survey rounds delivered in multiple iterations, interspersed with controlled opinion feedback [76], we were able to collect new ideas and correlate the panelists' opinions on our research items anonymously to improve the taxonomy [60,62,75]. Further, the method is highly flexible [77], accommodating many variations [78-80], and can be used to conduct evaluation studies [81], allowing us to adapt and modify the technique to evaluate our taxonomy. While the classic Delphi aims to generate opinions from experts to make forecasts [60,82], we wanted to collect and correlate the opinions of the taxonomy's users to improve and evaluate the TOPCOP taxonomy. We now outline the methodological approaches applied to our study.

Selection of the Panelists

There are no agreed standards on how to select the participants for Delphi studies [75,83]. We applied a criterion sampling strategy, which is a preferred approach in many Delphi studies [84]. Since the taxonomy should help health information managers to compare patient portals, eligible individuals had to hold a role within a health provider's organization where they would be actively involved in a patient portal's selection process. We applied the snowball method [85] to reach out to potential panelists involving patient portal vendors known from the taxonomy development phase [32]. The snowball method has been legitimized and used in many other Delphi studies [75,86,87]. Since there is no standard method for identifying the best number of individuals for inclusion in a Delphi study [86,87], we determined this number based on our research aim and availability of expertise as proposed by several researchers [88-90].

We selected 13 health information managers from Germany, Switzerland, and Austria as we wanted to include panelists from countries with diverse health systems and different progress in eHealth [31,33]. By sampling panelists from different countries but who speak the same language, we aimed to avoid a possible language bias as the survey questions and the complex explanations of the dimensions could otherwise be misinterpreted [91]. Since homogeneous panels tend to find consensus more quickly than heterogeneous panels [92,93], we wanted to enhance credibility through diversity, considering the broadest possible range of participants' experiences with patient portals and geographic diversity to depict the real situation of different health systems. The educational backgrounds of health information managers are not uniform but may be diverse [39]. We, therefore, included panelists with different educational backgrounds in our sample. In Table 1, we present the final panel.

Table 1. Study sample selection.

Panelist number	Country	Gender	AE[a]	PB[b]	Role	WE[c] (years)	EPP[d] (years)	INST[e]	SPP[f]
1	Austria	Male	MEng	Computer engineer	Head of department	27	5	HCP[g]	Yes
2	Austria	Male	BEng	Computer scientist	Head of department	15	5	HCP	Yes
3	Austria	Female	BEng	Medical informatics	Project manager	26	1	HCP	Yes
4	Austria	Male	PhD	Electronics engineer	Head of department	25	1	HCP	Yes
5	Austria	Male	MEng	Medical informatics	Head of department	12	1	HCP	Yes
6	Austria	Male	BSc	Bioengineering	Head of department	33	10	HCP	Yes
7	Switzerland	Male	BEng	Medical informatics	System engineer	10	16	HCP	Yes
8	Switzerland	Female	MSc	eHealth management	Researcher	13	3	HCO[h]	No
9	Switzerland	Male	MSc	Medical informatics	Head of department	28	2	HCP	Yes
10	Germany	Male	BSc	System engineer	Head of department	20	7	HCP	Yes
11	Germany	Male	PhD	Medical informatics	Researcher	10	6	HCP	Yes
12	Germany	Male	MSc	Medical informatics	Head of eHealth	13	8	HCO	No
13	Germany	Female	MD	Physician	Head of eHealth	26	8	HCP	Yes

[a]AE: academic education.

[b]PB: professional background.

[c]WE: work experience.

[d]EPP: experience with patient portals.

[e]INST: institution.

[f]SPP: would be involved in selecting a patient portal.

[g]HCP: health care provider.

[h]HCO: health care organization.

Determination of the Number of Survey Rounds for Evaluation

As per common agreement, the number of survey rounds is guided by the nature of the study and the level of consensus achieved among the participants during each iteration [75,76,94]. Our study was planned to be performed in 4 assessment cycles, guided by the elicitation of evidence and the achievement of group consensus. In the first 3 rounds, we aimed to collect ideas to improve the taxonomy and to achieve group consensus on every single dimension of the taxonomy related to its appropriateness for comparing patient portals. In round 4, the panelists were asked to evaluate the final taxonomy as a whole as proposed by Wiliam and Black [95], related to its intended use. The survey rounds were performed between January 2021 and April 2021.

Achievement of Consensus—the RAND/UCLA Appropriateness Method

The goal of the Delphi technique is to achieve a consensus of opinions from a group of individuals concerning a particular topic or task [87,96,97]. However, there is no general agreement on what statistical aggregation or method is best to determine consensus [98]. Since we wanted to assess the appropriateness of the taxonomy, we considered the concept of the RAND/UCLA Appropriateness Method, called RAM [99], most suitable to determine achievement of consensus in our study. While the RAM method is widely used to determine the appropriateness of health care services [100,101], we applied the model's consensus measure to evaluate the appropriateness of the taxonomy's dimensions for classifying and comparing patient portals.

The RAM method uses the median to measure the central tendency of the panelists' ratings, and ratings should be spread over a 1-9 rating scale [89,99]. The RAM method offers various conditions to constitute disagreement of opinions [99], from which we chose DS9, the strictest definition of disagreement [99]. DS9 means that a dimension is appropriate for comparing patient portals if group consensus with a median of 7-9 without disagreement is achieved. Considering all ratings, disagreement exists when at least one rating is a 1 and at least one is a 9. A dimension is considered uncertain for comparing patient portals if group consensus achieves a median of 4-6 or if there is any median with disagreement. A dimension is considered inappropriate for comparing patient portals if group consensus achieves a median of 1-3 without disagreement. The DS9 measure was applied for all assessments to determine achievement of group consensus. The DS9 measure is summarized in Table 2.

Table 2. The DS9 RAND/UCLA Appropriateness Measure with dispersion: considering all ratings, at least one is a 1, and at least one is a 9.

Appropriateness	Panel median	Dispersion condition
Appropriate	7-9	Without disagreement
Uncertain	4-6	Or any median with disagreement
Inappropriate	1-3	Without disagreement

Applying Anonymity to Express Opinions Freely

The complete study was conducted anonymously, which means that none of our panelists knew who participated in the survey and no interaction was possible between them. Anonymity allowed greater freedom for our panelists to express their views [102,103] and opinions freely as it avoids the problem of dominant contributors possibly influencing individual opinions [104,105].

Introductory Conversations to Enhance Adherence of Panelists and to Create a Common Understanding of the Research Topics

We were aware of the known problem that participants might drop out [77] due to the time-consuming commitment, unforeseen shortage of time, loss of interest, or distraction between the rounds, risking a poor response rate [83]. To promote motivation and strengthen adherence, we conducted an introductory conversation with each panelist separately, as proposed by Daniel and White [106], using the Zoom video conferencing tool [107]. The scope was to give the panelists the possibility to ask questions related to the aim of the study, the research process, and their role in the study and to create a common understanding on all topics.

Set-Up of the Online Survey

All interviews were carried out with online questionnaires using a commercial survey product [108]. The survey was piloted by 4 different persons other than the researchers. The survey contained quantitative and qualitative questions. The quantitative ratings served to assess the dimensions' appropriateness (rounds 1-3) and the final taxonomy as a whole (round 4). To assess the dimensions, we presented only 1 dimension with its characteristics per page (example Figure 2) and added a comprehensive definition of the existing characteristics and the newly proposed characteristics to assure that all panelists had the same understanding of the dimensions.

We provided a Likert scale ranging from 1 to 9 as proposed by RAM [99] for assessment. The open-ended qualitative questions provided in rounds 1-3 allowed the panelists to comment on their ratings if the rating fell into values between 1 and 6. Further, the qualitative questions allowed the panelists to improve the taxonomy by making proposals for new characteristics or dimensions. All proposals had to follow the knowledge-guiding principle of being suitable for promoting patient engagement. The comments and proposals were presented in the subsequent round for the panelists' reflection. Comments related to ratings, to changes of existing dimensions, or to new characteristics and dimensions were assigned accordingly. The comments were quoted verbatim with no changes made to the original. At the beginning of each survey round, we provided short guidelines on evaluating the research items and presented the results of the previous round. In round 1, we presented the initial TOPCOP taxonomy (Figure 1). In round 2, we presented the ratings from the first round (Multimedia Appendix 1) demonstrating achievement of group consensus. In round 3, we presented the results from the second round demonstrating for which new characteristics and dimensions group consensus was achieved or not achieved (Multimedia Appendix 2). In round 4, we presented the jointly improved, final taxonomy.

Figure 2. Example for the display of a single dimension for rating.

The 4 Rounds of the TOPCOP Taxonomy's Improvement and Evaluation

Round 1: Assessment of the Existing Dimensions and Proposals for Improvement of the Taxonomy

In the first round, the panelists were asked to assess the appropriateness of the existing TOPCOP taxonomy's dimensions for classifying and comparing patient portals. Further, they were asked to propose unsuitable or missing characteristics related to the existing dimensions and to suggest new dimensions to improve the taxonomy guided by their needs.

Round 2: Assessment of the Newly Proposed Characteristics and Dimensions of Round 1

In the second round, the panelists were first asked to assess the proposals of round 1 for new characteristics to refine existing dimensions. Each proposal was presented with all existing and all new characteristics (Multimedia Appendix 3). Related to the proposals of adding new characteristics to the existing dimensions or merge characteristics, we stressed that the panelists should evaluate the appropriateness of the new or merged characteristic of improving the existing dimension to compare patient portals. Then, the panelists were asked to assess the appropriateness of the new dimensions proposed in round 1 for classifying and comparing patient portals.

Round 3: Re-Evaluation of Dimensions Where Group Consensus Was Not Achieved in Previous Rounds

In round 3, the panelists were asked to re-evaluate those dimensions and characteristics proposed in round 1 but where no group consensus could be reached in round 2. From some panelists' comments, we understood that their assessments were guided by national legal requests rather than by evaluating a general area of application. We, therefore, added a note stressing that the scope of this study was to create a generally applicable taxonomy and that specific national requirements should not guide the rating. Since the panelists were to re-evaluate dimensions already assessed in round 2, we provided all panelists with their first rating compared to the group ratings (Figure 3) as recommended by RAM [99]. This was intended to help them better reflect on their rating considering the group opinion.

Figure 3. Example of a panelist's rating in comparison with the group ratings.

Round 4: Evaluation of the Final TOPCOP Taxonomy as a Whole

Since a taxonomy is complete and adequate when it satisfies the requirements of the purpose for which it was built [58], in round 4, the panelists were asked to evaluate the usefulness of the final taxonomy as a whole [109], related to its intended use. Round 4 was divided into 2 consecutive steps: first performing a case study and then assessing the taxonomy's usefulness.

First, evaluation is based on comparison [110]. Therefore, the panelists were requested to classify and compare 2 real-world patient portals with the final TOPCOP taxonomy to test its usefulness in a case study. We provided the panelists with 2 anonymized product descriptions from important software companies along with instructions on how to carry out the comparison. Both patient portals could be used for any care sector. However, one was a tethered patient portal while the other was an integrated patient portal. We selected these 2 patient portals because they differ in many characteristics, allowing the panelists to see the taxonomy's usefulness in comparing very different patient portals.

Second, after the case study was performed, the panelists were asked to assess the usefulness of the taxonomy as a whole. To investigate the panelists' opinions related to taxonomies for patient portals, we started with the following questions: (Q1) How important do you consider the need for a taxonomy for comparing patient portals? (Q2) How suitable do you consider patient engagement as a guiding concept for comparing patient portals?

Further, the health information managers were asked to make proposals for other guiding concepts that they considered useful to compare patient portals. Since we determined patient engagement as a guiding concept to distinguish among patient portals for the TOPCOP taxonomy, we aimed to collect alternative proposals suitable for future research: (Q3) What other guiding concepts may be appropriate for comparing patient portals?

To assess the taxonomy's usefulness as a whole [111], 6 research questions related to the performed case study were presented. Since understanding an artifact is a fundamental requirement for its usefulness, the panelists were first asked to evaluate whether the final taxonomy was understandable: (Q4) How understandable is the form and structure of the final taxonomy? Then, they were asked to assess whether the improved TOPCOP taxonomy is useful related to its intended use: (Q5) How useful is the final taxonomy for classifying patient portals following patient engagement? (Q6) How useful is the final taxonomy for comparing patient portals following patient engagement? (Q7) How useful is the final taxonomy for assisting you in better understanding patient portals based on characteristics supporting patient engagement? (Q8) How useful is the final taxonomy for creating an initial requirement profile for patient portals based on characteristics supporting patient engagement? (Q9) How useful is the final taxonomy for selecting patient portals offered on the market based on characteristics supporting patient engagement?

Question Q3 was set up as an open-ended question to collect the panelists' proposals in the best possible way [112]. To categorize the proposals, we analyzed the responses by applying the summarizing content analysis [113], an inductive analysis method proposed by Mayring [114]. All other items were assessed by applying the RAM approach [99].

Results

Results of Round 1: Assessment of the Existing Dimensions and Proposals for Improvement of the Taxonomy

The panelists were asked to evaluate the TOPCOP taxonomy by assessing every single dimension related to its

appropriateness for classifying and comparing patient portals. In Figure 4, we present the assessment for the 20 dimensions of the initial TOPCOP taxonomy (Figure 1) indicating the median for each dimension. All 13 panelists evaluated all 20 dimensions. Since each dimension's median ranged between 7 and 9 without disagreement, group consensus on the dimensions' appropriateness was achieved for all 20 dimensions [99].

Figure 4. Achieved consensus for the existing dimensions of the TOPCOP taxonomy after round 1. The grey column shows the median value without disagreement. All 13 panelists assessed all dimensions.

ASSESSED DIMENSIONS	RATINGS												
D1: Care Sector Target	7	7	7	8	8	9	9	9	9	9	9	9	9
D2: Portal Specialization	6	7	7	8	8	8	8	8	8	9	9	9	9
D3: Activity Monitoring	5	6	7	7	8	8	8	9	9	9	9	9	9
D4: Patient Target	5	5	7	7	7	7	8	8	8	9	9	9	9
D5: Appointment Booking	8	8	8	8	9	9	9	9	9	9	9	9	9
D6: Prescription Renewal	2	6	7	7	7	7	8	8	8	9	9	9	9
D7: Portal Customizability	5	6	7	7	7	7	8	8	9	9	9	9	9
D8: E-Consult	7	8	8	8	8	8	9	9	9	9	9	9	9
D9: System Notifications	6	7	8	8	8	8	8	8	9	9	9	9	9
D10: Patient Education	7	7	7	8	8	8	9	9	9	9	9	9	9
D11: Therapy Instructions	5	5	7	7	7	7	8	8	8	8	9	9	9
D12: Health Monitoring	7	7	7	8	8	9	9	9	9	9	9	9	9
D13: Visit Preparation	7	7	7	8	8	8	8	9	9	9	9	9	9
D14: Declaration of Will	6	6	7	7	7	8	8	8	8	9	9	9	9
D15: Second Opinion	4	7	7	7	7	7	7	7	7	8	9	9	9
D16: Study Sign-Up	5	6	7	7	7	7	7	7	7	8	8	9	9
D17: Record Access	6	7	7	9	9	9	9	9	9	9	9	9	9
D18: Records Management	3	5	7	8	8	8	8	9	9	9	9	9	9
D19: Health Data Amend	7	7	8	8	9	9	9	9	9	9	9	9	9
D20: Health Data Upload	7	8	8	8	9	9	9	9	9	9	9	9	9

Further, the panelists were asked to propose unsuitable and missing characteristics to refine the existing dimensions and to suggest new dimensions to improve the taxonomy.

Two panelists proposed refining dimension D6 Prescription Renewal by merging the characteristics "basic renewal" and "advanced renewal" to create the characteristic "with renewal" instead. They argued that differentiating the dimension into the initial 2 characteristics is confusing rather than strengthening

the distinguishability of patient portals. Seven panelists proposed 8 new characteristics to improve the existing dimensions D1, D2, D5, D6, D8, D9, D11, and D12 (Figure 5 shows the content of each dimension). Five panelists proposed the new dimensions Account Protection, App Expandability, Medical Specialty, Medication Summary, Portal Type, and Web Accessibility to enhance the taxonomy. We present all the proposals for improvement of round 1 in Figure 5.

Figure 5. Proposals from round 1 for new characteristics and dimensions to improve the taxonomy.

DIMENSIONS	CHARACTERISTICS			
D1: Care Sector Target	primary care	secondary care	tertiary care	generic
D2: Portal Specialization	universal	extended	disease-specific	
D3: Activity Monitoring	no insight	with insight		
D4: Patient Target	outpatient	in & outpatient		
D5: Appointment Booking	no booking	request	schedule	hybrid
D6: Prescription Renewal	no renewal	with renewal		
D7: Portal Customizability	not customizable	customizable		
D8: E-Consult	no e-consult	asynchronous	synchronous	both
D9: System Notifications	no notifications	notifications	reminder	alerts
D10: Patient Education	no education	non-personalized	personalized	
D11: Therapy Instructions	no instructions	non-protocol-based	protocol-based	
D12: Health Monitoring	no monitoring	self-reported	self-tracked	combined
D13: Visit Preparation	no preparation	with preparation		
D14: Declaration of Will	no registration	with registration		
D15: Second Opinion	no inquiry	with inquiry		
D16: Study Sign-Up	no sign-up	with sign-up		
D17: Record Access	no control	shared control	full control	
D18: Records Management	no management	with management		
D19: Health Data Amend	review	correct	delete	
D20: Health Data Upload	no upload	with upload		
D21: Account Protection	low	medium	strong	
D22: App Expandability	not expandable	expandable		
D23: Medical Specialty	generic	specialized		
D24: Medication Summary	no summary	with summary		
D25: Portal Type	tethered	integrated		
D26: Web Accessibility	not supported	supported		

Legend
Existing dimension
Existing characteristic
New characteristc for existing dimension
New dimension
New characteristic for new dimension

Results of Round 2: Assessment of the Newly Proposed Characteristics and Dimensions of Round 1

In round 2, the panelists had to assess the proposals from round 1 (Figure 5). They were asked to evaluate the merger of the characteristics of dimension D6. Further, they assessed the appropriateness of the suggested characteristics of dimensions D1, D2, D5, D8, D9, D11, and D12.

All 13 panelists assessed the proposed 8 characteristics and justified their rating whenever it fell between 1 and 6. As demonstrated in Figure 6, the median assessment for the appropriateness of all changes ranged between 7 and 9 without

disagreement. Therefore, all 8 characteristics were appropriate for improving the taxonomy and became part of the taxonomy [99].

The panelists were further requested to evaluate the appropriateness of the proposed dimensions D21 Account Protection, D22 App Expandability, D23 Medical Specialty, D24 Medication Summary, D25 Portal Type, and D26 Web Accessibility.

In Figure 7, we demonstrate that for all these dimensions, the median ranged between 7 and 8. However, the condition for disagreement [99] was fulfilled for dimensions D21 Account Protection and D26 Web Accessibility. Therefore, only dimensions D22, D23, D24, and D25 were considered appropriate for improving the taxonomy and became part of the taxonomy. Since no panelist made any proposal for changing an existing dimension or for a new dimension in round 2, only dimensions D21 and D26 became subject to re-evaluation in round 3.

In Figure 8, we present the taxonomy in progress after round 2 showing for which characteristics group consensus was achieved and for which dimensions no group consensus was achieved.

Figure 6. Achieved consensus by all 13 panelists on new characteristics for existing dimensions proposed in round 1. The grey column shows the median without disagreement. *New characteristics to improve the dimension.

DIMENSIONS	CHARACTERISTICS *	RATINGS												
D1: Care Sector Target	tertiary care	5	6	8	8	8	8	8	9	9	9	9	9	9
D2: Portal Specialization	extended	2	4	5	6	7	7	7	7	8	8	9	9	9
D5: Appointment Booking	hybrid	2	4	6	7	7	8	9	9	9	9	9	9	9
D6: Prescription Renewal	renewal	4	7	7	7	8	8	8	8	8	9	9	9	9
D8: E-Consult	both	2	6	7	7	8	8	8	8	9	9	9	9	9
D9: System Notifications	notifications	2	6	7	7	7	7	7	8	8	9	9	9	9
D11: Therapy Instructions	no instructions	7	7	7	8	8	8	8	8	9	9	9	9	9
D12: Health Monitoring	combined	2	7	7	8	8	8	9	9	9	9	9	9	9

Figure 7. Achieved consensus by all 13 panelists on new dimensions proposed in round 1.

ASSESSED NEW DIMENSIONS	RATINGS												
D21: Account Protection	1	1	1	2	3	5	7	7	7	8	9	9	9
D22: App Expandability	2	3	5	5	7	7	7	8	8	8	8	8	9
D23: Medical Specialty	2	3	3	4	4	5	7	7	8	8	8	8	9
D24: Medication Summary	7	7	7	8	8	8	8	8	8	9	9	9	9
D25: Portal Type	2	5	7	8	8	8	8	8	8	9	9	9	9
D26: Web Accessibility	1	1	3	7	7	8	8	9	9	9	9	9	9

Legend

Median without disagreement.

Median with disagreement: considering all ratings, at least one is a 1, and at least one is a 9.

Figure 8. The TOPCOP taxonomy in progress after round 2.

DIMENSIONS	CHARACTERISTICS			
D1: Care Sector Target	primary care	secondary care	tertiary care	generic
D2: Portal Specialization	universal	extended	disease-specific	
D3: Activity Monitoring	no insight	with insight		
D4: Patient Target	outpatient	in & outpatient		
D5: Appointment Booking	no booking	request	schedule	hybrid
D6: Prescription Renewal	no renewal	with renewal		
D7: Portal Customizability	not customizable	customizable		
D8: E-Consult	no e-consult	asynchronous	synchronous	both
D9: System Notifications	no notifications	notifications	reminder	alerts
D10: Patient Education	no education	non-personalized	personalized	
D11: Therapy Instructions	no instructions	non-protocol-based	protocol-based	
D12: Health Monitoring	no monitoring	self-reported	self-tracked	combined
D13: Visit Preparation	no preparation	with preparation		
D14: Declaration of Will	no registration	with registration		
D15: Second Opinion	no inquiry	with inquiry		
D16: Study Sign-Up	no sign-up	with sign-up		
D17: Record Access	no control	shared control	full control	
D18: Records Management	no management	with management		
D19: Health Data Amend	review	correct	delete	
D20: Health Data Upload	no upload	with upload		
D21: Account Protection	low	medium	strong	
D22: App Expandability	not expandable	expandable		
D23: Medical Specialty	generic	specialized		
D24: Medication Summary	no summary	with summary		
D25: Portal Type	tethered	integrated		
D26: Web Accessibility	not supported	supported		

Legend: Proposed characteristic with consensus. Proposed characteristic without consensus. Proposed dimension with consensus. Proposed dimension without consensus.

Results of Round 3: Re-Evaluation of Dimensions Where Group Consensus Was Not Achieved in Previous Rounds

In round 3, the panelists were asked to again assess the new dimensions D21 Account Protection and D26 Web Accessibility as group consensus was not achieved in round 2. All 13 panelists assessed both dimensions.

As demonstrated in Figure 9, a median of 8 without disagreement [99] was achieved for dimension D26. Dimension D26 was therefore appropriate and became part of the taxonomy. For dimension D21, a median of 7 was achieved. However, as at least one rating is a 1 and at least one rating is a 9, disagreement existed among the panelists [99]. All 5 panelists who assessed dimension D21 with values of 1 and 3 argued consistently that a patient portal must provide the highest data protection due to legal or patient requirements. Therefore, as strong account protection is a mandatory requirement, dimension D21 is not appropriate for distinguishing among patient portals. Comparing the ratings of dimension D21 for rounds 2 and 3 showed that, besides the fact that disagreement was re-confirmed in round 3, 3 ratings deteriorated (Figure 9), which means that group consensus converged even more strongly towards disapproval of dimension D21. To avoid the known risk of fatiguing the panelists with too many evaluation rounds [92], we did not launch another evaluation round. Since no group

consensus was achieved, dimension D21 was not integrated into the taxonomy.

After assigning the new dimensions to suitable aspects and organizing and numbering the dimensions accordingly, the final TOPCOP taxonomy resulted in 25 dimensions based on 65 characteristics assigned to 7 aspects and is presented in Figure 10. In Multimedia Appendix 4, we provide a detailed description of the dimensions and characteristics.

Figure 9. Achieved consensus by all 13 panelists on dimensions D21 and D26 after round 3 in comparison with the consensus in round 2.

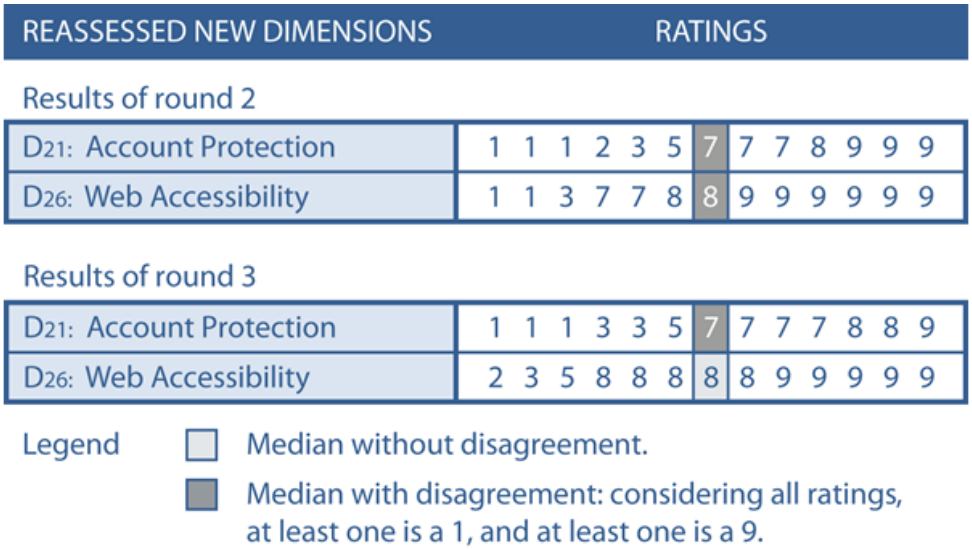


Figure 10. The final and user-evaluated TOPCOP taxonomy of patient portals.

ASPECTS	DIMENSIONS	CHARACTERISTICS			
Portal Design	D1: Portal Type	tethered		integrated	
	D2: Care Sector Target	primary care	secondary care	tertiary care	generic
	D3: Patient Target	outpatient		in & outpatient	
	D4: Portal Specialization	universal	extended	disease-specific	
	D5: Medical Specialty	generic		specialized	
	D6: Web Accessibility	not supported		supported	
	D7: App Expandability	not expandable		expandable	
	D8: Activity Monitoring	no insight		with insight	
Management	D9: Appointment Booking	no booking	request	schedule	hybrid
	D10: Prescription Renewal	no renewal		with renewal	
	D11: Portal Customizability	not customizable		customizable	
Communication	D12: E-Consult	no e-consult	asynchronous	synchronous	both
	D13: System Notifications	no notifications	notifications	reminder	alerts
Instruction	D14: Patient Education	no education	non-personalized	personalized	
	D15: Therapy Instructions	no instructions	non-protocol-based	protocol-based	
Self-Management	D16: Medication Summary	no summary		with summary	
	D17: Health Monitoring	no monitoring	self-reported	self-tracked	combined
	D18: Visit Preparation	no preparation		with preparation	
Self-Determination	D19: Declaration of Will	no registration		with registration	
	D20: Second Opinion	no inquiry		with inquiry	
	D21: Study Sign-Up	no sign-up		with sign-up	
Data Management	D22: Record Access	no control	shared control	full control	
	D23: Records Management	no management		with management	
	D24: Health Data Amend	review	correct	delete	
	D25: Health Data Upload	no upload		with upload	

Results of Round 4: Evaluation of the Final TOPCOP Taxonomy as a Whole

In round 4, the panelists were asked to assess the general need for a taxonomy, the appropriateness of patient engagement as a guiding concept, and the TOPCOP taxonomy's usefulness related to its intended purpose. All 13 panelists participated in round 4.

In Figure 11, we present the evaluations' results for the research questions Q1, Q2, and Q4–Q9. Since group consensus was achieved without disagreement [99] for all research questions, no further interview round was launched.

Figure 11. Achieved consensus by all 13 panelists on the research questions Q1, Q2, Q4–Q9. The grey column shows the median without disagreement. p: patient.

RESEARCH QUESTIONS	RATINGS
Q1: How important do you consider the need for a taxonomy for comparing p. portals?	6 6 7 7 7 8 **8** 9 9 9 9 9 9
Q2: How suitable do you consider patient engagement as a guiding concept for comparing p. portals?	3 7 8 8 8 8 **8** 8 8 8 8 9 9
Q4: How understandable is the form and structure of the final taxonomy?	4 7 7 7 8 8 **8** 8 8 8 8 8 9
Q5: How useful is the final taxonomy for classifying p. portals following patient engagement?	7 7 7 7 7 7 **8** 8 8 8 9 9 9
Q6: How useful is the final taxonomy for comparing p. portals following patient engagement?	2 7 7 7 7 7 **8** 8 8 8 8 9 9
Q7: How useful is the final taxonomy for assisting you in better understanding p. portals based on characteristics supporting patient engagement?	3 6 6 7 7 7 **7** 8 8 8 9 9 9
Q8: How useful is the final taxonomy for creating an initial requirement profile for p. portals based on characteristics supporting patient engagement?	6 7 7 7 7 7 **8** 8 8 8 9 9 9
Q9: How useful is the final taxonomy for selecting p. portals offered on the market based on characteristics supporting patient engagement?	6 7 7 7 8 8 **8** 8 8 8 9 9 9

The results shown in Figure 11 can be interpreted as follows. The panelists clearly agreed that there is a need for a taxonomy to distinguish among patient portals (median of 8 for Q1). Twelve panelists considered patient engagement to be an appropriate distinguishing concept for comparing patient portals (median of 8 for Q2). Only the panelist who assessed question Q2 with a rating of 3 proposed "System Architecture, Data Types, and Interoperability" as a more appropriate concept for comparing patient portals. Further, the panelists were asked to propose alternative distinguishing concepts appropriate for comparing patient portals (Q3). Since research question Q3 was an open-ended question, it is not part of Figure 11. Therefore, we present the proposed alternative concepts in Table 3.

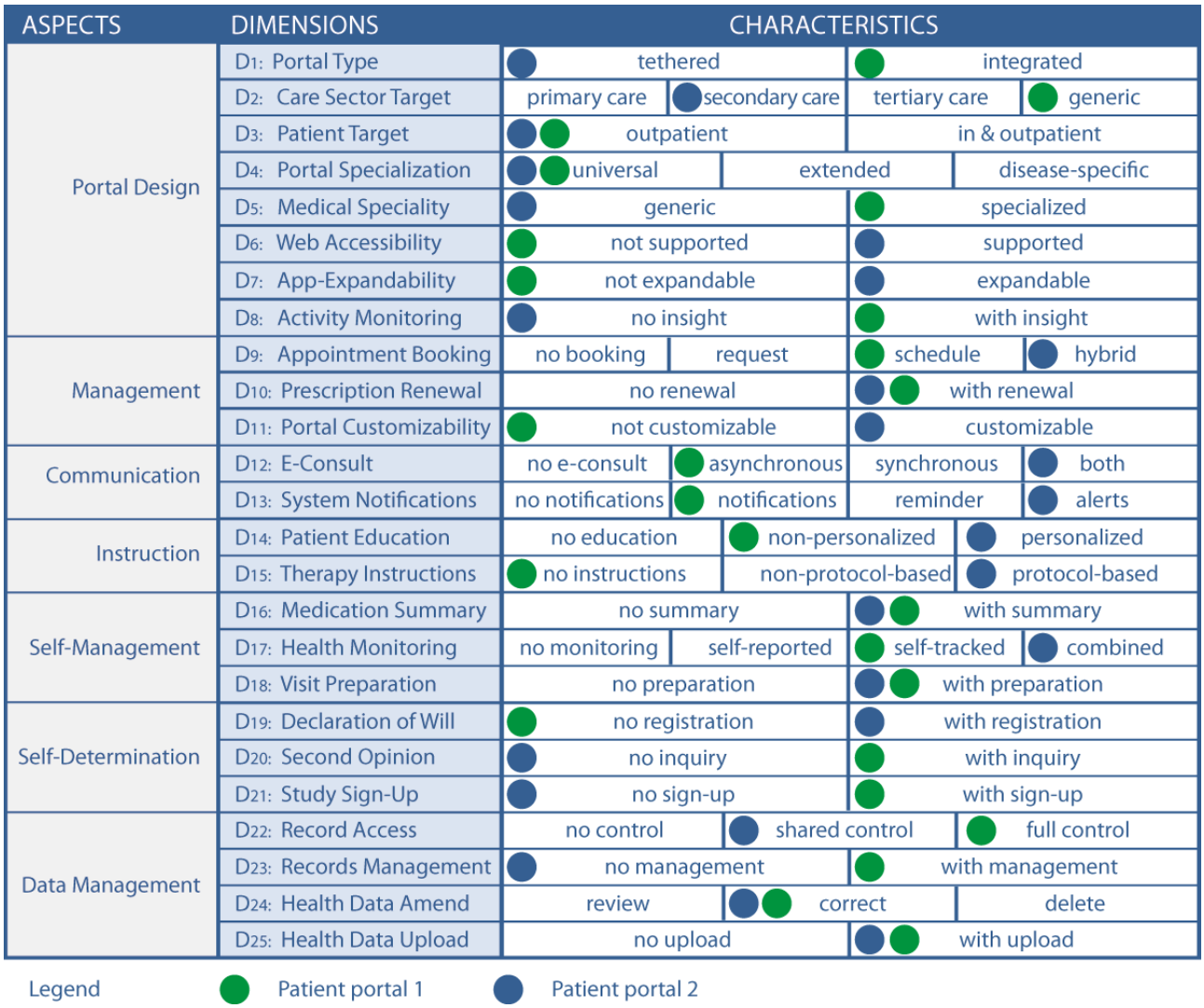
Table 3. Alternative distinguishing concepts proposed in round 4.

Proposed alternative distinguishing concepts (Q3)	Number of panelists proposing an alternative distinguishing concept
Comparison of patient portals based on characteristics promoting "Health Literacy"	1
Comparison of patient portals based on characteristics supporting "Improvement of Health Outcomes"	1
Comparison of patient portals based on characteristics related to "System Architecture, Data Types, and Interoperability"	3
Comparison of patient portals based on characteristics related to "Improvement of Work Efficiency and Cost Savings"	2

We continued interpreting the results presented in Figure 11, which relate to the case study and the assessment of the taxonomy's usefulness. With a median of 8, the panelists considered the form and structure of the final taxonomy to be understandable (Q4). However, 1 panelist who assessed Q4 with a rating of 4 argued that the taxonomy contains too many dimensions while, on the contrary, 1 panelist who assessed Q4 with a rating of 7 proposed refining the taxonomy with additional subcharacteristics to achieve a more accurate comparison of patient portals. All panelists considered the final taxonomy to be appropriate for classifying patient portals, giving ratings between 7 and 9 with a median of 8 (Q5). Further, they considered, with a median of 8, the taxonomy to be appropriate for comparing patient portals (Q6) and appropriate for better understanding of patient portals based on characteristics supporting patient engagement (median 7 for Q7).

Since the TOPCOP taxonomy is also intended to help health information managers select patient portals offered on the market, the panelists were requested to assess its usefulness in this regard. With a median of 8, the group consensus was achieved on both the taxonomy's usefulness for creating an initial requirement profile for patient portals (Q8) and selecting patient portals offered on the market based on characteristics supporting patient engagement (Q9). To sum up, by applying the TOPCOP taxonomy to compare 2 patient portals, we could indeed show that it is useful in contrasting and comparing patient portals from different vendors. In Figure 12, we show an example of a panelist's comparison. By marking each patient portal's characteristics, the differences and similarities of the 2 patient portals could be easily recognized.

Figure 12. Example of a comparison of 2 real patient portals carried out with the TOPCOP taxonomy by marking the respective characteristics.



Additional Findings: The Taxonomy-Evaluation-Delphi Approach (TED)

Evaluation is a challenging, essential, and crucial component of the research process [110,115]. One criterion for assessing artifacts such as taxonomies is by evaluating their usefulness related to their intended purpose [111,116]. However, there are only very few taxonomy-specific evaluation guidelines [117], but multiple evaluation approaches can be applied in health informatics [118-120]. Szopinski et al [49] analyzed the various approaches researchers applied to evaluate taxonomies in the information system's domain and demonstrated that the Delphi technique was hardly used. Analyzing 61 evaluation approaches, they found just 1 study where the Delphi technique was used

to evaluate the taxonomy [49]. In this study, panelists were asked to classify items into a deductively developed taxonomy and assess if the items were classified correctly [121]. In contrast, our modified Delphi approach aims to (1) first improve a conceptually and empirically created taxonomy [32] in multiple rounds by the users and (2) then evaluate the jointly created taxonomy by achieving users' consensus on the usefulness of the taxonomy. Therefore, the health information managers first compared the real-world patient portals with each other by performing a use case and then assessed the taxonomy related to its intended use. In Table 4, we describe the differences between the classic Delphi technique and our approach. The modified Delphi approach we used is, to the best of our knowledge, a new Delphi approach in health informatics for evaluating a taxonomy. We, therefore, call this approach the Taxonomy-Evaluation-Delphi (TED) approach.

Table 4. Comparison of the classic Delphi technique with the Taxonomy-Evaluation-Delphi (TED) approach.

Criteria/Delphi	Classic Delphi	TED approach
Objective	To make forecasts to plan ahead [60,82]	To collect new ideas to improve the taxonomy and to have the usefulness of a conceptually and empirically created taxonomy evaluated by the intended users
Approach	Obtain the most reliable consensus on the estimation of numerical quantity [105]	Obtain the most reliable consensus on the taxonomy's usefulness related to its intended purpose guided by the user's needs
Anonymity	No strict anonymity [60]	Strict anonymity
Consensus metric	Median without dispersion	Median with dispersion based on the RAM[a] method [99]
Panelists	Experts[b] with a deep understanding of the issues of concern [122]	The taxonomy's users with different levels of experience and understanding of the issue of concern
Number of rounds	Guided by the level of group consensus achieved [75,76,94]	Guided by elicitation of new evidence [95] and the level of group consensus achieved [75,76,94]
Procedure	Questionnaires and follow-up interviews [60]	Introductory conversations, online questionnaires, and performing a case study
Outset	Qualitative questions to collect initial knowledge to create/refine the research subject [60]	Quantitative questions to assess dimensions and characteristics, qualitative questions to collect new ideas to improve the taxonomy, and a case study to compare real patient portals as a basis for the taxonomy's evaluation
Result	Agreement on numerical quantities [60]	Improved and evaluated useful taxonomy based on the users' needs

[a]RAM: RAND/UCLA Appropriateness Method.

[b]There is ambiguity regarding the term expert concerning the Delphi technique as there is no unequivocal definition [82,96,123].

Discussion

Principal Findings

With this study, we were able to demonstrate that the need for a taxonomy to compare and classify patient portals exists among health information managers and that the concept of patient engagement to compare and select patient portals is considered appropriate.

Applying a modified Delphi approach, we improved the TOPCOP taxonomy based on the specific needs of the users. The final TOPCOP taxonomy consists of 25 dimensions with 65 characteristics, compared to 20 dimensions and 49 characteristics of the initial TOPCOP taxonomy.

We were able to demonstrate that the health information managers considered the final taxonomy to be useful in classifying and comparing patient portals. Further, we demonstrated that the final TOPCOP taxonomy supports the users in better understanding patient portals and assists them in selecting patient portals offered on the market. We were able to collect 4 alternative ideas on distinguishing concepts to compare patient portals that may serve for future research. As an additional outcome of our study, we created, to the best of our knowledge, a new Delphi approach in health informatics for evaluating a taxonomy.

Comparison With Prior Work

At present, there exists only a limited number of publications related to patient portal taxonomies. Ammenwerth et al [20] developed a taxonomy that aims to distinguish patient portals in a systematic review dealing with their effect on patient empowerment and health-related outcomes. Roehrs et al [124] developed a taxonomy that aims to identify open questions related to personal health record (PHR) data types, features, and architecture types. A PHR provides patients with web-based access to their health data that is under the control of the patient [124], while an EHR typically is under the control of the provider [7,19,22]. Fernández-Alemán et al [125] analyzed free web-based PHRs to identify their features and functions to better understand the PHR market. They created a framework of 4 dimensions intended to support patients in selecting a PHR that best fits their needs [125]. Scheplitz et al [126] created a framework for patient portal functionalities to record all possible functions to identify specification gaps related to software development. Walker et al [8] developed a framework to evaluate how well health information technology can support patient engagement by applying 5 engagement scoring levels.

These attempts only provide part of a potential patient portal taxonomy and are developed for different users and purposes. Since the user determines the intended purpose of a taxonomy

and the purpose guides the development by focusing on a specific phenomenon of interest, different users or purposes may lead to different taxonomies [46,51,55]. Further, a useful taxonomy must yield utility for a specific problem domain [58]. To sum up, the found taxonomies are not suitable to yield utility for health information managers for classifying and comparing patient portals based on characteristics appropriate for promoting patient engagement and understanding the differences and similarities. Therefore, the TOPCOP taxonomy was specifically developed for health information managers to compare and select patient portals offered on the market.

Limitations

Our approach to evaluating the TOPCOP taxonomy has some potential limitations.

First, panelists were selected from Germany, Austria, and Switzerland while the scope of the TOPCOP taxonomy is to support health information managers from any country. Since we were not able to attract participants from other countries, we aimed to assemble the panel as heterogeneously as possible, with different educational backgrounds and work experience, to achieve the best possible understanding of different viewpoints. Further, by selecting participants from 3 countries, we were able to map expertise from 3 different health care systems and integrate experiences with different levels of health care systems' digitization.

The second limitation relates to the panelists' experience with patient portals. Patient portals are not widely used in Germany, Austria, or Switzerland [32], and the experience with patient portals varied from little experience to much experience among the selected health information managers. Including participants with little experience with patient portals may lead to different results than if the participants had a deep understanding. As the taxonomy is not intended to only serve highly experienced but also inexperienced users, a composition of the panel that considers different levels of experience and understanding may increase the variety of viewpoints and the range of user needs related to the taxonomy. This variety may make the taxonomy even more useful [127].

The third limitation relates to the risk that the panelists may misunderstand what to evaluate. During the initial phone calls with potential participants, we noticed that some users assumed that the taxonomy's evaluation related to the suitability of functionalities for patient portals. However, the evaluation related to a dimension's appropriateness for classifying and comparing patient portals based on patient engagement. To ensure that there was no confusion, we explained the difference in individual introductory video conferences. Further, in the survey's introductory part, we outlined the scope of the evaluation and formulated the questionnaire's questions with unambiguous wording.

The fourth limitation is related to the Delphi technique itself. Delphi aims to obtain group consensus on opinions [68], but the achievement of consensus does not necessarily mean that the correct answer was found [84]. Besides, the composition of the panel may influence the research outcome [88,128]. To address these problems, we assembled the panel as heterogeneously as possible to integrate the broadest possible viewpoints and experience with patient portals. The selection of the panel was guided by the goal of achieving the best expertise available.

To determine the achievement of group consensus, we applied the RAND/UCLA concept appropriate for evaluation [101] and widely used to assess the appropriateness of health care services [100,129]. The survey was conducted anonymously to avoid the problem of dominant panelists possibly influencing individual opinions [102]. By applying all these measures, we believe that we were able to reduce any inherent bias in a possible method in the best way.

Practical Implication

The scope of the TOPCOP taxonomy is to serve health information managers with different degrees of knowledge related to patient portals and for various areas of application. The taxonomy may thus serve health information managers as a starting point to better understand the complex domain of patient portals since it describes the various aspects of patient portals. Further, 2 or more patient portals can be described by marking the respective characteristics. This shows the differences and similarities of the patient portals (Figure 12) and so supports the health information managers in classifying and comparing patient portals.

Since each health care institution may have different requirements related to a patient portal, the TOPCOP taxonomy can serve to create a requirement profile. By marking those characteristics in the taxonomy that best meet the needs of a health care institution, health information managers can create an initial requirement profile. This profile can then be used for a targeted search and selection of suitable portals offered on the market.

By providing a standardized terminology to describe various aspects of patient portals independent of clinical setting or country, the TOPCOP taxonomy is also useful for advancing research and evaluation of patient portals. It can, for example, be used to systematically describe patient portals as part of systematic reviews on their impact. The need for a taxonomy in this context has already been stressed in patient portal reviews [40].

Conclusions

The TOPCOP taxonomy aims to support health information managers in comparing and selecting patient portals. By providing a standardized terminology to describe various aspects of patient portals independent of clinical setting or country, the taxonomy will also be useful for advancing research and evaluation of patient portals. Since the health information managers contributed to the taxonomy's development, we were able to improve the taxonomy's quality and usefulness based on the users' needs.

The taxonomy consists of a manageable number of characteristics and dimensions and is therefore flexible for future changes. If needed, new dimensions can be added or removed according to future technological development. Further, due to its flexible form, the users can adjust the taxonomy to their

personal needs. The initial TOPCOP taxonomy was developed by analyzing patient portals from 15 countries worldwide. It was then improved by health information managers with various degrees of patient portal experience from 3 countries with different levels of health care digitization. We, therefore, consider our taxonomy suitable to compare and classify patient portals from any country. The taxonomy may also contribute to the progress of health care digitization as it may enhance human resources capacity and effectiveness.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Achieved consensus on existing dimensions after round 1.
[PDF File (Adobe PDF File), 400 KB - jmir_v23i10e30701_app1.pdf]

Multimedia Appendix 2
Consensus on new characteristics and dimensions after round 2 presented to the panelists in round 3.
[PDF File (Adobe PDF File), 402 KB - jmir_v23i10e30701_app2.pdf]

Multimedia Appendix 3
Presentation of a proposal of a new characteristic to refine an existing dimension.
[PDF File (Adobe PDF File), 284 KB - jmir_v23i10e30701_app3.pdf]

Multimedia Appendix 4
A detailed description of dimensions and characteristics.
[PDF File (Adobe PDF File), 190 KB - jmir_v23i10e30701_app4.pdf]

References

1. Beal LL, Kolman JM, Jones SL, Khleif A, Menser T. Quantifying Patient Portal Use: Systematic Review of Utilization Metrics. J Med Internet Res 2021 Feb 25;23(2):e23493 [FREE Full text] [doi: 10.2196/23493] [Medline: 33629962]
2. Nøst TH, Faxvaag A, Steinsbekk A. Participants' views and experiences from setting up a shared patient portal for primary and specialist health services- a qualitative study. BMC Health Serv Res 2021 Feb 24;21(171):2-9 [FREE Full text] [doi: 10.1186/s12913-021-06188-8] [Medline: 33627122]
3. Nazi KM, Turvey CL, Klein DM, Hogan TP. A Decade of Veteran Voices: Examining Patient Portal Enhancements Through the Lens of User-Centered Design. J Med Internet Res 2018 Jul 10;20(7):e10413 [FREE Full text] [doi: 10.2196/10413] [Medline: 29991468]
4. Wiljer D, Urowitz S, Apatu E, DeLenardo C, Eysenbach G, Harth T, et al. Patient accessible electronic health records: exploring recommendations for successful implementation strategies. J Med Internet Res 2008 Oct 31;10(4):e34 [FREE Full text] [doi: 10.2196/jmir.1061] [Medline: 18974036]
5. Vreugdenhil MMT, Ranke S, de Man Y, Haan MM, Kool RB. Patient and Health Care Provider Experiences With a Recently Introduced Patient Portal in an Academic Hospital in the Netherlands: Mixed Methods Study. J Med Internet Res 2019 Aug 20;21(8):13743 [FREE Full text] [doi: 10.2196/13743] [Medline: 31432782]
6. Stewart MT, Hogan TP, Nicklas J, Robinson SA, Purington CM, Miller CJ, et al. The Promise of Patient Portals for Individuals Living With Chronic Illness: Qualitative Study Identifying Pathways of Patient Engagement. J Med Internet Res 2020 Jul 17;22(7):e17744 [FREE Full text] [doi: 10.2196/17744] [Medline: 32706679]
7. Crameri K, Maher L, Van Dam P, Prior S. Personal electronic healthcare records: What influences consumers to engage with their clinical data online? A literature review. Health Inf Manag 2020 Jan 10:1833358319895369. [doi: 10.1177/1833358319895369] [Medline: 31918578]
8. Walker DM, Sieck CJ, Menser T, Huerta TR, Scheck McAlearney A. Information technology to support patient engagement: where do we stand and where can we go? J Am Med Inform Assoc 2017 Nov 01;24(6):1088-1094 [FREE Full text] [doi: 10.1093/jamia/ocx043] [Medline: 28460042]
9. Wass S, Vimarlund V. The Role of PAEHRs in Patient Involvement. J Med Syst 2018 Sep 25;42(11):210 [FREE Full text] [doi: 10.1007/s10916-018-1070-y] [Medline: 30255404]
10. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal Health Records: Definitions, Benefits, and Strategies for Overcoming Barriers to Adoption. Journal of the American Medical Informatics Association 2006 Mar 01;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.m2025]
11. Irizarry T, DeVito Dabbs A, Curran CR. Patient Portals and Patient Engagement: A State of the Science Review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]

12. Bouayad L, Ialynytchev A, Padmanabhan B. Patient Health Record Systems Scope and Functionalities: Literature Review and Future Directions. J Med Internet Res 2017 Nov 15;19(11):e388 [FREE Full text] [doi: 10.2196/jmir.8073] [Medline: 29141839]
13. Barbarito F, Pinciroli F, Barone A, Pizzo F, Ranza R, Mason J, et al. Implementing the lifelong personal health record in a regionalised health information system: the case of Lombardy, Italy. Comput Biol Med 2015 Apr;59:164-174. [doi: 10.1016/j.compbiomed.2013.10.021] [Medline: 24238969]
14. Risling T, Martinez J, Young J, Thorp-Froslie N. Evaluating Patient Empowerment in Association With eHealth Technology: Scoping Review. J Med Internet Res 2017 Sep 29;19(9):e329 [FREE Full text] [doi: 10.2196/jmir.7809] [Medline: 28963090]
15. Mossaed S, Leonard K, Eysenbach G. Patient Preferences and Perspectives on Accessing Their Medical Records. J Med Imaging Radiat Sci 2015 Jun;46(2):205-214. [doi: 10.1016/j.jmir.2014.11.001] [Medline: 31052095]
16. Prokosch H, Schüttler C, Schraudt M, Öfelein M, Maier C. Digital Patient Communication: Improving the Hospital-Patient Relationship. Stud Health Technol Inform 2019;259:3-9. [doi: 10.3233/978-1-61499-961-4-3] [Medline: 30923263]
17. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012 Nov 26;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]
18. Pinciroli F, Corso M, Fuggetta A, Masseroli M, Bonacina S, Marceglia S. Telemedicine and E-Health. IEEE Pulse 2011 May;2(3):62-70. [doi: 10.1109/mpul.2011.941524]
19. Tsai C, Eghdam A, Davoody N, Wright G, Flowerday S, Koch S. Effects of Electronic Health Record Implementation and Barriers to Adoption and Use: A Scoping Review and Qualitative Analysis of the Content. Life (Basel) 2020 Dec 04;10(12):A [FREE Full text] [doi: 10.3390/life10120327] [Medline: 33291615]
20. Ammenwerth E, Hoerbst A, Lannig S, Mueller G, Siebert U, Schnell-Inderst P. Effects of Adult Patient Portals on Patient Empowerment and Health-Related Outcomes: A Systematic Review. Stud Health Technol Inform 2019 Aug 21;264:1106-1110. [doi: 10.3233/SHTI190397] [Medline: 31438096]
21. Schaller M, Dornauer V, Hackl W, Lechleitner G, Überegger M, Ammenwerth E. Implementing National Electronic Health Records in Nursing Homes in Tyrol: A Nursing Management Perspective. Stud Health Technol Inform 2020 Jun 23;271:240-247. [doi: 10.3233/SHTI200102] [Medline: 32578569]
22. Rigby M, Georgiou A, Hyppönen H, Ammenwerth E, de Keizer N, Magrabi F, et al. Patient Portals as a Means of Information and Communication Technology Support to Patient- Centric Care Coordination - the Missing Evidence and the Challenges of Evaluation. A joint contribution of IMIA WG EVAL and EFMI WG EVAL. Yearb Med Inform 2015 Aug 13;10(1):148-159 [FREE Full text] [doi: 10.15265/IY-2015-007] [Medline: 26123909]
23. McCoy AB, Wright A, Eysenbach G, Malin BA, Patterson ES, Xu H, et al. State of the Art in Clinical Informatics: Evidence and Examples. Yearb Med Inform 2013 Mar 05;22(01):13-19. [doi: 10.1055/s-0038-1638827]
24. Ammenwerth E, Lannig S, Hörbst A, Mueller G, Schnell-Inderst P. Adult patient access to electronic health records. Cochrane Database Syst Rev 2017 Jul 27;6:CD012707 [FREE Full text] [doi: 10.1002/14651858.CD012707]
25. Alturkistani A, Greenfield G, Greaves F, Aliabadi S, Jenkins RH, Costelloe C. Patient Portal Functionalities and Uptake: Systematic Review Protocol. JMIR Res Protoc 2020 Jul 31;9(7):e14975. [doi: 10.2196/14975]
26. Nguyen OT, Alishahi Tabriz A, Huo J, Hanna K, Shea CM, Turner K. Impact of Asynchronous Electronic Communication-Based Visits on Clinical Outcomes and Health Care Delivery: Systematic Review. J Med Internet Res 2021 May 05;23(5):e27531 [FREE Full text] [doi: 10.2196/27531] [Medline: 33843592]
27. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J. What do we know about developing patient portals? a systematic literature review. J Am Med Inform Assoc 2016 Apr;23(e1):e162-e168 [FREE Full text] [doi: 10.1093/jamia/ocv114] [Medline: 26335985]
28. Nøhr C, Parv L, Kink P, Cummings E, Almond H, Nørgaard JR, et al. Nationwide citizen access to their health data: analysing and comparing experiences in Denmark, Estonia and Australia. BMC Health Serv Res 2017 Aug 07;17(1):534 [FREE Full text] [doi: 10.1186/s12913-017-2482-y] [Medline: 28784173]
29. Tavares J, Oliveira T. New Integrated Model Approach to Understand the Factors That Drive Electronic Health Record Portal Adoption: Cross-Sectional National Survey. J Med Internet Res 2018 Nov 19;20(11):e11032 [FREE Full text] [doi: 10.2196/11032] [Medline: 30455169]
30. Archer N, Fevrier-Thomas U, Lokker C, McKibbon KA, Straus SE. Personal health records: a scoping review. J Am Med Inform Assoc 2011;18(4):515-522 [FREE Full text] [doi: 10.1136/amiajnl-2011-000105] [Medline: 21672914]
31. Global diffusion of eHealth: making universal health coverage achievable: report of the third global survey on eHealth. World Health Organization. 2016 Dec 15. URL: https://www.who.int/publications/i/item/9789241511780 [accessed 2021-01-17]
32. Glöggler M, Ammenwerth E. Development and Validation of a Useful Taxonomy of Patient Portals Based on Characteristics of Patient Engagement. Methods Inf Med 2021 Jun 09;60(S 01):e44-e55 [FREE Full text] [doi: 10.1055/s-0041-1730284] [Medline: 34243191]
33. Ammenwerth E, Duftschmid G, Al-Hamdan Z, Bawadi H, Cheung NT, Cho K, et al. International Comparison of Six Basic eHealth Indicators Across 14 Countries: An eHealth Benchmarking Study. Methods Inf Med 2020 Dec;59(S 02):e46-e63 [FREE Full text] [doi: 10.1055/s-0040-1715796] [Medline: 33207386]

34. Lupiáñez-Villanueva F, Folkvord F, Faulí F, Altenhofer M, Hocking L, Harshfield A. Benchmarking deployment of eHealth among general practitioners - Final Report. Publications Office of the European Union. Luxembourg: European Union; 2018. URL: https://op.europa.eu/en/publication-detail/-/publication/d1286ce7-5c05-11e9-9c52-01aa75ed71a1 [accessed 2021-03-05]
35. Wildenbos GA, Horenberg F, Jaspers M, Peute L, Sent D. How do patients value and prioritize patient portal functionalities and usage factors? A conjoint analysis study with chronically ill patients. BMC Med Inform Decis Mak 2018 Nov 21;18(1):108 [FREE Full text] [doi: 10.1186/s12911-018-0708-5] [Medline: 30463613]
36. Shimada SL, Allison JJ, Rosen AK, Feng H, Houston TK. Sustained Use of Patient Portal Features and Improvements in Diabetes Physiological Measures. J Med Internet Res 2016 Jul 01;18(7):e179 [FREE Full text] [doi: 10.2196/jmir.5663] [Medline: 27369696]
37. Hazara AM, Durrans K, Bhandari S. The role of patient portals in enhancing self-care in patients with renal conditions. Clin Kidney J 2020 Feb;13(1):1-7 [FREE Full text] [doi: 10.1093/ckj/sfz154] [Medline: 32082547]
38. What do Health Information Managers Do? University of Wisconsin Health Information Management & Technology. 2021. URL: https://himt.wisconsin.edu/about-himt/what-him-managers-do/ [accessed 2021-02-20]
39. Bates M, Black C, Blair F, Davis L, Ingram S, Lane D, et al. Perceptions of health information management educational and practice experiences. Perspect Health Inf Manag 2014;11:1d [FREE Full text] [Medline: 25214821]
40. Ammenwerth E, Neyer S, Hörbst A, Mueller G, Siebert U, Schnell-Inderst P. Adult patient access to electronic health records. Cochrane Database Syst Rev 2021 Feb 26;2:CD012707. [doi: 10.1002/14651858.CD012707.pub2] [Medline: 33634854]
41. Ammenwerth E. Die Bewertung von Informationssystemen des Gesundheitswesens: Beiträge für ein umfassendes Informationsmanagement (UMIT-Schriftenreihe). Düren, Germany: Shaker Verlag; Sep 01, 2004.
42. Doty DH, Glick WH. Typologies As a Unique Form Of Theory Building: Toward Improved Understanding and Modeling. Acad Manag Rev 1994 Apr 01;19(2):230-251. [doi: 10.5465/amr.1994.9410210748]
43. Leiner F, Gaus W, Haux R, Knaup-Gregori P, editors. Medical Data Management: A Practical Guide, 1st edition. New York, NY: Springer Verlag; 2003:204.
44. Bailey KD. Typologies and Taxonomies: An Introduction to Classification Techniques. Thousand Oaks, CA: SAGE Publications, Inc; 1994:89.
45. Mrosek R, Dehling T, Sunyaev A. Taxonomy of health IT and medication adherence. Health Policy and Technology 2015 Sep;4(3):215-224. [doi: 10.1016/j.hlpt.2015.04.003]
46. Glass RL, Vessey I. Contemporary application-domain taxonomies. IEEE Softw 1995 Jul;12(4):63-76. [doi: 10.1109/52.391837]
47. Vegas S, Juristo N, Basili V. Maturing Software Engineering Knowledge through Classifications: A Case Study on Unit Testing Techniques. IIEEE Trans. Software Eng 2009 Jul;35(4):551-565. [doi: 10.1109/TSE.2009.13]
48. Bailey B, Martin G, Anderson T, editors. Taxonomies for the Development and Verification of Digital Systems. New York, NY: Springer US; 2005:180.
49. Szopinski D, Schoormann T, Kundisch D. Because your taxonomy is worth it: Towards a framework for taxonomy evaluation. 2019 Presented at: 27th European Conference on Information Systems (ECIS); June 8-14, 2019; Stock Uppsala, Sweden.
50. Lambe P. Organising Knowledge: Taxonomies, Knowledge and Organisational Effectiveness, 1st edition. Oxford, England: Chandos Publishing; Feb 28, 2007.
51. Nickerson RC, Varshney U, Muntermann J. A method for taxonomy development and its application in information systems. European Journal of Information Systems 2013;22(3):336-359 published online 19 June 2012. [doi: 10.1057/ejis.2012.26]
52. Remane G, Nickerson R, Hanelt A, Tesch J, Kolbe L. A Taxonomy of Carsharing Business Models. 2016 Presented at: 37th International Conference on Information Systems (ICIS); December 11-14, 2016; Dublin, Ireland.
53. De Los Reyes Arguelles E, Martinez-Goss M. Classical Taxonomy. Methods Microalgal Stud. Los Baños, Philippines: Philippine Science Letters and University of the Philippines Los Baños; Dec 2020:145-153.
54. Kwasnik BH. The Role of Classification in Knowledge Representation and Discovery. Library Trends 1999;48(1):22-47.
55. Nickerson RC, Varshney U, Muntermann J, Issac H. Taxonomy Development in Information Systems: Developing A Taxonomy Of Mobile Applications. Working Papers halshs-00375103. 2009. URL: https://ideas.repec.org/p/hal/wpaper/halshs-00375103.html [accessed 2020-07-15]
56. March ST, Smith GF. Design and natural science research on information technology. Decision Support Systems 1995 Dec;15(4):251-266. [doi: 10.1016/0167-9236(94)00041-2]
57. Venable J. A framework for Design Science research activities. In: 2006 Information Resources Management Association International Conference. Washington, DC: Idea Group Publishing; 2006.
58. Hevner AR, March ST, Park J, Ram S. Design Science in Information Systems Research. MIS Quarterly 2004;28(1):75-105. [doi: 10.2307/25148625]
59. Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S. A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems 2007;24(3):45-78. [doi: 10.2753/MIS0742-1222240302]
60. Dalkey N, Helmer O. An Experimental Application of the DELPHI Method to the Use of Experts. Management Science 1963 Apr;9(3):458-467. [doi: 10.1287/mnsc.9.3.458]

61. Strasser A. Delphi Method Variants in Information Systems Research: Taxonomy Development and Application. The Electronic Journal of Business Research Methods 2017;15(2):120-133 ISSN 1477-7029.
62. DeMuro PR, Ash J, Middleton B, Fletcher J, Madison CJ. A Quality, Benefit, Cost, and Financial Framework for Health Information Technology, E-Prescribing: A Delphi Study. Stud Health Technol Inform 2017;241:69-75 [FREE Full text] [doi: 10.3233/978-1-61499-794-8-69] [Medline: 28809185]
63. Brender J, Ammenwerth E, Nykänen P, Talmon J. Factors Influencing Success and Failure of Health Informatics Systems. Methods Inf Med 2006 Feb 06;45(1):125-136. [doi: 10.1055/s-0038-1634049]
64. Brender J, Nøhr C, McNair P. Research needs and priorities in Health Informatics - early results of a Delphi Study. In: Studies in Health Technology and Informatics. Amsterdam: IOS Press; 1999:191-196.
65. Nelson R, Staggers N. Health Informatics: An Interprofessional Approach. St. Louis, MI: Elsevier; Mar 31, 2015:678.
66. Fusfeld AR. Research Program on the Management of Science and Technology: The Delphi Technique, Survey and Comment. Cambridge, MA: Massachusetts Institute of Technology; Mar 1971.
67. Iivari J, Venable J. Action research and design science research - Seemingly similar but decisively dissimilar. : ECIS 2009 Proceedings. 73; 2009 Presented at: 17th European Conference on Information Systems (ECIS 2009); 2009; Verona, Italy URL: https://aisel.aisnet.org/ecis2009/73/
68. Flanagan T, Ashmore R, Banks D, MacInnes D. The Delphi method: Methodological issues arising from a study examining factors influencing the publication or non-publication of mental health nursing research. Mental Health Review Journal 2016 Jun 13;21(2):85-94. [doi: 10.1108/MHRJ-07-2015-0020]
69. Hanson CL, Oliver EJ, Dodd-Reynolds CJ, Pearsons A, Kelly P. A modified Delphi study to gain consensus for a taxonomy to report and classify physical activity referral schemes (PARS). Int J Behav Nutr Phys Act 2020 Dec 02;17(1):158 [FREE Full text] [doi: 10.1186/s12966-020-01050-2] [Medline: 33267840]
70. Ben-Chetrit E, Gattorno M, Gul A, Kastner DL, Lachmann HJ, Touitou I, et al. Consensus proposal for taxonomy and definition of the autoinflammatory diseases (AIDs): a Delphi study. Ann Rheum Dis 2018 Nov 12;77(11):1558-1565. [doi: 10.1136/annrheumdis-2017-212515] [Medline: 30100561]
71. Shoemaker LL, Lenker JA, Fuhrer MJ, Jutai JW, Demers L, DeRuyter F. Development and evaluation of a new taxonomy of mobility-related assistive technology devices. Am J Phys Med Rehabil 2010 Oct;89(10):795-808. [doi: 10.1097/PHM.0b013e3181f1bbcd] [Medline: 20855980]
72. Valentijn PP, Boesveld IC, van der Klauw DM, Ruwaard D, Struijs JN, Molema JJW, et al. Towards a taxonomy for integrated care: a mixed-methods study. Int J Integr Care 2015 Mar 04;15(1):e003 [FREE Full text] [doi: 10.5334/ijic.1513] [Medline: 25759607]
73. Chastin SFM, Schwarz U, Skelton DA. Development of a consensus taxonomy of sedentary behaviors (SIT): report of Delphi Round 1. PLoS One 2013 Dec 2;8(12):e82313 [FREE Full text] [doi: 10.1371/journal.pone.0082313] [Medline: 24312653]
74. Iqbal S, Pipon-Young L. The Delphi method. The Psychologist 2009;22(7):598-601.
75. Fink-Hafner D, Dagen T, Doušak M, Novak M, Hafner-Fink M. Delphi Method: Strengths and Weaknesses. Metodoloski zvezki 2019;16(2):1-19.
76. Linstone HA, Turoff M. The Delphi method : techniques and applications, 1st edition. Boston, MA: Addison-Wesley; 1975.
77. Donohoe HM, Needham RD. Moving best practice forward: Delphi characteristics, advantages, potential problems, and solutions. Int. J. Tourism Res 2009 Sep;11(5):415-437. [doi: 10.1002/jtr.709]
78. Erffmeyer RC, Erffmeyer ES, Lane IM. The Delphi Technique: An Empirical Evaluation of the Optimal Number of Rounds. Group & Organization Studies 1986 Mar;11(1-2):120-128. [doi: 10.1177/105960118601100110]
79. Hasson F, Keeney S. Enhancing rigour in the Delphi technique research. Technological Forecasting and Social Change 2011 Nov;78(9):1695-1704. [doi: 10.1016/j.techfore.2011.04.005]
80. Quyên DTN. Developing University Governance Indicators and their Weighting System Using a Modified Delphi Method. Procedia - Social and Behavioral Sciences 2014 Aug;141:828-833. [doi: 10.1016/j.sbspro.2014.05.144]
81. Erffmeyer RC. Decision-Making Formats: a Comparison on an Evaluative Task of Interacting Groups, Consensus Groups, the Nominal Group Technique, and the Delphi Technique. Louisiana State University. 1981. URL: https://core.ac.uk/download/pdf/217386485.pdf [accessed 2020-12-14]
82. du Plessis E, Human SP. The art of the Delphi technique: highlighting its scientific merit. Health SA Gesondheid 2007 Nov 17;12(4):a268. [doi: 10.4102/hsag.v12i4.268]
83. Hsu CC, Sandford BA. The Delphi Technique: Making Sense of Consensus. Pract Assessment, Research, and Evaluation 2007;12(10):1-8 [FREE Full text] [doi: 10.7275/pdz9-th90]
84. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs 2000;32(4):1008-1015. [doi: 10.1046/j.1365-2648.2000.t01-1-01567.x]
85. Goluchowicz K, Blind K. Identification of future fields of standardisation: An explorative application of the Delphi methodology. Technological Forecasting and Social Change 2011 Nov;78(9):1526-1541. [doi: 10.1016/j.techfore.2011.04.014]
86. Baker JA, Lovell K, Harris N, Campbell M. Multidisciplinary consensus of best practice for pro re nata (PRN) psychotropic medications within acute mental health settings: a Delphi study. J Psychiatr Ment Health Nurs 2007 Aug;14(5):478-484. [doi: 10.1111/j.1365-2850.2007.01112.x] [Medline: 17635256]

87. Skulmoski GJ, Hartman FT, Krahn J. The Delphi Method for Graduate Research. JITE:Research 2007;6:001-021. [doi: 10.28945/199]
88. Avella JR. Delphi Panels: Research Design, Procedures, Advantages, and Challenges. IJDS 2016;11:305-321. [doi: 10.28945/3561]
89. Brook RH, Chassin MR, Fink A, Solomon DH, Kosecoff J, Park RE. A method for the detailed assessment of the appropriateness of medical technologies. Int J Technol Assess Health Care 1986 Mar 10;2(1):53-63. [doi: 10.1017/s0266462300002774] [Medline: 10300718]
90. McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. Int J Clin Pharm 2016 Jun 5;38(3):655-662 [FREE Full text] [doi: 10.1007/s11096-016-0257-x] [Medline: 26846316]
91. Choi BCK, Pak AWP. A catalog of biases in questionnaires. Prev Chronic Dis 2005 Jan;2(1):A13 [FREE Full text] [Medline: 15670466]
92. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. PLoS One 2011 Jun 9;6(6):e20476 [FREE Full text] [doi: 10.1371/journal.pone.0020476] [Medline: 21694759]
93. Penna A, Grilli R, Filardo G, Mainini F, Zola P, Mantovani L, et al. Do different physicians' panels reach similar conclusions? A case study on practice guidelines for limited surgery in breast cancer. The European Journal of Public Health 1997 Dec 01;7(4):436-440. [doi: 10.1093/eurpub/7.4.436]
94. Akins RB, Tolson H, Cole BR. Stability of response characteristics of a Delphi panel: application of bootstrap data expansion. BMC Med Res Methodol 2005 Dec 01;5(1):37 [FREE Full text] [doi: 10.1186/1471-2288-5-37] [Medline: 16321161]
95. Wiliam D, Black P. Meanings and Consequences: a basis for distinguishing formative and summative functions of assessment? British Educational Research Journal 1996 Dec;22(5):537-548. [doi: 10.1080/0141192960220502]
96. Alarabiat A, Ramos I. The Delphi Method in Information Systems Research (2004-2017). EJBRM 2019 Jun 01;17(2):86-99. [doi: 10.34190/jbrm.17.2.04]
97. Schmidt R, Lyytinen K, Keil M, Cule P. Identifying Software Project Risks: An International Delphi Study. Journal of Management Information Systems 2001;17(4):5-36. [doi: 10.1080/07421222.2001.11045662]
98. Goodarzi Z, Abbasi E, Farhadian H. Achieving Consensus Deal with Methodological Issues in the Delphi Technique. International Journal of Agriculture Management and Development 2018;8(2):219-230. [doi: 10.22004/ag.econ.292533]
99. Fitch K, Bernstein SJ, Aguilar MD, Burnand B, LaCalle JR, Lázaro P, et al. The RAND/UCLA Appropriateness Method User's Manual. Santa Monica, CA: RAND Corporation; 2001:109.
100. Sanmartin C, Murphy K, Choptain N, Conner-Spady B, McLaren L, Bohm E, et al. Appropriateness of healthcare interventions: Concepts and scoping of the published literature. Int J Technol Assess Health Care 2008 Jul 4;24(03):342-349. [doi: 10.1017/s0266462308080458]
101. Santori G, Fontana I, Valente R, Ghirelli R, Valente U. Application of the RAND/UCLA Appropriateness Method to evaluate an information system for kidney/pancreas transplantation in adult recipients. Transplant Proc 2008 Jul;40(6):2021-2023. [doi: 10.1016/j.transproceed.2008.05.018] [Medline: 18675119]
102. Keeney S, Hasson F, McKenna HP. A critical review of the Delphi technique as a research methodology for nursing. International Journal of Nursing Studies 2001 Apr;38(2):195-200. [doi: 10.1016/s0020-7489(00)00044-4]
103. Kuusi O. Expertise in the future use of generic technologies: Epistemic and methodological considerations concerning Delphi studies (Acta Universitatis Oeconomicae Helsingiensis). Helsinki, Finland: Helsinki School of Economics and Business Administration; Jan 01, 1999:268.
104. Rowe G, Wright G. The Delphi technique as a forecasting tool: issues and analysis. International Journal of Forecasting 1999 Oct;15(4):353-375. [doi: 10.1016/S0169-2070(99)00018-7]
105. Dalkey NC. In: Dalkey NC, Rourke DL, Lewis R, Snyder D, editors. Studies in the quality of life: Delphi and decision-making. Lexington, MA: Lexington Books; Jan 01, 1972:161.
106. Daniel E, White A. The future of inter-organisational system linkages: findings of an international Delphi study. European Journal of Information Systems 2005;14(2):188-203. [doi: 10.1057/palgrave.ejis.3000529]
107. Video Conferencing, Cloud Phone, Webinars, Chat, Virtual Events. San Jose, California: Zoom Video Commun Inc URL: https://zoom.us/ [accessed 2020-12-11]
108. Online Umfrage erstellen. Switzerland: enuvo GmbH URL: https://www.umfrageonline.com/ [accessed 2021-01-20]
109. Dylan W. Integrating formative and summative functions of assessment. 2000 Presented at: Working Group 10 of the International Congress on Mathematics Education; July 31-August 6, 2000; Makuhari, Japan.
110. Bürkle T, Ammenwerth E, Prokosch H, Dudeck J. Evaluation of clinical information systems. What can be evaluated and what cannot? J Eval Clin Pract 2001 Nov;7(4):373-385. [doi: 10.1046/j.1365-2753.2001.00291.x] [Medline: 11737529]
111. Venable J, Pries-Heje J, Baskerville R. FEDS: a Framework for Evaluation in Design Science Research. European Journal of Information Systems 2016;25:77-89. [doi: 10.1057/ejis.2014.36]
112. Creswell JW, Creswell JD. Research Design: Qualitative, Quantitative and Mixed Methods Approaches, 5th edition. Los Angeles, CA: Sage Publications, Inc; Jan 02, 2018:304.
113. Mayring P. Qualitative Inhaltsanalyse: Grundlagen und Techniken, 12th edition. Weinheim and Basel: Beltz Verlag; 2015:152.

114. Mayring P. Einführung in die qualitative Sozialforschung, 6th edition. Weinheim and Basel: Beltz Verlag; 2016:170.
115. Tsai C, Koch S. Towards a Framework for National eHealth Evaluation and Monitoring: A Combined Top-Down and Bottom-Up Approach Using Sweden as Example. Stud Health Technol Inform 2019 Aug 21;264:954-958. [doi: 10.3233/SHTI190365] [Medline: 31438065]
116. Piirainen K, Gonzalez R. Seeking Constructive Synergy: Design Science and the Constructive Research Approach. In: vom Brocke J, Hekkala R, Ram S, Rossi M, editors. Design Science at the Intersection of Physical and Virtual Design. DESRIST 2013. Lecture Notes in Computer Science, vol 7939. Berlin, Heidelberg: Springer; 2013:59-72.
117. Szopinski D, Schoormann T, Kundisch D. Criteria as a Prelude for Guiding Taxonomy Evaluation. 2000 Presented at: 53rd Annual Hawaii International Conference on System Sciences; January 7-10, 2020; Wailea, Maui, HI. [doi: 10.24251/hicss.2020.622]
118. Nykänen P, Kaipio J. Quality of health IT evaluations. In: Ammenwerth E, Rigby M, editors. Evidence-Based Health Informatics. Amsterdam, Netherlands: IOS Press BV; 2016:291-303.
119. Ammenwerth E, Brender J, Nykänen P, Prokosch H, Rigby M, Talmon J. Visions and strategies to improve evaluation of health information systems. Reflections and lessons based on the HIS-EVAL workshop in Innsbruck. Int J Med Inform 2004 Jun 30;73(6):479-491. [doi: 10.1016/j.ijmedinf.2004.04.004] [Medline: 15171977]
120. Kitsiou S, Manthou V, Vlachopoulou M. A Framework for the Evaluation of Integration Technology Approaches in Healthcare. 2006 Presented at: The Fifth International Conference on Biological Information and Biomedical Engineering; October 2006; Ioannina - Epirus, Greece.
121. Nambisan S, Agarwal R, Tanniru M. Organizational Mechanisms for Enhancing User Innovation in Information Technology. MIS Quarterly 1999 Sep;23(3):365-395. [doi: 10.2307/249468]
122. Paraskevas A, Saunders MN. Beyond consensus: an alternative use of Delphi enquiry in hospitality research. Int J Contemp Hospitality Mngt 2012 Aug 17;24(6):907-924. [doi: 10.1108/09596111211247236]
123. Hussler C, Muller P, Rondé P. Is diversity in Delphi panelist groups useful? Evidence from a French forecasting exercise on the future of nuclear energy. Technological Forecasting and Social Change 2011 Nov;78(9):1642-1653. [doi: 10.1016/j.techfore.2011.07.008]
124. Roehrs A, da Costa CA, da Rosa Righi R, Farias De Oliveira KS. Personal Health Records: A Systematic Literature Review. J Med Internet Res 2017 Jan 06;19(1):e13 [FREE Full text] [doi: 10.2196/jmir.5876] [Medline: 28062391]
125. Fernández-Alemán JL, Seva-Llor CL, Toval A, Ouhbi S, Fernández-Luque L. Free Web-based personal health records: an analysis of functionality. J Med Syst 2013 Dec 13;37(6):9990. [doi: 10.1007/s10916-013-9990-z] [Medline: 24221916]
126. Scheplitz T, Benedict M, Esswein W. Patientenkompetenz durch Online-Portale – Eine Funktionsanalyse. 2018 Presented at: Multikonferenz Wirtschaftsinformatik (MKWI 2018); March 06-09, 2018; Lüneburg, Germany p. 744-755.
127. Rowe G, Wright G. The Delphi technique: Past, present, and future prospects — Introduction to the special issue. Technological Forecasting and Social Change 2011 Nov;78(9):1487-1490. [doi: 10.1016/j.techfore.2011.09.002]
128. Powell C. The Delphi technique: myths and realities. J Adv Nurs 2003 Feb;41(4):376-382. [doi: 10.1046/j.1365-2648.2003.02537.x] [Medline: 12581103]
129. Bell BG, Spencer R, Avery AJ, Campbell SM. Tools for measuring patient safety in primary care settings using the RAND/UCLA appropriateness method. BMC Fam Pract 2014 Jun 05;15(110):1-7 [FREE Full text] [doi: 10.1186/1471-2296-15-110] [Medline: 24902490]

Abbreviations

EHR: electronic health record
PHR: personal health record
RAM: RAND/UCLA Appropriateness Method
TED: Taxonomy-Evaluation-Delphi approach.
TOPCOP: Useful Taxonomy of Patient Portals based on Characteristics of Patient Engagement

Edited by G Eysenbach; submitted 25.05.21; peer-reviewed by M Lotto, S Hajesmaeel Gohari, L Chirchir; comments to author 15.06.21; revised version received 15.07.21; accepted 31.07.21; published 05.10.21.

Please cite as:
Glöggler M, Ammenwerth E
Improvement and Evaluation of the TOPCOP Taxonomy of Patient Portals: Taxonomy-Evaluation-Delphi (TED) Approach
J Med Internet Res 2021;23(10):e30701
URL: https://www.jmir.org/2021/10/e30701
doi:10.2196/30701
PMID:34403354

Original Paper

Health Care Providers' Acceptance of a Personal Health Record: Cross-sectional Study

Consuela Cheriece Yousef[1,2,3], PhD, MPH, PharmD; Teresa M Salgado[4], MPharm, PhD; Ali Farooq[5], DSc, MSc, MBA, MCS; Keisha Burnett[6], EdD, MS, SCT (ASCP), MB; Laura E McClelland[7], PhD; Laila Carolina Abu Esba[2,3,8], BSc, MSc, PharmD; Hani Solaiman Alhamdan[2,9], BSc, MSc; Sahal Khoshhal[2,3,10], PharmD; Ibrahim Fahad Aldossary[2,3,11], BSPharm, MSc; Omar Anwar Alyas[12]; Jonathan P DeShazo[7], MPH, PhD

[1]Pharmaceutical Care Department, Ministry of National Guard Health Affairs, Dammam, Saudi Arabia

[2]King Abdullah International Medical Research Center, Riyadh, Saudi Arabia

[3]King Saud bin Abdul-Aziz University for Health Sciences, Riyadh, Saudi Arabia

[4]Department of Pharmacotherapy & Outcome Science, School of Pharmacy, Virginia Commonwealth University, Richmond, VA, United States

[5]Department of Computing, University of Turku, Turku, Finland

[6]Department of Clinical Laboratory Sciences, Cytopathology Practice Program, University of Tennessee Health Science Center, Memphis, TN, United States

[7]Department of Health Administration, Virginia Commonwealth University, Richmond, VA, United States

[8]Pharmaceutical Care Department, Ministry of National Guard Health Affairs, Riyadh, Saudi Arabia

[9]Pharmaceutical Care Services, Ministry of National Guard Health Affairs, Jeddah, Saudi Arabia

[10]Pharmaceutical Care Department, Ministry of National Guard Health Affairs, Madinah, Saudi Arabia

[11]Pharmaceutical Care Department, Ministry of National Guard Health Affairs, Al Ahsa, Saudi Arabia

[12]Royal College of Surgeons in Ireland-Medical University of Bahrain, Busaiteen, Bahrain

Corresponding Author:
Consuela Cheriece Yousef, PhD, MPH, PharmD
Pharmaceutical Care Department
Ministry of National Guard Health Affairs
P.O. Box 4616
Dammam, 31412
Saudi Arabia
Phone: 966 8532555
Email: consuela_73@hotmail.com

Abstract

Background: Personal health records (PHRs) are eHealth tools designed to support patient engagement, patient empowerment, and patient- and person-centered care. Endorsement of a PHR by health care providers (HCPs) facilitates patient acceptance. As health care organizations in the Kingdom of Saudi Arabia begin to adopt PHRs, understanding the perspectives of HCPs is important because it can influence patient adoption. However, no studies evaluated HCPs' acceptance of PHRs in the Kingdom of Saudi Arabia.

Objective: The aim of this study was to identify predictors of HCPs' acceptance of PHRs using behavioral intention to recommend as a proxy for adoption.

Methods: This cross-sectional study was conducted among HCPs (physicians, pharmacists, nurses, technicians, others) utilizing a survey based on the Unified Theory of Acceptance and Use of Technology. The main theory constructs of performance expectancy, effort expectancy, social influence, facilitating conditions, and positive attitude were considered independent variables. Behavioral intention was the dependent variable. Age, years of experience, and professional role were tested as moderators between the main theory constructs and behavioral intention using partial least squares structural equation modeling.

Results: Of the 291 participants, 246 were included in the final analysis. Behavioral intention to support PHR use among patients was significantly influenced by performance expectancy (β=.17, *P*=.03) and attitude (β=.61, *P*<.01). No moderating effects were present.

Conclusions: This study identified performance expectancy and attitude as predictors of HCPs' behavioral intention to recommend PHR to patients. To encourage HCPs to endorse PHRs, health care organizations should involve HCPs in the implementation and provide training on the features available as well as expected benefits. Future studies should be conducted in other contexts and include other potential predictors.

(J Med Internet Res 2021;23(10):e31582) doi:10.2196/31582

KEYWORDS

personal health records; patient portals; Ministry of National Guard Health Affairs; UTAUT; eHealth; Middle East

Introduction

Overview

A wide range of eHealth technologies has become available over the past 2 decades as countries have introduced eHealth initiatives to support the goals for patient engagement and person-centered care [1]. Legislation around the world advocates for patients to have electronic access to their health information through personal health records (PHRs) [2]. PHRs are an eHealth tool to increase patient engagement and empowerment by allowing individuals to keep track of their personal health information. The Markle Foundation defined PHRs as "an Internet-based set of tools that allows people to access and coordinate their lifelong health information and make appropriate parts of it available to those who need it" [3]. Person-centered care and patient engagement are considered pillars of any high-functioning health care system, and PHRs can contribute to both [4,5]. While various terms have been used interchangeably with PHR in the literature (eg, patient portal, patient web portal, computerized patient portal, patient accessible electronic health record [EHR], tethered PHR, electronic PHR), the broader term of PHR will be used predominantly throughout this paper.

PHR adoption has been associated with a wide range of benefits, including better patient–provider relationships, improvements in patient engagement, better medication adherence, positive health outcomes (eg, blood pressure and glycemic control), and increased organizational efficiencies [6]. As the benefits of PHR adoption are achieved, health care costs potentially decrease as individuals become empowered to take better control of their health and rely less on interactions with the health care system [6]. However, multiple studies have shown low adoption rates [7-9]. Even though the 2009 Health Information Technology for Economic and Clinical Health (HITECH) Act and its Meaningful Use criteria accelerated PHR access in the United States [10], only 15%-30% of patients use PHRs while 90% of health care systems offer them [11]. Outside of the United States, a systematic review showed adoption rates of around 0.13% in the United Kingdom and 5% in other European countries [7].

Various barriers to PHR adoption have been identified [7,9,12,13]. In the systematic review by Niazkhani et al [13], the barriers were characterized as patient demographic factors (eg, age, gender); environment/medical practice (eg, providers' communication about PHRs, physician resistance); technological (eg, perceived PHR usefulness, perceived PHR complexity); and chronic disease characteristics (eg, patients' feeling of control over the disease, number of comorbidities). Health care providers' (HCPs) attitudes are a major contributing factor in patients' adoption of PHRs [14-16]. HCPs play a key role in supporting and engaging patients through their attitudes, behavior, and endorsement of services [17]. Although studies have shown a high level of patient interest in PHRs [5,18-20], there has been a disconnect between interest and uptake. This is partially due to HCPs' reticence toward the acceptance and promotion of their use [5,21,22].

Researchers around the world have studied HCPs' attitudes and perceptions of PHRs. Nazi [22] explored the experiences and perspectives of HCPs (physicians, nurses, and pharmacists) related to patients' use of the My HealtheVet PHR in the United States and found that many HCPs had limited familiarity with the PHR features, contributing to its underutilization [22]. The author identified the following 8 factors to be key in the implementation, adoption, and use of PHRs: (1) showing the relevance of PHRs; (2) increasing the perceived value by focusing on unique services; (3) providing education and training; (4) integrating PHRs into the existing technology; (5) aligning PHR functions with the workflow; (6) offering incentives to individuals or teams; (7) making information accessible; and (8) supporting asynchronous and bidirectional communication.

A study in Finland, which included a wide range of HCPs (eg, nurses, social workers, dentists, physicians, physical therapists, and psychologists), found that the most important factors influencing HCPs' support for a national patient portal were expected positive influences on their work, the usability of the portal, and benefits for the patients [17]. However, only few (13%) respondents felt they had received adequate information about the portal. The authors recommended HCPs be informed about PHR benefits to garner their support. In Canada, Wiljer et al [23] endorsed institutional strategies such as "continuous organizational reassurance," education, and a physician champion to stimulate a paradigm shift to patient-centered care for successful PHR implementation. In a Swedish study of oncology HCPs (nurses and physicians), the authors compared HCPs working in outpatient clinics with those working in primary care units [24]. A greater proportion of HCPs in primary care believed there were benefits of patients using PHRs such as better adherence (50% vs. 35%), greater ability to clarify important information (50% vs. 26%), and improved patient communication (36% vs. 20%) [25].

In the Kingdom of Saudi Arabia, enhancing patient-centered care through patient involvement with technology is an objective of The National Transformation Program, a component of Vision 2030. The Ministry of National Guard Health Affairs (MNGHA)

implemented the MNGHA Care PHR in 2018. No studies have evaluated HCPs' acceptance of PHRs in the country.

The aim of this study was to identify a set of factors that affect the intention to recommend the use of MNGHA Care PHR among HCPs. To promote patient engagement and patient-centered care, a better understanding of how HCPs perceive PHRs is needed.

Theoretical Background

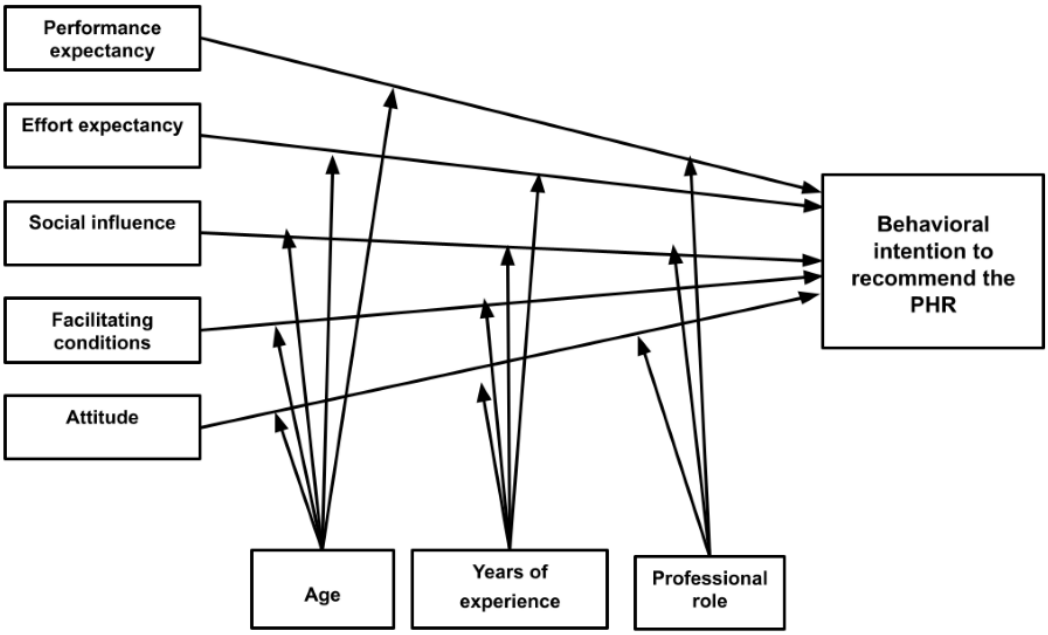
In 2003, Venkatesh et al [26] developed the Unified Theory of Acceptance and Use of Technology (UTAUT) to provide a comprehensive framework to explain acceptance, intention, and usage of information technology in organizations. It is an integration of 8 theories—theory of reasoned action, technology acceptance model (TAM), motivational model, theory of planned behavior (TPB), combined TAM–TPB, model of personal computer utilization, diffusion of innovation theory, and social cognitive theory [26]. The core constructs of performance expectancy, effort expectancy, social influence, and facilitating conditions directly act on behavioral intention and, ultimately, predict the use of the technology. Gender, age, voluntariness, and experience are moderators in the framework. The model explained approximately 77% of the variance in behavioral intention and 52% of the variance in technology use [26]. Since its development, UTAUT has been used to explain technology acceptance in different user groups in a wide range of contexts with various technologies, strengthening the generalizability [27]. UTAUT has also been used broadly in other health care areas, including telemedicine [28,29], electronic medical/health records [30-34], electronic documentation systems [35], picture archiving and communication systems [36], and health information systems [37,38].

Research Model and Hypotheses

Most studies have not examined the full UTAUT with the moderation effects but rather the main effects alone, combined with a subset of the moderators, or with new constructs or mechanisms [39]. Venkatesh et al [39] proposed that future research should use UTAUT as the baseline model to transform the theory from static to dynamic. New endogenous mechanisms or new moderation mechanisms are the most common types of extensions [39]. While UTAUT includes the technological dimension (performance expectancy and effort expectancy) and organizational/environmental dimension (social influence and facilitating conditions), the individual dimension is not included. Nonetheless, individual traits (attitude, personal innovativeness, computer self-efficacy) may significantly predict the acceptance of technology [27,40,41]. Constructs representing individual traits are frequently used as endogenous mechanisms to extend UTAUT.

The research model for this study includes the 4 core UTAUT constructs: performance expectancy, effort expectancy, social influence, and facilitating conditions (Figure 1). The construct of attitude was added as an individual characteristic. Unlike the original UTAUT model, we did not include behavior in the proposed model because we were unable to objectively assess use. Instead, we measured intention to recommend PHR, using it as a proxy for HCPs' acceptance. Behavioral intention is frequently a proxy for actual technology adoption in the literature [42-44].

Figure 1. Adapted UTAUT model. PHR: personal health record; UTAUT: Unified Theory of Acceptance and Use of Technology.



Proposed differences between this model and the original UTAUT model are shown in Table 1. The moderators chosen for this study were age, years of experience, and professional role. Previous literature indicated that age was inversely

associated with eHealth adoption. For example, electronic medical record use was inversely associated with physician age [45]. A potential explanation is that, in the initial stages of technology use, older users are believed to be more influenced by experience, and ease of use is more important [26]. Next, years in practice has been associated with acceptance of eHealth [45]. As the number of years since medical school graduation increased, physicians became less likely to accept eHealth technologies [45]. There have also been differences in eHealth acceptance by professional role [45]. Nonphysicians used advanced EHR features less than physicians, and specialists (eg, obstetrician/gynecologists) were less likely to use an EHR in their practices [45]. Voluntariness of use and gender were dropped as moderators in the proposed model. PHR use is not mandatory; therefore voluntariness of use is not relevant [26]. In the health care context, professional role takes precedence over gender and no differences in acceptance by gender were expected [34].

Table 1. Original Unified Theory of Acceptance and Use of Technology (UTAUT) versus adapted UTAUT for health care providers.

Construct	Original UTAUT[a] moderators				Adapted UTAUT moderators		
	Gender	Age	Experience	Voluntariness	Age	Years of experience	Professional role
PE[b] → BI[c]	✓	✓			✓		✓
EE[d] → BI	✓	✓	✓		✓	✓	
SI[e] → BI	✓	✓	✓	✓	✓	✓	✓
BI → Use[f]							
FC[g] → Use		✓	✓				
FC → BI					✓	✓	
ATT[h] → BI					✓	✓	✓

[a]UTAUT: Unified Theory of Acceptance and Use of Technology.

[b]PE: performance expectancy.

[c]BI: behavioral intention.

[d]EE: effort expectancy.

[e]SI: social influence.

[f]Use: actual usage.

[g]FC: facilitating conditions.

[h]ATT: attitude.

This study tested the following hypotheses:

- H1: Performance expectancy positively influences behavioral intention to recommend the PHR
- H2: Effort expectancy positively influences behavioral intention to recommend the PHR
- H3: Social influence positively influences behavioral intention to recommend the PHR
- H4: Facilitating conditions positively influence behavioral intention to recommend the PHR
- H5: Attitude positively influences behavioral intention to recommend the PHR
- H6: Age, years of experience, and professional role selectively moderate the relationships between the main constructs and behavioral intention to recommend the PHR

Methods

Study Design

A cross-sectional study utilizing a survey was conducted at a large, integrated health care system in the Kingdom of Saudi Arabia. The survey was administered to HCPs across the organization to assess acceptance of the PHR. Since 2018, patients have had access to the MNGHA Care PHR, which includes the following features: scheduling appointments, requesting medical reports and prescription refills, viewing radiology reports, checking laboratory results, and receiving vaccination reminders [46]. Additionally, personal health information such as weight, blood pressure, blood sugar, and exercise details can be uploaded. Finally, MNGHA Care contains links to health education information and a self-assessment feature permitting patients to enter information related to pain control, performance status, and quality of life.

Setting and Participants

The study population consisted of HCPs from MNGHA hospitals and primary health care centers in Dammam, Riyadh, Jeddah, Madinah, Al Ahsa, and Qassim, including physicians, dentists, pharmacists, nurses, physical and occupational therapists, optometrists, technicians (pharmacy, medical imaging, medical and pathology laboratory, dental), paramedics, and dietitians.

Instrument and Data Collection

Data were collected using an anonymous self-administered online survey between April and May 2021. The initial version of the survey included 63 items adapted from previously published technology acceptance surveys used in health care in 3 parts [26,47-51]. The first part captured demographic characteristics including age, gender, region, facility type,

profession, specialty area (for physicians), years in profession, years at MNGHA, and nationality. The second part contained 4 general PHR questions: (1) Have you heard of MNGHA Care?; (2) Do you have an MNGHA Care account?; (3) Have you used MNGHA Care yourself?; and (4) Have you recommended patients to use MNGHA? This section also included Likert-scale statements associated with PHR acceptance along with an open-ended question and a checklist. The third section related to acceptance of secure messaging and included Likert-scale statements, an open-ended question, and a checklist.

The instrument was built on QuestionPro [52] and pilot tested with 7 HCPs (2 physicians, 3 pharmacists, and 2 nurses) working within MNGHA. The QuestionPro survey link and a cover letter explaining the purpose of the study were emailed to these 7 HCPs to obtain feedback regarding survey length, clarity, and flow of the questionnaire. After comments were compiled, 12 items were removed, and some were modified to improve clarity and to decrease survey length. The final version of the survey included 51 items and was approved by the institutional review boards at the Virginia Commonwealth University and King Abdullah International Medical Research Center.

For this study, the focus was on parts 1 and 2 of the instrument. However, the open-ended question and checklist from part 2 are not included in this paper. Responses to the PHR acceptance items were provided on a 5-point Likert scale from strongly disagree (1) to strongly agree (5). Acceptance was operationalized as the intention to recommend patients use the PHR using the statement "I will probably recommend patients use MNGHA Care in the future" [49].

Performance expectancy was defined as the degree to which the HCP believes a PHR will be beneficial in the health care delivery process [20]. It was measured with the following 4 items:

1. MNGHA Care is a useful tool to help patients feel more involved in their care [47,53].
2. I believe MNGHA Care helps patients to better manage their health [48].
3. MNGHA Care will increase patient satisfaction with their health care [48].
4. MNGHA Care can improve the quality of patient care [51].

Effort expectancy is the degree of ease associated with use of the PHR [20]. It was measured with the following 3 items:

1. Information in MNGHA Care should be easy for our patients to understand [48,53].
2. I believe most patients have the skills needed to use MNGHA Care [47].
3. I think it is not difficult for our patients to learn to use MNGHA Care [47].

Social influence is the degree to which an individual perceives important others believe the PHR should be used [20]. It was measured with the following 2 items:

1. I believe our patients support the use of MNGHA Care [48].
2. In general, the organization has supported the use of MNGHA Care [26].

Facilitating conditions was defined as the degree to which an individual believes an organizational and technical infrastructure exists to support use of the PHR [20]. It was measured with the following 3 items:

1. I have enough information about MNGHA Care [26,48].
2. There is technical help for patients who use MNGHA Care [26].
3. I know the goals of MNGHA Care [26].

Attitude was defined as positive feelings related to patients using the PHR [54]. It was measured with the following 4 items:

1. MNGHA Care is a valuable tool [26,47].
2. It is a good idea for patients to use MNGHA Care [26,47].
3. MNGHA Care is a positive advancement in this digital age [47].
4. I believe MNGHA Care will be used by many patients [47].

Although behavioral intention and social influence used less than 3 items, partial least squares structural equation modeling (PLS) supports using single-item measures [55] and earlier research using PLS has used less than 3 items for measuring constructs [56,57].

Sampling

A snowball and convenience sampling strategy was used to recruit HCPs from across the organization. HCPs were initially recruited through the hospital's email list in combination with WhatsApp as it is a widely used social media platform for professional communication. They were asked to forward the survey to other HCPs. Follow-up reminders were also sent out. The target sample size for this study was 200 HCPs, which has been considered a fair sample size for statistical analysis with structural equation modeling [58]. To encourage participation, there was a random drawing for twenty five 37.5 Saudi Arabian Riyal (US $10) Amazon gift cards.

Statistical Analyses

Descriptive statistics were analyzed using SPSS version 25 (IBM) [59]. PLS was used to test the research model using SmartPLS version 3.0 [60]. The advantage of PLS is the ability to estimate complex research models without distributional assumptions [61]. Compared with traditional SEM, PLS has greater statistical power, which means that there is a higher likelihood of identifying significant relationships if they are actually present in the population [61]. Furthermore, PLS has been widely used in empirical studies of technology acceptance, including with UTAUT [26,27,34] and with PHR acceptance [50,62]. To test the research model, a measurement model was used to evaluate construct reliability, indicator reliability, convergent validity, and discriminant validity. A structural model was tested after ensuring reliability and validity.

Results

Demographic Characteristics

Overall, 291 HCPs participated in the survey. However, after removing the data for missing values, a usable sample of 246 was used for further analysis. Table 2 presents the demographic characteristics. Most were 40-49 years old (95/246, 38.6%),

female (158/246, 64.2%), non-Saudi (132/246, 53.7%), nurses (106/246, 43.1%), in Riyadh (81/246, 32.9%), over 10 years of experience (167/246, 67.9%) and over 10 years at MNGHA (128/246, 52.0%), and hospital based (228/246, 92.7%).

Table 2. Demographic characteristics (N=246).

Variables	Values, n (%)
Age	
20-29 years	37 (15.0)
30-39 years	77 (31.3)
40-49 years	95 (38.6)
50 years and above	37 (15.0)
Gender	
Male	88 (35.8)
Female	158 (64.2)
Nationality	
Saudi	114 (46.3)
Non-Saudi	132 (53.7)
Health care provider	
Physician	40 (16.3)
Pharmacist	57 (23.2)
Nurse	106 (43.1)
Technician	33 (13.4)
Other	10 (4.1)
Years in profession	
Less than 5 years	33 (13.4)
5-10 years	46 (18.7)
Greater than 10 years	167 (67.9)
Years working at MNGHA[a]	
<1 year	13 (5.3)
1-4 years	40 (16.3)
5-10 years	65 (26.4)
>10 years	128 (52.0)
Location	
Dammam	46 (18.7)
Madinah	35 (14.2)
Al Ahsa	51 (20.7)
Jeddah	33 (13.4)
Riyadh	81 (32.9)
Type of facility	
Hospital	228 (92.7)
Primary health care clinic	18 (7.3)

[a]MNGHA: Ministry of National Guard Health Affairs.

General PHR Use Characteristics

The majority of HCPs were aware of MNGHA Care (225/246, 91.5%), had an account (213/246, 86.6%), used MNGHA Care (202/246, 82.1%), and recommended it to patients (198/246, 80.5%).

Measurement Model

The measurement model testing results are summarized in Table 3. After removing missing data, the usable sample for hypothesis testing was 246. The variance inflation factor of all items was below the threshold of 5, showing no evidence of multicollinearity. Item loadings were all above 0.40 and in the range of 0.70-0.93. Composite reliability was above the threshold of 0.70, showing good internal consistency. Moreover, the average variance extracted (AVE) of the constructs were greater than 0.50 and in the range of 0.55-0.81, indicating convergent validity.

Discriminant validity was tested using the Fornell–Larcker criterion. The square roots of the corresponding AVE are shown in italics, with each construct's AVE higher than its highest correlation with any other construct (Table 4). Results in Tables 3 and 4 provide evidence of the validity and reliability of the constructs used in the model.

Table 3. Measurement model statistics.

Construct and items	Mean	SD	VIF[a]	Loadings	CR[b]	AVE[c]
Performance expectancy (PE)	4.09	0.73			0.95	0.81
PE1			2.526	0.87		
PE2			3.792	0.92		
PE3			3.711	0.92		
PE4			3.462	0.90		
Effort expectancy (EE)	3.75	0.67			0.79	0.55
EE1			1.099	0.81		
EE2			1.473	0.70		
EE3			1.465	0.72		
Facilitating conditions (FC)	3.60	0.78			0.88	0.71
FC1			1.756	0.85		
FC2			1.547	0.77		
FC3			2.023	0.90		
Social influence (SI)	3.82	0.69			0.85	0.74
SI1			1.3	0.84		
SI2			1.3	0.88		
Attitude (ATT)	4.08	0.63			0.94	0.80
ATT1			4.171	0.93		
ATT2			3.603	0.92		
ATT3			3.486	0.91		
ATT4			2.029	0.83		
Behavioral intention (BI)						
BI	4.18	0.68		1	1	1

[a]VIF: variance inflation factor.

[b]CR: composite reliability.

[c]AVE: average variance extracted.

Table 4. Discriminant validity of the constructs.[a]

Constructs	1	2	3	4	5
Attitude	*0.896*				
Effort expectancy	0.697	*0.742*			
Facilitating conditions	0.596	0.570	*0.843*		
Performance expectancy	0.742	0.708	0.527	*0.901*	
Social influence	0.646	0.671	0.645	0.602	*0.860*

[a]Square roots of the corresponding average variance extracted are shown in italics.

Structural Model

The R^2 was used to assess the structural model. Overall, the model explained 70% of the variance in the intention to recommend the PHR among HCPs (Figure 2). Table 5 presents the structural model results, while Table 6 presents the test for moderating effects.

Figure 2. Structural model showing path coefficients (β) and coefficient of determination (R^2) (*P<.05, **P<.01). PHR: personal health record.

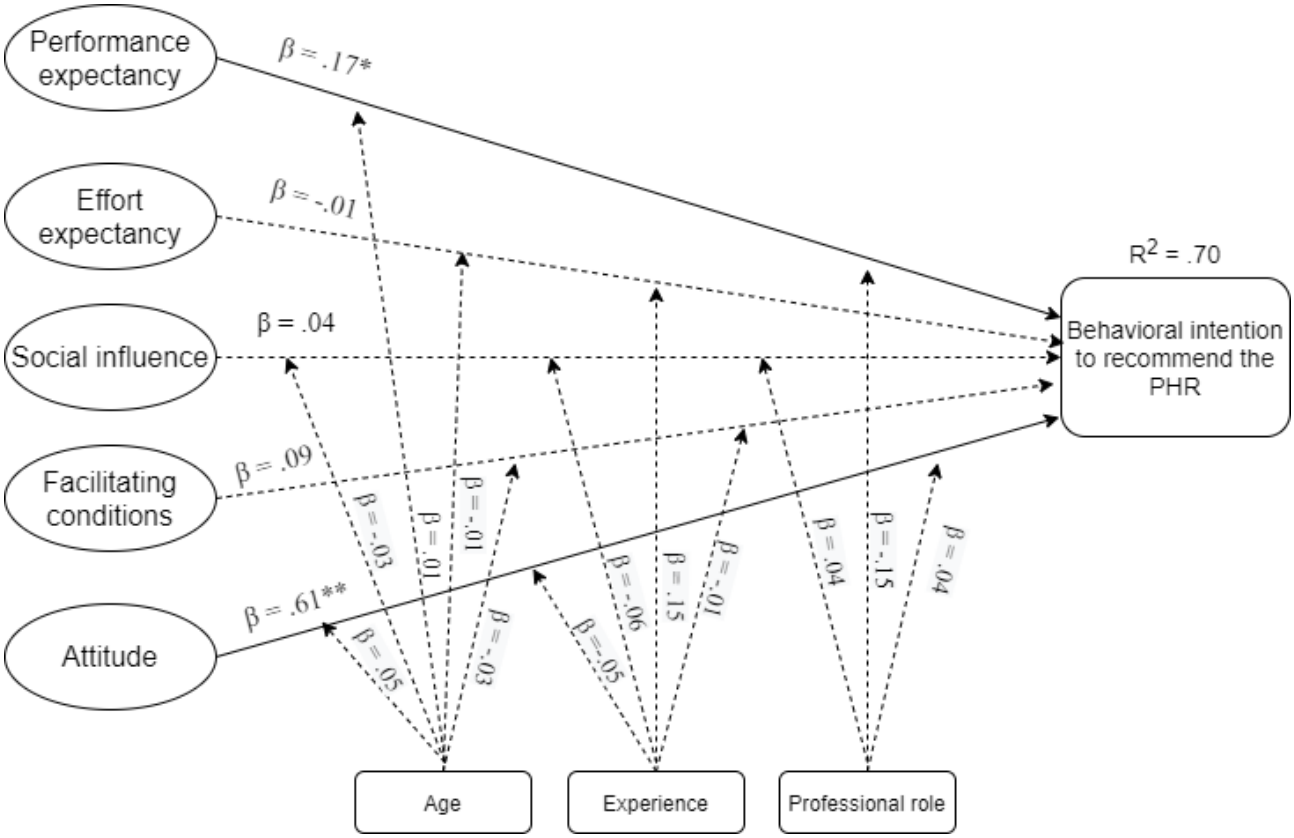


Table 5. Structural model results.

Structural model	β	*t*-statistics[a] (*df*)	*P* value	f^2
PE[b] → BI[c]	.17	2.132 (499)	.03[d]	0.035
EE[e] → BI	–.01	0.166 (499)	.87	0
SI[f] → BI	.04	0.473 (499)	.63	0.002
FC[g] → BI	.09	1.241 (499)	.21	0.013
ATT[h] → BI	.61	6.385 (499)	<.01	0.369

[a]Two tailed.
[b]PE: performance expectancy.
[c]BI: intention to recommend PHR.
[d]P<.05.
[e]EE: effort expectancy.
[f]SI: social influence.
[g]FC: facilitating conditions.
[h]ATT: attitude.

Table 6. Moderation analysis results.

Analysis	β	*t*-statistics[a] (*df*)	*P* value	f^2
Moderation of age				
PE[b] × AGE[c] → BI[d]	.01	0.118 (499)	.91	0
EE[e] × AGE → BI	–.01	0.159 (499)	.87	0
FC[f] × AGE → BI	–.03	0.360 (499)	.72	0.001
SI[g] × AGE → BI	.05	0.633 (499)	.53	0.003
ATT[h] × AGE → BI	–.03	0.307 (499)	.76	0.001
Moderation of experience				
EE × EXP[i] → BI	.15	1.688 (499)	.09	0.016
SI × EXP → BI	–.06	0.609 (499)	.54	0.003
FC × EXP → BI	–.01	0.205 (499)	.84	0
ATT × EXP → BI	–.05	0.597 (499)	.55	0.003
Moderation of professional role				
PE × HCP[j] → BI	–.15	1.598 (499)	.11	0.023
SI × HCP → BI	.04	0.620 (499)	.54	0.003
ATT × HCP → BI	.04	0.441 (499)	.66	0.002

[a]Two tailed.
[b]PE: performance expectancy.
[c]AGE: age.
[d]BI: intention to recommend PHR.
[e]EE: effort expectancy.
[f]FC: facilitating conditions.
[g]SI: social influence.
[h]ATT: attitude.
[i]EXP: experience.
[j]HCP: health care provider.

Discussion

Principal Findings

To the best of our knowledge, this is the first study to examine factors that influence HCPs' intention to recommend PHRs to patients in the Kingdom of Saudi Arabia. Prior studies in the country evaluated the challenges in implementing PHRs and identified HCP resistance as a barrier [63,64]. Although HCPs are not the primary users of PHRs, their endorsement can positively influence patient engagement with this technology [12]. While some providers find promoting the PHR to be an additional burden, those providers who present a PHR to their patients as a tool to supplement their care can facilitate patient adoption [9]. Our study found a high level of awareness among HCPs, with 88.2% (217/246) having an account and 82.1% (202/246) recommending patients use the PHR. In our previous study in patients, HCPs and hospital staff were primarily responsible for recommending the PHR in 58.7% of patients who reported using MNGHA Care [18].

Predictors of patient adoption of PHRs may differ from those that affect HCPs to endorse a PHR [7,12]. Therefore, the research model for HCPs was slightly different from the one used for patients [18]. The proposed theoretical model explained 70% of the variance in HCPs' behavioral intention to recommend PHRs to patients. Performance expectancy and attitude were significantly associated with behavioral intention to recommend the PHR. Much of the literature has shown performance expectancy as the strongest predictor of intention to use technology among HCPs [30,38,47]. In patient and consumer studies of PHRs, performance expectancy has also been a positive predictor of adoption [50,65-67]. However, the attitude was the strongest predictor of behavioral intention in our study. Other studies on PHR adoption have also found attitude to be an important predictor [47,51].

Our findings did not support the hypothesis that age, years of experience, and professional role moderate behavioral intention. Several studies have shown that older and more experienced HCPs are more resistant to health information technology and are less comfortable with using technology [35]. Physicians also have been found to be less enthusiastic about the introduction of eHealth services [33]. In our study, most had over 10 years of experience as an HCP (190/289, 65.7%) and more than 10 years in MNGHA (149/289, 51.6%). Furthermore,

while Moll and Cajander [25] found differences in attitudes of HCPs who worked in primary care units compared with outpatient clinics, most HCPs in this study were from the hospital (265/284, 93.3%), limiting the ability to draw comparisons.

Implications for Theory

This research adds to the literature on HCPs' acceptance of PHR using an adapted UTAUT model. To our knowledge, this is the first study to extend UTAUT with the construct of attitude in the context of HCPs' acceptance of PHR. Only few studies evaluating HCPs' acceptance of PHRs have used theory [68]. This study revealed that the adapted UTAUT model was a good predictive model of HCPs' behavioral intention to recommend PHR use. While our model found that performance expectancy and attitude individually influence behavioral intention, it may also be the case that attitude mediates the relationship between performance expectancy and behavioral intention, as proposed by Dwivedi et al [40].

The original UTAUT explained 76% of the variance in behavioral intention. No studies on PHR adoption have used the original UTAUT model [50,65,66,69,70]. The advantage of the adapted model is a similar predictive power while parsimoniously eliminating the construct "use behavior" and the moderator "voluntariness" in the original model. Although the model explained 70% of the variance in behavioral intention and provided support for the proposed theoretical model, other factors may be important for HCPs' acceptance of PHR. In the health care setting, UTAUT has been criticized for its focus on general technology acceptance factors and the inability to completely explain health information technology adoption [71]. Therefore, it is recommended that UTAUT be adapted to fit the health care context by incorporating health behavior theories, privacy and security issues, and negative factors that inhibit technology adoption [71].

Implications for Practice

This study provides practical contributions based on the proposed relationships and supports the need to focus on strategies to enhance perceived usefulness and a positive attitude toward the PHR in HCPs. While some patients view self-management as a burden and prefer the status quo [9], others will respond to HCP's endorsement of the use of PHRs. Several studies identified HCP recommendation as an important factor in patients' choosing to use PHRs [2,22,23,72,73]. Lyles et al [11] found one-on-one training to be the most effective intervention in PHR implementation in the United States. Providing short educational sessions to individuals or teams can facilitate acceptance among HCPs [48]. These training sessions could be conducted by each department. Training HCPs on the features available supports successful implementation by increasing skills and knowledge. Campaigns can also be directed at promoting awareness among HCPs. Through these interventions, HCPs will perceive the usefulness of PHRs and develop more positive attitudes regarding the benefits. Consequently, they will be more inclined to recommend PHRs to patients. Through their interactions with HCPs, patients will perceive PHRs as useful and are more likely to adopt them [66].

Limitations

There are several limitations to this study. While cross-sectional studies are useful for examining associations, a causal relationship cannot be established [74]. Snowball and convenience sampling, both nonprobability sampling strategies, were used to select participants, limiting generalizability; however, participants from multiple sites were selected to attain good representation across MNGHA. Self-administered online surveys are associated with various biases, including social desirability response bias, self-selection, and nonresponse bias [75]. To minimize social desirability response bias, participants had the option of not answering any question that made them uncomfortable. To minimize nonresponse bias, HCPs were contacted multiple times and offered an incentive to encourage a high response rate.

Recommendations for Future Research

Future studies should evaluate the proposed model in other contexts. This study involved a large integrated health care organization. Research in other organizations within the country and in this part of the world will increase the generalizability of our findings. Research should also be conducted in individual HCP groups. Differences in PHR acceptance have been observed based on a variety of characteristics, including age, gender, professional role, and practice setting. Future researchers should focus on HCP group–specific interventions. Finally, while this study used an open-ended question and checklist (analyzed separately) to achieve greater depth, one-on-one interviews would provide valuable data on the motivation of HCPs and nuances within this context.

Conclusion

This study examined factors affecting HCPs' behavioral intention to recommend PHRs to patients in the Kingdom of Saudi Arabia. The proposed model accounted for 70% of the variance in behavioral intention, indicating significant predictive power. Performance expectancy and attitude were significant predictors of HCPs' behavioral intention to support PHR use. Our results suggest that health care organizations should focus on strategies associated with these factors to improve HCP support and decrease barriers to patient use of PHRs. Future research should test this model and explore other predictors in order to develop successful interventions to encourage the adoption and continued use of the PHR among patients.

Acknowledgments

We thank the health care providers who participated in this study and acknowledge Dr Sherine Esmail and Mr Mansoor Ahmed Khan for their assistance in participant recruitment.

Conflicts of Interest

None declared

References

1. Petrovskaya O, Lau F, Antonio M. Synthesising evidence on patient portals: a protocol for an umbrella review. BMJ Open 2019 Mar 30;9(3):e024469 [FREE Full text] [doi: 10.1136/bmjopen-2018-024469] [Medline: 30928933]
2. Gagnon M, Payne-Gagnon J, Breton E, Fortin J, Khoury L, Dolovich L, et al. Adoption of Electronic Personal Health Records in Canada: Perceptions of Stakeholders. Int J Health Policy Manag 2016 Jul 01;5(7):425-433 [FREE Full text] [doi: 10.15171/ijhpm.2016.36] [Medline: 27694670]
3. Markle Foundation. Connecting for Health: A Public-Private Collaborative. The Personal Health Working Group Final Report. 2003. URL: https://www.markle.org/sites/default/files/final_phwg_report1.pdf [accessed 2021-10-05]
4. Berwick D, Snair M, Nishtar S. Crossing the Global Health Care Quality Chasm: A Key Component of Universal Health Coverage. JAMA 2018 Oct 02;320(13):1317-1318. [doi: 10.1001/jama.2018.13696] [Medline: 30177995]
5. Shah SD, Liebovitz D. It Takes Two to Tango: Engaging Patients and Providers With Portals. PM R 2017 May;9(5S):S85-S97. [doi: 10.1016/j.pmrj.2017.02.005] [Medline: 28527507]
6. Tapuria A, Porat T, Kalra D, Dsouza G, Xiaohui S, Curcin V. Impact of patient access to their electronic health record: systematic review. Inform Health Soc Care 2021 Jun 02;46(2):192-204. [doi: 10.1080/17538157.2021.1879810] [Medline: 33840342]
7. Abd-Alrazaq AA, Bewick BM, Farragher T, Gardner P. Factors that affect the use of electronic personal health records among patients: A systematic review. Int J Med Inform 2019 Jun;126:164-175. [doi: 10.1016/j.ijmedinf.2019.03.014] [Medline: 31029258]
8. Fraccaro V, Balatsoukas B, Peek VDV. Patient portal adoption rates: A systematic literature review and meta-analysis. MEDINFO 2017: Precision Health Care Through Informatics. Hangzhou, China: International Medical Informatics Association; 2017. URL: https://livrepository.liverpool.ac.uk/3043795/1/patient portal.pdf [accessed 2021-10-05]
9. Zhao JY, Song B, Anand E, Schwartz D, Panesar M, Jackson GP, et al. Barriers, Facilitators, and Solutions to Optimal Patient Portal and Personal Health Record Use: A Systematic Review of the Literature. AMIA Annu Symp Proc 2017;2017:1913-1922 [FREE Full text] [Medline: 29854263]
10. Ammenwerth E. From eHealth to ePatient: The Role of Patient Portals in Fostering Patient Empowerment. ejbi 2018;14(2):20-23. [doi: 10.24105/ejbi.2018.14.2.4]
11. Lyles C, Nelson E, Frampton S, Dykes P, Cemballi A, Sarkar U. Using Electronic Health Record Portals to Improve Patient Engagement: Research Priorities and Best Practices. Annals of Internal Medicine 2020 Jun 02;172(11_Supplement):S123-S129. [doi: 10.7326/m19-0876]
12. Thompson M, Reilly J, Valdez R. Work system barriers to patient, provider, and caregiver use of personal health records: A systematic review. Appl Ergon 2016 May;54:218-242. [doi: 10.1016/j.apergo.2015.10.010] [Medline: 26851482]
13. Niazkhani Z, Toni E, Cheshmekaboodi M, Georgiou A, Pirnejad H. Barriers to patient, provider, and caregiver adoption and use of electronic personal health records in chronic care: a systematic review. BMC Med Inform Decis Mak 2020 Jul 08;20(1):153 [FREE Full text] [doi: 10.1186/s12911-020-01159-1] [Medline: 32641128]
14. Agarwal R, Anderson C, Zarate J, Ward C. If we offer it, will they accept? Factors affecting patient use intentions of personal health records and secure messaging. J Med Internet Res 2013 Feb 26;15(2):e43 [FREE Full text] [doi: 10.2196/jmir.2243] [Medline: 23470453]
15. Assadi V, Hassanein K. Consumer Adoption of Personal Health Record Systems: A Self-Determination Theory Perspective. J Med Internet Res 2017 Jul 27;19(7):e270 [FREE Full text] [doi: 10.2196/jmir.7721] [Medline: 28751301]
16. Liu L, Miguel Cruz A, Rios Rincon A, Buttar V, Ranson Q, Goertzen D. What factors determine therapists' acceptance of new technologies for rehabilitation – a study using the Unified Theory of Acceptance and Use of Technology (UTAUT). Disabil Rehabil 2015;37(5):447-455. [doi: 10.3109/09638288.2014.923529] [Medline: 24901351]
17. Kujala S, Hörhammer I, Kaipio J, Heponiemi T. Health professionals' expectations of a national patient portal for self-management. Int J Med Inform 2018 Sep;117:82-87 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.06.005] [Medline: 30032968]
18. Yousef CC, Thomas A, Alenazi AO, Elgadi S, Abu Esba LC, AlAzmi A, et al. Adoption of a Personal Health Record in the Digital Age: Cross-Sectional Study. J Med Internet Res 2020 Oct 28;22(10):e22913 [FREE Full text] [doi: 10.2196/22913] [Medline: 32998854]
19. Yousef CC, Salgado TM, Farooq A, Burnett K, McClelland LE, Thomas A, et al. Predicting Patients' Intention to Use a Personal Health Record Using an Adapted Unified Theory of Acceptance and Use of Technology Model: Secondary Data Analysis. JMIR Med Inform 2021 Aug 17;9(8):e30214 [FREE Full text] [doi: 10.2196/30214] [Medline: 34304150]
20. Patel V, Abramson E, Edwards A, Cheung M, Dhopeshwarkar RV, Kaushal R. Consumer attitudes toward personal health records in a beacon community. Am J Manag Care 2011 Apr;17(4):e104-e120 [FREE Full text] [Medline: 21774099]
21. Kaelber DC, Jha AK, Johnston D, Middleton B, Bates DW. A research agenda for personal health records (PHRs). J Am Med Inform Assoc 2008;15(6):729-736 [FREE Full text] [doi: 10.1197/jamia.M2547] [Medline: 18756002]

22. Nazi KM. The personal health record paradox: health care professionals' perspectives and the information ecology of personal health record systems in organizational and clinical settings. J Med Internet Res 2013 Apr 04;15(4):e70 [FREE Full text] [doi: 10.2196/jmir.2443] [Medline: 23557596]
23. Wiljer D, Urowitz S, Apatu E, DeLenardo C, Eysenbach G, Harth T, Canadian Committee for Patient Accessible Health Records. Patient accessible electronic health records: exploring recommendations for successful implementation strategies. J Med Internet Res 2008 Oct 31;10(4):e34 [FREE Full text] [doi: 10.2196/jmir.1061] [Medline: 18974036]
24. Wass S, Vimarlund V. Same, same but different: Perceptions of patients' online access to electronic health records among healthcare professionals. Health Informatics J 2019 Dec;25(4):1538-1548 [FREE Full text] [doi: 10.1177/1460458218779101] [Medline: 29874962]
25. Moll J, Cajander Å. Oncology health-care professionals' perceived effects of patient accessible electronic health records 6 years after launch: A survey study at a major university hospital in Sweden. Health Informatics J 2020 Jun;26(2):1392-1403 [FREE Full text] [doi: 10.1177/1460458219881007] [Medline: 31621459]
26. Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
27. Venkatesh V, Thong JYL, Xu X. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly 2012;36(1):157. [doi: 10.2307/41410412]
28. Kohnke A, Cole ML, Bush R. Incorporating UTAUT Predictors for Understanding Home Care Patients' and Clinician's Acceptance of Healthcare Telemedicine Equipment. Journal of Technology Management & Innovation 2014 Jul;9(2):29-41. [doi: 10.4067/s0718-27242014000200003]
29. Alaboudi A, Atkins A, Sharp B, Alzahrani M, Balkhair A, Sunbul T. Perceptions and attitudes of clinical staff towards telemedicine acceptance in Saudi Arabia. New York, NY: IEEE; 2016 Presented at: IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA); November 29 to December 2, 2016; Agadir, Morocco p. 1-8. [doi: 10.1109/aiccsa.2016.7945714]
30. Kim S, Lee K, Hwang H, Yoo S. Analysis of the factors influencing healthcare professionals' adoption of mobile electronic medical record (EMR) using the unified theory of acceptance and use of technology (UTAUT) in a tertiary hospital. BMC Med Inform Decis Mak 2016 Jan 30;16:12 [FREE Full text] [doi: 10.1186/s12911-016-0249-8] [Medline: 26831123]
31. Nematollahi M, Moosavi A, Lazem M, Aslani N, Kafashi M, Garavand A. Factors Affecting in Adoption and Use of Electronic Medical Record Based on Unified Theory of Acceptance and Use of Technology in Iran. Shiraz E-Med J 2017 Aug 27;18(9):e57582. [doi: 10.5812/semj.57582]
32. Mammen A, Weeks R. Electronic Medical Record (EMR) technology acceptance by healthcare professionals in South Africa. New York, NY: IEEE; 2014 Presented at: Proceedings of PICMET '14 Conference: Portland International Center for Management of Engineering and Technology; Infrastructure and Service Integration; July 27-31, 2014; Kanazawa, Japan p. 3539-3548.
33. Hossain A, Quaresma R, Rahman H. Investigating factors influencing the physicians' adoption of electronic health record (EHR) in healthcare system of Bangladesh: An empirical study. International Journal of Information Management 2019 Feb;44:76-87. [doi: 10.1016/j.ijinfomgt.2018.09.016]
34. Venkatesh V, Sykes T, Zhang X. 'Just what the doctor ordered': A revised UTAUT for EMR system adoption and use by doctors. New York, NY: IEEE; 2011 Presented at: 2011 44th Hawaii International Conference on System Sciences; January 4-7, 2011; Kauai, HI p. 1-10. [doi: 10.1109/hicss.2011.1]
35. Ibrahim S, Donelle L, Regan S, Sidani S. Predicting Registered Nurses' Behavioural Intention to Use Electronic Documentation System in Home Care: Application of an Adapted Unified Theory of Acceptance and Use of Technology Model. Nurs Leadersh (Tor Ont) 2019 Jun 28;32(2):48-70. [doi: 10.12927/cjnl.2019.25961] [Medline: 31613214]
36. Jahanbakhsh M, Nazemi Z, Mohammadi F, Hasanzadeh A. A study of picture archiving and communication system adoption in one hospital: Applying the unified theory of acceptance and use of technology model. J Educ Health Promot 2018;7:103 [FREE Full text] [doi: 10.4103/jehp.jehp_149_17] [Medline: 30159349]
37. Bawack RE, Kala Kamdjoug JR. Adequacy of UTAUT in clinician adoption of health information systems in developing countries: The case of Cameroon. Int J Med Inform 2018 Jan;109:15-22. [doi: 10.1016/j.ijmedinf.2017.10.016] [Medline: 29195701]
38. Ifinedo P. Technology acceptance by health professionals in Canada: an analysis with a modified UTAUT model. New York, NY: IEEE; 2012 Presented at: 2012 45th Hawaii International Conference on System Sciences; January 4-7, 2012; Maui, HI. [doi: 10.1109/hicss.2012.556]
39. Venkatesh V, Thong J, Xu X. Unified Theory of Acceptance and Use of Technology: A Synthesis and the Road Ahead. JAIS 2016 May;17(5):328-376. [doi: 10.17705/1jais.00428]
40. Dwivedi Y, Rana N, Jeyaraj A, Clement M, Williams M. Re-examining the Unified Theory of Acceptance and Use of Technology (UTAUT): Towards a Revised Theoretical Model. Inf Syst Front 2017 Jun 8;21(3):719-734. [doi: 10.1007/s10796-017-9774-y]
41. Rosen P. The Effect of Personal Innovativeness in the Domain of Information Technology on the Acceptance and Use of Technology: A Working Paper. 2004 Jan. URL: https://www.researchgate.net/publication/228868534_The_effect_of_

personal_innovativeness_in_the_domain_of_information_technology_on_the_acceptance_and_use_of_technology [accessed 2021-10-05]
42. Wu J, Du H. Toward a better understanding of behavioral intention and system usage constructs. European Journal of Information Systems 2017 Dec 19;21(6):680-698. [doi: 10.1057/ejis.2012.15]
43. Williams MD, Rana NP, Dwivedi YK. The unified theory of acceptance and use of technology (UTAUT): a literature review. J Ent Info Manage 2015 Apr 13;28(3):443-488. [doi: 10.1108/JEIM-09-2014-0088]
44. Balau M. Exploring the Link Between Intention and Behavior in Consumer Research. 2018 May. URL: https://www.researchgate.net/publication/325972981_Exploring_the_Link_between_Intention_and_Behavior_in_Consumer_Research [accessed 2021-10-05]
45. Li J, Talaei-Khoei A, Seale H, Ray P, Macintyre CR. Health Care Provider Adoption of eHealth: Systematic Literature Review. Interact J Med Res 2013;2(1):e7 [FREE Full text] [doi: 10.2196/ijmr.2468] [Medline: 23608679]
46. MNGHA Organization Profile. Ministry of National Guard - Health Affairs. 2018. URL: https://ngha.med.sa/English/AboutNGHA/Pages/profile.aspx [accessed 2019-09-06]
47. Chung M, Ho C, Wen H. Predicting intentions of nurses to adopt patient personal health records: A structural equation modeling approach. Comput Methods Programs Biomed 2016 Nov;136:45-53. [doi: 10.1016/j.cmpb.2016.08.004] [Medline: 27686702]
48. Hennemann S, Beutel ME, Zwerenz R. Ready for eHealth? Health Professionals' Acceptance and Adoption of eHealth Interventions in Inpatient Routine Care. J Health Commun 2017 Mar 19;22(3):274-284. [doi: 10.1080/10810730.2017.1284286] [Medline: 28248626]
49. Hoogenbosch B, Postma J, de Man-van Ginkel JM, Tiemessen NA, van Delden JJ, van Os-Medendorp H. Use and the Users of a Patient Portal: Cross-Sectional Study. J Med Internet Res 2018 Sep 17;20(9):e262 [FREE Full text] [doi: 10.2196/jmir.9418] [Medline: 30224334]
50. Tavares J, Oliveira T. Electronic Health Record Patient Portal Adoption by Health Care Consumers: An Acceptance Model and Survey. J Med Internet Res 2016 Mar;18(3):e49 [FREE Full text] [doi: 10.2196/jmir.5069] [Medline: 26935646]
51. Khaneghah P, Miguel-Cruz A, Bentley P, Liu L, Stroulia E, Ferguson-Pell M. Users' attitudes towards personal health records. Appl Clin Inform 2017 Dec 16;07(02):573-586. [doi: 10.4338/aci-2015-12-ra-0180]
52. QuestionPro. URL: https://www.questionpro.com/ [accessed 2021-10-10]
53. Dünnebeil S, Sunyaev A, Blohm I, Leimeister JM, Krcmar H. Determinants of physicians' technology acceptance for e-health in ambulatory care. Int J Med Inform 2012 Nov;81(11):746-760. [doi: 10.1016/j.ijmedinf.2012.02.002] [Medline: 22397989]
54. Davis FD, Bagozzi RP, Warshaw PR. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
55. Ringle C, Sarstedt M, Straub D. Editor's Comments: A Critical Look at the Use of PLS-SEM in "MIS Quarterly". MIS Quarterly 2012;36(1):iii. [doi: 10.2307/41410402]
56. Farooq A, Ahmad F, Khadam N, Lorenz B, Isoaho J. The impact of perceived security on intention to use e-learning among students. 2020 Presented at: IEEE 20th International Conference on Advanced Learning Technologies (ICALT) 2020; July 6-9, 2020; Tartu, Estonia p. 360-364. [doi: 10.1109/icalt49669.2020.00115]
57. Farooq A, Laato S, Islam A, Isoaho J. Understanding the impact of information sources on COVID-19 related preventive measures in Finland. Technology in Society 2021 May;65:101573. [doi: 10.1016/j.techsoc.2021.101573]
58. Kline RB. Principles and Practice of Structural Equation Modeling. New York, NY: Guilford Publications; 2015.
59. IBM Corporation. IBM SPSS Statistics for Windows. Armonk, NY: IBM Corporation; 2017. URL: https://www.ibm.com/in-en/products/spss-statistics [accessed 2021-10-05]
60. Ringle C, Wende S, Becker J. SmartPLS 3.0. Hamburg, Germany: SmartPLS; 2014. URL: https://www.smartpls.com [accessed 2021-10-05]
61. Hair JF, Risher JJ, Sarstedt M, Ringle CM. When to use and how to report the results of PLS-SEM. EBR 2019 Jan 14;31(1):2-24. [doi: 10.1108/ebr-11-2018-0203]
62. Tavares J, Oliveira T. Electronic Health Record Portal Adoption: a cross country analysis. BMC Med Inform Decis Mak 2017 Jul 05;17(1):97 [FREE Full text] [doi: 10.1186/s12911-017-0482-9] [Medline: 28679423]
63. Al-Sahan A, Saddik B. Perceived challenges for adopting the Personal Health Record (PHR) at Ministry of National Guard Health Affairs (MNGHA)- Riyadh. Online J Public Health Inform 2016;8(3):e205 [FREE Full text] [doi: 10.5210/ojphi.v8i3.6845] [Medline: 28210426]
64. Alanazi A, Al Anazi Y. The challenges in personal health record adoption. Journal of Healthcare Management 2019;64(2):104-109. [doi: 10.1097/jhm-d-17-00191]
65. Abd-Alrazaq A, Bewick B, Farragher T, Gardner P. Factors Affecting Patients' Use of Electronic Personal Health Records in England: Cross-Sectional Study. J Med Internet Res 2019 Jul 31;21(7):e12373 [FREE Full text] [doi: 10.2196/12373] [Medline: 31368442]
66. Abd-Alrazaq A, Alalwan A, McMillan B, Bewick B, Househ M, Al-Zyadat A. Patients' Adoption of Electronic Personal Health Records in England: Secondary Data Analysis. J Med Internet Res 2020 Oct 07;22(10):e17499 [FREE Full text] [doi: 10.2196/17499] [Medline: 33026353]

67. Hsieh H, Kuo Y, Wang S, Chuang B, Tsai C. A Study of Personal Health Record User's Behavioral Model Based on the PMT and UTAUT Integrative Perspective. Int J Environ Res Public Health 2016 Dec 23;14(1):8 [FREE Full text] [doi: 10.3390/ijerph14010008] [Medline: 28025557]
68. Macintosh CI. Healthcare Provider Attitudes and Experiences with Patient Use of Personal Health Records: Evaluation and Modification of the Unified Theory of Acceptance and Use of Technology. 2017. URL: https://tinyurl.com/3yrmptsh [accessed 2021-10-10]
69. Park H, Kim KI, Soh J, Hyun Y, Jang S, Lee S, et al. Factors Influencing Acceptance of Personal Health Record Apps for Workplace Health Promotion: Cross-Sectional Questionnaire Study. JMIR Mhealth Uhealth 2020 Jun 04;8(6):e16723 [FREE Full text] [doi: 10.2196/16723] [Medline: 32496202]
70. Alsahafi Y, Gay V, Khwaji A. Factors affecting the acceptance of integrated electronic personal health records in Saudi Arabia: The impact of e-health literacy. Health Inf Manag 2020 Nov 28:1833358320964899. [doi: 10.1177/1833358320964899] [Medline: 33249857]
71. Alaiad A, Alsharo M, Alnsour Y. The Determinants of M-Health Adoption in Developing Countries: An Empirical Investigation. Appl Clin Inform 2019 Oct 30;10(5):820-840 [FREE Full text] [doi: 10.1055/s-0039-1697906] [Medline: 31667819]
72. Ryan BL, Brown JB, Terry A, Cejic S, Stewart M, Thind A. Implementing and Using a Patient Portal: A qualitative exploration of patient and provider perspectives on engaging patients. J Innov Health Inform 2016 Jul 04;23(2):848 [FREE Full text] [doi: 10.14236/jhi.v23i2.848] [Medline: 27869582]
73. Vreugdenhil M, Ranke S, de Man Y, Haan M, Kool R. Patient and Health Care Provider Experiences With a Recently Introduced Patient Portal in an Academic Hospital in the Netherlands: Mixed Methods Study. J Med Internet Res 2019 Aug 20;21(8):13743 [FREE Full text] [doi: 10.2196/13743] [Medline: 31432782]
74. Hulley S. Designing Clinical Research. Philadelphia, PA: Lippincott Williams & Wilkins; 2007:85-87.
75. Timmins F. Nursing Research Generating and Assessing Evidence for Nursing Practice. Nurse Education in Practice 2013 Nov;13(6):e29. [doi: 10.1016/j.nepr.2013.04.001]

Abbreviations

EHR: electronic health record
HCP: health care provider
HITECH: Health Information Technology for Economic and Clinical Health
MNGHA: Ministry of National Guard Health Affairs
PHR: personal health record
TAM: technology acceptance model
TPB: theory of planned behavior
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by G Eysenbach; submitted 26.06.21; peer-reviewed by H Sun, J Moll, N Archer; comments to author 26.08.21; revised version received 25.09.21; accepted 26.09.21; published 21.10.21.

Please cite as:
Yousef CC, Salgado TM, Farooq A, Burnett K, McClelland LE, Abu Esba LC, Alhamdan HS, Khoshhal S, Aldossary IF, Alyas OA, DeShazo JP
Health Care Providers' Acceptance of a Personal Health Record: Cross-sectional Study
J Med Internet Res 2021;23(10):e31582
URL: https://www.jmir.org/2021/10/e31582
doi:10.2196/31582
PMID:34569943

Original Paper

Human-Technology Interaction Factors Associated With the Use of Electronic Personal Health Records Among Younger and Older Adults: Secondary Data Analysis

Yan Luo[1], MSW; Krystal Dozier[1], MSW; Carin Ikenberg[1], MSW

School of Social Work, The University of Alabama, Tuscaloosa, AL, United States

Corresponding Author:
Yan Luo, MSW
School of Social Work
The University of Alabama
2106 Forest Lake Dr
Tuscaloosa, AL, 35401
United States
Phone: 1 205 886 8781
Email: yluo30@crimson.ua.edu

Abstract

Background: An electronic personal health record (ePHR), also known as a personal health record (PHR), has been broadly defined as an electronic application through which individuals can access, manage, and share their health information in a secure and confidential environment. Although ePHRs can benefit individuals as well as caregivers and health care providers, the use of ePHRs among individuals continues to remain low.

Objective: The current study aims to examine the relationship between human-technology interaction factors and ePHR use among adults and then to compare the different effects of human-technology interaction factors on ePHR use between younger adults (18-54 years old) and older adults (55 years of age and over).

Methods: We analyzed data from the Health Information National Trends Survey (HINTS 5 cycle 3) collected from US adults aged 18 years old and over in 2019. Descriptive analysis was conducted for all variables and each item of ePHR use. Bivariate tests (Pearson correlation coefficient for categorical variable and *F* test for continuous variables) were conducted over 2 age groups. Finally, after adjustments were made for sociodemographics and health care resources, a weighted multiple linear regression was conducted to examine the relationship between human-technology interaction factors and ePHR use.

Results: The final sample size of 1363 (average age 51.19) was divided into 2 age groups: 18 to 54 years old and 55 years old and older. The average level of ePHR use was low (mean 2.76, range 0-8). There was no significant difference in average ePHR use between the 2 age groups. Including clinical notes was positively related to ePHR use in both groups: 18 to 54 years old (β=.*28*, *P*=.005), 55 years old and older (β=.15, *P*=.006). Although accessing ePHRs using a smartphone app was only associated with ePHR use among younger adults (β=.29*; P*<.001), ease of understanding health information in ePHRs was positively linked to ePHR use only among older adults (β=.13; *P*=.003).

Conclusions: This study found that including clinical notes was positively related to ePHR use in both age groups, which suggested that including clinical notes as a part of ePHRs might improve the effective use of ePHRs among patients. Moreover, accessing ePHRs using a smartphone app was associated with higher ePHR use among younger adults while ease of understanding health information in ePHRs was linked to higher ePHR use among older adults. The design of ePHRs should provide the option of being accessible through mobile devices to promote greater ePHR use among young people. For older adults, providers could add additional notes to explain the health information recorded in the ePHRs.

(J Med Internet Res 2021;23(10):e27966) doi:10.2196/27966

KEYWORDS

electronic personal health records; human-technology interaction factors; clinical notes; smartphone app; ease of understanding

Introduction

Electronic Personal Health Record and Its Functions

An electronic personal health record (ePHR), also known as a personal health record (PHR), was broadly defined by the Markle Foundation (2004) to be an electronic application through which individuals can access, manage, and share their health information in a secure and confidential environment [1]. Unlike the electronic health record that is managed by health care providers, an ePHR is managed by individuals [2]. Pagliari and colleagues [3] summarized 7 potential functions of ePHRs: (1) access to health care providers' electronic clinical records (eg, history, drugs, test results); (2) personal health organizer or diary (eg, clinics, doctors, tests, dates, nonprescribed treatments, scanned documents); (3) self-management support (eg, care plans, graphing of symptoms, passive biofeedback, tailored instructive or motivational feedback, decision aids, or reminders); (4) secure patient-provider communication for scheduling appointments, reordering prescriptions, or seeking advice (eg, patient-doctor email); (5) links to static or informative information about illnesses, treatments, or self-care; (6) links to sources of support; and (7) collective data on symptom or health behavior data by self-report or objective monitoring through electronic devices.

Benefits of Using ePHRs

Using ePHRs allows individuals to access and coordinate their health information and to share appropriate parts to those who need it [1]. The use of ePHRs can benefit individuals as well as caregivers and health care providers [2,3]. For individuals, ePHRs provide them with credible health information, data, and guidance on potential ways to self-manage diseases and improve health, which facilitates collaborative disease tracking and improved communication between individuals (or their caregivers) and health care providers [2,3]. Moreover, ePHRs provide health care providers with more data on individuals, which allows the provider to make informed decisions, as well as improve the efficiency of care by empowering individuals' active involvement in health care and enabling PHR-mediated electronic communication [2,3]. For payers and purchasers of health care, the use of ePHRs has the potential to lower costs on chronic disease management, medications, and wellness programs [2]. Several studies have been conducted to evaluate some of the benefits of using ePHRs [4]. For example, a clinical trial testing the effects of ePHRs on advance care planning delivery in primary care settings revealed that using ePHRs improved advanced care planning documentation and quality, especially among patients between 50 and 60 years of age [5]. Another study evaluated the impact of using a decision module through ePHRs to inform cancer screening and demonstrated that participants' decision on cancer screening can be proactively facilitated through an ePHR decision module [6]. Aside from primary care and preventive care settings, ePHRs also play a positive role in mental health care settings. In a study that compared scores of Patient Health Questionnaire 9 between participants who used ePHRs with their collaborative care managers and those who did not use ePHRs, Pecina and colleagues [7] suggested that ePHR users had a higher number of contacts with care managers and showed higher depression remission.

ePHR Use and Age Disparity

Although ePHR use has been promoted and health care providers have offered most of their patients access to ePHRs [8], the use of ePHRs among individuals continues to remain low. Using data from the Health Information National Trends Surveys (HINTS) in 2008, 2011, and 2013, a study predicted that the ePHR adoption rate would exceed 75% by 2020 [9]. Although the most recent data of the ePHR use rate in 2020 are not available, a study analyzing data from the HINTS data set in 2018 reported that the use of ePHRs in the United States was only 31.4% [10]. The relationship between age and ePHR use has been documented in previous studies, which indicate that younger age is related to higher ePHR use and that patients who are younger are more likely to use ePHRs [11,12]. Pagliari and colleagues [3] pointed out that older adults had poor technical skills that might cause access disparities regarding use of ePHRs. When encountering technology, older adults face physical or cognitive challenges [13-15]. Age-related changes in functional abilities, such as sight loss, hearing loss, decreased kinesthetic ability, and decreased psychomotor and cognitive skills pose barriers for older adults to use technology [13,14,16]. Previous studies have documented barriers to adopting ePHRs among older adults, including a lack of confidence in the ability to use technology [17], concerns related to privacy [18], problems with access to computers or devices and the PHR system [15], and low health literacy or computer literacy [15]. Examining the human-technology interaction factors associated with ePHR use among individuals in different age groups might inspire tailored ePHR design and training regarding ePHR use among people from different age groups.

After reviewing 97 studies regarding factors that affect the use of ePHRs among patients, a systematic review identified 3 human-technology interaction factors that affect ePHR use: perceived usefulness (positively), internet access (positively), and privacy and security concerns (negatively) [19]. Previous studies have also suggested that other human-technology interaction factors including perceived ease of use [20,21], difficulty getting onto the system [22], and response costs [23] are associated with the use of ePHRs; however, Abd-Alrazaq and colleagues [19] believe that more evidence is needed to draw a firm conclusion regarding these factors.

This Study

Broadly, the use of ePHRs provides benefits not only to individuals and caregivers, but also health care providers. Thus, the US Department of Health and Human Services has made investments and efforts to improve ePHR use [8]. Despite the low rate of ePHR use and the digital divide between different age groups, previous studies have not investigated the different effects of human-technology interaction factors on ePHR use between different age groups. Controlling for sociodemographics and health care resources, the current study aims to examine the relationship between human-technology interaction factors and ePHR use among adults and to then compare the different effects of human-technology interaction factors on ePHR use between younger adults (18-54 years old) and older adults (55

years old and older). Although 65 years is widely used as a cutoff point for older adults, using 55 years as the cutoff point in this study was based on previous literature on technology use among older adults. In the United States, the National Telecommunications and Information Administration (2011) used 55 years as a cutoff point and reported that older Americans aged 55 years and older had the lowest adoption rate of broadband [24]. Moreover, a European project, "ICT 4 the Elderly", developed to improve older adults' digital skills, also defined older adults using 55 years as a cutoff point [25]. Other literature that has studied the use of health information technology among older adults has used 55 as a cutoff point as well [10,26,27].

Methods

Data Collection

In this study, we used the most recent iteration of the HINTS 5 (cycle 3) [28] collected from US adults aged 18 years old and over in 2019. HINTS is a national representative data set from the National Cancer Institute, and it routinely collects data about the American public's knowledge of, attitudes toward, and use of cancer- and health-related information. Since 2003, HINTS has been used by researchers to understand health communication through the internet in the information age among American adults. Two-stage sampling strategy and two-sampling strata (high- and low-minority strata) were applied during the data collection phase. Random samples of household addresses were selected in the first stage, and 1 adult within each sampled household was randomly selected in the second stage. All selected households received a total of 4 mailings: an initial mailing with a US $2 incentive, a reminder postcard, and 2 follow-up mailings. Participants were provided with 2 toll-free phone numbers (for English and Spanish calls) if they had questions, concerns, or requests for the Spanish survey. Each returned questionnaire was scanned, verified, cleaned, and edited. The final sample yielded 4448 potential respondents with a response rate of 30.2% (4448/14,730) and 3370 completed questionnaires. Only participants who had accessed their ePHRs at least 1 time in the past 12 months were included in this study. The final sample size of 1363 was divided into 2 age groups: 18 to 54 years old and 55 years old and older.

Dependent Variable

The dependent variable of interest was ePHR use. Participants who accessed their ePHRs at least once in the past 12 months were asked if in the past 12 months they used their online medical record to do any of the following: request a refill of medications; look up test results; request correction of inaccurate information; securely message health care provider and staff; download health information to a computer or mobile device, such as a cell phone or tablet; add health information, such as health concerns, symptoms, and side effects, to share with a health care provider; and help make decisions about how to treat an illness or condition. Each item was answered with a yes or no response by respondents (0=no, 1=yes). The eighth item of ePHR use pertained to sending health information electronically. Participants were asked if they had electronically sent their medical information to another health care provider, to a family member or another person involved with their care, or to a service or app that could help manage and store their health information. This response was also answered with a yes or no response for each option. Participants who selected yes on one of the options were coded as yes on sending health information electronically, while participants who selected no on all 3 options were coded as no on sending health information electronically (0=no, 1=yes). The total ePHR use score was obtained by summing up all 8 items and was analyzed as a continuous variable (range from 0 to 8).

Sociodemographics and Health Care Resources

Sociodemographics

Sociodemographic variables including gender (0=male, 1=female), urbanity (0=rural, 1=urban), and educational attainment (0=below bachelor's degree, 1=bachelor's degree and above) were included.

Health Care Resources

Having a regular health care provider (0=no, 1=yes) and frequency of visiting health care providers in the past 12 months (0=0-3 times, 1=4 times and above) were included. Having family or friends to talk to about health was also included and analyzed as a dichotomous variable (0=no, 1=yes).

Human-Technology Interaction Factors

Including Clinical Notes

Respondents were asked the following: "Do any of your online medical records include clinical notes (health provider's notes that describe a visit)?", with responses yes, no, and "don't know." After responses of no and "don't know" were combined into 1 category, a dichotomous variable was obtained (0=no/don't know, 1=yes).

Ease of Understanding

To determine ease of understating, a 4-point scale was used for participant responses to the following question: "How easy or difficult was it to understand the health information in your online medical records?" Ease of understanding was analyzed as a continuous variable ranging from 0 to 3 (0=very difficult, 1=somewhat difficult, 2=somewhat easy, 3=very easy).

Access via Smartphone App

Respondents were asked the following question: "Did you use a smartphone health app to access your online medical record?", with responses categorized as yes, no, and "don't know." Responses of no and "don't know" were grouped into 1 category; thus, a dichotomous variable was used for the accessibility of a smartphone app (0=no/don't know, 1=yes).

Statistical Analysis

Three researchers in this study devised the statistical analysis plan, and the statistical analyses were conducted by YL. The results and interpretation were reviewed by KD and CI. As the complex sampling procedure was applied in the HINTS data collection, the data analysis in this study was conducted using STATA/SE 5.1 (StataCorp), which allowed for incorporating the jackknife replicate weights to assess variation estimation. Descriptive analysis was conducted for all variables and each

item of ePHR use. Bivariate tests (Pearson correlation coefficient for categorical variables and *F* test for continuous variables) were conducted over the 2 age groups. Finally, after sociodemographics and health care resource factors were adjusted for, a weighted multiple linear regression was conducted to examine the relationship between human-technology interaction factors and ePHR use. The final sample weight was used to obtain population estimates, and 50 jackknife replicate weights were used to obtain variation estimates. Listwise deletion of participants was also applied in all analyses.

Results

Description of Sociodemographics, Health Care Resources, and Human-Technology Interaction Factors

The average age of all participants was 51.18 years. According to Table 1, more than half of the participants were female (762/1266, 57.12% weighted), and less than half of the participants had a bachelor's degree or above (790/1334, 39.91% weighted). The majority of participants were from urban areas (1245/1363, 89.75% weighted). In terms of health care resources, more than three-quarters of the participants had a regular health care provider (1091/1339, 78.35% weighted), and around half of the participants visited health care providers more than 4 times in the past 12 months (690/1346, 49.43%). Most of participants reported that they had friends or family to talk to about health (1167/1332, 85.62% weighted). With regard to human-technology interaction factors of ePHRs, about 40% of participants reported that they accessed their ePHRs using a smartphone app (436/1290, 39.56% weighted), and half of the participants said their ePHRs included clinical notes (650/1278, 50.34% weighted). Participants tended to report that it was easy to understand health information in ePHRs (mean 2.31, range 0-3). Table 1 also shows the significant differences in having a regular provider and accessing ePHRs using a smartphone app between the 2 age groups ($P<.001$). More older adults (55 years old and older) reported having a regular health provider, while more younger adults (18-54 years old) reported accessing ePHRs using a smartphone app.

Table 1. Description of sociodemographics, health care resources, and human-technology interaction factors (N=1363).

Characteristic	All	18-54 years old	55 years old and older	*P* value
Sociodemographics, n (%)				
Gender				
Male	504 (42.88)	190 (41.4)	309 (45.7)	.27
Female	762 (57.12)	342 (58.6)	414 (54.3)	
Urbanity				.99
Rural	118 (10.25)	46 (10.2)	68 (10.3)	
Urban	1245 (89.75)	505 (89.8)	706 (89.7)	
Education				.10
Below Bachelor's degree	544 (60.09)	182 (58.0)	354 (64.2)	
Bachelor's degree and above	790 (39.91)	369 (42.1)	416 (35.8)	
Health care resources, n (%)				
Regular health care provider				<.001
No	248 (21.65)	150 (27.0)	86 (11.4)	
Yes	1091 (78.35)	395 (73.0)	676 (88.6)	
Frequency of visiting health care provider				.47
0-3 times in the past 12 months	656 (50.57)	297 (51.6)	334 (48.5)	
4 times and above in the past 12 months	690 (49.43)	253 (48.4)	421 (51.5)	
Having friends/family to talk to about health				.41
No	165 (14.38)	68 (13.5)	91 (16.2)	
Yes	1167 (85.62)	481 (86.5)	666 (83.8)	
Human-technology interaction factors, n (%)				
Accessing ePHRs[a] using smartphone app				<.001
No/don't know	854 (60.44)	305 (54.8)	528 (71.8)	
Yes	436 (39.56)	238 (45.2)	193 (28.2)	
ePHRs include clinical notes				.40
No/don't know	628 (49.66)	265 (51.1)	345 (46.8)	
Yes	650 (50.34)	275 (48.9)	369 (53.2)	
Ease of understanding ePHRs health information (range 0-3), mean	2.31	2.33	2.28	.42

[a]ePHRs: electronic personal health records.

Description of ePHR Use

Participants' ePHR use is reported in Table 2. The average level of ePHR use was low (mean 2.76, range 0-8). There was no significant difference in average ePHR use between the 2 age groups. Table 2 also shows the rate on each item of ePHR use. Specifically, the majority of participants used ePHRs to look up test results (1081/1277, 84.59% weighted). Around half of the participants used ePHRs to request a refill of medications (596/1276, 46.57% weighted) and securely message health care provider and staff (686/1278, 52.96% weighted). About one-quarter of participants used ePHRs to download health information to a computer or mobile device (292/1276, 25.88% weighted), add health information to share with health care providers (307/1278, 23.58% weighted), and help make a decision about how to treat an illness or condition (324/1274, 24.77% weighted). A small percentage of participants used ePHRs to request correction of inaccurate information (104/1263, 7.58% weighted) and electronically send health information (108/1268, 9.62% weighted). Significant differences between the 2 age groups were found related to using ePHRs to "download health information to a computer or mobile device," indicating that more younger adults used ePHRs to download health information to computers or mobile devices (P=0.04).

Table 2. Description of the use of electronic personal health records among participants (N=1363).

Participant use of ePHR[a] in the past 12 months	Total, n (%)	18-54 years old, n (%)	55 years old and above, n (%)	*P* values[b]
1. Request refill of medications	596 (46.57)	218 (44.0)	370 (51.8)	.13
2. Look up test results	1081 (84.59)	462 (84.2)	605 (85.3)	.75
3. Request correction of inaccurate information	104 (7.58)	39 (8.3)	63 (6.1)	.34
4. Securely message health care provider and staff (for example, email)	686 (52.96)	302 (54.9)	377 (49.0)	.10
5. Download your health information to your computer or mobile device, such as a cell phone or tablet	292 (25.88)	147 (28.7)	144 (20.1)	*.04* [c]
6. Add health information to share with your health care provider, such as health concerns, symptoms, and side effects	307 (23.58)	134 (24.0)	173 (22.8)	.74
7. Help you make a decision about how to treat an illness or condition	324 (24.77)	130 (24.7)	188 (25.0)	.94
8. Electronically send health information	108 (9.62)	65 (10.6)	41 (7.7)	.26

[a]ePHR: electronic personal health record.

[b]*F* test was used for all items.

[c]Italics indicate *P*<.05.

Weighted Multiple Linear Regression on Human-Technology Interaction Factors in Predicting ePHR Use Between 2 Age Groups

The regression analysis included 494 participants between 18 and 54 years old and 610 participants older than 55 years old. According to Table 3, at least 2 of the human-technology interaction factors of ePHRs were associated with the use of ePHRs among participants in both age groups. Including clinical notes was positively related to ePHR use in those 18 to 54 years old (β=.28; *P*=.005) and those 55 years old and older (β=.15; *P*=.006). Although accessing ePHRs using a smartphone app was only associated with ePHR use among younger adults (β=.29; *P*<.001), ease of understanding health information in ePHRs was positively linked to ePHR use only among older adults (β=.13; *P*=.003). Other than some human-technology interaction factors, having a regular health care provider and having friends or family to talk to about health were positively associated with the use of ePHRs among younger adults.

Table 3. Weighted multiple linear regression on human-technology interaction factors predicting electronic personal health record use between 2 age groups.

Predictor of use of ePHRs[a]	Standardized coefficient for 18-54 years old (β)	*P* values	Standardized coefficient for 55 years old and older (β)	*P* values
Sociodemographics				
Gender: female (ref[b]=male)	.01	.89	–.06	.27
Urbanity: urban (ref=rural)	.08	.40	.04	.76
Education: bachelor's degree and above (ref=below bachelor's degree)	–.05	.44	.11	.44
Health care resources				
Regular health care provider: yes (ref=no)	.14	.04	.12	.08
Frequency of visiting health care provider: 4 times and above (ref=0-3 times)	.02	.31	.05	.36
Having friends/family to talk to about health: yes (ref=no)	.12	.005	.10	.19
Human-technology interaction factors				
Accessing ePHRs using smartphone app: yes (ref=no/don't know)	.29	<.001	.10	.14
ePHRs include clinical notes: yes (ref=no/don't know)	.28	.005	.15	.006
Ease of understanding ePHR health information (range:0-3)	–.01	.89	.13	.003

[a]ePHRs: electronic personal health records.

[b]ref: reference.

Discussion

Principal Results and Comparison to Prior Work

Analyzing the most recent iteration of the HINTS collected in 2019, this study aimed to examine the relationship between human-technology interaction factors and ePHR use among adults and then to compare its different effects between younger adults (18-54 years old) and older adults (55 years old and older) while controlling for sociodemographics and health care resources.

The Level of ePHR Use Among Younger Adults and Older Adults

This study found that the average level of ePHR use was low (mean 2.76, range 0-8). This is in line with Hong and colleagues' [10] study that reported the use of ePHRs in the United States to be 31.4%. However, while Hong et al measured ePHR use by asking participants whether they had accessed ePHRs in the past 12 months (yes or no), our study only included participants who had accessed their ePHRs at least once in the past 12 months and measured the use level of different ePHR functions (eg, request refills of medications, look up test results, message health care provider and staff). This suggested that even among participants who accessed ePHRs, the use of ePHR functions is still low. The study also found that there was no significant difference in average ePHR use between the 2 age groups, which contradicts the findings of Greenberg et al [11] and McInnes et al [12], who reported younger age to be related to higher ePHR use. Including performance expectancy, effort expectancy, social influence, and facilitating conditions as independent variables, 2 studies conducted by Abd-Alrazaq and colleagues [29,30] found that age moderated the effects of performance expectancy, effort expectancy, and facilitating conditions on intention to use ePHRs. The moderating effect of age might be able to explain the nonsignificant finding of age difference in our study, which suggests that future studies are needed to explore the moderating effect of age using the current data set and measurements.

Human-Technology Interaction Factors Associated With ePHR Use

In terms of human-technology interaction factors associated with ePHR use, this study found that including clinical notes was positively related to ePHR use in both age groups. Previous studies examining the relationship between including clinical notes and ePHR use rates were not found. Nonetheless, in a survey evaluating veterans' access to an ePHR program called My HealtheVet Pilot, participants reported the highest rates (585/657, 89%) on using patient records including clinical notes or lab test results, and participants perceived that viewing medical records including clinical notes was the most useful feature of the ePHR programs [31]. In a qualitative study exploring participants' views on the My Health*e*Vet Pilot, participants identified that clinical notes promoted active patient participation by helping them prepare for the clinical visit, gain insight about their health and treatment plans, and gain insight into the providers' perspectives [32].

Regarding the different effects of human-technology interaction factors on ePHR use between the 2 age groups, accessing ePHRs using a smartphone app was significantly associated with ePHR use among younger adults while ease of understanding health information in ePHRs was significantly linked to ePHR use among older adults.

In terms of accessing ePHRs using a smartphone app, our findings are consistent with Bell et al's [33] findings that indicate accessing ePHRs through a mobile app to be associated with higher ePHR use. However, a conflicting finding was found in 2 previous studies: using the ePHRs only via a mobile device was related to infrequent use of ePHRs [34,35]. This discrepancy might be the result of samples with different characteristics in different studies being used. The study from Bell et al [33] was conducted among adults after elective orthopedic surgery, the study from Graetz et al [34] was conducted among adult patients with diabetes, and the study from Jung et al [35] was conducted with adults in South Korea. Moreover, previous studies showed that younger participants are more likely to use ePHRs only via a mobile device [33,34], which was also found in our study. The bivariate analysis of our study also indicated that younger adults were more likely to download health information to computers or mobile devices, such as a cell phone or tablet, which highlighted the significant role of mobile devices in ePHR use among younger adults.

Another human-technology interaction factor, ease of understanding health information in ePHRs, was found to be significantly linked to ePHR use among older adults but not younger adults. This finding is in line with Abd-Alrazaq et al's [29,30] studies, which suggested that perceived ease of use is positively associated with the intention to use ePHRs, with this relationship being stronger among older patients. This difference might be explained by the lower health literacy among older adults compared to their younger counterparts [36].

Conclusions

The purpose of this cross-sectional study was to examine the relationship between human-technology interaction factors and ePHR use among adults and then compare its different effects between younger adults (18-54 years old) and older adults (55 years old and older). The study found that the average level of ePHR use was low and that there was no significant difference in average use of ePHRs between the 2 age groups. Regarding the human-technology interaction factors, including clinical notes was positively related to ePHR use in both age groups, and accessing ePHRs using a smartphone app was positively associated with ePHR use among younger adults, while ease of understanding health information in ePHRs was a positive factor for ePHR use among older adults. The current study showed that there is a significant relationship between human-technology interaction factors and ePHR use and that the human-technology interaction factors associated with ePHR use vary across different age groups. In order to broadly promote the use of ePHRs, the design of ePHRs should take significant human-technology interaction factors into consideration, and the education or training regarding ePHR use should be provided for both health care providers and patients, especially for older adults.

Limitations

There are several limitations of this study. First, this study used a cross-sectional data set that was not able to examine causality between human-technology interaction factors and ePHR use. Second, the this study only included participants who were offered accesses to their ePHRs and accessed their ePHRs at least once in the past 12 months. Only 34% of US adults reported that they were offered access to their ePHRs [37], and, of those patients who were offered ePHR access, only 30% of patients actually accessed their ePHRs at least once in a year [11]. In order to promote meaningful use of ePHRs and maximize the benefit of ePHRs for patients, future studies may explore the factors that affect offering ePHR access to patients and patients' not accessing ePHRs even with access being granted. Finally, ePHR use in this study was measured by 8 self-reported items regarding the purposes for which participants used ePHRs, which might not have accurately recorded the actual use of ePHRs among participants. Future studies may consider using data including the frequencies and times that participants login to their ePHR accounts.

Implications for Practice and Future Research

Despite these limitations, this study is the first of its kind to examine the association between human-technology interaction factors and ePHR use among US adults and to compare its different effects between younger adults (18-54 years old) and older adults (55 years old and older). The findings of this study provide implications for practice and future research. This study found that including clinical notes was positively related to ePHR use in both age groups, which suggests that including clinical notes as a part of ePHRs might improve the effective use of ePHRs among patients. Although clinical notes can serve as a fundamental feature for ePHRs, participants in Woods et al's [32] study also demonstrated difficulties in seeing clinical notes, such as the use of derogatory terms, stress when seeing detailed personal information, and challenging conversations with providers. Although our study only examined "including clinical notes" as a single item, future studies are needed to explore patients' preferences on the type of clinical notes that should be included in ePHRs. This will also maximize the meaningful use of clinical notes. Moreover, this study found that accessing ePHRs using a smartphone app was associated with higher ePHR use among younger adults while ease of understanding health information in ePHRs was linked to higher ePHR use among older adults. The design of ePHRs should provide the option of being accessible through mobile devices to promote greater ePHR use among young people. For older adults, providers could add additional notes to explain the health information recorded in the ePHRs.

Empirical evidence has demonstrated that ePHRs provide consumers with easy and convenient access to their health data [1]. As the landscape of personal health care delivery changes due to increased technological advancements, there will be continued use of ePHRs. By addressing the concerns related to clarity in clinical notes for older adults and a simpler app platform for younger adults, ePHRs can increase access to health care data for both younger and older adults. With this increased use and access, it is important to highlight the benefits of using ePHRs in rural communities. Rural communities are often racially diverse, older, and tend to have lower incomes with limited access to health care [38]. In rural communities, telehealth is being used to address inequities in health care. Coupled with telehealth options, ePHRs can provide greater access to health data for individuals who reside in rural communities. This ease of access is also critical during times of prolonged crises, such as a pandemic. Since the coronavirus outbreak in March 2020 in the United States, the country has dealt with unprecedented circumstances in the medical field as medical staff continue to serve patient's routine and emergent health care needs. The use of and access to health care data through ePHRs has allowed patients to stay in touch with their providers while allowing protective social distancing measures to remain in place, especially for older adults who are at higher risk. Properly educating physicians and consumers on the benefits of ePHRs and how to use ePHRs to access data at any time will increase communication between the physicians and consumers. This will also aid consumers in adjusting to changes in health care delivery as it allows them to continue to feel connected to their health care provider during such a critical time in health care. During a pandemic, voluntary participation in data sharing via ePHRs would allow health authorities access to critical data on medical diagnoses that indicate who is at an elevated risk for additional negative impacts from COVID-19 [39]. This access could allow for valuable protective measures to be extended for at-risk populations and keep health authorities apprised of the success or failure of proactive measures to protect these higher-risk groups.

Authors' Contributions

YL contributed to the study conception and design, data analysis and interpretation, and manuscript preparation. KD contributed to manuscript preparation and editing. CI contributed to manuscript review and editing.

Conflicts of Interest

None declared.

References

1. Connecting Americans to their Health care: Final Report in Working Group on Policies for Electronic Information Sharing between Doctors and Patients. Markle Foundation. 2004. URL: https://www.markle.org/sites/default/files/CnctAmerHC_fullreport.pdf [accessed 2021-09-22]

2. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. Journal of the American Medical Informatics Association 2006 Mar 01;13(2):121-126. [doi: 10.1197/jamia.m2025]
3. Pagliari C, Detmer D, Singleton P. Potential of electronic personal health records. BMJ 2007 Aug 16;335(7615):330-333. [doi: 10.1136/bmj.39279.482963.ad]
4. Lyles CR, Nelson EC, Frampton S, Dykes PC, Cemballi AG, Sarkar U. Using electronic health record portals to improve patient engagement: research priorities and best practices. Annals of Internal Medicine 2020 Jun 02;172(11 Supplement):S123-S129. [doi: 10.7326/m19-0876]
5. Bose-Brill S, Feeney M, Prater L, Miles L, Corbett A, Koesters S. Validation of a novel electronic health record patient portal advance care planning delivery system. J Med Internet Res 2018 Jun 26;20(6):e208 [FREE Full text] [doi: 10.2196/jmir.9203] [Medline: 29945860]
6. Krist AH, Woolf SH, Hochheimer C, Sabo RT, Kashiri P, Jones RM, et al. Harnessing information technology to inform patients facing routine decisions: cancer screening as a test case. Ann Fam Med 2017 May;15(3):217-224 [FREE Full text] [doi: 10.1370/afm.2063] [Medline: 28483886]
7. Pecina J, North F, Williams MD, Angstman KB. Use of an on-line patient portal in a depression collaborative care management program. J Affect Disord 2017 Jan 15;208:1-5. [doi: 10.1016/j.jad.2016.08.034] [Medline: 27736688]
8. Health information technology: HH should assess the effectiveness of its efforts to enhance patient access to and use of electronic health information. United States Government Accountability Office. 2017. URL: https://www.gao.gov/assets/gao-17-305.pdf [accessed 2021-09-22]
9. Ford EW, Hesse BW, Huerta TR. Personal health record use in the United States: forecasting future adoption levels. J Med Internet Res 2016 Mar 30;18(3):e73 [FREE Full text] [doi: 10.2196/jmir.4973] [Medline: 27030105]
10. Hong Y, Jiang S, Liu P. Use of patient portals of electronic health records remains low from 2014 to 2018: results from a national survey and policy implications. Am J Health Promot 2020 Jul;34(6):677-680. [doi: 10.1177/0890117119900591] [Medline: 32030989]
11. Greenberg AJ, Falisi AL, Finney Rutten LJ, Chou WS, Patel V, Moser RP, et al. Access to electronic personal health records among patients with multiple chronic conditions: a secondary data analysis. J Med Internet Res 2017 Jun 02;19(6):e188 [FREE Full text] [doi: 10.2196/jmir.7417] [Medline: 28576755]
12. Keith McInnes D, Shimada SL, Rao SR, Quill A, Duggal M, Gifford AL, et al. Personal health record use and its association with antiretroviral adherence: survey and medical record data from 1871 US veterans infected with HIV. AIDS Behav 2013 Nov;17(9):3091-3100. [doi: 10.1007/s10461-012-0399-3] [Medline: 23334359]
13. Fischer SH, David D, Crotty BH, Dierks M, Safran C. Acceptance and use of health information technology by community-dwelling elders. Int J Med Inform 2014 Sep;83(9):624-635 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.06.005] [Medline: 24996581]
14. Xie B. Older adults, computers, and the Internet: Future directions. Gerontechnology 2003;2(4):289-305. [doi: 10.4017/GT.2003.02.04.002.00] [Medline: 110496150]
15. Kim E, Stolyar A, Lober WB, Herbaugh AL, Shinstrom SE, Zierler BK, et al. Challenges to using an electronic personal health record by a low-income elderly population. 2006 Presented at: 2006 AMIA Annual Symposium; Nov 11-15, 2006; Washington DC, USA.
16. Czaja S, Sharit J, Charness N, Fisk A, Rogers W. The Center for Research and Education on Aging and Technology Enhancement (CREATE): a program to enhance technology for older adults. Gerontechnology 2001;1(1):50-59. [doi: 10.4017/gt.2001.01.01.005.00]
17. Logue MD, Effken JA. An exploratory study of the personal health records adoption model in the older adult with chronic illness. Inform Prim Care 2012;20(3):151-169 [FREE Full text] [doi: 10.14236/jhi.v20i3.21] [Medline: 23710840]
18. Price MM, Pak R, Müller H, Stronge A. Older adults' perceptions of usefulness of personal health records. Univ Access Inf Soc 2012 Mar 8;12(2):191-204. [doi: 10.1007/s10209-012-0275-y]
19. Abd-Alrazaq AA, Bewick BM, Farragher T, Gardner P. Factors that affect the use of electronic personal health records among patients: A systematic review. Int J Med Inform 2019 Jun;126:164-175. [doi: 10.1016/j.ijmedinf.2019.03.014] [Medline: 31029258]
20. Morton A. Examining acceptance of an integrated personal health record (PHR). In: ProQuest Dissertations Publishing. University of Maryland, Baltimore: ProQuest Dissertations Publishing; 2011:33-42.
21. Sherifi D. Perceived Usefulness and Perceived Ease of Use Impact on Patient Portal Use. In: ProQuest Dissertations Publishing. Minneapolis, Minnesota: Walden University; 2018:77-87.
22. Turner A, Osterhage K, Joe J, Hartzler A, Lin L, Demiris G. Use of patient portals: personal health information management in older adults. Stud Health Technol Inform 2015;216:978. [Medline: 26262280]
23. Dontje K, Corser WD, Holzman G. Understanding patient perceptions of the electronic personal health records. The Journal for Nurse Practitioners 2014 Nov;10(10):824-828. [doi: 10.1016/j.nurpra.2014.09.009]
24. Digital nation: expanding internet usage. National Telecommunications and Information Administration. 2011. URL: https://www.ntia.doc.gov/files/ntia/publications/ntia_internet_use_report_february_2011.pdf [accessed 2021-09-22]

25. ICT4 the Elderly. The Digital Opportunities Foundation. URL: https://www.digitale-chancen.de/content/sdcprojekte/index.cfm/action.show/key.116/secid.144/secid2.191 [accessed 2021-09-22]
26. Paul G, Stegbauer C. Is the digital divide between young and elderly people increasing? FM 2005 Oct 03;10(10). [doi: 10.5210/fm.v10i10.1286]
27. Tirado-Morueta R, Aguaded-Gómez JI, Hernando-Gómez. The socio-demographic divide in Internet usage moderated by digital literacy support. Technology in Society 2018 Nov;55:47-55. [doi: 10.1016/j.techsoc.2018.06.001]
28. Health Information National Trends Survey 5 (HINTS 5) Cycle 3 Methodology Report. HINTS. 2021. URL: https://hints.cancer.gov/docs/Instruments/HINTS5_Cycle3_MethodologyReport.pdf [accessed 2021-08-01]
29. Abd-Alrazaq A, Bewick BM, Farragher T, Gardner P. Factors affecting patients' use of electronic personal health records in England: cross-sectional study. J Med Internet Res 2019 Jul 31;21(7):e12373 [FREE Full text] [doi: 10.2196/12373] [Medline: 31368442]
30. Abd-Alrazaq A, Alalwan AA, McMillan B, Bewick BM, Househ M, Al-Zyadat AT. Patients' adoption of electronic personal health records in england: secondary data analysis. J Med Internet Res 2020 Oct 07;22(10):e17499 [FREE Full text] [doi: 10.2196/17499] [Medline: 33026353]
31. Nazi K, Hogan T, McInnes D, Woods S, Graham G. Evaluating patient access to electronic health records: results from a survey of veterans. Medical Care 2013;51:S52-S56. [doi: 10.1097/mlr.0b013e31827808db]
32. Woods SS, Schwartz E, Tuepker A, Press NA, Nazi KM, Turvey CL, et al. Patient experiences with full electronic access to health records and clinical notes through the My HealtheVet Personal Health Record Pilot: qualitative study. J Med Internet Res 2013 Mar 27;15(3):e65 [FREE Full text] [doi: 10.2196/jmir.2356] [Medline: 23535584]
33. Bell K, Warnick E, Nicholson K, Ulcoq S, Kim SJ, Schroeder GD, et al. Patient adoption and utilization of a web-based and mobile-based portal for collecting outcomes after elective orthopedic surgery. Am J Med Qual 2018;33(6):649-656. [doi: 10.1177/1062860618765083] [Medline: 29562769]
34. Graetz I, Huang J, Brand RJ, Hsu J, Yamin CK, Reed ME. Bridging the digital divide: mobile access to personal health records among patients with diabetes. Am J Manag Care 2018 Jan;24(1):43-48 [FREE Full text] [Medline: 29350505]
35. Jung SY, Lee K, Hwang H, Yoo S, Baek HY, Kim J. Support for sustainable use of personal health records: understanding the needs of users as a first step towards patient-driven mobile health. JMIR Mhealth Uhealth 2017 Mar 23;5(2):e19 [FREE Full text] [doi: 10.2196/mhealth.6021] [Medline: 28232300]
36. Kutner M, Greenburg E, Jin Y, Paulsen C. The health literacy of America's adults: results from the 2003 National Assessment of Adult Literacy. National Center for Education Statistics. 2006. URL: https://files.eric.ed.gov/fulltext/ED493284.pdf [accessed 2021-09-22]
37. Peacock S, Reddy A, Leveille S, Walker J, Payne T, Oster N, et al. Patient portals and personal health information online: perception, access, and use by US adults. J Am Med Inform Assoc 2017 Apr 01;24(e1):e173-e177 [FREE Full text] [doi: 10.1093/jamia/ocw095] [Medline: 27413120]
38. Nelson D, Lang-Lindsey K. Rural health care and telehealth: the importance of social work departments at HBCUs in developing a competent workforce in the rural South. Journal of Community Engagement and Scholarship 2020;12(3):39-46 [FREE Full text]
39. Stoeger K, Schmidhuber M. The use of data from electronic health records in times of a pandemic-a legal and ethical assessment. J Law Biosci 2020;7(1):lsaa041 [FREE Full text] [doi: 10.1093/jlb/lsaa041] [Medline: 32793374]

Abbreviations

ePHR: electronic personal health record
HINTS: Health Information National Trends Survey
PHR: personal health record

Edited by R Kukafka; submitted 15.02.21; peer-reviewed by P Zou, A Naser, A Kanaan, R Menhas; comments to author 01.03.21; revised version received 23.03.21; accepted 06.05.21; published 26.10.21.

Please cite as:
Luo Y, Dozier K, Ikenberg C
Human-Technology Interaction Factors Associated With the Use of Electronic Personal Health Records Among Younger and Older Adults: Secondary Data Analysis
J Med Internet Res 2021;23(10):e27966
URL: https://www.jmir.org/2021/10/e27966
doi:10.2196/27966
PMID:34698646

Original Paper

Implementation Strategies to Improve Engagement With a Multi-Institutional Patient Portal: Multimethod Study

Jamie Keiko Fujioka[1], MPH; Julia Bickford[2], PhD; Jennifer Gritke[3], MA; Vess Stamenova[1], PhD; Trevor Jamieson[1,4,5], MBI, MD; R Sacha Bhatia[1,5,6], MBA, MD; Laura Desveaux[1,6], PT, PhD

[1]Women's College Hospital Institute for Health Systems Solutions and Virtual Care, Toronto, ON, Canada

[2]Northern Health Authority, Prince George, BC, Canada

[3]London Health Sciences Centre, London, ON, Canada

[4]Unity Health Toronto, Toronto, ON, Canada

[5]Department of Medicine, University of Toronto, Toronto, ON, Canada

[6]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Laura Desveaux, PT, PhD
Women's College Hospital Institute for Health Systems Solutions and Virtual Care
76 Grenville St
Room 6319
Toronto, ON, M5S 1B2
Canada
Phone: 1 416 323 6400
Email: laura.desveaux@wchospital.ca

Abstract

Background: Comprehensive multi-institutional patient portals that provide patients with web-based access to their data from across the health system have been shown to improve the provision of patient-centered and integrated care. However, several factors hinder the implementation of these portals. Although barriers and facilitators to patient portal adoption are well documented, there is a dearth of evidence examining how to effectively implement multi-institutional patient portals that transcend traditional boundaries and disparate systems.

Objective: This study aims to explore how the implementation approach of a multi-institutional patient portal impacted the adoption and use of the technology and to identify the lessons learned to guide the implementation of similar patient portal models.

Methods: This multimethod study included an analysis of quantitative and qualitative data collected during an evaluation of the multi-institutional MyChart patient portal that was deployed in Southwestern Ontario, Canada. Descriptive statistics were performed to understand the use patterns during the first 15 months of implementation (between August 2018 and October 2019). In addition, 42 qualitative semistructured interviews were conducted with 18 administrative stakeholders, 16 patients, 7 health care providers, and 1 informal caregiver to understand how the implementation approach influenced user experiences and to identify strategies for improvement. Qualitative data were analyzed using an inductive thematic analysis approach.

Results: Between August 2018 and October 2019, 15,271 registration emails were sent, with 67.01% (10,233/15,271) registered for an account across 38 health care sites. The median number of patients registered per site was 19, with considerable variation (range 1-2114). Of the total number of sites, 55% (21/38) had ≤30 registered patients, whereas only 2 sites had over 1000 registered patients. Interview participants perceived that the patient experience of the portal would have been improved by enhancing the data comprehensiveness of the technology. They also attributed the lack of enrollment to the absence of a broad rollout and marketing strategy across sites. Participants emphasized that provider engagement, change management support, and senior leadership endorsement were central to fostering uptake. Finally, many stated that regional alignment and policy support should have been sought to streamline implementation efforts across participating sites.

Conclusions: Without proper management and planning, multi-institutional portals can suffer from minimal adoption. Data comprehensiveness is the foundational component of these portals and requires aligned policies and a key base of technology infrastructure across all participating sites. It is important to look beyond the category of the technology (ie, patient portal) and consider its functionality (eg, data aggregation, appointment scheduling, messaging) to ensure that it aligns with the underlying

strategic priorities of the deployment. It is also critical to establish a clear vision and ensure buy-ins from organizational leadership and health care providers to support a cultural shift that will enable a meaningful and widespread engagement.

(J Med Internet Res 2021;23(10):e28924) doi:10.2196/28924

KEYWORDS

patient portal; electronic health record; patient health record; digital health

Introduction

Background

Effective health care requires patient-centered and integrated health services coordinated across multidisciplinary teams and delivery settings [1,2].

Efficient communication and health information exchange between patients, providers, and caregivers across settings is currently lacking, jeopardizing patient safety and health system costs [3-5]. To address this gap, health systems are increasingly investing in patient portals that enable timely and comprehensive information access through a single channel by patients [6,7]. Broadly speaking, patient portals offer secure electronic access to clinical information collected by one or more health care institutions [8,9]. They are related to but distinct from personal health records, which store health data compiled, managed, and updated by patients [10].

Comprehensive patient portals that enable access to data from multiple health services (ie, internal and external laboratories, diagnostic images, and discharge reports) have been shown to reduce duplicate testing [11] and enhance communication [12,13]. The primary argument for patient portals is not that they intend to create value for money; rather, they foster patient empowerment and self-management and improve patients' engagement in their care [14]. The downstream anticipated impact of this includes improved health outcomes [12,13,15-18] and reduced health resource usage [19-22]. However, suboptimal implementation threatens the potential realization of benefits. For instance, approximately 90% of the US health care systems and providers offer patient portals; however, only 15% to 30% of patients report using these resources [23,24]. Australia has noted similar adoption rates, with only 22% of citizens registered on their national portal since its launch in 2012 [25]. Several factors hinder adoption and use, including privacy concerns [18], low motivation to enroll [26], design and usability problems [18,27,28], and issues related to health or technical literacy [28] and equitable access [29]. In addition, the implementation of a comprehensive multi-institutional portal requires system interoperability and interorganizational coordination, which is often difficult to achieve across fragmented health services [7,15].

Although barriers and facilitators to patient portal adoption are well documented [9,14,17,30], there is a dearth of evidence on how the implementation process interacts with individual and organizational contexts [13]. This makes it difficult to discern the specific processes or factors that underpin the success or failure of the implementation. In response to this knowledge gap, we conducted an evaluation of a multi-institutional patient portal deployment to understand the experiences of patients, health care providers, and administrators.

Objectives

The specific objectives are to explore how the implementation approach impacts the adoption and usage of a patient portal and to identify lessons learned when implementing multi-institutional patient portals that transcend traditional boundaries and disparate systems.

Methods

Study Design

This multimethod study included an analysis of routinely collected usage data that were extracted from the portal by the vendor to understand use patterns during the first 15 months of implementation (between August 2018 and October 2019). In addition, qualitative semistructured interviews were conducted with patients (and their informal caregivers when applicable), health care providers, and administrative stakeholders to understand how the implementation approach influenced patient and provider experiences and overall usage. Ethics approval was received from the Research Ethics Board of the Women's College Hospital (REB approval no. 2019-0035-E). The reporting of this study was guided by the Consolidated Criteria for Reporting Qualitative Research checklist [31].

Study Setting

Southwestern Ontario (SWO) is a region of Ontario, Canada, that includes rural, suburban, and urban populations. The approximately 3.6 million residents of SWO represent 30% of Ontario's population [32]. Over 45,000 health care providers in SWO securely access publicly stored patient information through a regional webportal (ClinicalConnect) [32]. ClinicalConnect is a provider-facing viewer that consolidates data from 72 acute care hospitals, 4 home and community care organizations, and 4 regional cancer care programs in SWO [32]. It enables access to 4 provincial data repositories that house diagnostic imaging reports, drug information, laboratory results, and acute care information.

In 2017, Hamilton Health Sciences (HHS), a hospital network consisting of 7 hospitals and a regional cancer center, received funding from Canada Health Infoway, a not-for-profit organization funded by the federal government, to deploy a multi-institutional patient portal for residents in the region [33]. This initiative represents one of the largest deployments of a patient-facing digital health access channel in Ontario. To develop the portal, HHS collaborated with Sunnybrook Health Sciences Centre, a Toronto hospital that adapted its hospital-developed patient portal (MyChart) for the SWO region. To enable access to information across multiple systems,

MyChart was integrated with ClinicalConnect rather than being directly tethered to individual institutional electronic patient record systems. Organizations in SWO contributing data to ClinicalConnect were required to sign agreements to enable their data to flow to MyChart for patient access.

The MyChart implementation was launched in August 2018, with a target of 65,000 registered users by December 2019. MyChart was rolled out in stages, as site agreements were finalized with the participating organizations. Patient enrollment was only available in person at each site at the time of this study but was later made available via the internet in March 2020 because of the COVID-19 pandemic. Participating sites were responsible for independently determining what data would be uploaded and with how much delay and for creating implementation processes, such as onboarding of the patient and provider to the portal. Participating sites also had the flexibility to control the availability of portal features (eg, direct messaging) for patients accessing care at their sites (a full list of functionalities is given in Textbox 1). As of January 20, 2020, there were 48 sites contributing data to the regional MyChart and 38 sites actively offering the portal to their patients out of a possible 57 sites that had signed data-sharing agreements in SWO.

Textbox 1. Key functions of the regional MyChart patient portal.

Regional MyChart patient portal functionality

- Access a subset of their clinical information from any location at any time.
- Record and manage certain personal health information electronically.
- Delegate viewing of their record to caregivers and providers who accept a patient's request to receive access through their own MyChart accounts.
- Send direct messages to authorize MyChart account delegates (ie, clinicians), although very few organizations leveraged this feature.

Participant Recruitment

Overview

Initial recruitment used purposive and snowball sampling of providers and administrative staff at 2 early adopter organizations—HHS and Huron Perth Healthcare Alliance. Patients were recruited through convenience sampling, and recruitment posters were posted in waiting rooms and hospital staff notified eligible patients of the study. We used a maximum variation sampling strategy with the aim of recruiting a representative patient and provider sample. In particular, age, sex, health care condition, and geographic location (urban or rural) were considered. Provider recruitment targeted different health care professions and diverse clinical areas. Interviews were also sought with administrative stakeholders representing a variety of organizational roles and responsibilities in relation to implementation. Owing to slower than anticipated adoption of the portal, recruitment was expanded beyond HHS and Huron Perth Healthcare Alliance to include all partner organizations in SWO. Interview participants were asked to refer colleagues or contacts that could provide relevant insights following the completion of their interview.

Data Collection

Sunnybrook Health Sciences Centre provided aggregate, deidentified use data between August 2018, when MyChart was first implemented, and October 2019, the most recent month at the time of analysis. Data elements included the number of registered users, total number of log-ins per user, number of users who granted delegate access, number of users who recorded personal health information independently, and the page views for the various sections. In addition, they provided a summary of the number of users who logged in more than once and more than five times during October 2019.

Semistructured qualitative interviews, which lasted 30 to 60 minutes, were conducted over telephone between April and December 2019 with administrative implementation stakeholders, patients, caregivers, and health care providers who were informed of the study objectives. The interview guides included questions on implementation strategy, barriers and enablers to patient portal use, impact on the Quadruple Aim framework (patient experience, provider experience, health care costs, and health outcomes) [34], and suggestions for improvement. Interviews were conducted by members of the research team (JB and JG) who had no relationship with eligible providers or patient or caregiver participants. Some of the administrative stakeholders were familiar with the interviewers given their roles in the SWO community. All interviews were audio recorded and transcribed verbatim by a third party. Verbal consent was obtained before the interviews, and participants were given a US $13 gift card as a token of appreciation.

Data Analysis

Aggregate quantitative data were analyzed using descriptive summary statistics (eg, calculating frequencies and averages) with Excel software (Microsoft Corporation) to understand use patterns, including which features of the portal were most frequently used and how often the portal was accessed.

Qualitative data were analyzed using an inductive thematic analysis approach, which aimed to identify recurring patterns across a data set [35]. Two researchers (JB and JG) developed a preliminary codebook by independently coding the first 3 interview transcripts. The 2 researchers then discussed the codes, clarified and resolved discrepancies, and created a project codebook. The codebook was then discussed with the research team to ensure alignment with the study objectives and applied to additional transcripts as interviews were completed. Throughout this process, JB and JG met periodically to iteratively refine the codebook to reflect new codes, merge related codes, and resolve discrepancies. Once coding was complete, codes were synthesized into preliminary themes that were mapped back to the study objectives. Refinements and specifications of thematic categories and subcategories and

relationships between themes were determined through in-depth discussion and negotiated consensus between members of the research team (JB, JG, JKF, and LD). NVivo 12 software (QSR International) was used to assist with coding and analysis.

Results

Quantitative Use Data Findings

Between August 2018 and October 2019, a total of 15,271 registration emails were sent, with 67.01% (10,233/15,271) of patients registered for an account across 38 sites. The median number of patients registered per site was 19, with considerable variation (range 1-2114). Of the total number of sites, 55% had less than or equal to 30 registered patients (21/38), whereas only 2 sites had over 1000 registered patients. Among the registered patients, 92.00% (9,414/10,233) logged in at least once during the 15-month evaluation period. At the time of this study (October 2019), only 23.69% (2,424/10,233) had logged in more than once during the preceding month, with less than 3.83% (392/10,233) accessing the portal more than five times. High-traffic information pages included radiology exams and laboratories, discharge summaries, pathology results, and medications (Table 1). Only 2.98% (305/10,233) of registered users appointed a delegate, and less than 1% (73/10,233) actively shared information with their providers. A small number of patients (140/10,233, 1.37%,) independently recorded data in their patient health records via the portal.

Table 1. Top information page views accessed between August 2018 and October 2019—views of the lists associated with each category.

Category	Page views (August 2018 to October 2019)
Radiology	46,268
Laboratory results	43,799
Discharge summaries	18,377
Pathology results	16,884
Medications	14,689
Allergies	13,486
Microbiology results	11,771
Record summary	11,311
Home and community care	8507
Blood bank tests	8223

Qualitative Data Findings

A total of 42 individual interviews were conducted with stakeholders from 17 organizations within the SWO region to understand how the implementation unfolded and the factors underlying the adoption rates. The majority of participants (11/42, 26%) were recruited from HHS, followed by the Heron Perth Health Alliance (7/42, 17%) and the London Health Sciences Centre (5/42, 12%). The interview participants included 18 administrative stakeholders involved in the MyChart implementation, 16 patients, 7 health care providers, and 1 informal caregiver (Table 2). Administrative stakeholders included individuals involved in clinical leadership, clerical roles, patient experience teams, and information technology (IT). Among the patients and caregiver interviewed, 76.5% (13/17) had used MyChart, 11.8% (2/17) had registered but had not accessed it because of log-in challenges, and 11.8% (2/17) did not have access because of challenges with registration. The average age of patient or caregiver participants was 54 years, and all indicated that they were managing a chronic illness.

Table 2. Characteristics of the interview participants (N=42).

Interview participants	Participants, n (%)
Administrative stakeholders, (n=42)	18 (43)
Patients or caregivers, (n=42)	17 (40)
Patients (n=17)	16 (94)
Caregivers (n=17)	1 (6)
Gender (n=17)	
Female	11 (65)
Male	6 (35)
Registered MyChart user (n=17)	
Yes—have registered log-in and used it	13 (76)
Yes—have registered but have not used it	2 (12)
No	2 (12)
Geographic area	
Urban (n=32)	8 (47)
Rural (n=17)	4 (23)
Suburban (n=17)	3 (18)
Small town (n=17)	2 (12)
Education (n=17)	
Bachelor's degree	77 (41)
College diploma or certificate	6 (35)
Some college	2 (12)
Postgraduate education	1 (6)
High school	1 (6)
Providers (n=42)	7 (17)
Health profession (n=7)	
Physicians	4 (57)
Nurse or nurse practitioner	2 (29)
Nonnursing allied health professional	1 (14)

Five key themes emerged. They described how the implementation process influenced adoption rates and lessons learned that can be leveraged to increase the uptake of similar patient portal models.

Optimize the Patient Experience by Prioritizing Data Comprehensiveness

All interview participants unanimously endorsed the importance of leveraging technology to enable access to a comprehensive patient record, which they believed would reduce health system fragmentation by streamlining access to information. Patients perceived the benefits of this to include timely, efficient, and remote access to their health information, which enabled them to feel more prepared to manage their health concerns. Generally, patients had positive perceptions of MyChart because of its user-friendliness and ability to provide them access to pertinent health information (eg, test results, radiology reports). In addition, they felt that their relationships with their providers could be improved because the patient portal fostered increased transparency, knowledge, and empowerment. However, several data gaps were identified that limited perceived value, such as details related to home and community care (eg, care plans) and diagnostic images (not just reports). Particular emphasis was placed on the value of accessing clinic notes:

> *What's missing are the clinical notes...So, you're sitting with the physician say prior to chemo, that he's reviewed A, B, C with you. You know, and sometimes, and especially if it's chemo, there is an unfortunate thing called chemo brain or chemo fog and you think you've asked questions, but you know, three days later you're like, did I ask that? So, to go back and double check that I indeed did ask that question [helps me make sure] I don't want to waste the physician's time [by following up].* [Participant 40, Patient]

Participants suggested that data comprehensiveness could be further improved by expanding the scope of the regional data

to include information currently unavailable in MyChart but available in other provincial repositories, such as certain drug and immunization records.

In addition, educational materials (eg, information on how to interpret laboratory results) were perceived to be valuable in helping patients when interpreting clinical information and would likely improve engagement with the portal. Administrative stakeholders emphasized that core elements that enable access to comprehensive information, such as interoperability and integration with electronic medical records, should be prioritized during implementation. This aligns with patient motivations for engaging with the portal and its overall objective as a service within the system. Although extra features (ie, secure messaging with providers and appointment reminders) that require additional workflows and provider engagement were desired by patients, administrative stakeholders perceived these should be considered based on local needs and implemented in a graduated manner to keep the implementation in scope and logistically feasible.

Enhance Adoption by Using a Broad Rollout Strategy Instead of Targeting Specific Departments

Increasing patient awareness of the portal and implementing a clear and simple registration process with the aid of hands-on support were identified as adoption facilitators across participant groups. Patients generally perceived the registration and onboarding process to be easy to follow because this was facilitated through the aid of registration clerks. The low uptake in this study was attributed to inadequate marketing and promotion. Many sites used a targeted rollout strategy in select clinical areas instead of an organization-wide approach, which failed to harness broader marketing strategies and ultimately led to patient confusion. This approach was also highlighted as a potential threat to achieving health equity:

> *I know that some hospitals choose just kind of departments or areas to kind of trial it out. I really don't recommend that approach. We thought about it and I just said you know what, it has to be, for any success, we're gonna have to do a big bang theory and do it at all points of registration. So, that would be my recommendation even for large sites. I know it seems daunting, but I think the success is far outweighs for the patient because I don't think it's fair when a patient goes to say chemo and they don't have access to their chart, but the next time they go to DI [diagnostic imaging], they're asked about MyChart.* [Participant 13, Administrative Stakeholder]

Most administrative and provider participants felt that a broad rollout strategy would have been more effective, with sites that implemented this strategy describing more effective advertising and communication efforts. In addition, participants across stakeholder groups suggested leveraging diverse patient registration approaches to improve uptake, such as a combination of onsite and web-based enrollment options.

Providers Should Be Engaged to Understand and Mitigate Concerns

Gaining clinical buy-ins and alleviating provider concerns around open access to health information was a common challenge acknowledged across participant groups. Providers were specifically resistant to enabling access to their clinical notes because of potential misinterpretation by patients and liability concerns. In contrast, patients expressed a desire to have open access to all their health information. A related tension arose between the patient's desire for real-time access to information and provider preferences for upload delays to allow time for review and patient communication (ie, when information contained a new diagnosis). Many providers anticipated an increased workload as they expected they would need to assist with patient onboarding, help patients interpret medical jargon and resolve patient misinterpretation of their clinical information. However, these concerns were not realized during the study period. To foster provider acceptance, administrative stakeholders suggested that the patient portal should be framed as a patient service with the clear objective of promoting patient-centered care instead of as a clinical tool. Additional enablers to provider endorsement of the technology included clear, upfront communication about its purpose and visible clinical champions who could allay concerns among peers. In the event that additional features were activated (ie, secure messaging), administrative and provider participants stated that additional efforts should be made to engage providers to consider whether and how to align them with existing workflows:

> *So when we're trying to get their [providers'] attention, we're trying to explain to them that this is a patient service, not a clinical project...it's something you're going to offer your patients...and you don't have to do anything else.* [Participant 1, Administrative Stakeholder]

Change Management Support and Senior Leadership Endorsement Is Central to Early Success

The implementation of MyChart required additional administrative and logistical activities to recruit and onboard multiple sites. Interviews with administrative stakeholders revealed that organization size influenced deployment, with smaller hospitals reporting greater success in rolling out the portal and encouraging adoption than larger organizations. Competing priorities and resource constraints introduced implementation challenges in larger hospitals, which were often attributed to the upfront time and resources required to develop an operational model (ie, developing privacy and security agreements, identifying appropriate age and criteria of consent, and ensuring organizations have the technical requirements to contribute). As such, adoption was more successful when dedicated and protected resources were available to support upfront change management and implementation requirements:

> *[T]his is actually transformational change in the way we approach the health care encounter. That it's not just about rolling out access to a portal. And therefore, you need to have dedicated resources. I'm going to be frank and candid. It's not the sort of work*

> *that can just be done off the side of people's desks and it's almost what – it feels like we weren't resourced properly to foster success.* [Participant 14, Provider]

Another challenge to MyChart deployment was the need to gather stakeholder buy-in from multiple organizations. Within organizations, administrative stakeholders linked slower implementation to a lack of strong and overt senior leadership support. Conversely, rapid portal uptake, engagement, and onboarding were attributed to visible senior leadership endorsement. This was because of the senior leaders' ability to encourage high organizational motivation and interest to both contribute to the portal and enroll patients:

> *The CEO was like, let's go. They were visibly – like they were popping by our booth a few times a day and I think that just – like the optics, it looks good, it shows that they're interested. They're approving of this initiative...every single executive came and signed up first thing and they were excited and they were telling staff about it and prompting staff to get registered.* [Participant 20, Administrative Stakeholder]

Regional Alignment and Policy Supports Are Required to Streamline Implementation Efforts

Administrative stakeholders discussed how reconciling differential IT infrastructure to achieve system interoperability was a major barrier to expanding the portal across organizations. Before implementation, HHS had to work with potential enrollment sites to standardize data (ie, test names and medical terminology) for data filters to work consistently. Administrative stakeholders recommended the application of province-wide data standards to resolve these issues.

In addition, implementation was impeded by the need for organizations to gather consensus on complex policy considerations, such as identifying age and capacity to consent and operationalizing proxy access to delegate users. Although organizations appreciated that they had the autonomy to make these decisions to adapt to local needs, participants recommended the development of provincial guidelines and best practice approaches to inform organizational policies:

> *It's a bigger conversation about how we partner with families and we don't have well established policies or limitations.* [Participant 24, Administrative Stakeholder]

To foster a cohesive and well-integrated digital health information ecosystem, participants highlighted that the implementation of multi-institutional portals should be considered in tandem with institutional or private third-party offerings. As many hospitals within Ontario have developed their own in-house, institutionally tethered portals, 1 participant stated that there needs to be a clearer provincial strategy for promoting patient access to their health data:

> *One of the criteria is to basically have shared data with patients and have them as active participants in their health and so MyChart is clearly part of that, but I think we're going to lose people if everyone has different systems and if they have ten different apps. There needs to be a coordinated push on behalf of the ministry or governing bodies to, not mandate, but strongly suggest less and less alternatives because ultimately...you know, I've been to all over Ontario. Ottawa, Toronto, Thunder Bay and I don't want to have to keep signing up for new systems wherever I am. So, I think we need an Ontario-based system.* [Participant 27, Administrative Stakeholder]

Discussion

Principal Findings

Multi-institutional patient portals, such as MyChart, that collect and house information across organizations within a given region can enable more effective patient management by streamlining communication, access to clinical information, and service coordination [22]. The downstream benefits of such portals include increased patient engagement [14] and a reduction in emergency care and hospital admissions [20,22]. Although randomized controlled trials assessing patient portals are lacking, some randomized controlled trials have shown that patient portals can reduce hospital readmissions [36], reduce office visit rates, and result in greater adherence to treatment in comparison with control groups [16]. Our findings suggest that data comprehensiveness, organization-wide deployment, provider engagement, and senior leadership endorsement are central to achieving these aims.

Despite the high perceived value of multi-institutional patient portals, MyChart was adopted suboptimally. The population in SWO is approximately 3.6 million [32], implying that less than 0.3% of individuals within this region registered to use the portal at the time of this analysis. In comparison, the adoption of patient portals in other jurisdictions is estimated to reach 5% to 10% of the targeted population per year, with uptake in larger scale (ie, national or regional) implementations lagging behind smaller, more targeted deployments [37]. The slow growth of larger scale patient portals is attributed to challenges in obtaining alignment and system interoperability across fragmented health services [7,37]. In the absence of a cohesive regional implementation strategy, participating organizations of MyChart had to independently establish processes for marketing patient enrollment and training, resulting in variations in adoption and possibly heterogeneous patient experience. Further, a lack of dedicated resources, change management support, senior leadership endorsement, and clinical buy-in impeded success in several organizations. Strategies to improve change management processes and facilitate senior leadership and clinician support include clearly articulating the value of the technology, building consensus on key decisions and operational processes built around a strategic vision, and investing in the required infrastructure and resources, such as interoperable systems and staff training, to foster success [38].

Our findings highlight strategies for the successful implementation of patient portals. First, the implementation strategy needs to align with the core objectives of the technology. In this case, the portal should focus on enabling access to comprehensive clinical information as an initial

priority, given that this is the core functionality of interest and would provide the most benefit to patients. Other features (ie, direct messaging and appointment booking) can be explored in consultation with end users once a plan for comprehensive data access is established and successfully operationalized. Second, implementation processes should mirror patient experience. A blanketed rollout approach across an entire organization, rather than limiting deployment to certain departments, was perceived to be more conducive to facilitating information access that transcends traditional health system boundaries. Organizations that implemented the portal uniformly across the organization were able to more effectively advertise and communicate to patients about the portal. Conversely, segmented implementation exacerbates fragmented access to information, limits patient and clinician awareness and shared understanding of the purpose of the technology. This can create a disconnect for patients as they interact with different services within and across organizations. It is important to note that introducing functionality across an organization does not necessarily imply a one-size-fits-all approach as some flexibility is required to adapt to specific population needs [38]. For instance, special considerations regarding sharing clinic notes, proxy consent, and age of consent should be made for pediatric, mental health, and geriatric patients because of concerns regarding their capacity and autonomy [39]. However, interviewees raised the important point that these issues were better governed by universal best practices rather than by individual organizational idiosyncrasies.

Patient portals and other digital technologies can create value for organizations and health systems but only if the surrounding sociocultural factors are considered [40]. Effective leadership and clinical endorsement of technology can reduce behavioral resistance to change [41]. Patient portals may also precipitate changing dynamics between patients and providers, which underpins the cultural shift toward patient-centered care. Similar to other studies [42,43], we found that tensions between paternalistic and patient-centered medicine need to be resolved to facilitate widespread portal use. In tandem, individuals must be equipped with the necessary time, resources, and tools to carry out activities required for adoption, such as onsite training, enrollment, and technical support [43].

At the organizational and system levels, identifying and developing the right infrastructure is an essential component of strategic planning [38]. Multi-institutional patient portals require standard policy and technical infrastructure to enable data sharing that is consistent across sites. This includes identifying guidelines and best practices to establish access policies (ie, proxy access, age of consent, and data delays). The lack of basic integration and interoperability across institutional boundaries impedes the data comprehensiveness required for portals to function effectively in line with their goals [7,15,42]. The exponential growth of digital technologies across health systems implies the need for a degree of interoperability in alignment with more integrated health care [40,42].

Future research should focus on effectively describing and evaluating the implementation strategies that surround multi-institutional patient portals (eg, the use of champions, patient and provider training, addressing beliefs, etc) to identify effective strategies for promoting uptake. In addition, further studies should focus on evaluating the effectiveness of patient portals against their stated aims, including increased patient engagement in care, improved knowledge, and improved patient experience.

Limitations

The generalizability of our findings is limited to the early stages of implementation. Our sample may have been biased toward early adopters of technology or participants with limited user engagement. Most patient and provider participants had limited interactions with the patient portal, and their perspectives may not reflect the experiences that emerge alongside a more mature patient portal model. Despite this, all participants believed in the value of multi-institutional patient portals that offer patients comprehensive access to their health information in contrast to institutionally tethered offerings that do not centralize health information from across the health system. It is important to note that participants with limited user engagement did not endorse the current operationalization in its entirety; rather, they described the features and functions of a future state patient portal that would provide value to patients. Considerable benefits would be gained from evaluating the factors associated with sustained engagement in such a model. Although most interviews were conducted with individuals who had experience with the portal, the level of knowledge of and exposure to the portal varied. Owing to slower-than-anticipated adoption, our recruitment approach did not seek to discern between high and low adopters, as there were few *high* adopters at the time of the study. Further evaluations of MyChart should examine if there are characteristic differences (eg, based on patient population, region, or institution) between high and low users. Perspectives from diverse and often hard-to-reach patients were not included (eg, newcomers to Canada, non-English speakers, individuals experiencing housing insecurity); they may experience barriers to accessing technology. Consequently, further work is needed to engage with these populations and determine the impact on access and patient engagement from an equity perspective. Despite this, our study provides useful strategies to inform implementation planning at organizational and system levels.

Conclusions

Although multi-institutional patient portals can enable efficient access to clinical information from across the health system, successful implementation can be affected by several factors. Without proper management and planning, portals can suffer from minimal adoption from patients and poor support from providers. Data comprehensiveness is the foundational component of patient portals and requires aligned policies and a key base of IT infrastructure across all participating sites. It is important to look beyond the category of the technology (ie, patient portal) and consider its functionality (eg, data aggregation, appointment scheduling, messaging) to ensure that it aligns with the underlying strategic priorities of the deployment. Finally, it is critical to establish a clear vision and ensure buy-in from organizational leadership and health care providers to support a culture shift that will enable meaningful and widespread engagement.

Acknowledgments

The authors would like to thank Sarina Cheng and Emily Grillo at Sunnybrook Health Sciences Centre for providing aggregate usage data on the regional MyChart patient portal to help inform the results of this manuscript. They would also like to thank Mark Berry and Elizabeth Murray at Hamilton Health Sciences for providing information and resources on the regional MyChart patient portal in Southwestern Ontario. This study was funded by the Ontario Ministry of Health.

Conflicts of Interest

None declared.

References

1. Serbanati LD, Ricci FL, Mercurio G, Vasilateanu A. Steps towards a digital health ecosystem. J Biomed Inform 2011 Aug;44(4):621-636 [FREE Full text] [doi: 10.1016/j.jbi.2011.02.011] [Medline: 21362497]
2. Chouvarda IG, Goulis DG, Lambrinoudaki I, Maglaveras N. Connected health and integrated care: Toward new models for chronic disease management. Maturitas 2015 Sep;82(1):22-27. [doi: 10.1016/j.maturitas.2015.03.015] [Medline: 25891502]
3. Bodenheimer T. Coordinating care--a perilous journey through the health care system. N Engl J Med 2008 Mar 06;358(10):1064-1071. [doi: 10.1056/NEJMhpr0706165] [Medline: 18322289]
4. Jones CD, Vu MB, O'Donnell CM, Anderson ME, Patel S, Wald HL, et al. A failure to communicate: A qualitative exploration of care coordination between hospitalists and primary care providers around patient hospitalizations. J Gen Intern Med 2015 Apr 15;30(4):417-424 [FREE Full text] [doi: 10.1007/s11606-014-3056-x] [Medline: 25316586]
5. Samal L, Dykes PC, Greenberg JO, Hasan O, Venkatesh AK, Volk LA, et al. Care coordination gaps due to lack of interoperability in the United States: a qualitative study and literature review. BMC Health Serv Res 2016 Apr 22;16(1):143 [FREE Full text] [doi: 10.1186/s12913-016-1373-y] [Medline: 27106509]
6. Otte-Trojel T, Rundall TG, de Bont A, van de Klundert J. Can relational coordination help inter-organizational networks overcome challenges to coordination in patient portals? Int J Healthc Manag 2016 Jan 10;10(2):75-83 [FREE Full text] [doi: 10.1080/20479700.2015.1101911]
7. Otte-Trojel T, de Bont A, Aspria M, Adams S, Rundall TG, van de Klundert J, et al. Developing patient portals in a fragmented healthcare system. Int J Med Inform 2015 Oct;84(10):835-846. [doi: 10.1016/j.ijmedinf.2015.07.001] [Medline: 26190049]
8. van Kuppenveld SI, van Os-Medendorp H, Tiemessen NA, van Delden JJ. Real-time access to electronic health record via a patient portal: is it harmful? A retrospective observational study. J Med Internet Res 2020 Feb 06;22(2):e13622 [FREE Full text] [doi: 10.2196/13622] [Medline: 32044753]
9. Zhao JY, Song B, Anand E, Schwartz D, Panesar M, Jackson GP, et al. Barriers, facilitators, and solutions to optimal patient portal and personal health record use: a systematic review of the literature. AMIA Annu Symp Proc 2017;2017:1913-1922 [FREE Full text] [Medline: 29854263]
10. Dameff C, Clay B, Longhurst CA. Personal health records: More promising in the smartphone era? J Am Med Assoc 2019 Jan 29;321(4):339-340. [doi: 10.1001/jama.2018.20434] [Medline: 30633300]
11. Wakefield BJ, Turvey C, Hogan T, Shimada S, Nazi K, Cao L, et al. Impact of patient portal use on duplicate laboratory tests in diabetes management. Telemed J E Health 2020 Oct 01;26(10):1211-1220. [doi: 10.1089/tmj.2019.0237] [Medline: 32045320]
12. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J. How outcomes are achieved through patient portals: a realist review. J Am Med Inform Assoc 2014 Jul;21(4):751-757 [FREE Full text] [doi: 10.1136/amiajnl-2013-002501] [Medline: 24503882]
13. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: Evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
14. Kooij L, Groen WG, van Harten WH. Barriers and facilitators affecting patient portal implementation from an organizational perspective: qualitative study. J Med Internet Res 2018 May 11;20(5):e183 [FREE Full text] [doi: 10.2196/jmir.8989] [Medline: 29752253]
15. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J. What do we know about developing patient portals? A systematic literature review. J Am Med Inform Assoc 2016 Apr;23(e1):162-168. [doi: 10.1093/jamia/ocv114] [Medline: 26335985]
16. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012 Nov 26;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]
17. Irizarry T, DeVito DA, Curran CR. Patient portals and patient engagement: A state of the science review. J Med Internet Res 2015 Jun;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]

18. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015 Feb 20;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
19. Zhong X, Liang M, Sanchez R, Yu M, Budd PR, Sprague JL, et al. On the effect of electronic patient portal on primary care utilization and appointment adherence. BMC Med Inform Decis Mak 2018 Oct 16;18(1):84 [FREE Full text] [doi: 10.1186/s12911-018-0669-8] [Medline: 30326876]
20. Nicolás I, Lê Cook B, Flores M, Rodriguez M, Hernández Rodríguez C, Sillero P, et al. The impact of a comprehensive electronic patient portal on the health service use: an interrupted time-series analysis. Eur J Public Health 2019 Jun 01;29(3):413-418. [doi: 10.1093/eurpub/cky257] [Medline: 30544169]
21. Kim D, Kagel JH, Tayal N, Bose-Brill S, Lai A. The effects of doctor-patient portal use on health care utilization rates and cost savings. SSRN J 2020:2775261. [doi: 10.2139/ssrn.2775261]
22. Reed ME, Huang J, Brand RJ, Neugebauer R, Graetz I, Hsu J, et al. Patients with complex chronic conditions: Health care use and clinical events associated with access to a patient portal. PLoS One 2019 Jun 19;14(6):e0217636 [FREE Full text] [doi: 10.1371/journal.pone.0217636] [Medline: 31216295]
23. Health information technology: HHS should assess the effectiveness of its efforts to enhance patient access to and use of electronic health information. U.S. Government Accountability Office. 2017 Mar 15. URL: https://www.gao.gov/products/gao-17-305 [accessed 2021-09-24]
24. Lyles CR, Nelson EC, Frampton S, Dykes PC, Cemballi AG, Sarkar U. Using electronic health record portals to improve patient engagement: research priorities and best practices. Ann Internal Med 2020 Jun 02;172(11_Supplement):123-129. [doi: 10.7326/m19-0876]
25. Bidargaddi N, van Kasteren Y, Musiat P, Kidd M. Developing a third-party analytics application using australia's national personal health records system: case study. JMIR Med Inform 2018 Apr 24;6(2):e28 [FREE Full text] [doi: 10.2196/medinform.7710] [Medline: 29691211]
26. Goel MS, Brown TL, Williams A, Cooper AJ, Hasnain-Wynia R, Baker DW. Patient reported barriers to enrolling in a patient portal. J Am Med Inform Assoc 2011 Dec;18 Suppl 1:8-12 [FREE Full text] [doi: 10.1136/amiajnl-2011-000473] [Medline: 22071530]
27. Greenhalgh T, Hinder S, Stramer K, Bratan T, Russell J. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. Br Med J 2010 Nov 16;341:c5814 [FREE Full text] [doi: 10.1136/bmj.c5814] [Medline: 21081595]
28. Baldwin JL, Singh H, Sittig DF, Giardina TD. Patient portals and health apps: Pitfalls, promises, and what one might learn from the other. Healthc (Amst) 2017 Sep;5(3):81-85 [FREE Full text] [doi: 10.1016/j.hjdsi.2016.08.004] [Medline: 27720139]
29. Antonio M, Petrovskaya O, Lau F. Is research on patient portals attuned to health equity? A scoping review. J Am Med Inform Assoc 2019 Aug 01;26(8-9):871-883 [FREE Full text] [doi: 10.1093/jamia/ocz054] [Medline: 31066893]
30. Mishuris RG, Stewart M, Fix GM, Marcello T, McInnes DK, Hogan TP, et al. Barriers to patient portal access among veterans receiving home-based primary care: a qualitative study. Health Expect 2015 Dec 12;18(6):2296-2305 [FREE Full text] [doi: 10.1111/hex.12199] [Medline: 24816246]
31. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
32. Digital Health Services. eHealth Ontario. URL: https://ehealthontario.on.ca/en/health-care-professionals/digital-health-services [accessed 2021-06-01]
33. HHS obtains funding for patient portal in SW Ontario. Canadian Healthcare Technology. 2017. URL: https://www.canhealth.com/2017/09/20/hhs-to-extend-clincalconnect-to-patients/ [accessed 2021-06-01]
34. Bodenheimer T, Sinsky C. From triple to quadruple aim: Care of the patient requires care of the provider. Ann Fam Med 2014 Nov;12(6):573-576 [FREE Full text] [doi: 10.1370/afm.1713] [Medline: 25384822]
35. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
36. Creber RM, Grossman L, Ryan B, Qian M, Polubriaginof F, Restaino S, et al. Engaging hospitalized patients with personalized health information: a randomized trial of an inpatient portal. J Am Med Inform Assoc 2019 Feb 01;26(2):115-123 [FREE Full text] [doi: 10.1093/jamia/ocy146] [Medline: 30534990]
37. Gheorghiu B, Hagens S. Use and maturity of electronic patient portals. Stud Health Technol Inform 2017;234:136-141. [doi: 10.3233/978-1-61499-742-9-136]
38. Cresswell KM, Bates DW, Sheikh A. Ten key considerations for the successful implementation and adoption of large-scale health information technology. J Am Med Inform Assoc 2013 Jun;20(e1):9-13 [FREE Full text] [doi: 10.1136/amiajnl-2013-001684] [Medline: 23599226]
39. Mehta S, Jamieson T, Ackery A. Helping clinicians and patients navigate electronic patient portals: Ethical and legal principles. Can Med Asso J 2019 Oct 07;191(40):1100-1104 [FREE Full text] [doi: 10.1503/cmaj.190413] [Medline: 31591096]

40. Devlin AM, McGee-Lennon M, O'Donnell CA, Bouamrane M, Agbakoba R, O'Connor S, Dallas Evaluation Team. Delivering digital health and well-being at scale: Lessons learned during the implementation of the dallas program in the United Kingdom. J Am Med Inform Assoc 2016 Jan;23(1):48-59 [FREE Full text] [doi: 10.1093/jamia/ocv097] [Medline: 26254480]
41. Lorenzi NM, Riley RT. Managing change: an overview. J Am Med Inform Assoc 2000;7(2):116-124 [FREE Full text] [doi: 10.1136/jamia.2000.0070116] [Medline: 10730594]
42. Prey JE, Polubriaginof F, Kuperman GJ, Tiase V, Collins SA, Vawdrey DK. International perspectives on sharing clinical data with patients. Int J Med Inform 2016 Feb;86:135-141. [doi: 10.1016/j.ijmedinf.2015.11.007] [Medline: 26631956]
43. Avdagovska M, Ballermann M, Olson K, Graham T, Menon D, Stafinski T. Patient portal implementation and uptake: qualitative comparative case study. J Med Internet Res 2020 Jul 27;22(7):e18973 [FREE Full text] [doi: 10.2196/18973] [Medline: 32716308]

Abbreviations

HHS: Hamilton Health Sciences
IT: information technology
SWO: Southwestern Ontario

Edited by R Kukafka; submitted 18.03.21; peer-reviewed by M Levine, P Elkin, C Slightam; comments to author 08.04.21; revised version received 27.07.21; accepted 12.08.21; published 28.10.21.

Please cite as:
Fujioka JK, Bickford J, Gritke J, Stamenova V, Jamieson T, Bhatia RS, Desveaux L
Implementation Strategies to Improve Engagement With a Multi-Institutional Patient Portal: Multimethod Study
J Med Internet Res 2021;23(10):e28924
URL: https://www.jmir.org/2021/10/e28924
doi:10.2196/28924
PMID:34709195

Original Paper

Factors Affecting the Extent of Patients' Electronic Medical Record Use: An Empirical Study Focusing on System and Patient Characteristics

Lavlin Agrawal[1], MS; Theophile Ndabu[1*], MS; Pavankumar Mulgund[1*], PhD; Raj Sharman[1*], PhD

The State University of New York at Buffalo, School of Management, Buffalo, NY, United States

[*]these authors contributed equally

Corresponding Author:
Lavlin Agrawal, MS
The State University of New York at Buffalo, School of Management
160 Jacobs Management Center
Buffalo, NY, 14260-4000
United States
Phone: 1 (716) 645 3204
Email: lavlinag@buffalo.edu

Abstract

Background: Patients' access to and use of electronic medical records (EMRs) places greater information in their hands, which helps them better comanage their health, leading to better clinical outcomes. Despite numerous benefits that promote health and well-being, patients' acceptance and use of EMRs remains low. We study the impact of predictors that affect the use of EMR by patients to understand better the underlying causal factors for the lower use of EMR.

Objective: This study aims to examine the critical system (eg, performance expectancy and effort expectancy) and patient characteristics (eg, health condition, issue involvement, preventive health behaviors, and caregiving status) that influence the extent of patients' EMR use.

Methods: We used secondary data collected by Health Information National Trends Survey 5 cycle 3 and performed survey data analysis using structural equation modeling technique to test our hypotheses. Structural equation modeling is a technique commonly used to measure and analyze the relationships of observed and latent variables. We also addressed common method bias to understand if there was any systematic effect on the observed correlation between the measures for the predictor and predicted variables.

Results: The statistically significant drivers of the extent of EMR use were performance expectancy (β=.253; *P*<.001), perceived behavior control (β=.236; *P*<.001), health knowledge (β=–.071; *P*=.007), caregiving status (β=.059; *P*=.013), issue involvement (β=.356; *P*<.001), chronic conditions (β=.071; *P*=.016), and preventive health behavior (β=.076; *P*=.005). The model accounted for 32.9% of the variance in the extent of EMR use.

Conclusions: The study found that health characteristics, such as chronic conditions and patient disposition (eg, preventive health behavior and issue involvement), directly affect the extent of EMR use. The study also revealed that issue involvement mediates the impact of preventive health behaviors and the presence of chronic conditions on the extent of patients' EMR use.

(J Med Internet Res 2021;23(10):e30637) doi:10.2196/30637

KEYWORDS

electronic medical record; patient safety; caregiver; chronic conditions; HINTS dataset; patient technology acceptance model

Introduction

Background

An electronic medical record (EMR), also called the online medical record system, is a kind of software that stores clinical information such as medication lists, laboratory results, physician observations, immunizations, allergies, and discharge information [1]. Due to the impetus provided by the Health Information Technology for Economic and Clinical Health Act, EMR usage by providers and hospital administrators surged significantly, leading to improved documentation, data availability, and streamlined order entry to decrease prescription errors [2].

Although physician adoption and use of EMRs have been widely investigated [3-10], patients' use of EMR warrants further research. Patient adoption and use of EMRs represent a different phenomenon contrasted with physician adoption and use of EMR. For example, patients are not subject to organizational pressures prevalent in physician adoption and use decisions. Other factors differentiating the two contexts derive from the fact that patients may not be familiar with the technology to access EMR, and their understanding of clinical terms may be limited.

It is crucial to increase patient use of EMR for various consequential reasons, specifically patient empowerment. According to the World Health Organization, patient empowerment is a process through which people gain greater control over decisions and actions affecting their health [11]. Patients' access to and use of medical records empowers them to take a more active role in managing their health [12] by placing more information in their hands, which can improve clinical outcomes. Further, a patient portal built on top of EMR offers several benefits to patients, including continuous monitoring of health information, improved interactions with providers, better patient engagement in health management, scheduling appointments, and messaging physicians [13].

This research investigates the factors that influence patients' adoption and use of EMRs using an extended version of the patient technology acceptance model (PTAM). Specifically, we focus on the effect of salient patient characteristics such as health conditions, issue involvement, preventive health behaviors, and caregiving status on the adoption and use of EMR systems because they remain understudied.

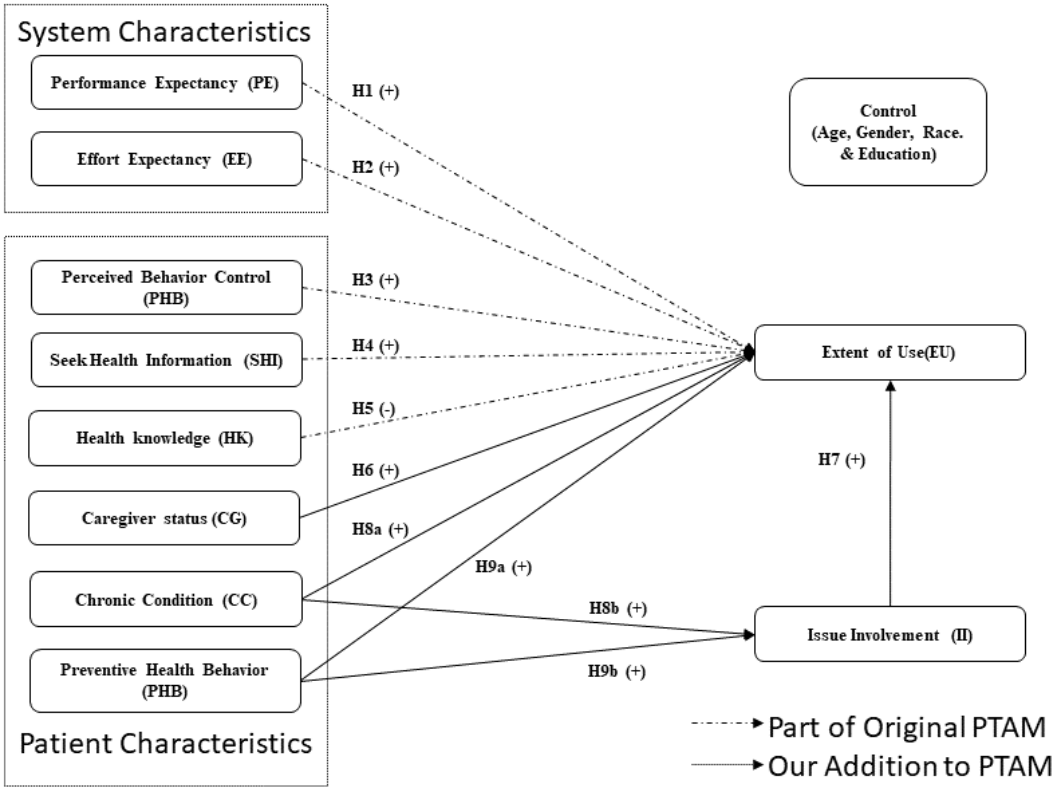
Hypotheses and Proposed Model

Overview

From a theoretical perspective, the unified theory of acceptance and use of technology has been employed to understand technology adoption and use in general [14]. It was adapted to the health care context with the addition of patient-centered factors (psychomotor, visual, and cognitive aspects) to study patients' adoption of technology [15] and was called PTAM. It considers perceived usefulness, perceived ease of use, perceived behavior control, subjective norm, and patient characteristics (psychomotor, visual, and cognitive aspects) as main predictors of the adoption and use of health information technologies by patients. PTAM was originally developed by Or et al [15] for the general context of patient adoption of technology. Since EMR is a specific technology for storing medical records, we had to adapt it to our context.

In this study, we not only considered factors such as performance expectancy (perceived usefulness), effort expectancy (perceived ease of use), perceived behavior control, but also extended the model with patient characteristics that are not part of the original PTAM, such as health condition, preventive health behavior, issue involvement, and patients' caregiving status. Following that, we introduced issue involvement as a mediator between the extent of EMR use and (1) chronic conditions and (2) preventive health behavior. Age, gender, education, and race were used as control variables in our study. Our proposed research model is illustrated in Figure 1. This adaptation contributes to the development of a theoretical foundation that could be used to improve our understanding of patients' EMR use.

Figure 1. Research model. PTAM: patient technology acceptance model



Performance Expectancy

Davis [16] defined perceived usefulness as one of the key predictors of new system adoption. Venkatesh et al [14] extended this notion of perceived usefulness by defining performance expectancy as the degree to which a person feels that using a system will help them perform a job more efficiently. In keeping with this understanding, we refer to performance expectancy as the degree to which the patient believes that using EMRs helps them monitor their health. Venkatesh et al [14] theorized that performance expectancy drives the intention to use information systems. Several researchers have also identified performance expectancy as one of the critical predictors of eHealth acceptance and use [15,17-21]. Because EMRs improve patient engagement and empower patients to access their health information anytime and anywhere [22], we propose the following hypothesis:

- H1: Performance expectancy is positively related to the extent of EMR use.

Effort Expectancy

Extending the ease of use construct [16], Venkatesh et al [14] defined effort expectancy as the degree of comfort associated with system use. Consistent with Venkatesh et al [14], we define effort expectancy as the degree of ease associated with understanding the health information in the online medical record. Venkatesh et al [14] suggested that effort expectancy has a positive effect on use intentions. Many researchers have also identified effort expectancy as one of the critical predictors of health adoption and use [15,19-21]. Studies have confirmed that ease of use is an essential predictor of intended use. Therefore, we propose the following hypothesis:

- H2: Effort expectancy is positively related to the extent of EMR use.

Perceived Behavioral Control

Or et al [15] defined perceived behavioral control as an individual's perception of their ability to do something (ie, self-efficacy). Many researchers have suggested that self-efficacy directly determines intent to use, especially online and mobile applications [23]. Lack of self-efficacy with computers and the internet is one of the most frequently identified barriers to adopting and using patient portals [24,25]. Turner et al [26] confirmed that the lack of comfort with computers is one of the common barriers to patient adoption of a portal. Thus, we posit that competency with technology is more likely to generate confidence in using EMRs. We hypothesize the following:

- H3: Perceived behavioral control is positively related to the extent of EMR use.

Seek Health Information

Seek health information (SHI) refers to individuals' urge to look for health-related information. Wilson and Lankton [27] argued that individuals seeking health information are more likely to adopt eHealth applications because such applications increase the availability of health information and reduce the effort needed to access that information. Or et al [15] extended the same concept and theorized that individuals who need to review health information are more likely to accept and use technology. As EMRs can help individuals get their health information and health history, and based on prior studies regarding health-information-seeking behavior, we propose the following hypothesis:

- H4: Seeking health information is positively related to the extent of EMR use.

Health Knowledge

Or et al [15] defined health knowledge as the knowledge that individuals feel they have about their health condition. Fowles et al [28] reported that sharing medical records with individuals has a modest positive impact on their knowledge. Wilson and Lankton [27] stated that an individual with little knowledge about their health is more likely to accept the eHealth tools managed by providers. Therefore, we hypothesize the following:

- H5: Health knowledge is negatively related to the extent of EMR use.

Caregiving

Caregiving implies providing paid or unpaid support and making medical decisions for a patient when appropriate [29,30]. In this study, we consider only an unpaid caregiver (generally family members or friends) who is currently caring for or making health care decisions for someone with a medical condition, behavioral or physical disability, or other condition.

King et al [31] provided evidence that caregivers use assistive health technologies (ie, any product, hardware, or software used to increase, maintain, or improve the functional capabilities of individuals with disabilities) to better care for children with a physical disability. Studies have also suggested that caregiver status strongly influences portal use, especially for caregivers who provide care for patients with chronic health conditions [32,33]. Caregivers' exposure to EMRs enhances their proficiency in using EMRs and makes them more likely to use EMRs themselves. Thus, we propose the following hypothesis:

- H6: Caregiving individuals are more likely to use EMRs extensively.

Issue Involvement

Issue involvement refers to how personally relevant people find an issue [34]. Abdelhamid et al [35] define issue involvement in the health care domain as "how relevant a specific health issue is to a patient." A more involved patient frequently visits providers, has several diseases, or has a severe health condition [36]. Consistent with Angst and Agarwal [36], we consider a patient with more physician visits (measured in our study as the number of physician visits in the past 12 months) as more involved with issues. Prior studies have demonstrated a positive relationship between issue involvement and the use of eHealth products [35,36]. Ross et al [37] argued that issue involvement has a significant positive impact on the adoption and use of EMRs. They found that EMRs better prepare patients for their upcoming visits with physicians by enhancing their knowledge of their medical condition, increasing their sense of control, and allowing them to seek clarification about treatment. Accordingly, we hypothesize the following:

- H7: Issue involvement is positively related to the extent of EMR use.

Chronic Conditions

Wagner et al [38] and Kruse et al [39] advocated for patients' use of health care systems and available resources to self-manage their health, especially for chronic health conditions. Studies have also suggested that patients with chronic conditions are more likely to use available eHealth applications to be better informed and manage their health [40-42]. A literature review [18] confirmed that patients with comorbidities are more likely to use electronic personal health record systems. Therefore, we posit that EMRs help patients track their improvement or deterioration in health and make informed decisions to better take care of themselves. Hence, we hypothesize that patients with existing chronic conditions are more likely to use EMRs.

Broemeling et al [43] demonstrated that a person with a chronic condition is more likely to visit a physician regularly. We, therefore, hypothesize that chronic conditions affect issue involvement (ie, frequency of physician visits). A higher number of chronic conditions may motivate patients to check their conditions, diagnostics, or prescriptions more closely, leading to greater EMR use. Those patients may also want to ensure that their records are correct to improve patient safety. Hence, we hypothesize that the extent of the chronic condition increases issue involvement and the need for frequent doctor visits.

- H8a: The presence of chronic conditions is positively related to the extent of EMR use.
- H8b: The presence of chronic conditions is positively related to issue involvement.

Preventive Health Behavior

Kasl et al [44] defined preventive health behavior as "any activity undertaken by a person who believes himself to be healthy for preventing disease or detecting disease in an asymptomatic stage." People with such drive are likely to monitor their health conditions through their EMRs. Studies have suggested that individuals use the available resources and skills to engage in preventive health behavior [35,45]. These resources may include accessing their records in EMR systems and seeking the help of physicians.

In psychology, motivation is described as a reason that drives action [46]. Thus, we posit that health motivators—in this case, preventive health behavior—influence people to engage in behaviors that improve their health outcomes and encourage frequent EMR use. This reasoning also finds resonance with earlier studies on the benefits of EMRs and the quality of health outcomes [47-49]. In this study, we consider a person to be

involved in preventive health behavior if they exercise and eat fruits and vegetables as recommended by the US Centers for Disease Control and Prevention (CDC). This understanding is similar to the operationalization by Hart et al [50].

Näslund [51] concluded that an individual engaging in preventive health behavior would have more doctor visits; this tendency is more pronounced in women. Grembowski et al [52] argued that individuals with preventive health behavior are more likely to initiate preventive care and early treatment. Other studies have suggested that health information technology plays a significant role in self-management [53,54]. Therefore, we hypothesize that individuals practicing preventive health behavior are more likely to visit their physicians often and use EMRs.

- H9a: Preventive health behavior is positively related to the extent of EMR use.
- H9b: Preventive health behavior is positively related to issue involvement.

Figure 1 and Table 1 summarize the hypothesis and definitions of the variables used in this model.

Table 1. Summary of hypothesis and variables.

Hypothesis	Variable	Defined in this study as	Relates
DV[a] : extent of EMR[b] use (EU)[c]			
H1	Performance expectancy (PE)	Degree to which the patient believes that using EMRs help them monitor their health	Positively
H2	Effort expectancy (EE)	Degree of ease associated with understanding the health information in the online medical record	Positively
H3	Perceived behavioral control (PBC)	Individual's perception of their ability to use electronic means	Positively
H4	Seek health information (SHI)	Whether an individual looked for information about health or medical topic from any source	Positively
H5	Health knowledge (HK)	If an individual is confident about ability to take good care of health	Negatively
H6	Caregiving status (CG)	If an individual is providing unpaid care to a patient	Positively
H7	Issue involvement (II)	Frequency of doctor visits in last 12 months	Positively
H8a	Chronic conditions (CC)	Number of chronic conditions an individual has	Positively
H9a	Preventive health behavior (PHB)	An individual is considered to have preventive health behavior if they exercise, eat fruits and vegetables as recommended by CDC[d]	Positively
DV: Issue involvement (II)[e]			
H8b	Chronic conditions (CC)	Number of chronic conditions an individual has	Positively
H9b	Preventive health behavior (PHB)	An individual is considered to have preventive health behavior if they exercise, eat fruits and vegetables as recommended by CDC	Positively

[a]DV: dependent variable.

[b]EMR: electronic medical record

[c]Number of times the online medical record has been accessed in the last 12 months.

[d]CDC: Centers for Disease Control and Prevention.

[e]Frequency of doctor visits in the last 12 months.

Methods

Data Source

We used data collected between January and May of 2019 by the National Cancer Institute (NCI) for Health Information National Trends Survey (HINTS) 5 cycle 3 to test our hypotheses. NCI administered a paper-based questionnaire and an online questionnaire to survey participants with an overall response rate of 30.3%. This survey was completed by 5438 participants. These data are publicly available and can be accessed at the HINTS website [55].

We filtered the data to include only those respondents who had used EMRs at least once during the previous 12 months. The resulting sample size was 2110. Data did not include outliers. Additionally, missing values on critical variables were less than 5%.

Measurements

The main dependent variable, the extent of EMR by the individual, was measured with a single item. System characteristics variables, performance expectancy and effort expectancy, were also measured with single items. Single items are acceptable if the question does not leave room for interpretation [56] and is used in information systems research that uses structural equation modeling (SEM) in the health care domain [35,36].

The patient characteristic, issue involvement, was measured with a single item. Other patient-related characteristics such as caregiving status, seek health information, health knowledge,

and perceived behavioral control were each measured with a binary choice question. There were 6 binary-choice questions for chronic conditions. The number of responses for chronic conditions was summed for analysis.

A formative measure of preventive health behavior was constructed using 3 items: the number of cups of fruit each day, the number of cups of vegetables per day, and the number of days per week with moderate exercise. According to the CDC [57], eating 1½ to 2 cups of fruit per day and 2 to 3 cups of vegetables per day is a healthy eating pattern. The CDC also recommends physical activity at least 2 days per week [58]. Based on these recommendations, we calculated the score for preventive health behavior as the sum of the responses to each item. Gender, age, race, income, and education were used as controls in the model. Please refer to Multimedia Appendix 1 for a detailed questionnaire, scale, and how they were used in this study.

Statistical Analysis

In this paper, we used SEM to conduct a path analysis. Although SEM is predominantly used to model latent variables, it is also applied to conduct path analysis in a mediation model, and in our study, we have 2 mediating relationships. First, issue involvement mediates the relationship between chronic conditions and the extent of EMR use. Second, issue involvement also mediates the relationship between preventive health behavior and the extent of EMR use. Therefore, we use SEM to test the model similar to prior scholars [59-61]. We used SEM with robust diagonally weighted least squares (DWLS) to test the hypotheses. DWLS is ideal for ordinal outcome variables [62-64]. We ran our model in R (version 4.0.2; R Core Team) using the “lavaan-survey” package.

Results

Descriptive Statistics

Table 2 shows the descriptive statistics of the survey respondents. The survey included questions about the extent of participants’ EMR use. Other questions focused on our model variables, including performance expectancy, effort expectancy, perceived behavioral control, seek health information, health knowledge, caregiving, chronic conditions, preventive health behavior, and issue involvement.

Table 2. Descriptive statistics.

Characteristics	Sample size, n (%)
Total responses	2110 (100)
Extent of EMR[a] use (EU)	
1 to 2 times	896 (42.46)
3 to 5 times	679 (32.18)
6 to 9 times	280 (13.27)
10 or more times	255 (12.09)
Performance expectancy (PE)	
Don't use	126 (5.97)
Not at all useful	26 (1.23)
Not very useful	145 (6.87)
Somewhat useful	831 (39.38)
Very useful	950 (45.02)
Effort expectancy (EE)	
Very difficult	22 (1.04)
Somewhat difficult	184 (8.72)
Somewhat easy	979 (46.4)
Very easy	883 (41.85)
Number of chronic conditions (CC)	
0	696 (32.99)
1	678 (32.13)
2	445 (21.09)
3	209 (9.91)
4	68 (3.22)
5	13 (0.62)
6	1 (0.05)
Issue involvement (II)	
None	86 (4.08)
1 time	225 (10.66)
2 times	390 (18.48)
3 times	336 (15.92)
4 times	354 (16.78)
5-9 times	438 (20.76)
10 or more times	272 (12.89)
Caregiver (CG)	
Yes	383 (18.15)
No	1682 (79.72)
Health Knowledge (HK)	
Not confident at all	15 (0.71)
A little confident	57 (2.7)
Somewhat confident	435 (20.62)
Very confident	1030 (48.82)
Completely confident	552 (26.16)

Characteristics	Sample size, n (%)
Perceived behavioral control (PHB)	
Yes	1701 (80.62)
No	385 (18.25)
Seek health information (SHI)	
Yes	1923 (91.14)
No	164 (7.77)
Preventive health behavior (PHB)	
0	515 (24.41)
1	729 (34.55)
2	537 (25.45)
3	329 (15.59)
Gender	
Male	815 (38.63)
Female	1259 (59.67)
Education	
High school or less	231 (10.95)
More than high school	1843 (87.35)
Race	
White	1596 (75.64)
Black	249 (11.8)
Others	148 (7.01)
Income, USD	
Less than $20,000	171 (8.1)
$20,000 to < $35,000	172 (8.15)
$35,000 to <$50,000	241 (11.42)
$50,000 to <$75,000	382 (18.1)
$75,000 or more	957 (45.36)
Age (years)	
Min	18
Max	97
Mean	54.21
SD	16.14

[a]EMR: electronic medical record.

Reliability and Validity

Table 3 shows the correlations between all the variables. Correlation coefficients are important as a high correlation among independent variables indicates a potential bias in coefficients due to multicollinearity. In this data set, the highest correlation is 0.41 between perceived expectancy and effort expectancy. None of the correlations were greater than 0.5, and they were within the acceptable threshold of 0.6 [65], so multicollinearity was not a concern in this analysis. Table 3 also provides the means and standard deviations for the principal variables.

Table 3. Correlation matrix.

	Mean (SD)	EU	EE	PE	PBC	SHI	HK	CG	II	CC	PHB
Extent of EMR use (EU)	1.97 (1.03)	1.00									
Effort Expectancy (EE)	3.33 (0.67)	0.12	1.00								
Performance Expectancy (PE)	4.22 (1.00)	0.26	0.41	1.00							
Perceived behavioral control (PBC)	0.83 (0.38)	0.24	0.10	0.23	1.00						
Seek Health Information (SHI)	0.93 (0.26)	0.07	–0.02	0.02	0.11	1.00					
Health Knowledge (HK)	3.99 (0.80)	–0.04	0.28	0.13	0.04	–0.01	1.00				
Caregiving Status (CG)	0.19 (0.39)	0.06	0.00	0.02	–0.02	–0.01	–0.01	1.00			
Issue Involvement (II)	3.44 (1.71)	0.37	–0.07	0.02	0.08	0.10	–0.14	–0.01	1.00		
Chronic Conditions (CC)	1.20 (1.13)	0.14	–0.08	–0.03	0.02	0.00	–0.27	–0.01	0.26	1.00	
Preventive Health Behavior (PHB)	1.33 (1.00)	0.06	0.08	0.08	0.05	0.05	0.21	0.02	–0.04	–0.17	1.00

Variance Inflation Factor

We used variance inflation factor (VIF) statistics to determine if data is suffering from multicollinearity. Multicollinearity refers to the linear relationship between 2 or more predictor variables [66]. VIF indicates the increase in the variance of a regression coefficient as a result of multicollinearity. Table 4 shows the VIF for each variable. The VIFs for all variables were well below 5.0, suggesting that the data did not suffer from multicollinearity [67].

Table 4. Variance inflation factors.

Variable	EE[a]	PE[b]	PBC[c]	SHI[d]	HK[e]	CG[f]	II[g]	CC[h]	PHB[i]
VIF[j]	1.29	1.26	1.08	1.03	1.20	1.00	1.10	1.16	1.07

[a]EE: effort expectancy.
[b]PE: performance expectancy.
[c]PBC: perceived behavioral control.
[d]SHI: seek health information.
[e]HK: health knowledge.
[f]CG: caregiving status.
[g]II: issue involvement.
[h]CC: chronic conditions.
[i]PHB: preventive health behavior.
[j]VIF: variance inflation factors.

Common Method Variance

Because the data were self-reported and collected through a single survey, the data may suffer from common method variance (CMV), which hampers the relationship between the variables [68]. Therefore, we assessed CMV bias using a marker variable technique [69]. A marker variable is a variable that is theoretically unrelated to one or more of the principal variables measured in the study and typically has a low correlation with the central variables.

Table 5 shows the correlation between the principal variables and marker variables. The theoretically unrelated construct “enjoy time in sun” (ETS) was used as a marker variable. The correlation between the marker variable ETS and other principal variables was low, meeting the threshold below 0.1 [69], except seek health information, which had a correlation of –0.11 with ETS. Similar findings were obtained using “morning-night person” as a marker variable (see Multimedia Appendix 1). The low correlation of the marker variable with the variables in the model indicates the absence of CMV.

Table 5. Correlation with marker variables.

	EU[a]	EE[b]	PE[c]	PBC[d]	SHI[e]	HK[f]	CG[g]	II[h]	CC[i]	PHB[j]
With marker variable "enjoy time in sun"										
Correlation	–0.04	0.03	0.05	–0.03	–0.11	0.03	0.01	-0.07	–0.06	0.01
P value	.07	.18	.02	.23	<.001	.20	.66	.002	.01	.77
With marker variable "morning-night person"										
Correlation	0.03	–0.04	–0.02	–0.01	0.01	–0.09	0.01	0.05	0.03	–0.09
P value	.27	.09	.33	.62	.57	<.001	.78	.03	.27	<.001

[a]EU: extent of EMR use.

[b]EE: effort expectancy.

[c]PE: performance expectancy.

[d]PBC: perceived behavioral control.

[e]SHI: seek health information.

[f]HK: health knowledge.

[g]CG: caregiving status.

[h]II: issue involvement.

[i]CC: chronic conditions.

[j]PHB: preventive health behavior.

Data Analysis

Overview

Since the NCI administered a paper-based questionnaire and an online questionnaire to survey participants, we regressed the dependent variable "extent of EMR use" on the mode of survey administration. We found that the relationship between the two was not significant, which means that the mode of survey administration did not affect the extent of EMR use.

Further, we ran our model in R using the "lavaan-survey" package. The overall fit statistics (χ^2=78.461; *P*<.001; comparative fit index=0.784, Tucker–Lewis index=0.982, root mean square error of approximation=0.056, root mean square residual=0.000, and goodness-of-fit statistic=0.935) of the structural model indicated a good model fit [70]. The SEM results are shown in Table 6. Table 7 presents the mediation analysis results for issue involvement with chronic conditions and preventive disease behavior.

Table 6. Results of structural equation modeling.

Variables	Standard estimates	CI lower	CI upper	*P* value	Significant
DV[a]: Issue involvement (II)					
Chronic conditions (CC)	0.237	0.163	0.261	<.001	Yes
Preventive health behavior	0.004	–0.047	0.055	.87	No
DV: Extent of EMR[b] use (EU)					
Performance expectancy (PE)	0.253	0.219	0.340	<.001	Yes
Effort expectancy (EE)	0.047	–0.009	0.170	.08	No
Perceived behavioral control (PBC)	0.236	0.544	0.868	<.001	Yes
Seek health information (SHI)	0.027	–0.099	0.341	.28	No
Health knowledge (HK)	–0.073	–0.176	–0.028	.01	Yes
Caregiving status (CG)	0.060	0.037	0.309	.01	Yes
Issue involvement (II)	0.353	0.343	0.430	<.001	Yes
Chronic conditions (CC)	0.071	0.013	0.127	.02	Yes
Preventive health behavior (PHB)	0.076	0.026	0.145	.01	Yes
Male	–0.091	–0.321	–0.097	<.001	Yes
Age	0.055	0.003	0.122	.04	Yes
High School or More	0.017	–0.134	0.264	.52	No
Black	0.032	–0.146	0.367	.40	No
White	0.056	–0.055	0.370	.15	No
Income	0.010	–0.037	0.054	.71	No

[a]DV: dependent variable.

[b]EMR: electronic medical record.

Table 7. Mediation results of structural equation modeling.

Mediation analysis	Standard estimates	CI lower	CI upper	*P* value	Significant
Chronic condition (CC)					
Direct	0.071	0.012	0.126	.02	Yes
Indirect through Issue Involvement	0.084	0.062	0.103	<.001	Yes
Total	0.155	0.099	0.205	<.001	Yes
Preventive health behavior (PHB)					
Direct	0.076	0.026	0.144	.005	Yes
Indirect through Issue Involvement	0.001	–0.018	0.021	.89	No
Total	0.077	0.031	0.142	.002	Yes

System Characteristics

H1 theorized a positive relationship between performance expectancy and the extent of EMR use. Our analysis revealed a positive and statistically significant path coefficient between performance expectancy and the extent of EMR use (βPE=.253; *P*<.001; see Table 6). This indicates that higher performance expectancy results in higher EMR use, thus supporting H1. In line with prior research [15,18-20], we theorized a positive relationship between effort expectancy and the extent of EMR use. The path coefficient was positive but statistically nonsignificant (βEE=.047; *P*=0.77; see Table 6). Therefore, additional research is warranted to examine the results further.

Patient Characteristics

H3 anticipated a positive relationship between perceived behavioral control and the extent of EMR use. The path coefficient was positive and statistically significant (βPBC=.24; *P*<.001; see Table 6), suggesting that patients' higher perceived behavioral control results in higher EMR use, supporting H3. This result is consistent with the findings of prior studies relating to perceived behavioral control in other domains.

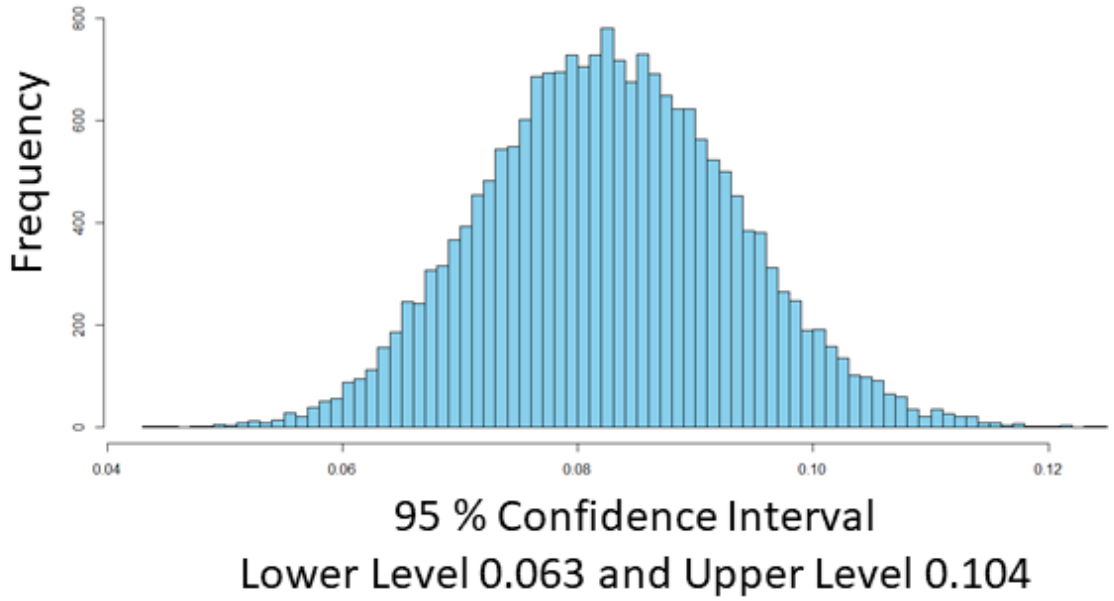
H4 theorized a positive relationship between SIH and the extent of EMR use. Although the path coefficient was positive, it was statistically nonsignificant (βSHI=.028; *P*=.28; see Table 6). Therefore, H4 was not supported. H5 theorized a negative

relationship between health knowledge and the extent of EMR use. The path coefficient was negative and statistically significant (βHK=–.071; *P*=.01; see Table 6). Therefore, H5 was supported. H6 theorized a positive relationship between caregiving status and the extent of EMR use. The path coefficient was positive and statistically significant (βCG=.06; *P*=.01; see Table 6), thus supporting H6.

H7 proposed a positive relationship between issue involvement and the extent of EMR use. The path coefficient was positive and statistically significant (βII=.356, *P*<.001; see Table 6), suggesting that higher issue involvement results in higher EMR use. Thus, H7 was supported. H8a argued that a positive relationship exists between the extent of chronic conditions and the extent of EMR use. The path coefficient was positive and significant (βCC=.071; *P*=.02; see Table 6), suggesting that a patient with more chronic conditions is more likely to use EMRs frequently. Thus, H8a was supported. H8b predicted a positive relationship between the extent of chronic conditions and issue involvement. The path from chronic conditions to issue involvement was positive and statistically significant (βCC=.238; *P*<.001; see Table 6). The mediation analysis suggests that issue involvement partially mediates the effect of chronic conditions on the extent of EMR use.

Monte Carlo simulation, also known as the Monte Carlo method or a multiple probability simulation, is a mathematical technique used to estimate the possible outcomes of an uncertain event [71]. We used Monte Carlo simulation to draw a probability distribution of the indirect effect of chronic conditions on the extent of EMR use. Figure 2 provides the probability distribution of the indirect effect of chronic conditions on the extent of EMR use. As the Monte Carlo CI [72] of the indirect effect did not contain zero (CI=0.063-0.104), the mediation of issue involvement between chronic conditions and the extent of EMR use is supported (Table 7). The mediation effect of issue involvement accounted for 48.9% of the impact of chronic conditions on the extent of EMR use.

Figure 2. Distribution of indirect effect of chronic conditions on the extent of EMR use. EMR: electronic medical record.



H9a argued that a positive relationship exists between preventive health behavior and the extent of EMR use. The path coefficient was positive and significant (βPHB=.076; *P*=.005; see Table 6), suggesting that a patient with preventive health behavior is more likely to use EMRs frequently. Thus, H9a was supported. H9b predicted a positive relationship between preventive health behavior and issue involvement. The path coefficient was positive but statistically nonsignificant (βPHB=.001; *P*=.89; see Table 6). Thus, H9b was not supported, which excludes the possibility of any mediation.

The study results suggest that, among the patient characteristics, issue involvement (βII=.356; *P*<.001) is the most important factor, followed by perceived behavior control (βPBC=.236; *P*<.001). Figure 3 shows the research model with the path coefficients and their significance, and Table 8 summarizes the study results.

Figure 3. Model results. PTAM: patient technology acceptance model

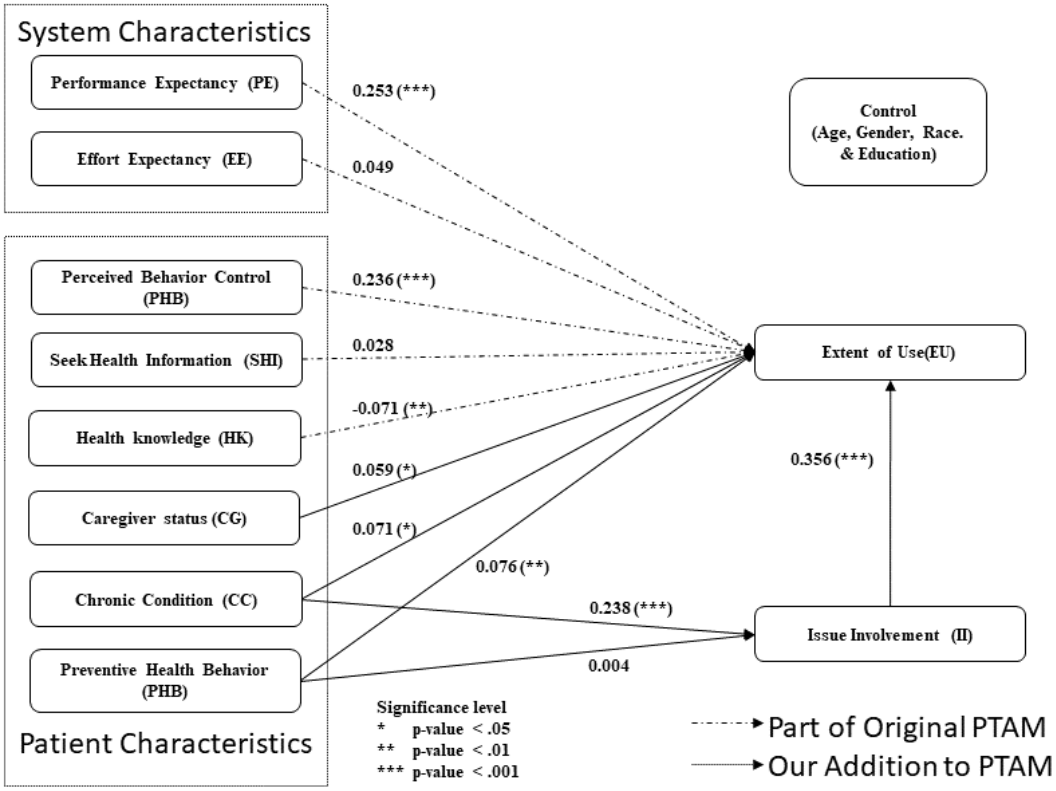


Table 8. Summary of study results.

Hypothesis	Independent variable	Result
DV[a] : Extent of EMR[b] use (EU)		
H1	Performance expectancy (PE)	Supported
H2	Effort expectancy (EE)	Not Supported
H3	Perceived behavioral control (PBC)	Supported
H4	Seek health information (SHI)	Not Supported
H5	Health knowledge (HK)	Supported
H6	Caregiving status (CG)	Supported
H7	Issue involvement (II)	Supported
H8a	Chronic conditions (CC)	Supported
H9a	Preventive health behavior (PHB)	Supported
DV: Issue Involvement (II)		
H8b	Chronic conditions (CC)	Supported
H9b	Preventive health behavior (PHB)	Not Supported

[a]DV: dependent variable.

[b]EMR: electronic medical record.

Discussion

Theoretical Implications

Our study extends the line of research on the PTAM [15] to explore patients' use of EMRs and examine the impact of factors that have not been studied by prior research. In line with prior studies [15,18,19], our statistical analysis showed that performance expectancy is a critical system characteristic that influences the patient adoption and use of EMRs. Further, this study finds perceived behavioral control [24-26] and caregiving status [32,33] significant variables for patient adoption and use of EMR. The number of chronic conditions and issue involvement also significantly impact the extent of EMR use by patients.

Our study also finds preventive health behavior to be a significant factor that impacts the extent of EMR use. However, seek health information is not statistically significant. A typical EMR captures the patient's medical history, including conditions, treatment decisions, medications, procedures, allergies, progress notes, and immunization records [73]. However, as noted by previous studies [15,27], it has not evolved into a source of medical information for patients who usually seek information from providers or the internet. Therefore, it is not surprising that seek health information emerged as an insignificant factor in determining the extent of EMR use. We also find effort expectancy nonsignificant, which is counterintuitive since several studies have found it critical in determining the extent of use [14].

Further, the study results suggest that issue involvement is the most critical patient characteristic, followed by perceived behavioral control as reflected by the model coefficients. Issue involvement refers to an individual's involvement with their health care issues and reflects their motivation to manage their health-related decisions. Therefore, it is natural that issue involvement emerged as a vital patient characteristic. Perceived behavioral control is another critical determinant of the extent of EMR use.

The study of the effect of chronic conditions and preventive health behaviors on the extent of EMR use is one of the most salient contributions of this article from a theoretical perspective. Unlike others, patients suffering from chronic conditions engage in continuous health monitoring, frequent interactions with medical providers, and an ongoing adjustment of medications. Such patients also require interactions with medical specialists, necessitating the frequent transfer of medical information among several physicians. Given the complexity of care and the patients' frequent interactions with providers, the statistical significance of chronic conditions as a determinant of EMR use is intuitive. Finally, the study highlights the value of caregiving in sensitizing and educating people about their health. Caregivers witness the challenges patients face and develop an empathetic understanding that increases their awareness and motivates them to adopt better health practices, including keeping track of patients' health information using EMRs.

In a nutshell, the contributions of this study include insights into how patients' characteristics and health conditions, along with their perceived system characteristics, influence the extent of EMR use. Our model adds (1) patient characteristics, such as caregiver status and preventive health practices, and (2) health conditions, such as chronic conditions and issue involvement, to the PTAM framework.

Practical Implications

Understanding the factors that influence the extent of EMR use by patients can be crucial in developing processes and systems that can enhance their adoption and usage. Given the significance of perceived behavioral control, we can institute inventions such as developing high-quality training modules and end-user support services. In addition to demonstrating the product features, training modules can also educate users on the potential value and utility of EMRs, thereby enhancing performance expectancy. The results of this study also suggest that practitioners and providers should dedicate efforts to educating and training patients about the benefits of EMR use. Also, we should promote success stories and best practices of patients using EMRs through case studies. Further, since chronically ill patients are more likely to use EMRs, patient engagement interventions should be directed at them. During the design and development phases, EMRs should also consider the role of caregivers.

Study Limitations

This study has several limitations. First, the HINTS data relied on self-reported information, so there is potential for CMV [68]. Using the marker variable technique [69], we evaluated that data are not suffering from CMV. Second, the study is based on secondary data and could only use variables present in the data. Certain key variables, such as social norms that may interest a general audience, were not included as these variables were not captured in the survey. Social norms, commonly defined as typical behaviors expected from people, are significant in original PTAM; consequently, the absence of social norms in this study might have inflated some of the estimates. However, since the patient adoption and use of EMRs is a relatively new phenomenon, the social norms around adopting and using EMRs are not well-established. Likely, its impact may not have been significant. Future studies should examine the impact of social factors and analyze their role in the extent of patients' EMR use. Third, the operationalization of chronic conditions was limited to only 6 major chronic conditions: diabetes, hypertension, heart disease, lung disease, depression, and cancer. To overcome these limitations, researchers should examine factors that affect patients' EMR use through longitudinal studies that include key variables such as social norms in addition to the variables in the current study.

Conclusions

Our study contributes to both theory and practice. First, we described how the phenomenon of patient adoption of EMRs is different from physician adoption of EMRs. Second, to understand the factors affecting patients' EMR use, we adapted the PTAM to the context of EMR use. This resulted in the addition of several new patient characteristics (eg, chronic conditions, preventive health behavior, issue involvement, and caregiving status) that influence the extent of EMR use. Thus,

our study contributes to the literature on health information systems. We also found that effort expectancy had no significant effect on the extent of patients' EMR use. We found that health characteristics, such as chronic conditions, preventive health behaviors, caregiving status, health knowledge, and issue involvement directly affect the extent of EMR use. Our analysis also revealed that issue involvement has a mediating effect on the impact of the extent of the chronic condition on EMR use.

EMR enables patients to track their health care history and understand the progress or deterioration in their health conditions. It also provides an opportunity for patients to examine their medical records and get the erroneous medical record corrected. Hence, improving EMR use contributes to patients' greater control over decisions and actions and adds to the larger goal of patient empowerment.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Operationalization of Constructs (source HINTS 5 cycle 3).
[DOCX File , 17 KB - jmir_v23i10e30637_app1.docx]

References

1. NCI dictionary of cancer terms. Cancer.gov. URL: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/electronic-medical-record [accessed 2021-05-15]
2. Gerber D, Laccetti A, Chen B, Yan J, Cai J, Gates S, et al. Predictors and intensity of online access to electronic medical records among patients with cancer. Journal of oncology practice 2014:e307-e312 [FREE Full text] [doi: 10.1200/jop.2013.001347]
3. Janssen A, Donnelly C, Elder E, Pathmanathan N, Shaw T. Electronic medical record implementation in tertiary care: factors influencing adoption of an electronic medical record in a cancer centre. BMC Health Services Research. URL: https://doi.org/10.1186/s12913-020-06015-6 [accessed 2021-05-15]
4. Rahal R, Mercer J, Kuziemsky C, Yaya S. Factors affecting the mature use of electronic medical records by primary care physicians: a systematic review. BMC Medical Informatics and Decision Making. URL: https://doi.org/10.1186/s12911-021-01434-9 [accessed 2021-05-15]
5. Tandon A. Practical AffordanceMR Use Within Outpatient Consulting on Women?s Health. 2020 Dec 03 Presented at: IFIP Joint Working Conference on the Future of Digital Work: The Challenge of Inequality 2020 Dec 10 (pp.). Springer, Cham; 12-10-2020; Hyderabad, India, p. 180-193 URL: https://doi.org/10.1007/978-3-030-64697-4_14 [doi: 10.1007/978-3-030-64697-4_14]
6. Adler-Milstein J, Holmgren A, Kralovec P, Worzala C, Searcy T, Patel V. Electronic health record adoption in US hospitals: the emergence of a digital advanced use divide. Journal of the American Medical Informatics Association 2017:1142-1148 [FREE Full text] [doi: 10.1093/jamia/ocx080]
7. Barrows RC, Clayton PD. Privacy, Confidentiality, and Electronic Medical Records. Journal of the American Medical Informatics Association 1996 Mar 01;3(2):139-148. [doi: 10.1136/jamia.1996.96236282]
8. Rantz M, Hicks L, Petroski G, Madsen R, Alexander G, Galambos C, et al. Cost, staffing and quality impact of bedside electronic medical record (EMR) in nursing homes. Journal of the American Medical Directors Association 2010:485-493 [FREE Full text] [doi: 10.1016/j.jamda.2009.11.010]
9. Sykes T, Venkatesh V, Rai A. Explaining physicians' use of EMR systems and performance in the shakedown phase. Journal of the American Medical Informatics Association 2011:125-130 [FREE Full text] [doi: 10.1136/jamia.2010.009316]
10. Yamamoto LG, Khan ANGA. Challenges of electronic medical record implementation in the emergency department. Pediatr Emerg Care 2006 Mar;22(3):184-91; quiz 192. [doi: 10.1097/01.pec.0000203821.02045.69] [Medline: 16628105]
11. Health promotion glossary. Health Promotion Glossary. 1998. URL: https://www.who.int/publications/i/item/WHO-HPR-HEP-98.1 [accessed 2021-05-15]
12. Ennis-O'Connor M. What is patient empowerment Internet. Powerfulpatients. URL: https://powerfulpatients.org/2018/05/22/what-does-it-mean-to-be-an-empowered-patient/ [accessed 2018-05-15]
13. Healthit.gov. What is a patient portal?. URL: https://www.healthit.gov/faq/what-patient-portal [accessed 2021-05-15]
14. Venkatesh V, Morris M, Davis G, Davis F. User acceptance of information technology: Toward a unified view. MIS quarterly 2003:425-478 [FREE Full text] [doi: 10.2307/30036540]
15. Or C, Karsh B, Severtson D, Burke L, Brown R, Brennan P. Factors affecting home care patients' acceptance of a web-based interactive self-management technology. J Am Med Inform Assoc 2011;18(1):51-59 [FREE Full text] [doi: 10.1136/jamia.2010.007336] [Medline: 21131605]
16. Davis F. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS quarterly 1989:319-340 [FREE Full text] [doi: 10.2307/249008]

17. Mukhopadhyay S, Basak R, Carpenter D, Reithel B. Patient use of online medical records: an application of technology acceptance framework. Information & Computer Security 2019 [FREE Full text] [doi: 10.1108/ics-07-2019-0076]
18. Niazkhani Z, Toni E, Cheshmekaboodi M, Georgiou A, Pirnejad H. Barriers to patient, provider, and caregiver adoption and use of electronic personal health records in chronic care: a systematic review. BMC Medical Informatics and Decision Making 2020:a [FREE Full text] [doi: 10.1186/s12911-020-01159-1]
19. Tavares J, Oliveira T. Electronic health record portal adoption: a cross country analysis. BMC medical informatics and decision making 2017 [FREE Full text] [doi: 10.1186/s12911-017-0482-9]
20. de Veer AJE, Peeters J, Brabers A, Schellevis F, Rademakers J, Francke A. Determinants of the intention to use e-Health by community dwelling older people. BMC Health Serv Res 2015 Mar 15;15:103-109 [FREE Full text] [doi: 10.1186/s12913-015-0765-8] [Medline: 25889884]
21. E-Health individual adoption-empirical model based on UTAUT 2(Doctoral dissertation). Goulão AP. URL: https://doi.org/http://hdl.handle.net/10362/13760 [accessed 2021-05-15]
22. Goel MS, Brown TL, Williams A, Cooper AJ, Hasnain-Wynia R, Baker DW. Patient reported barriers to enrolling in a patient portal. Journal of the American Medical Informatics Association 2011 Dec 01;18(Supplement 1):i8-i12 [FREE Full text] [doi: 10.1136/amiajnl-2011-000473]
23. Yu C. Factors affecting individuals to adopt mobile bankingmpirical evidence from the UTAUT model. Journal of electronic commerce research May 1 2012;13(2):E.
24. Lober WB, Zierler B, Herbaugh A, Shinstrom SE, Stolyar A, Kim EH, et al. Barriers to the use of a personal health record by an elderly population. AMIA Annu Symp Proc 2006:514-518 [FREE Full text] [Medline: 17238394]
25. Luque A, Van Keken A, Winters P, Keefer M, Sanders M, Fiscella K. Barriers and facilitators of online patient portals to personal health records among persons living with HIV: formative research. JMIR research protocols 2013:e8 [FREE Full text] [doi: 10.2196/resprot.2302]
26. Turner K, Clary A, Hong Y, Tabriz A, Shea C. Patient Portal Barriers and Group Differences: Cross-Sectional National Survey Study. Journal of medical Internet research 2020:e18870 [FREE Full text] [doi: 10.2196/18870]
27. Wilson E, Lankton N. Modeling patients' acceptance of provider-delivered e-health. Journal of the American Medical Informatics Association 2004:241-248 [FREE Full text] [doi: 10.1197/jamia.m1475]
28. Fowles JB, Kind AC, Craft C, Kind EA, Mandel JL, Adlis S. Patients' Interest in Reading Their Medical Record. Arch Intern Med 2004 Apr 12;164(7):793 [FREE Full text] [doi: 10.1001/archinte.164.7.793]
29. Pearlin L, Mullan J, Semple S, Skaff M. Caregiving and the stress process: An overview of concepts and their measures. The gerontologist 1990:A [FREE Full text] [doi: 10.1093/geront/30.5.583]
30. Hileman JW, Lackey NR, Hassanein RS. Identifying the needs of home caregivers of patients with cancer. Oncol Nurs Forum 1992 Jun;19(5):771-777. [Medline: 1608841]
31. King G, Maxwell J, Karmali A, Hagens S, Pinto M, Williams L, et al. Connecting families to their health record and care team: the use, utility, and impact of a client/family health portal at a children?s rehabilitation hospital. Journal of medical Internet research 2017:e97 [FREE Full text] [doi: 10.2196/jmir.6811]
32. Byczkowski T, Munafo J, Britto M. Variation in use of Internet-based patient portals by parents of children with chronic disease. Archives of pediatrics and adolescent medicine 2011:405-411 [FREE Full text] [doi: 10.1001/archpediatrics.2011.55]
33. Irizarry T, DeVito Dabbs A, Curran C. Patient Portals and Patient Engagement: A State of the Science Review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
34. de Graaf A, van den Putte B, de Bruijn GJ. Effects of Issue Involvement and Framing of a Responsible Drinking Message on Attitudes, Intentions, and Behavior. J Health Commun 2015 Aug;20(8):989-994 [FREE Full text] [doi: 10.1080/10810730.2015.1018623] [Medline: 26132603]
35. Abdelhamid M, Gaia J, Sanders G. Putting the focus back on the patient: how privacy concerns affect personal health information sharing intentions. Journal of medical Internet research 2017:e169 [FREE Full text] [doi: 10.2196/jmir.6877]
36. Angst C, Agarwal R. Adoption of electronic health records in the presence of privacy concerns: The elaboration likelihood model and individual persuasion. MIS quarterly 2009:339-370 [FREE Full text] [doi: 10.2307/20650295]
37. Ross S, Todd J, Moore L, Beaty B, Wittevrongel L, Lin C. Expectations of patients and physicians regarding patient-accessible medical records. J Med Internet Res 2005 May 24;7(2):e13 [FREE Full text] [doi: 10.2196/jmir.7.2.e13] [Medline: 15914460]
38. Wagner EH, Davis C, Schaefer J, Von Korff M, Austin B. A survey of leading chronic disease management programs: are they consistent with the literature? Manag Care Q 1999;7(3):56-66. [Medline: 10620960]
39. Kruse C, Argueta D, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. Journal of medical Internet research 2015:a [FREE Full text] [doi: 10.2196/jmir.3703]
40. Milani R, Bober R, Lavie C. The role of technology in chronic disease care. Progress in cardiovascular diseases 2016:579-583 [FREE Full text] [doi: 10.1016/j.pcad.2016.01.001]
41. Urowitz S, Wiljer D, Dupak K, Kuehner Z, Leonard K, Lovrics E, et al. Improving diabetes management with a patient portal: qualitative study of a diabetes self-management portal. Journal of medical Internet research 2012:e158 [FREE Full text] [doi: 10.2196/jmir.2265]
42. Osborn CY, Mayberry LS, Mulvaney SA, Hess R. Patient web portals to improve diabetes outcomes: a systematic review. Curr Diab Rep 2010 Dec;10(6):422-435 [FREE Full text] [doi: 10.1007/s11892-010-0151-1] [Medline: 20890688]

43. Broemeling A, Watson D, Prebtani F. Population patterns of chronic health conditions, co-morbidity and health care use in Canada: implications for policy and practice. Health care quarterly 2008:70-76 [FREE Full text] [doi: 10.12927/hcq.2008.19859]
44. Kasl S, Cobb S. Health behavior, illness behaviorsick role behavior: I. Health and illness behavior 1966:246-266 [FREE Full text] [doi: 10.1080/00039896.1966.10664365]
45. Jayanti R, Burns A. The antecedents of preventive health care behavior: An empirical study. Journal of the academy of marketing science 1998 [FREE Full text] [doi: 10.1177/0092070398261002]
46. Grant A. Does intrinsic motivation fuel the prosocial fire? Motivational synergy in predicting persistence, performance, and productivity. Journal of applied psychology 2008:48 [FREE Full text] [doi: 10.1037/0021-9010.93.1.48]
47. Lee J, Kuo Y, Goodwin J. The effect of electronic medical record adoption on outcomes in US hospitals. BMC health services research 2013:1-7 [FREE Full text] [doi: 10.1186/1472-6963-13-39]
48. Øvretveit J, Scott T, Rundall T, Shortell S, Brommels M. Improving quality through effective implementation of information technology in health care. International Journal for Quality in Health Care 2007:259-266 [FREE Full text] [doi: 10.1093/intqhc/mzm031]
49. Miller R, Sim I. Physicians' use of electronic medical records: barriers and solutions. Health affairs 2004:116-126 [FREE Full text] [doi: 10.1377/hlthaff.23.2.116]
50. Hart PD, Benavidez G, Erickson J. Meeting Recommended Levels of Physical Activity in Relation to Preventive Health Behavior and Health Status Among Adults. J Prev Med Public Health 2017 Jan;50(1):10-17 [FREE Full text] [doi: 10.3961/jpmph.16.080]
51. Näslund GK. Relationships between health behavior, knowledge, and beliefs among Swedish blue-collar workers. Scand J Soc Med 1997 Jun;25(2):100-110 [FREE Full text] [doi: 10.1177/140349489702500207] [Medline: 9232720]
52. Grembowski D, Patrick D, Diehr P, Durham M, Beresford S, Kay E, et al. Self-efficacy and health behavior among older adults. Journal of health and social behavior 1993:89-104 [FREE Full text] [doi: 10.2307/2137237]
53. Solomon M. Information technology to support self-management in chronic care. Disease Management and Health Outcomes 2008:391-401 [FREE Full text] [doi: 10.2165/0115677-200816060-00004]
54. El-Gayar O, Timsina P, Nawar N, Eid W. A systematic review of IT for diabetes self-management: are we there yet? International journal of medical informatics 2013:637-652 [FREE Full text] [doi: 10.1016/j.ijmedinf.2013.05.006]
55. What's new? National Cancer Institute. URL: https://hints.cancer.gov/ [accessed 2021-08-11]
56. Wanous J, Reichers A, Hudy M. Overall job satisfaction: how good are single-item measures? Journal of applied Psychology 1997:247 [FREE Full text] [doi: 10.1037/0021-9010.82.2.247]
57. Only 1 in 10 adults get enough fruits or vegetables. Cdc.gov. URL: https://www.cdc.gov/media/releases/2017/p1116-fruit-vegetable-consumption.html [accessed 2021-05-15]
58. CDC. Physical activity recommendations for different age groups. Cdc.gov. URL: https://www.cdc.gov/physicalactivity/basics/age-chart.html [accessed 2021-05-15]
59. MacKinnon D, Cheong J, Pirlott A. Statistical mediation analysis. In: American Psychological Association. Washington DC: American Psychological Association; 2012.
60. Hox J, Bechger T. An introduction to structural equation modeling. Family Science Review 1998;11:354-373.
61. Santibáñez-Andrade G, Castillo-Argüero S, Vega-Peña E, Lindig-Cisneros R, Zavala-Hurtado J. Structural equation modeling as a tool to develop conservation strategies using environmental indicators: The case of the forests of the Magdalena river basin in Mexico City. Ecological Indicators 2015:124-136 [FREE Full text] [doi: 10.1016/j.ecolind.2015.02.022]
62. Li C. The performance of ML. DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables 2016:369 [FREE Full text] [doi: 10.1037/met0000093]
63. Shi D, DiStefano C, McDaniel H, Jiang Z. Examining chi-square test statistics under conditions of large model size and ordinal data. Structural Equation Modeling: A Multidisciplinary Journal 2018:924-945 [FREE Full text] [doi: 10.1080/10705511.2018.1449653]
64. Shi D, Maydeu-Olivares A. The effect of estimation methods on SEM fit indices. Educational and Psychological Measurement 2019:421-445 [FREE Full text] [doi: 10.1177/0013164419885164]
65. Hair J, Anderson R, Tatham R, Black W. Multivariate data with readings. In: US America: Prentice Hall Inc. Hoboken, New Jersey: Prentice Hall; 1995.
66. Alin A. Multicollinearity. Wiley Interdisciplinary Reviews: Computational Statistics 2010:370-374 [FREE Full text] [doi: 10.1002/wics.84]
67. Ringle, Christian M. Wende, Sven, and Becker, Jan-Michael. SmartPLS 3. 2015. URL: http://www.smartpls.com [accessed 2021-05-15]
68. Podsakoff P, MacKenzie S, Lee J, Podsakoff N. Common method biases in behavioral research: a critical review of the literature and recommended remedies. Journal of applied psychology 2003:879-903 [FREE Full text] [doi: 10.1037/0021-9010.88.5.879]
69. Lindell M, Whitney D. Accounting for common method variance in cross-sectional research designs. Journal of applied psychology 2001:114 [FREE Full text] [doi: 10.1037/0021-9010.86.1.114]

70. Hu L, Bentler P. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural equation modeling: a multidisciplinary journal 1999:1-55 [FREE Full text] [doi: 10.1080/10705519909540118]
71. Harrison R. Introduction to monte carlo simulation. 2010 Presented at: AIP conference proceedings (Vol. 1204, No. 1, pp.). American Institute of Physics; 01-5-2010; Bratislava (Slovakia) p. 17-21. [doi: 10.1063/1.3295638]
72. Rich B, Lepine J, Crawford E. Job engagement: Antecedents and effects on job performance. Academy of management journal 2010 [FREE Full text] [doi: 10.5465/amj.2010.51468988]
73. Experityhealth. com. URL: https://www.experityhealth.com/resources/emr-vs-ehr-electronic-records/ [accessed 2021-05-15]

Abbreviations

CDC: Centers for Disease Control and Prevention
CMV: common method variance
DWLS: diagonally weighted least squares
EMR: electronic medical record
ETS: enjoy time in the sun
HINTS: Health Information National Trends Survey
NCI: National Cancer Institute
PTAM: patient technology acceptance model
SEM: structural equation modeling
VIF: variance inflation factor

Edited by R Kukafka; submitted 22.05.21; peer-reviewed by L Martinengo, N Harahap; comments to author 28.06.21; revised version received 13.08.21; accepted 12.09.21; published 28.10.21.

Please cite as:
Agrawal L, Ndabu T, Mulgund P, Sharman R
Factors Affecting the Extent of Patients' Electronic Medical Record Use: An Empirical Study Focusing on System and Patient Characteristics
J Med Internet Res 2021;23(10):e30637
URL: https://www.jmir.org/2021/10/e30637
doi:10.2196/30637
PMID:34709181

Original Paper

A Tool to Assess the Trustworthiness of Evidence-Based Point-of-Care Information for Health Care Professionals (CAPOCI): Design and Validation Study

Gerlinde Lenaerts[1], MSc, PhD; Geertruida E Bekkering[1,2], MSc, PhD; Martine Goossens[1], MSc; Leen De Coninck[1,3], MSc; Nicolas Delvaux[3], MD, PhD; Sam Cordyn[4], MSc; Jef Adriaenssens[5], MSc, PhD; Bert Aertgeerts[1,3], MD, PhD; Patrik Vankrunkelsven[1,2,3], MD, PhD

[1]Belgian Centre for Evidence-Based Medicine, Leuven, Belgium

[2]Cochrane Belgium, Leuven, Belgium

[3]Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium

[4]Federation of the White and Yellow Cross of Flanders, Brussels, Belgium

[5]Belgian Health Care Knowledge Centre, Brussels, Belgium

Corresponding Author:
Gerlinde Lenaerts, MSc, PhD
Belgian Centre for Evidence-Based Medicine
Kapucijnenvoer 7
Leuven, 3000
Belgium
Phone: 32 16 37 72 73
Fax: 32 16 33 74 80
Email: gerlinde.lenaerts@cebam.be

Abstract

Background: User-friendly information at the point of care for health care professionals should be well structured, rapidly accessible, comprehensive, and trustworthy. The reliability of information and the associated methodological process must be clear. There is no standard tool to evaluate the trustworthiness of such point-of-care (POC) information.

Objective: We aim to develop and validate a new tool for assessment of trustworthiness of evidence-based POC resources to enhance the quality of POC resources and facilitate evidence-based practice.

Methods: We designed the Critical Appraisal of Point-of-Care Information (CAPOCI) tool based on the criteria important for assessment of trustworthiness of POC information, reported in a previously published review. A group of health care professionals and methodologists (the authors of this paper) defined criteria for the CAPOCI tool in an iterative process of discussion and pilot testing until consensus was reached. In the next step, all criteria were subject to content validation with a Delphi study. We invited an international panel of 10 experts to rate their agreement with the relevance and wording of the criteria and to give feedback. Consensus was reached when 70% of the experts agreed. When no consensus was reached, we reformulated the criteria based on the experts' comments for a next round of the Delphi study. This process was repeated until consensus was reached for each criterion. In a last step, the interrater reliability of the CAPOCI tool was calculated with a 2-tailed Kendall tau correlation coefficient to quantify the agreement between 2 users who piloted the CAPOCI tool on 5 POC resources. Two scoring systems were tested: a 3-point ordinal scale and a 7-point Likert scale.

Results: After validation, the CAPOCI tool was designed with 11 criteria that focused on methodological quality and author-related information. The criteria assess authorship, literature search, use of preappraised evidence, critical appraisal of evidence, expert opinions, peer review, timeliness and updating, conflict of interest, and commercial support. Interrater agreement showed substantial agreement between 2 users for scoring with the 3-point ordinal scale (τ=.621, P<.01) and scoring with the 7-point Likert scale (τ=.677, P<.01).

Conclusions: The CAPOCI tool may support validation teams in the assessment of trustworthiness of POC resources. It may also provide guidance for producers of POC resources.

(J Med Internet Res 2021;23(10):e27174) doi:10.2196/27174

KEYWORDS

evidence-based medicine; evidence-based practice; point-of-care systems; health care quality; information science; practice guidelines as a topic

Introduction

Evidence-based medicine (EBM) aims to integrate the experience of the health care professional, the values of the patient, and the best available scientific information to guide clinical decision making. Back in 1996, Sackett defined EBM as "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients." The practice of EBM means integrating individual clinical expertise with the best available external clinical evidence from systematic research [1,2]. However, keeping up to date with the best available evidence is time-consuming, and the format of a systematic review is impractical for quickly answering clinical questions [3]. As more and more health professions worldwide adopted the EBM concept, the name shifted in the last decade to evidence-based practice (EBP). Nowadays, clinical guidelines are the gold standard in EBP to guide the clinical decision process, but for many health problems and health care professions, there are no guidelines available [4]. Technological progress allows rapid and easy access to an enormous amount of information, but the trustworthiness is often unclear. Point-of-care (POC) information is defined as high-quality information needed by health care professionals when they interact with the patient; this information should be well structured, quick and easily accessible and, most importantly, relevant and reliable [5,6]. Many authors have highlighted the importance of trustworthiness or quality of web-based POC information [7-12]. For systematic reviews and clinical guidelines, well-developed critical appraisal tools are available (eg, AMSTAR [A Measurement Tool to Assess Systematic Reviews] [13,14] and AGREE II [Appraisal of Guidelines for Research and Evaluation II] [15]). However, these tools are not appropriate for the evaluation of POC information.

Trustworthy POC information sources for health care professionals require a robust methodological process for searching, appraisal, and synthesis of the best available evidence in a reproducible way. Health information for professionals differs at this point from health information for the lay public, and consequently the initiatives to evaluate health information for patients (eg, e-Health Code of Ethics [16], Health on the Net [17], Journal of the American Medical Association [18], and DISCERN [19]) are not suitable for the evaluation of POC information for professionals. A standard tool to evaluate the trustworthiness of POC information that is not a guideline is therefore essential.

This study built on the results of a systematic review [20] that searched for existing tools to assess the trustworthiness of POC information. The content of existing tools was analyzed, and the tools were examined for validity and reliability. However, a tool that is complete, usable, and validated could not be found. Therefore, the aim of this study was to develop a new tool for assessment of trustworthiness of evidence-based POC resources to enhance the quality of POC resources and support health care professionals to have access to reliable EBP information. This paper describes the development process and associated validity of this new tool, which we named the Critical Appraisal of Point of Care Information (CAPOCI) tool.

Methods

Systematic Review

In a first step, a systematic review was performed to search for already existing tools to assess trustworthiness of POC resources. This systematic review has been published as a separate paper [20]. We aimed to describe and analyze the content of these tools by documenting the general characteristics, purpose for which a tool was developed, and criteria and scoring systems that were used. We also checked whether the included tools were examined for validity and reliability. Seventeen tools were included in the review. The tools encompassed a variety of criteria important for assessment of trustworthiness of POC information. Only two tools were assessed for both reliability and validity, but they lacked some essential criteria for assessment of trustworthiness of health care information for use at the point of care, pointing to the need to develop a new tool.

Formulation of Criteria for the CAPOCI Tool

In a second step, two methodologists (GB and GL) started from the results of the systematic review to derive relevant criteria for the CAPOCI tool. They related the relevance of criteria to frequency of occurrence in the existing tools and contribution of the criteria to the trustworthiness of POC information according to the reviewers. A working group of health care professionals and methodologists (ie, the authors of this paper) then discussed the relevance, applicability, and accurate wording of the different criteria. The further refinement of the items for the CAPOCI tool was the result of an iterative process of discussion and pilot testing until consensus was reached for all criteria, resulting in a tool with 9 criteria.

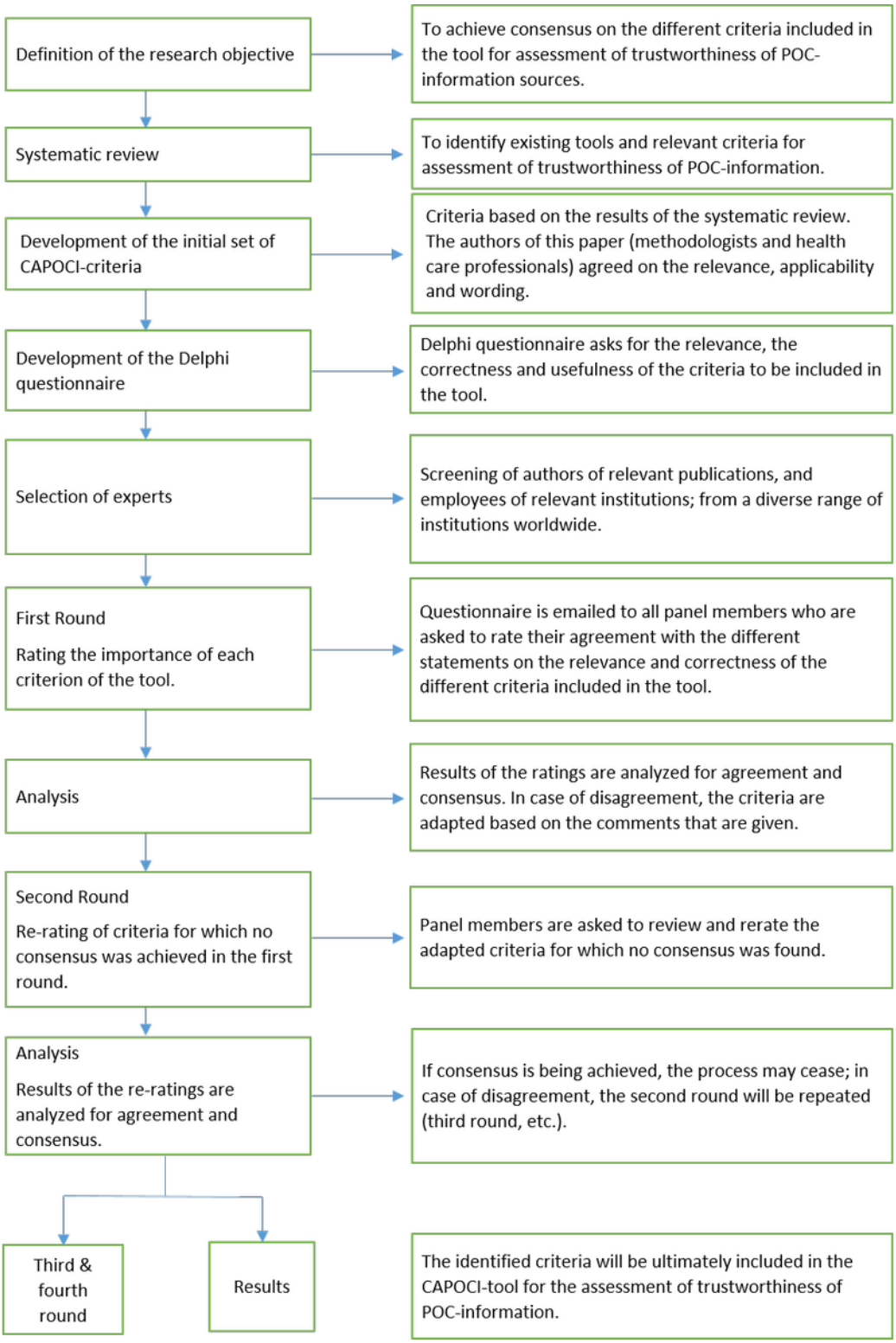
Content Validation

The content validity of the CAPOCI tool was tested on an international panel of experts (Multimedia Appendix 1) with the RAND modified Delphi method [21]. Panel members were selected based on their authorship of publications in the domain and their affiliation to relevant institutions. All panel members received a payment of €250 (US $306) for their time investment. The whole process was in writing and anonymous, and only the study coordinator (GL) knew the identity of the panel members.

All panel members received a study protocol and an electronic questionnaire. They were asked to rate the relevance and wording (accuracy and correctness) of the different CAPOCI criteria. The panel members rated their agreement with each item as strongly agree, agree, disagree, or strongly disagree. If the answer was disagree or strongly disagree, an explanation was required.

The questionnaires were returned to the study coordinator. Consensus on a criterion was achieved when 70% of the panelists rated a criterion as strongly agree or agree. Based on the justification formulated in case of disagreement, the criterion was reformulated and again presented to all panel members in a next round. The process was repeated until consensus was reached for all criteria. Figure 1 shows an overview of the development process of the CAPOCI tool.

Figure 1. Overview of the development process of the Critical Appraisal of Point of Care Information (CAPOCI) tool. POC: point-of-care.



Scoring Methods for the CAPOCI Tool

The results of the systematic review did not identify the most appropriate scoring system for the CAPOCI tool. Therefore, we decided to test 2 scoring systems. The first scoring system rated each criterion on a 3-point ordinal scale with the following possible ratings: (1) fulfilled, (2) minor remark, and (3) major remark. The conditions for awarding one of the 3 ratings were discussed and defined by the authors of this paper. The second

scoring system rated each criterion on a 7-point Likert scale, where a score of 1 indicated fully disagree and a score of 7 indicated fully agree.

Interrater Reliability And Statistical Analysis

The CAPOCI tool was tested for interrater reliability. To this end, two methodologists (GB and GL) assessed 5 POC resources used in Belgian health care with the CAPOCI tool, using the 2 scoring systems described above.

The statistical analyses were performed using SPSS (version 24, IBM Corp). A 2-tailed Kendall tau correlation coefficient was calculated to quantify the agreement between the 2 reviewers for the entire CAPOCI tool per scoring system. Level of significance was set at .01. In addition, a descriptive analysis was performed for each criterion separately. We counted how often a criterion was scored differently by the 2 reviewers for the 5 evaluated POC resources. For scoring with the Likert scale, a difference in score of more than 1 was considered a different score. The reason for interrater disagreement was documented.

Results

Delphi Study

Based on the results of the systematic review, the CAPOCI tool was initially defined with 9 criteria. These 9 criteria were then tested in the Delphi study. Ten experts from a diverse range of organizations and institutions reputable in EBM worldwide participated (see Multimedia Appendix 1). One expert dropped out after the first round.

The 9 criteria were assessed as relevant by all the experts after the first round, but refinement of the wording was needed. After 4 rounds of the Delphi study, a final version of the CAPOCI tool was developed including 11 criteria, where the criteria related to authorship and experience of authors, literature search and surveillance were split up. Table 1 summarizes the CAPOCI criteria as defined by the expert panel, together with the conditions for rating a criterion as fulfilled, minor remark, or major remark.

Table 1. The criteria of the CAPOCI tool and the conditions for scoring on a 3-point ordinal scale.

CAPOCI[a] criteria	Conditions to rate as fulfilled	Conditions to rate as minor remark	Conditions to rate as major remark
1. Authorship: The authors must be referenced on the website but do not need to be identified for each individual topic (clicking and searching may be necessary).	Name and affiliations of all authors are mentioned.	Only a general description is available (eg, of the editorial board).	There is no information available on the authors.
2. Expertise of the authors: The author team is qualified in the specific domain and can demonstrate their expertise on request.	The expertise of the author team is demonstrated.	The expertise of the author team is unclear.	There is no information available on the expertise of the author team.
3a. Literature search and surveillance: A systematic search strategy was used to search for source information.	A systematic search strategy has been used to search for source information. This search strategy is described in detail in the EBP[b] source.	The description is not sufficiently detailed to be able to assess; there are inaccuracies in the methodological process.	Literature search seems to be implemented, but there is no description of the process or there is no information on how the literature search was done.
3b. Literature search and surveillance: Systematic methods were used for selection of the evidence from the search.	Systematic methods have been used to select the evidence from the results of the literature search. These methods are described in detail.	The description is not sufficiently detailed to be able to assess; there are inaccuracies in the methodological process.	A systematic selection process seems implemented, but there is no description of the process or there is no information on how this selection was done.
4. Critical appraisal of the evidence: A critical appraisal has been implemented to assess the validity of the evidence used. The critical appraisal must be scientifically robust and transparent. The critical appraisal assessment has informed the interpretation of the evidence.	An adequate critical assessment of the quality of scientific evidence has been performed, and the procedure has been described in a transparent way. The critical assessment serves as a basis for the interpretation of the evidence.	The description is not sufficiently detailed to be able to assess; there are inaccuracies in the methodological process.	It is unclear whether a critical assessment of study data has taken place.
5. Use of the best available evidence: The content of the EBP source should be based on the best available evidence, specific to the clinical question. Well-designed and conducted evidence synthesis documents, when available, are preferred above primary studies.	The content of the EBP source is based on the best available evidence, specific to the clinical question. If available, well-designed and conducted evidence synthesis documents are preferred over primary studies.	The description is not sufficiently detailed to be able to assess; there are inaccuracies in the methodological process.	It is unclear whether the authors prioritize evidence synthesis documents over primary studies.
6. Citation of expert opinions: When expert opinions are cited, this must be clearly indicated in order to distinguish it from empirical evidence. Experts should be listed along with their professional designation, organization, and a conflicts of interest statement.	It is clearly stated when expert opinions are cited to distinguish it from empirical evidence. There is a description of the expertise of the experts, along with their professional affiliations, including a declaration of possible conflicts of interest.	The description is not sufficiently detailed to be able to assess. The expertise of the experts is unclear or the affiliations and declaration of conflicts of interest are lacking.	It is unclear whether expert opinions are cited or the distinction between expert opinion and empirical evidence is unclear.
7. Review process: The scientific quality and the clinical applicability of the EBP source is assessed by peer reviewers.	There is a detailed description of the review process of the scientific quality and clinical applicability of the EBP source.	Only a general description of the review process is available (eg, “information was reviewed by external reviewers”).	There is no information available about the review process.
8. Timeliness and updating: The frequency of updates is determined by the speed of developments in the field and is documented in the methodology. The content of the EBP source is checked and updated when new information is available. The date of first publication, date of the last update, and data on the next planned update are clearly displayed in the EBP source.	The EBP source is frequently updated in accordance with the developments in the field. The frequency of the updates is documented in the methodology. The date of first publication and last update can be found in the source, as well as information on the next planned update.	Updates are performed but not sufficiently frequently, which means that the content may be out of date.	Insufficient information about updates; date of last update not displayed.

CAPOCI[a] criteria	Conditions to rate as fulfilled	Conditions to rate as minor remark	Conditions to rate as major remark
9. Conflicts of interest: There is a formal policy on declaring and managing financial and nonfinancial conflicts of interest of the authors and other stakeholders. Possible conflicts of interest are reported.	Procedure for conflicts of interest has been implemented and documented (conflicts of interest should not be explicitly stated on the website, but the information must be able to be submitted to the assessor).	Conflict of interest procedure seems implemented but not reported.	No information about conflict of interest procedure available (conflicts of interest are not checked or reported).
10. Commercial support: It is clearly described to what extent commercial support was accepted for developing the content of the EBP source. The financier has no substantive input and therefore no influence on the result or the content of the EBP source. When advertisements on websites are a source of income, this must be clearly stated on the site. A short description of the advertising policy is published on the site. Advertisements and other promotional material must be presented in such a way that visitors can clearly distinguish between editorial content.	If commercial support is accepted, this is clearly and publicly announced and there is no influence of the financier on the content or the result of the EBP source.	Not applicable.	There is insufficient information to judge.

[a]CAPOCI: Critical Appraisal of Point of Care Information.

[b]EBP: evidence-based practice.

Interrater Reliability

Assessment of 5 POC resources with the CAPOCI tool (Multimedia Appendix 2) showed a substantial agreement between the reviewers (GB and GL) of .621 ($P<.01$) for scoring with the 3-point ordinal scale and .677 ($P<.01$) for scoring with the 7-point Likert scale, as calculated by the Kendall tau correlation coefficient.

The descriptive analysis of agreement between the reviewers for each criterion separately showed that both scoring systems lead to similar results (Table 2). There were no interrater differences in the scoring of criterion 5 (use of the best available evidence) for the 5 evaluated POC resources. For the other criteria, the most common reason for disagreement was a difference in rigor of application of the criterion by the 2 reviewers. For the scoring of criterion 6 (citation of expert opinion), the information found in the POC resources was differently interpreted by both reviewers, leading to a different score. Criterion 8 (timeliness and updating) and 9 (conflict of interest) were once scored differently because the necessary information was not always retrieved in the POC resource by both reviewers.

Table 2. Descriptive analysis of interrater agreement between 2 reviewers using the CAPOCI criteria on 5 POC resources. Scoring was done on a 3-point scale (fulfilled, minor remark, major remark) and on a 7-point Likert scale.

CAPOCI[a] tool	Scoring with 3-point scale; number of interrater disagreements (/5)	Scoring with 7-point Likert scale; number of interrater disagreements (>1) (/5)	Reason for interrater disagreement		
			Difference in rigor of application	The information in the POC[b] resource was differently interpreted by the reviewers	The information was found/not found in the POC resource by the reviewers
Criterion 1	1	0	✓	—[c]	—
Criterion 2	1	1	✓	—	—
Criterion 3a	1	1	✓	—	—
Criterion 3b	1	2	✓	—	—
Criterion 4	1	1	✓	—	—
Criterion 5	0	0	—	—	—
Criterion 6	2	1	—	✓	—
Criterion 7	2	2	✓	—	—
Criterion 8	1	2	✓	—	✓
Criterion 9	1	1	—	—	✓
Criteron 10	1	1	✓	—	—

[a]CAPOCI: Critical Appraisal of Point of Care Information.

[b]POC: point-of-care.

[c]Not applicable.

Discussion

Principal Findings

We developed a uniform, comprehensive, and validated tool specifically designed for the evaluation of POC information. The CAPOCI tool allows the systematic evaluation of the trustworthiness of POC resources, including the methodological process of searching, appraising, and synthesizing the best available evidence for a specific health topic.

Content of the CAPOCI Tool

Our systematic review showed that items for assessing POC resources could be divided into 4 main domains: author-related information, criteria related to evidence-based methodology, criteria related to website quality, and criteria related to website design and usability [20]. The rigor and specificity with which each of these domains should be evaluated is an interesting discussion. For the development of the CAPOCI tool, we focused on the first 2 domains as these directly relate to the trustworthiness of the information. The tool can be used for different types of POC resources (eg, interactive websites as well as online PDFs). Website quality and design and usability of web-based POC resources can be evaluated by existing tools for assessment of readability [22], design evaluation, and accessible health information contents [23,24].

The rigorous reporting of author-related information is a first important point when we focus on trustworthiness of (web-based) POC information. In this internet era, where personal opinions are easily expressed, it is crucial to have one or more authors who take responsibility for the content of a POC resource. Clarity and transparency about authorship must always go hand in hand with transparent and consistent reporting of conflicts of interest. Users of EBP resources need to be sure that the provided information reflects the best evidence and potential influence by competing interests has been minimized. The disclosure of conflicts of interests in guideline development has received much attention after reports on undisclosed financial conflicts of interest in guideline panels [25]. In response to this, the American College of Physicians Clinical Guidelines Committee recently published their policy for disclosure of interests and management of conflicts of interest in clinical guidelines [26] and the Cochrane Collaboration announced a new conflict of interest policy in October 2020 to strengthen user confidence [27].

To fulfill the claim of being evidence-based, a POC resource should be built on a robust methodological process. First, an explicit methodology for literature search and surveillance helps to avoid biases. A literature search should be systematic and well-documented for transparency and auditability. However, a systematic search with nonsystematic article selection can negate the effort for systematic search and surveillance. Therefore, a systematic strategy for article selection should be worked out as well, including a clear definition of inclusion and exclusion criteria for the selection of source information. These are basic principles of comprehensiveness in literature search that are important and widely adopted for guideline development [28], but comprehensiveness is also essential for other POC resources.

Once the information is selected, it should be critically appraised to assess the validity of the evidence it provides and to prevent inclusion of biased results. Some sources such as Cochrane systematic reviews are conducted following a strict protocol, and therefore they can be considered as trustworthy. But there are an increasing number of systematic reviews available in which the quality is often suboptimal [29,30]. When a POC resource is derived from a clinical practice guideline, the guideline should be critically appraised preferably with the AGREE II instrument [15]. Despite international efforts for quality standards of guidelines, development processes of guidelines still vary substantially, and many guidelines do not meet basic quality criteria [31].

When the criteria for literature and surveillance and critical appraisal of evidence are met, the criterion concerning the use of best available evidence will probably be fulfilled as well. We added this criterion to the tool to clarify that the content production process of a POC resource should be based preferentially on well-designed and conducted evidence synthesis documents rather than single studies. The best available evidence can be explained in different ways. The 6S pyramidal model of Haynes is often used as a conceptual framework for searching information resources for EBM [32], suggesting that the best available evidence is always higher up in the pyramid. The 6S model was later reworked to the evidence-based health care pyramid 5.0 for accessing preappraised evidence and guidance [33]. This pyramidal model adds systematically derived recommendations as a major type of information and simplifies the overall framework to 5 major layers of information types. Although these models give a good perspective on the difference in information types for EBM while focusing on preappraisal of information, it might be rather arbitrary since some resources may overflow between layers and may be difficult to attribute to one specific layer [6]. Furthermore, the best available evidence is also related to the clinical question to which the POC resource wants to provide an answer and the associated ideal study design. In epidemiological studies, the best available evidence will be observational data; in intervention studies, the best available evidence will be experimental data; and in studies of human experience, the best available evidence will be qualitative data. Levels of evidence frameworks such as the Oxford Centre for Evidence-Based Medicine table of evidence [34] are useful in the process of finding the appropriate evidence for a specific research question, but these frameworks predominantly address questions related to quantitative studies, while qualitative research is sometimes the best and most appropriate type of information to inform policy and planning decisions. Therefore, the framework with most relevance to the stakeholder group and most appropriate for the clinical question that is being asked should be considered when deciding on the best available evidence for the content of a POC resource.

For some clinical questions concerning rare diseases, best practice guidance, or treatments that are strongly context-dependent (eg, resistance for antibiotics), the evidence may be limited. In these situations, obtaining evidence from experts can be efficient, and experts may be the only or main source of evidence. However, the use of expert opinion should be clearly distinguished from empirical data in a POC resource [35]. To this end, the use of in-text referencing to published study data and clear statements on the use of expert opinion will contribute to more transparency and add to the trustworthiness of the POC resource.

Peer review, procedures for updating, and clear policies on conflict of interest and commercial support are basic principles for EBP that are present in all critical appraisal tools for clinical trials, systematic reviews, or guidelines [13,14,36]. Because of the POC aspect, a peer-review policy of a POC resource should not only address the scientific quality but also the clinical applicability of the information. To increase the transparency and auditability of the POC resource, the review process should be documented. For the same reason, procedures for updating should be documented. Systematic searches for new inputs should be executed with a frequency appropriate for the developments in the field of interest. An update should also be based upon either higher level of evidence or higher quality evidence; there is no point in adding new evidence if it is of lower quality or does not add empirical rigor.

POC resources are often web-based, which makes them very interesting for biomedical companies for online advertising. It is possible that a POC resource depends on commercial funding. In that case, a strict policy that ensures that the funder or advertiser has no influence on the content of the POC resource is mandatory. The distinction between content and advertising must always be very clear to the user.

Scoring Systems

We tested 2 scoring systems for the CAPOCI tool, both showing good reliability. The purpose of the evaluation of a POC resource can be a determining factor in the choice of scoring system. The 7-point Likert scale allows more nuance. It is also used in the AGREE II instrument and therefore probably familiar for users. It allows for a quantitative comparison when rating and comparing different POC sources.

The 3-point scale may be better suited when different reviewers must formulate a final judgement on a POC resource (eg, for granting a quality label) in consensus. The predefined conditions for the 3 categories also allow structured feedback to developers for POC resources on the criteria where additional information is required.

Use of the CAPOCI Tool

The CAPOCI tool is valid and reliable instrument that can be used by health care providers, researchers, and decision and policy makers for the evaluation of the trustworthiness of POC resources. Developers of POC resources can also use the tool as a guide in the development process.

Although the results of the interrater variability tests showed substantial agreement between reviewers for the CAPOCI tool, the descriptive analysis showed a difference in rigor of application between reviewers for 8 of the 11 criteria. In addition, different reviewers might interpret information in the POC resource differently, and the necessary information for evaluation was sometimes not retrieved. Therefore, a robust use of the CAPOCI tool includes an evaluation by at least 2

independent reviewers who reach a joint final judgment after discussion in order to reduce subjectivity and inaccuracy in the evaluation. Furthermore, reviewers should have experience with EBP methodology or should be trained by an experienced methodologist. Pilot testing is recommended to mitigate differences in rigor of application between reviewers and can be part of training for the good use of the CAPOCI tool.

Strengths and Limitations

The CAPOCI tool is based on a rigorous development process starting from a systematic review to consider and cover all possible criteria for the assessment of trustworthiness of POC information. The author group who defined the initial CAPOCI criteria consisted of methodologists and health care professionals, each with broad experience in guideline development and/or guideline validation.

Of all POC tools that were previously analyzed in our systematic review, only 2 were validated and tested for reliability [20]. However, these are essential requirements for tools used for evaluation in the context of EBP that were adopted in the CAPOCI tool.

Reliability testing was performed by reviewers who were involved in the development of the tool, which might have influenced the results. We will use this tool among different methodologists at our institute. The criteria and scoring methods will be refined where needed based on the experience and feedback of the users. Adding examples to the different criteria may also contribute to a better understanding and ease of use for less experienced users. Furthermore, content validation was done with a panel of 10 experts and reliability testing with only 2 reviewers, which can be considered as a limitation of this study. Using more reviewers or a more extended expert panel would add to the quality of validity and reliability testing.

Conclusion

With the development of the CAPOCI tool, we filled a gap in the evaluation of POC information. The CAPOCI tool facilitates the assessment of trustworthiness in POC resources. It may also provide guidance for producers of POC resources. Wide use of the CAPOCI tool may improve the quality and reliability of POC resources over time and may take EBP in daily practice to a higher level.

Acknowledgments

The authors would like to acknowledge the Belgian National Institution for Health and Disability Insurance (RIZIV) for funding this project. We also want to thank the panel members who participated in the Delphi study, for their valuable contribution to the development of the Critical Appraisal of Point-of-Care Information (CAPOCI) tool.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Members of the international panel of experts who participated in the Delphi study.
[DOCX File , 14 KB - jmir_v23i10e27174_app1.docx]

Multimedia Appendix 2
Scoring of five point-of-care resources with the Critical Appraisal of Point of Care Information tool.
[XLSX File (Microsoft Excel File), 14 KB - jmir_v23i10e27174_app2.xlsx]

References

1. Sackett DL. Evidence-based medicine. Semin Perinatol 1997 Feb;21(1):3-5. [doi: 10.1016/s0146-0005(97)80013-4] [Medline: 9190027]
2. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996 Jan 13;312(7023):71-72 [FREE Full text] [doi: 10.1136/bmj.312.7023.71] [Medline: 8555924]
3. Alper BS, Hand JA, Elliott SG, Kinkade S, Hauan MJ, Onion DK, et al. How much effort is needed to keep up with the literature relevant for primary care? J Med Libr Assoc 2004 Oct;92(4):429-437 [FREE Full text] [Medline: 15494758]
4. O'Connor PJ. Adding value to evidence-based clinical guidelines. JAMA 2005 Aug 10;294(6):741-743. [doi: 10.1001/jama.294.6.741] [Medline: 16091578]
5. Kwag KH, González-Lorenzo M, Banzi R, Bonovas S, Moja L. Providing doctors with high-quality information: an updated evaluation of web-based point-of-care information summaries. J Med Internet Res 2016 Jan 19;18(1):e15 [FREE Full text] [doi: 10.2196/jmir.5234] [Medline: 26786976]
6. Moja L, Banzi R. Navigators for medicine: evolution of online point-of-care evidence-based services. Int J Clin Pract 2011 Jan;65(1):6-11. [doi: 10.1111/j.1742-1241.2010.02441.x] [Medline: 21155939]
7. Wyatt J, Vincent S. Selecting computer-based evidence sources. Ann Oncol 1999 Mar;10(3):267-273 [FREE Full text] [doi: 10.1023/a:1008397018017] [Medline: 10355569]
8. Wyatt JC. Evidence-based health informatics and the scientific development of the field. Stud Health Technol Inform 2016;222:14-24. [Medline: 27198088]

9. Gagliardi A, Jadad AR. Examination of instruments used to rate quality of health information on the internet: chronicle of a voyage with an unclear destination. BMJ 2002 Mar 09;324(7337):569-573 [FREE Full text] [doi: 10.1136/bmj.324.7337.569] [Medline: 11884320]
10. Banzi R, Liberati A, Moschetti I, Tagliabue L, Moja L. A review of online evidence-based practice point-of-care information summary providers. J Med Internet Res 2010 Jul 07;12(3):e26 [FREE Full text] [doi: 10.2196/jmir.1288] [Medline: 20610379]
11. Campbell JM, Umapathysivam K, Xue Y, Lockwood C. Evidence-based practice point-of-care resources: a quantitative evaluation of quality, rigor, and content. Worldviews Evid Based Nurs 2015 Dec 02;12(6):313-327. [doi: 10.1111/wvn.12114] [Medline: 26629973]
12. Risk A, Dzenowagis J. Review of internet health information quality initiatives. J Med Internet Res 2001 Dec 26;3(4):E28 [FREE Full text] [doi: 10.2196/jmir.3.4.e28] [Medline: 11772543]
13. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol 2009 Oct;62(10):1013-1020. [doi: 10.1016/j.jclinepi.2008.10.009] [Medline: 19230606]
14. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017 Sep 21;358:j4008 [FREE Full text] [doi: 10.1136/bmj.j4008] [Medline: 28935701]
15. Appraisal of Guidelines for Research & Evaluation (AGREE) II Instrument. 2009. URL: http://www.agreetrust.org/resource-centre/agree-reporting-checklist/ [accessed 2021-07-26]
16. e-Health Ethics Initiative. e-Health Code of Ethics (May 24). J Med Internet Res 2000;2(2):E9 [FREE Full text] [doi: 10.2196/jmir.2.2.e9] [Medline: 11720928]
17. Boyer C, Selby M, Appel RD. The Health On the Net Code of Conduct for medical and health web sites. Stud Health Technol Inform 1998;52 Pt 2:1163-1166. [Medline: 10384641]
18. Winker MA, Flanagin A, Chi-Lum B, White J, Andrews K, Kennett RL, et al. Guidelines for medical and health information sites on the internet: principles governing AMA web sites. American Medical Association. JAMA 2000 Mar 22;283(12):1600-1606. [doi: 10.1001/jama.283.12.1600] [Medline: 10735398]
19. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb 01;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
20. Lenaerts G, Bekkering GE, Goossens M, De Coninck L, Delvaux N, Cordyn S, et al. Tools to assess the trustworthiness of evidence-based point-of-care information for health care professionals: systematic review. J Med Internet Res 2020 Jan 17;22(1):e15415 [FREE Full text] [doi: 10.2196/15415] [Medline: 31951213]
21. Fitch K, Bernstein SJ, Aguilar MD, LaCalle JR, Lázaro P, van het Loo M, et al. RAND/UCLA Appropriateness Method User's Manual. Santa Monica: RAND Corporation; 2001:1-126.
22. McInnes N, Haglund BJA. Readability of online health information: implications for health literacy. Inform Health Soc Care 2011 Dec 18;36(4):173-189. [doi: 10.3109/17538157.2010.542529] [Medline: 21332302]
23. Battineni G, Baldoni S, Chintalapudi N, Sagaro GG, Pallotta G, Nittari G, et al. Factors affecting the quality and reliability of online health information. Digit Health 2020 Aug 30;6:2055207620948996 [FREE Full text] [doi: 10.1177/2055207620948996] [Medline: 32944269]
24. The LIDA instrument: minervation validation instrument for health care web sites, version 1.2. Minervation Ltd. 2007. URL: http://www.minervation.com/wp-content/uploads/2011/04/Minervation-LIDA-instrument-v1-2.pdf [accessed 2021-09-07]
25. Editorial. Managing conflicts of interests in clinical guidelines. Lancet 2019 Aug 31;394(10200):710 [FREE Full text] [doi: 10.1016/S0140-6736(19)32000-8] [Medline: 31478486]
26. Qaseem A, Wilt TJ, Clinical Guidelines Committee of the American College of Physicians, Forciea M, Kansagara D, Crandall C, et al. Disclosure of interests and management of conflicts of interest in clinical guidelines and guidance statements: methods from the Clinical Guidelines Committee of the American College of Physicians. Ann Intern Med 2019 Sep 03;171(5):354-361 [FREE Full text] [doi: 10.7326/M18-3279] [Medline: 31426089]
27. Nejstgaard C, Bero L, Hróbjartsson A, Jørgensen A, Jørgensen K, Le M, et al. Association between conflicts of interest and favourable recommendations in clinical guidelines, advisory committee reports, opinion pieces, and narrative reviews: systematic review. BMJ 2020 Dec 09;371(12):m4234 [FREE Full text] [doi: 10.1136/bmj.m4234] [Medline: 33298430]
28. Developing NICE Guidelines: The Manual. London: National Institute for Health and Care Excellence; 2020 Oct 15. URL: https://www.nice.org.uk/media/default/about/what-we-do/our-programmes/developing-nice-guidelines-the-manual.pdf [accessed 2021-07-26]
29. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009 Jul 21;339(jul21 1):b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]

30. Gao Y, Cai Y, Yang K, Liu M, Shi S, Chen J, et al. Methodological and reporting quality in non-Cochrane systematic review updates could be improved: a comparative study. J Clin Epidemiol 2020 Mar;119:36-46. [doi: 10.1016/j.jclinepi.2019.11.012] [Medline: 31759063]
31. Qaseem A, Forland F, Macbeth F, Ollenschläger G, Phillips S, van der Wees P, Board of Trustees of the Guidelines International Network. Guidelines International Network: toward international standards for clinical practice guidelines. Ann Intern Med 2012 Apr 03;156(7):525-531 [FREE Full text] [doi: 10.7326/0003-4819-156-7-201204030-00009] [Medline: 22473437]
32. Dicenso A, Bayley L, Haynes RB. Accessing pre-appraised evidence: fine-tuning the 5S model into a 6S model. Evid Based Nurs 2009 Oct 24;12(4):99-101. [doi: 10.1136/ebn.12.4.99-b] [Medline: 19779069]
33. Alper BS, Haynes RB. EBHC pyramid 5.0 for accessing preappraised evidence and guidance. Evid Based Med 2016 Aug 20;21(4):123-125. [doi: 10.1136/ebmed-2016-110447] [Medline: 27325531]
34. OCEBM Levels of Evidence Working Group. Oxford Centre for Evidence-Based Medicine 2011 Levels of Evidence. Oxford: Oxford Centre for Evidence-Based Medicine; 2011. URL: https://www.cebm.ox.ac.uk/files/levels-of-evidence/cebm-levels-of-evidence-2-1.pdf [accessed 2020-10-27]
35. Schünemann HJ, Zhang Y, Oxman AD, Expert Evidence in Guidelines Group. Distinguishing opinion from evidence in guidelines. BMJ 2019 Jul 19;366:l4606. [doi: 10.1136/bmj.l4606] [Medline: 31324659]
36. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, AGREE Next Steps Consortium. AGREE II: advancing guideline development, reporting, and evaluation in health care. Prev Med 2010 Nov;51(5):421-424. [doi: 10.1016/j.ypmed.2010.08.005] [Medline: 20728466]

Abbreviations

AGREE II: Appraisal of Guidelines for Research and Evaluation II
AMSTAR: A Measurement Tool to Assess Systematic Reviews
CAPOCI: Critical Appraisal of Point of Care Information
EBM: evidence-based medicine
EBP: evidence-based practice
POC: point-of-care

Edited by R Kukafka; submitted 15.01.21; peer-reviewed by C Urquhart, D Tao; comments to author 25.04.21; revised version received 08.05.21; accepted 13.07.21; published 05.10.21.

Please cite as:
Lenaerts G, Bekkering GE, Goossens M, De Coninck L, Delvaux N, Cordyn S, Adriaenssens J, Aertgeerts B, Vankrunkelsven P
A Tool to Assess the Trustworthiness of Evidence-Based Point-of-Care Information for Health Care Professionals (CAPOCI): Design and Validation Study
J Med Internet Res 2021;23(10):e27174
URL: https://www.jmir.org/2021/10/e27174
doi:10.2196/27174
PMID:34609314

Review

Psychometric Properties of the Internet Gaming Disorder Scale–Short-Form (IGDS9-SF): Systematic Review

Lok Y J Poon[1], BSc; Hector W H Tsang[1,2], PhD; Tsan Y J Chan[1], BSc; Sze W T Man[1], BSc; Lok Y Ng[1], BSc; Yi L E Wong[1], BSc; Chung-Ying Lin[3,4,5,6*], PhD; Chi-Wen Chien[1*], PhD; Mark D Griffiths[7], PhD; Halley M Pontes[8], PhD; Amir H Pakpour[9], PhD

[1]Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hung Hom, Hong Kong
[2]Mental Health Research Centre, The Hong Kong Polytechnic University, Hung Hom, Hong Kong
[3]Institute of Allied Health Sciences, College of Medicine, National Cheng Kung University, Tainan, Taiwan
[4]Department of Public Health, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan
[5]Department of Occupational Therapy, College of Medicine, National Cheng Kung University, Tainan, Taiwan
[6]Biostatistics Consulting Center, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan
[7]International Gaming Research Unit, Department of Psychology, Nottingham Trent University, Nottingham, United Kingdom
[8]Department of Organizational Psychology, Birkbeck, University of London, London, United Kingdom
[9]Department of Nursing, School of Health and Welfare, Jönköping University, Jönköping, Sweden
[*]these authors contributed equally

Corresponding Author:
Chung-Ying Lin, PhD
Institute of Allied Health Sciences
College of Medicine
National Cheng Kung University
1 University Road
Tainan, 701
Taiwan
Phone: 886 6 2353535 ext 5106
Fax: 886 6 2367981
Email: cylin36933@gmail.com

Abstract

Background: The Internet Gaming Disorder Scale–Short-Form (IGDS9-SF) is among the best with regard to its psychometric properties. Therefore, clinical psychologists are likely guided to use the IGDS9-SF if they want to assess or screen the disordered gaming in their practice. However, the information, especially psychometric evidence, concerning the IGDS9-SF has not been fully examined and summarized.

Objective: This systematic review evaluated the psychometric properties of different language versions of the IGDS9-SF and assessed its methodological quality in order to improve the clinicians' understanding of the IGDS9-SF and facilitate its use.

Methods: Systematic literature searches were carried out using *Embase*, *MEDLINE*, *PsycINFO*, *PubMed*, *ScienceDirect*, *Scopus*, and *Web of Science*. The review included English-language studies of any research design that have reported at least one psychometric property of the IGDS9-SF, as defined by the COnsensus-based Standards for the selection of health status Measurement INstrument (COSMIN), and have aimed at testing the psychometric properties of the IGDS9-SF.

Results: In total, 21 studies comprising 15 language versions of the IGDS9-SF were included. Overall, the IGDS9-SF showed adequate internal consistency (although some items did not have satisfactory item-total correlation [IT]), excellent criterion validity, and the ability to distinguish different subgroups with measurement invariance being supported across gender and age. In terms of factor structure, the IGDS9-SF was shown to have a unidimensional factor structure across all 21 studies.

Conclusions: Although there is insufficient evidence regarding the responsiveness and properties of the IGDS9-SF using item response theory, the existing evidence supports its use in assessing disordered gaming among individuals.

(J Med Internet Res 2021;23(10):e26821) doi:10.2196/26821

KEYWORDS

psychometrics; IGDS9-SF; gaming addiction; gaming disorder; problematic gaming; internet; gaming; internet gaming

Introduction

In the era of technology, internet use has become one of the essential components of everyone's life [1]. However, internet use can be a potential hazardous tool for a minority of individuals, particularly adolescents and emerging adults [2]. Many websites and mobile applications, including online games, provide highly interactive features and services (eg, forums) that attract millions of users worldwide. Users, especially younger people, may therefore be unable to fully control their internet use as they encounter challenges to their self-control, alongside the addictive features of these applications, which tend to be associated with poor self-control levels among dysregulated and disordered users [3,4]. Therefore, they may end up spending a substantial amount of time within different virtual environments because they want to engage in social interaction on the internet [2]. In other words, internet use (such as online communication and gaming) may be a facilitator of reinforcing social relationships due to peoples' preference for online social interaction [5].

Despite the many positive outcomes associated with online gaming, such as decreased loneliness and promotion of psychological well-being [6], the negative and dysfunctional effects of online gaming were observed in a minority of individuals in a systematic review and meta-analysis (N=226,247 from 53 studies across 17 countries, including European, American, and Asian countries, with different populations, eg, adolescents, gamers, and the general population, with a prevalence of 3.05% with a 95% confidence interval between 2.38% and 3.91%) [7]. Its negative impacts on both physical and mental health (eg, poor sleep quality, musculoskeletal discomfort, and increased psychological distress) have been widely reported in recent years [8,9]. Consequently, internet gaming disorder (IGD) was incorporated into the fifth revision of the American Psychiatric Association's (APA) *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) as a tentative disorder and behavioral addiction [10].

Additionally, the 11th revision of the *International Classification of Diseases* (ICD-11) also formally recognized IGD and categorized it as an addictive disorder [11]. However, diagnosis in the ICD-11 relies upon clinical and functional impairment by manifesting pathological aspects rather than biological concepts, such as withdrawal and tolerance, as suggested in the DSM-5 in relation to IGD [11].

Ever since the emergence of the first internet addiction scale in 1998 [12], many psychometric assessment tools assessing the addictive effects of problematic internet use have been developed. More recently, self-report psychometric tests assessing IGD have been developed to assess the following nine IGD criteria proposed in the DSM-5 [10]: (1) preoccupation or obsession with gaming; (2) withdrawal symptoms when unable to engage in gaming; (3) tolerance, leading to necessity of spending more time in gaming for satisfying the urge of gaming; (4) inability to control participation in gaming; (5) not engaging in other hobbies and entertainment because of gaming; (6) persisting in excessive gaming irrespective of the psychosocial problems; (7) lying to family members or others in respect to the time spent on gaming; (8) using gaming to escape negative mood states; and (9) compromising occupation/education/significant relationships due to the involvement in gaming [13].

Currently, there are at least seven psychometric assessment tools for IGD that have been developed using DSM-5 criteria [14-17]. Among these seven instruments, two instruments rate responses using a Likert-type scale, two instruments rate items using either a Likert-type scale or a dichotomous scale, and one instrument begins rating items with a Likert-type scale and then converts to a dichotomous scale. More specifically, the 20-item Internet Gaming Disorder Test (IGDT-20) and the 9-item Internet Gaming Disorder Scale–Short-Form (IGDS9-SF) both use a 5-point Likert scale from 1 (never) to 5 (very often) and use the scale for scoring [18,19]. The 10-item Internet Gaming Disorder Test (IGDT-10) uses a 3-point Likert scale (ie, 0=never, 1=sometimes, and 2=often), which is then converted to a dichotomous score (ie, 0=never, 1=sometimes and often) [10]. The 27-item Internet Gaming Disorder Scale (IGDS) [20] and the IGDS9-SF [18] use either a 6-point Likert scale from 0 (never) to 5 (every day or almost every day) or a dichotomous scale (0=no and 1=yes).

Among the aforementioned psychometric instruments, the IGDS9-SF emerges as a robust and concise psychometric tool for assessing individuals with IGD, with a recent study identifying this tool as having great support in relation to its sound psychometric properties and significant advantages in comparison to most existing tools for IGD [21]. The IGDS9-SF includes all nine IGD criteria proposed by the APA in the DSM-5, with the features of conciseness and brief administration time, which is of great pragmatic utility in busy clinical settings when screening for the risk of IGD. Moreover, the psychometric properties of the IGDS9-SF have been widely assessed, including structural validity, internal consistency, cross-cultural validity/measurement invariance, reliability, measurement error, criterion validity, convergent validity, and discriminative or known-group validity.

In the current literature, the IGDS9-SF has been translated into 17 languages: Chinese, with three sublanguages of traditional Chinese in Hong Kong [22-24], traditional Chinese in Taiwan [22-24], and simplified Chinese in mainland China [25]; Albanian [26]; Italian [26]; English [26]; European and South American Portuguese [27,28]; Slovenian [29]; Persian [30]; Polish [31]; Spanish [32-34]; Turkish [35]; German [36]; Czech [37]; Malay [38]; and Korean [39]. As a screening tool, the IGDS9-SF can help clinicians in assessing IGD severity and the detrimental health impacts on the individual's life with reasonable accuracy in a time-efficient way. Therefore, it has increasingly been psychometrically examined and used widely in epidemiological studies [40].

Although the psychometric properties of the IGDS9-SF have been examined among different populations, to the best of our knowledge, there is no systematic review of the IGDS9-SF reporting its psychometric characteristics in depth. Dispersed information in the extant literature regarding the IGDS9-SF with varying sample sizes and across different countries makes it timely to investigate whether clinicians should be adopting the IGDS9-SF for assessment of IGD. More specifically, there are a number of questions that are best answered by carrying out a systematic review of the psychometric properties of the IGDS9-SF.

First, the IGDS9-SF may have different psychometric features across different language versions, and it is unclear whether the existing psychometric evidence for different language versions of the IGDS9-SF is equivalent. Second, prior psychometric testing studies on the IGDS9-SF need to be evaluated for their methodological quality.

Without formal assessment of the quality of previous studies on the IGDS9-SF, the results pertaining to its psychometric properties may be biased. For the sake of improving clinicians' understanding and facilitating the use of IGDS9-SF across other contexts beyond research settings, this systematic review incorporated different items of evidence concerning the psychometric features of the IGDS9-SF across a wide range of populations. More specifically, if the psychometric properties of the IGDS9-SF are found to be supported across different contexts, clinicians can use the IGDS9-SF criteria to exchange their expert opinions using the same signs, symptoms, and components. For example, a clinician in Taiwan can use IGDS9-SF scores to assess the IGD severity level of a disordered gamer, and this information can be well understood and correctly interpreted by clinicians in other countries.

Furthermore, when carrying out psychological assessment, clinicians are required by professional governing bodies (eg, the American Psychological Association, the British Psychological Society, the Australian Psychological Society) to adopt valid and reliable psychometric tools to support the adoption of evidence-based practices. Therefore, this review is of importance to clinicians working with disordered gamers.

Methods

This review followed the recommended flow of the Preferred Reporting Items of Systematic Reviews and Meta-analyses (PRISMA) guidelines [41]. The evaluation of the psychometric properties of the IGDS9-SF was conducted with reference to the COnsensus-based Standards for the selection of health status Measurement INstrument (COSMIN) guidelines for systematic reviews of patient-reported outcome measures (PROMs) [42]. The review protocol was registered in the international prospective register of systematic reviews (PROSPERO; registration no. CRD42020198376).

Search Strategy

A literature search was carried out using *Embase*, *MEDLINE*, *PsycINFO*, *PubMed*, *ScienceDirect*, *Scopus*, and *Web of Science* (these seven databases were used because they are commonly used databases in this field of psychology) to retrieve relevant studies published up to March 31, 2020, with the following search strategy: IGDS9-SF [All fields] OR IGDS-SF9 [All fields] OR IGD-SF [All fields] OR Internet Gaming Disorder Scale-Short-form [All fields] OR Internet Gaming Disorder Scale Short form [All fields] OR Internet Gaming Disorder Scale - 9-Item Short Form [All fields] OR 9-item Internet Gaming Disorder Scale - Short Form [All fields]. Slight modifications were made to the search strategy in order to optimize the search within each database (Multimedia Appendix 1).

A total of 2533 journal articles were identified. Duplicates (n=200) were removed using EndNote. The titles and abstracts of the remaining journal articles (n=2333) were screened for eligibility by two authors independently (ie, the same two authors who screened all 2333 journal articles). Of these, 2286 articles did not focus on the IGDS9-SF and were removed. Full texts of all potential articles were then retrieved (n=47) and screened using the same procedure. Of these, 26 were removed because of the following reasons: (1) the study did not test the psychometric properties of the IGDS9-SF (n=22), (2) the study was a non-peer-reviewed conference paper (n=3), or (3) the study was a review paper (n=1). The remaining 21 studies were evaluated and analyzed in this systematic review.

Study Selection

The review included only English-language studies of all types of research design under the condition that they (1) reported at least one psychometric property of the IGDS9-SF (eg, internal consistency, reliability, measurement error, content validity, construct validity, criterion validity, or responsiveness), as defined by COSMIN, and (ii) aimed at testing the psychometric properties of the IGDS9-SF. Exclusion criteria were nonrefereed studies, review studies, conference proceedings, dissertations, commentaries, editorials, or letters to journal editors. The aforementioned inclusion and exclusion criteria were applied to all paper titles and abstracts, and screening of full texts.

Evaluation of Psychometric Properties

The psychometric properties of the included studies were evaluated by five authors (TYJC, SWTM, LYN, LYJP, and YLEW), and each study was independently assessed by any two of the five authors using the COSMIN Risk-of-Bias checklist [43] according to the user manual of the COSMIN methodology for systematic reviews of PROMs. Following this, the corresponding author (C-YL) verified the evaluation results made by the five authors. The checklist comprised eight assessment properties: structural validity (ie, the property assessing whether the IGDS9-SF has a unidimensional structure), internal consistency (ie, the property assessing whether the nine items in the IGDS9-SF assess the same underlying construct, ie, IGD), cross-cultural validity/measurement invariance (ie, the property assessing whether the IGDS9-SF is interpreted similarly across different subgroups/cultures), reliability (ie, the property assessing whether the IGDS9-SF can be reproduced), measurement error (ie, the property reporting the error that cannot be assessed using the IGDS9-SF), criterion validity (ie, the property assessing the association between the IGDS9-SF and the gold standard assessing the same construct of IGD), hypothesis testing for

construct validity (ie, the property assessing the association between the IGDS9-SF and a tool assessing a similar construct of IGD [ie, concurrent validity] or a tool assessing a construct different from IGD [ie, discriminant validity]), and responsiveness (ie, the property assessing whether the IGDS9-SF can effectively detect the improvement of IGD when effective treatment is given). It was noted that ΔCFI of >–0.01 would be used for evaluation of measurement invariance (where CFI is the comparative fit index) [44]. In the evaluation of criterion, concurrent, and convergent validity, coefficient values greater than or equal to 0.5 indicated strong correlation, values between 0.3 and 0.5 indicated fair correlation, and values between 0.1 and 0.3 indicated poor correlation [45].

Ranging from 3 to 13 items, each property was scored on a 4-point scale with four predefined options: very good (V), adequate (A), doubtful (D), and inadequate (I) [43]. The overall score of a psychometric property was graded based on the worst-score-counts principle [43]. Psychometric properties that were not available in the published study were marked as not applicable (NA). In this review, structural validity, internal consistency, cross-cultural validity/measurement invariance, reliability, measurement error, criterion validity, concurrent validity, convergent validity, and discriminant validity were synthesized, evaluated, and reported. Furthermore, additional psychometric properties, including the floor and ceiling effects and item-total correlation (IT), were also reported.

Assessments of the Quality of Statistical Findings

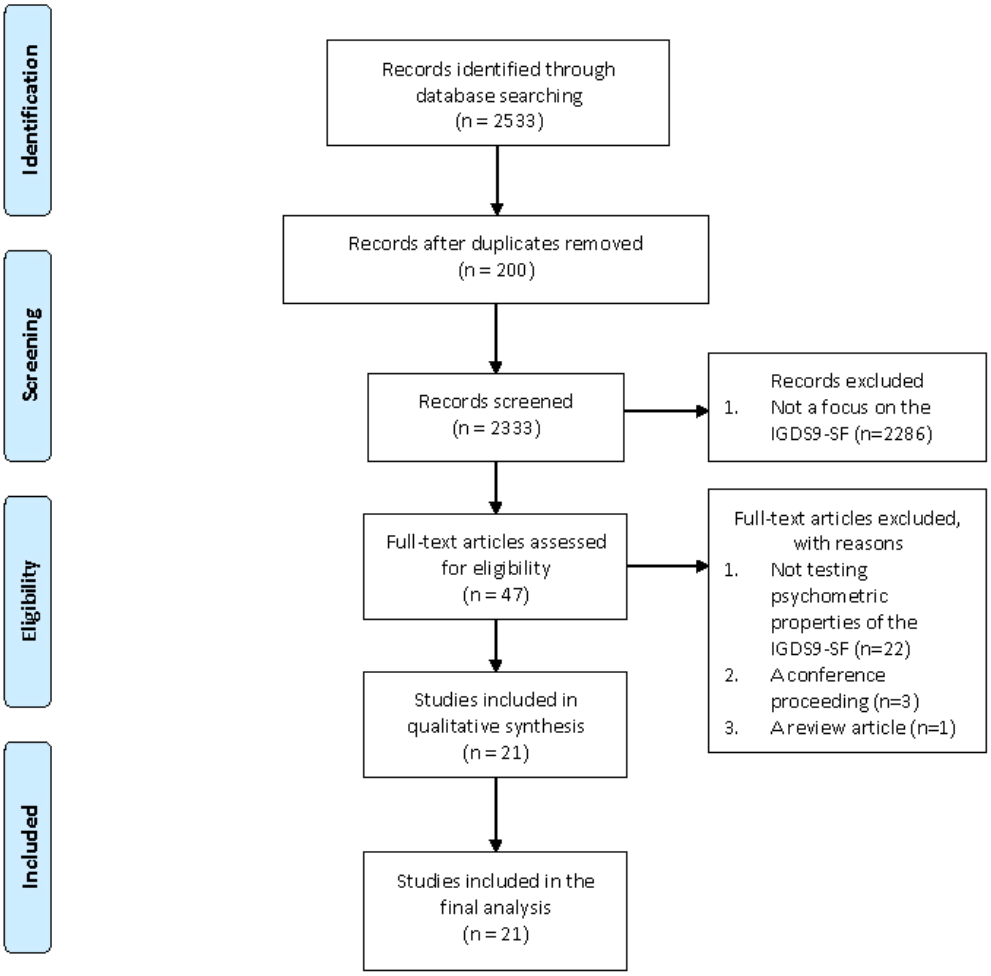
The quality of the statistical results of each reported measurement property of the IGDS9-SF was rated against the updated criteria for good measurement properties based on Terwee et al [46] and Prinsen et al [47]. Each criterion was rated as sufficient (+), insufficient (–), or indeterminate (?), with each result being compared against the criterion and reported in the results table.

Results

Study Selection

Of the 2533 identified studies, 200 were duplicates and 2333 were screened for abstracts (Figure 1). A total of 47 studies met the inclusion criteria and underwent subsequent full-text screening, of which 26 studies were further excluded due to the following reasons: the focus was not on the IGDS9-SF (n=22), they were conference proceedings (n=3), or they were review studies that compared different instruments without detailed information and quality assessment of each instrument (n=1). Therefore, a total of 21 studies on 15 language versions of the IGDS9-SF (ie, English, European Portuguese, South American Portuguese, Spanish, Albanian, Italian, Turkish, Slovenian, Polish, Persian, Malay, Korean, and Chinese, with three sublanguages of traditional Chinese in Hong Kong, traditional Chinese in Taiwan, and simplified Chinese in mainland China) were found to be relevant and included in the qualitative synthesis.

Figure 1. Flowchart of study selection [41]. IGDS9-SF: Internet Gaming Disorder Scale–Short-Form.



Demographic Features

In terms of sample characteristics across all eligible studies, nine studies included gamers [19,26,31,38,39,48-51], two studies included people from gaming halls [26,52], one study comprised e-sports tournament (Electronic Sports League [ESL]) amateur e-sport players [48], and two studies included adults from the general community who played massively multiplayer online (MMO) games [38,51]. In addition, nine studies included university students [22-24,26,28,35,39,48,52,53], two studies included high school students [28,30], and two studies included students without specifying the education level [26,52]. Furthermore, one study included students learning English in a private institution and students in a private teaching institution [36]. The remaining studies included students in the sixth to ninth grades [29], students in the eighth grade [29], primary students in the fourth to sixth grades [24], and vocational training students [31]. The characteristics and demographics of the included studies are summarized in Table 1. Although these studies recruited participants with different demographic characteristics, the findings on the properties of the IGDS9-SF were similar, and the psychometric properties are summarized in Table 2. Detailed information about the psychometric properties is reported descriptively in subsequent sections.

Table 1. Characteristics of included studies and psychometric properties of the IGDS9-SF[a] not included in COSMIN[b].

Author/country	Study design	Population	Sample size	Age (years)	Testing period	Others (not included in COSMIN)
Chen et al [24]/Hong Kong and Taiwan	Cross-sectional	University students	Hong Kong 304 (67.4% female) Taiwan 336 (50.3% female)	Hong Kong 24.18 ± 4.92 Taiwan 20.51 ± 1.22	3 months (for testing temporal invariance)	Skewness=0.21-2.41 Kurtosis=−1.06-6.74
Beranuy et al [32]/Spain	Cross-sectional	Vocational training students	535 (21.5% female)	18.35 ± 2.13	N/A[c]	IT[d]=0.47-0.67
de Palo et al [26]/Albania, USA, UK, and Italy	Cross-sectional	Albania and Italy: People from schools, universities, and gaming halls USA and UK: English-speaking gamers from popular online gaming forums	Overall 1411 (36.4% female) Albania 228 (50.9% female) USA 237 (21.7% female) UK 275 (13.9% female) Italy 671 (45.4% female)	Overall 25.94 ± 8.91 Albania 31.38 ± 10.97 USA 29.09 ± 10.72 UK 29.50 ± 9.48 Italy 21.62 ± 3.9	N/A	No item showed absolute values of skewness >2 or values of kurtosis >7
Evren et al [48]/Turkey	Cross-sectional	University students, active internet gamers, and ESL Turkey amateur e-sport players	457 (62.4% female)	N/A	N/A	IT=0.663-0.826
Gomez et al [49]/USA	Cross-sectional	Internet gamers	868 (39.7% female)	Overall 28.64 ± 8.79 Male 27.94 ± 7.95 Female 20.09 ± 9.29	N/A	N/A
Leung et al [22]/Taiwan and Hong Kong	Cross-sectional	University students	Hong Kong 306 (67.6% female) Taiwan 336 (50.3% female)	Hong Kong 24.08 ± 5.06 Taiwan 20.51 ± 1.22	N/A	IT=0.68-0.85
Monacis et al [52]/Italy	Cross-sectional	Students from Italian schools, universities, and gaming hall	687 (45.4% female)	21.62 ± 3.90	N/A	No item showed absolute values of skewness >2 or values of kurtosis >7
Pontes and Griffiths [19]/English-speaking countries	Cross-sectional	English-speaking gamers from 58 different countries	1060 (14.9% female)	27 ± 9.02	N/A	Floor effect=3.8%-5.6% Ceiling effect=0.2%-0.8%
Pontes and Griffiths [27]/Portugal	Cross-sectional	Students in sixth, seventh, eighth, and ninth grades of a major located in the Algarve	509 (47.9% female)	13 ± 1.64	N/A	No item showed absolute values of skewness >3 or values of kurtosis >9
Pontes et al [50]/USA, India, and UK	Cross-sectional	English-speaking gamers from English-speaking online gaming forums that are popular among gamers	USA 405 (38% female) India 336 (32.4% female) UK 272 (49.3% female)	USA 32.57 ± 11.33 India 30.37 ± 8.90 UK 41.61 ± 14.03	N/A	N/A
Pontes et al [29]/Slovenia	Cross-sectional	Students in eighth grade	1071 (49.8% female)	13.44 ± 0.59	N/A	No item showed absolute values of skewness >3 or values of kurtosis >9
Schivinski et al [31]/Poland	Cross-sectional	Gamers	3377 (17.4% female)	20 ± 4.3	N/A	Skewness=–0.08-1.51 Kurtosis=–0.91-1.19

Author/country	Study design	Population	Sample size	Age (years)	Testing period	Others (not included in COSMIN)
Severo et al [28]/Brazil	Cross-sectional	High school and college students	555 (42.5% female)	20.3 ± 5.4	N/A	No item showed absolute values of skewness >3 or values of kurtosis >9 IT=0.342-0.668
Stavropoulos et al [54]/USA and Australia	Longitudinal	Emerging adults from the general community who played massively multiplayer online (MMO) games	Australia 61 (26.2% female) USA 120 (40.2% female)	Australia 22.53 ± 3.04 USA 22.35 ± 2.82	60-90 days	N/A
Wu et al [30]/Iran	Cross-sectional	High school adolescents	2363 (35.2% female)	15.6 ± 1.2	2 weeks (for testing test-retest reliability)	IT=0.54-0.74 Floor effect=0.8% Ceiling effect=1.8%
Yam et al [23]/Hong Kong	Cross-sectional	University students	307 (67.6% female)	21.64 ± 8.11	N/A	IT=0.527-0.724 Floor effect=21% Ceiling effect=0%
Stavropoulos et al [51]/Australia, USA, and UK	Cross-sectional	Internet gamers	Australia 171 (23.4% female) USA 463 (42.1% female) UK 281 (13.9% female) Total 915 (55.1 female)	Australia 25.72 ± 5.52 USA 25.23 ± 2.76 UK 29.49 ± 9.47 Total 15.54 ± 0.65	N/A	N/A
Arıcak et al [35]/Turkey	Cross-sectional	Group 1: Students learning in English in a private institution Group 2: University students learning in English Group 3: Students from fifth grade to final year of university Group 4: Students in a private teaching institution	Group 1, 35 (54% female) Group 2, 33 (42% female) Group 3, 455 (46% female) Group 4, 64 (48% female)	Group 1, 12.50 ± 1.20 Group 2, 23.94 ± 1.52 Group 3, 15.83 ± 4.16 Group 4, 13.84 ± 1.59	Group 4: 2 weeks (for testing test- retest reliability)	Floor effect=9% Ceiling effect=0%
Chen et al [25]/mainland China	Cross-sectional	Primary school children in fourth to sixth grades	1108 (51.7% female)	10.37 ± 0.95	N/A	IT=0.55-0.76 Floor effect=24.6% Ceiling effect=0%
Kim and Ko [39]/Korea	Cross-sectional	Korean internet game users from major online gaming forums, universities, counseling centers, and libraries located in the greater Seoul area, Gyeonggi and Chungcheong Provinces of Korea	594 (29.6% female)	23.5 ± 6.29	N/A	IT=0.49-0.68

Author/country	Study design	Population	Sample size	Age (years)	Testing period	Others (not included in COSMIN)
T'ng and Pau [38]/Malaysia	Cross-sectional	Youth who played MOBA[e]	1050 (25.1% female)	21.96 ± 2.37	N/A	No item showed absolute values of skewness >2 or values of kurtosis >7 Floor effect=0.1% Ceiling effect=1.7%

[a]IGDS9-SF: Internet Gaming Disorder Scale–Short-Form.

[b]COSMIN: COnsensus-based Standards for the selection of health status Measurement Instrument.

[c]N/A: not available.

[d]IT: item-total correlation

[e]MOBA: multiplayer online battle arena

Table 2. Summarized psychometric properties of the IGDS9-SF[a] in the analyzed studies.

Psychometric property included in COSMIN[b]	N	Meth qual[c]	Result (rating[d])
Structural validity	19,049	I	• One factor (+)
Internal consistency	19,049	V	• .810-.963 (+)
Cross-cultural validity	7352	I	• Age (+) full invariance • Gender (?) partial invariance • Time on gaming (+) partial invariance • Country (–) partial invariance
Reliability	2962	D	• ICC[e]=.94 • Pearson correlation=.756-.87
Measurement error	2962	D	• 0.16-2.27 (?)
Criterion validity	457	V	• r=.988 (+)
Concurrent validity	12,323	V	• Absolute r=.00-.556 • Absolute β=.103-.663
Convergent validity	6149	V	• Absolute r=.06-.827
Discriminative validity	1142	D	• Significant difference found in age and gender

[a]IGDS9-SF: Internet Gaming Disorder Scale–Short-Form.

[b]COSMIN: Consensus-based Standards for the selection of health status Measurement Instrument.

[c]COSMIN score after removing the sample size item from the rating: V, very good; A, adequate; D, doubtful; I, inadequate; N, not applicable.

[d]Quality score of the measurement property: +, sufficient; -, insufficient; ?, indeterminate.

[e]ICC: intraclass correlation coefficient.

Structural Validity

All studies reported the structural validity of the IGDS9-SF for the 15 versions (n=19,049). Of the 21 studies, 14 demonstrated very good methodological quality and had a positive rating for the quality of statistical findings; 6 studies with good to excellent methodological quality showed an indeterminate rating on the quality of statistical findings due to the absence of the standardized root-mean-square residual (SRMR; Supplementary Table S1 in Multimedia Appendix 2); and 1 study related to the Turkish version and with poor methodological quality had an indeterminate rating on the quality of statistical findings of structural validity (CFI=.987, root-mean-square error of approximation [RMSEA]=.064).

In addition, one study performed Rasch analysis to test the person separation reliability, person separation index, item separation reliability, and item separation index of the Persian version of the IGDS9-SF. The results showed doubtful methodological quality and a positive rating for the quality of statistical findings (person separation reliability=.86, person separation index=2.50, item separation reliability=1.00, item separation index=28.79). The study also reported an acceptable range of infit (0.79-1.37) and outfit mean square (0.74-1.34), as well as the range of item difficulties (–1.06-1.57).

Internal Consistency

All studies evaluated the internal consistency of the IGDS9-SF for the 15 versions (n=19,049). All studies had very good methodological quality and showed a positive rating for the quality of statistical findings concerning internal consistency (Cronbach α=.810-.963 and person separation reliability=.86; see Supplementary Table S1 in Multimedia Appendix 2).

Cross-Cultural Validity/Measurement Invariance

Of the 21 studies, 9 examined the measurement invariance of the IGDS9-SF across different factors (n=7352; see Supplementary Table S2 in Multimedia Appendix 2). The measurement invariance across age and gender was found to be fully or partially supported in these studies. Furthermore, one study on the Persian version found that partial invariance is supported across hours spent online gaming per week.

In addition, four studies reported the measurement invariance across multiple countries, and one study supported full invariance across the traditional Chinese (Hong Kong and Taiwan) versions. Partial metric and scalar invariance were supported across Albania, the United States, the United Kingdom, and Italy but neither across the United States, the United Kingdom, and India nor across the United States, the United Kingdom, and Australia.

Furthermore, one study examined the differential item functioning (DIF) contrast across gender and time spent gaming using Rasch analysis; three DIF items were found, namely Item 4, “Fail to control or cease gaming activities” concerning gender (DIF contrast=–0.55); item 4 concerning time spent online gaming per week (DIF contrast=–0.67); and item 9, “Jeopardize or lose an important thing because of gaming activity” concerning time spent online gaming per week (DIF contrast=.61).

Reliability

Reliability of the IGDS9-SF was evaluated in four studies on four versions (n=2962). The studies investigated the test-retest reliability of the IGDS9-SF (ie, traditional Chinese [Hong Kong and Taiwan], Persian, and Turkish versions), and all demonstrated fair methodological quality. All studies also demonstrated a positive rating for the quality of statistical findings (Supplementary Table S2 in Multimedia Appendix 2), indicating the high reliability of the IGDS9-SF.

Measurement Error

The above-mentioned four studies on four versions also examined the measurement errors in using the IGDS9-SF (n=2962; see Supplementary Table S3 in Multimedia Appendix 2). Of the four studies, three demonstrated doubtful methodological quality in traditional Chinese (Hong Kong and Taiwan) and Turkish versions, while the remaining study on the Persian version showed very good methodological quality. The standard error of measurement (SEM) was reported in two studies on the traditional Chinese (Hong Kong and Taiwan) and Persian versions. The SEM range of the four included studies was from 0.16 to 2.27. Given that only one study on the traditional Chinese (Hong Kong and Taiwan) versions reported the minimal important change (MIC) in the IGDS9-SF (smallest real difference=.44), the quality of this statistical finding remains indeterminate (Supplementary Table S3 in Multimedia Appendix 2).

Criterion Validity

One study on the Turkish version evaluated criterion validity and showed very good methodological quality (Supplementary Table S3 in Multimedia Appendix 2). This study reported a strong correlation between the IGDS9-SF and the 27-item IGDS (r=.988) and demonstrated a positive rating for the quality of statistical findings.

Concurrent Validity

Of the 21 studies, 11 using 11 versions evaluated the concurrent validity of the IGDS9-SF (Supplementary Table S4 in Multimedia Appendix 2). Very good methodological quality was noted in all 11 studies. Given the coefficient range (either the Pearson r correlation or the standardized regression coefficient) below 0.7, the rating for the quality of their statistical results was negative.

Construct Validity (Convergent and Discriminant)

Of the 21 studies, 10 evaluated the convergent validity of the IGDS9-SF. Very good methodological quality was demonstrated in seven studies: traditional Chinese (Hong Kong and Taiwan), simplified Chinese, Spanish, Italian, European Portuguese, South American Portuguese, and Korean versions of the IGDS9-SF. Another three studies on the traditional Chinese (Hong Kong and Taiwan), English, and Turkish versions showed inadequate methodological quality. All 10 studies had a positive rating for the statistical quality (Supplementary Table S4 in Multimedia Appendix 2).

In addition, two studies evaluated the discriminative validity of the IGDS9-SF (Supplementary Table S5 in Multimedia Appendix 2). Both studies on the Turkish and Italian versions of the IGDS9-SF demonstrated very good methodological quality and had a positive rating for the statistical quality. Furthermore, two studies on the Italian and Turkish versions demonstrated significant gender differences on the IGDS9-SF score, with males having a higher level of IGD than females: $t(451)=5.73$, P=.001, d=.54 versus $t(676.317)=6.61$, P<.001. The study on the Italian version also showed significant age differences on the score, with young adults obtaining higher scores than old adults: $t(648.267)=10.03$ and P<.001.

Other Psychometric Properties

Floor and Ceiling Effects

Floor and ceiling effects of the IGDS9-SF were reported in six studies on the English, traditional Chinese (Hong Kong), simplified Chinese, Persian, Turkish, and Malay versions. The study on the English version of the IGDS9-SF reported more gamers at floor-level scores (3.8%-5.6%) than at the ceiling level (0.2%-0.8%). The study on the Persian version of the IGDS9-SF reported a 1.8% ceiling effect and a 0.8% floor effect, whereas studies on the traditional Chinese (Hong Kong) and simplified Chinese versions reported no ceiling effect but a relatively high floor effect (21%-24.6%). Another study on the Turkish version did not report any ceiling effect of the IGDS9-SF but reported a 9% floor effect. A study on the Malay

version of the IGDS9-SF reported a 1.7% ceiling effect and a 0.1% floor effect. Except for the studies on the traditional and simplified Chinese versions, the distribution of scoring on the IGDS9-SF in the related studies indicated acceptable floor and ceiling effects (ie, an effect less than 15%) [46].

IT

IT was reported in eight studies and ranged from 0.342 to 0.86 (Table 1). The items with the lowest IT in the Spanish version of the IGDS9-SF were item 7 ("Have you deceived any of your significant others because of the amount of your gaming activity?") and item 8 ("Do you play to temporarily escape or relieve a negative mood?") (IT=.47), while the lowest IT was in the traditional Chinese (Taiwan) version, which was found for item 5 ("Have you lost interest in previous hobbies because of your engagement with gaming?") and item 7 ("Have you deceived any of your significant others because the amount of your gaming activity?") (IT=.74).

Two studies on the traditional Chinese (Hong Kong) and Turkish versions also had the lowest IT correlation on item 7 (IT=.663-.68). Item 9 ("Have you jeopardized an important relationship, job, or educational or career opportunity because of your gaming activity?") demonstrated the lowest IT in the South American Portuguese, traditional Chinese (Hong Kong), and simplified Chinese versions (IT=.342-.55), while the item with the lowest IT in the Persian version was item 4 ("Do you systematically fail when trying to control or cease your gaming activity?") (IT=.54).

Discussion

Principal Findings

The quality of methodology and the quality of statistical findings for each psychometric property of the IGDS9-SF were evaluated for all 21 eligible studies. The evidence regarding the psychometric properties of the IGDS9-SF was summarized based on the existing evidence. In general, the IGDS9-SF demonstrated good internal consistency, although some items did not have satisfactory IT, especially items 7 (Spanish, Chinese, South American Portuguese, Turkish and Persian), 8 (Spanish, Chinese, and South American Portuguese), and 9 (Chinese, South American Portuguese, Turkish, and Persian). However, the IGDS9-SF has excellent criterion validity, as evidenced by the strong correlation with the 27-item IGDS [20]. Furthermore, the IGDS9-SF can distinguish different subgroups when assessing disordered gaming, with measurement invariance supported across both gender and age. In addition, the structure validity of the IGDS9-SF has been verified as a unidimensional structure among all 21 studies. Based on the aforementioned findings, the IGDS9-SF can best be used for clinicians to assess an individual's IGD severity level.

Internal Consistency and IT

All nine items of the IGDS9-SF demonstrated satisfactory IT (greater than 0.4) [22,23,25,30,32,48], except for items 7, 8, and 9 in one specific study [28]. Several plausible reasons for the unsatisfactory ITs are discussed here. The deception criterion (ie, item 7) may be influenced by the living status and external attitudes (eg, parents' perception) on gaming behaviors [55,56] and therefore may be less associated with other IGDS9-SF items. According to previous findings, parents of a child living with them and indulging in excessive online gaming are usually aware of their child's problematic gaming behavior and thus deception may not be a central IGD symptom [40]. This idea is supported by studies showing that the deception criterion is not associated with higher IGD severity [10]. If the severity of IGD in the population included is not high, a low IT may be reported. Furthermore, the deception criterion has shown low diagnostic accuracy in discriminating disordered gamers from nonproblematic gamers in previous studies due to the acceptability and accessibility of the internet [55] as well as the age of population investigated [40].

In relation to compromising occupation/education or a significant relationship due to the involvement in gaming (ie, item 9), a simpler criterion that only covers occupational/educational aspects is more appropriate than concurrently conflating relationships and occupational/educational loss [56]. For instance, the item may have greater clinical utility if it is split into two (ie, one reflecting the compromising of occupation/education and the other reflecting the compromising of significant relationships). Moreover, because the majority of the studies reviewed recruited individuals in full-time education, these particular participants might not feel that they have compromised their education, when responding to this item.

The behavior of escaping from adverse moods by gaming (ie, item 8) is reported as a criterion that is unable to differentiate between disordered and nondisordered gamers [57]. Internet use as a form of escapism could be considered as a coping strategy of a nondisordered gamer as much as that of a disordered gamer [58]. It is further supported by other studies showing that item 8 has the lowest specificity and diagnostic accuracy in discriminating disordered gamers and nondisordered gamers [20,40].

Validity

The IGDS9-SF demonstrated a strong correlation with the 27-item IGDS [48], showing good criterion validity as expected. Like the IGDS9-SF, the 27-item IGDS was developed in accordance with the nine IGD criteria in the DSM-5. Therefore, the criterion validity of the IGDS9-SF is illustrated by the strong correlation observed with another psychometric test developed using the same DSM-5 criteria (ie, the 27-item IGDS) to measure IGD.

In addition to criterion validity, the concurrent validity of the IGDS9-SF was supported by the positive correlation between the IGDS9-SF score and the hours spent on online activities (eg, smartphone use, social media use, and gaming), with the correlations between the IGDS9-SF and gaming frequency being the most significant [19,22,24,25,28-31]. This implies that the frequency of gaming is positively associated with the severity of IGD.

In addition, a positive association was found between IGDS9-SF scores and three psychological symptoms (ie, depression, anxiety, and stress) [25,30]. This finding indicates that individuals with IGD might have a higher chance of suffering

from mental distress, as evidenced by recent longitudinal findings [59]. One of the potential reasons underpinning this phenomenon might be the fact that excessive use of the internet can result in social withdrawal when an individual engages in gaming with features encouraging social disconnection (eg, features that increase social comparison and rumination [60,61]), which may lead to poorer psychological well-being among a minority of individuals [62]. This idea is also aligned with the time displacement hypothesis, which posits that the development and maintenance of social relationships require continuous commitment toward understanding, learning about other individuals, and communicating with them [63].

Although the IGDS9-SF is positively associated with other similar measures (eg, Mobile Phone-Related Experiences Questionnaire [CERM] and Online Gambling Disorder Questionnaire [OGD-Q]) [22-25,32,52], it only shows low-to-moderate correlation in most of the studies reviewed (r=.06-.440) [22-24,32]. The aforementioned measurements mainly focus on the conflicts and problems of smartphone use, social media use, and online gambling. These results show that the association of IGD with online gaming is greater than that with social media use or mobile phone use, which reaffirms that the IGDS9-SF has adequate psychometric properties when assessing the specific construct of IGD. One study showed an expected negative correlation between the IGDS9-SF and KIDSCREEN-27 [32]—an instrument assessing the quality of life—further supporting the notion that more severe IGD levels will likely lead to a poorer quality of life, such as physical or psychological well-being (eg, [59,64]).

Factor Structure and Measurement Invariance

All 21 studies demonstrated the unidimensional structure of the IGDS9-SF and were supported by both confirmatory factor analysis and Rasch analysis. All factor loadings reported across the studies reviewed were satisfactory. Moreover, measurement invariance across gender [22,25,30,52] and age [32,52] was fully supported in most studies. However, measurement invariance across countries [26,50,51] and time spent gaming [30] was only partially supported. This phenomenon might be associated with the key cultural differences between the individualistic and collectivist countries. Research also suggests that cultural differences might be a plausible reason affecting the response patterns of psychometric instruments [65].

Item Response Theory and Rasch Analysis

In addition to the most commonly used classical test theory (CTT) in the included studies, only two included studies used item response theory (or Rasch analysis) [31,49] for a better understanding of the psychometric properties of the IGD criteria beyond CTT. There is currently insufficient evidence on Rasch findings regarding the IGDS9-SF, given the scarcity of studies, and therefore future studies using Rasch-based models are required to provide further psychometric information about the IGDS9-SF.

Strengths and Limitations

This systematic review had several strengths. First, a comprehensive search strategy was adopted to identify and evaluate potential studies. Second, the review evaluated studies with an accepted set of criteria, that is, both methodological quality (COSMIN) and statistical quality. Moreover, all 21 eligible studies included for evaluation in the systematic review had the strength of having good sample sizes (ie, N>100) based on the COSMIN risk-of-bias checklist. Therefore, this review took advantage of the good sample sizes of the studies evaluated to increase the accuracy in the summarized psychometric properties of the IGDS9-SF.

This review also had potential limitations. In the study selection stage, the language was limited to only English and each included study was assessed by only two authors (and these authors were not necessarily the same two for each paper). More specifically, although consensus was achieved in accordance with COSMIN guidelines, the authors who rated the quality of each study may have used slightly varied rating styles. Nevertheless, the COSMIN guidelines provided clear and concrete instructions for the evaluation of the studies, so the use of different authors in the study evaluation made it unlikely that this led to serious bias. Additionally, evaluation of the results was verified by the corresponding author, who is an expert psychometrician, which further minimized the possibility of evaluation bias. Some psychometric properties of the IGDS9-SF (eg, responsiveness) were not evaluated in any of the eligible studies. More specifically, responsiveness as an important psychometric property to understand whether an instrument is sensitive in detecting change was not carried out in any of the 21 studies. Further studies investigating responsiveness are therefore required. Finally, this systematic review did not use the Google Scholar database to supplement the literature search concerning the psychometric properties of the IGDS9-SF. Therefore, some studies may have been missed in this systematic review.

Implications for Clinical Practice and Future Directions

The IGDS9-SF was designed both as a brief screening tool to assist clinicians in assessing IGD severity as well as for use in epidemiological studies [19]. Given that disordered gaming has been recognized by both the APA [13] and the World Health Organization [66], clinicians and other mental health professionals need to have a robust psychometric tool to assess IGD. Therefore, this review provides strong evidence enabling practitioners to better understand the psychometric features of the IGDS9-SF.

In sum, the IGDS9-SF is a time-efficient and psychometrically sound tool that can help clinicians screen for potential patients who may need more detailed clinical evaluation regarding their gaming behaviors. The use of the IGDS9-SF may also help in busy clinical settings by saving evaluation time in diagnosing and assessing IGD in patients. Indeed, a recent review of all screening instruments currently available for disordererd gaming (n=32) reported that the IGDS9-SF was among the best tools with regard to its psychometric properties when compared with all the others similar tools [21]. In terms of administration, the IGDS9-SF takes only a few minutes for individuals to complete it by themselves. This can substantially reduce the time for clinicians to diagnose and assess IGD symptoms. To this end, clinical studies have suggested adopting a cutoff of 32 points

when diagnosing between disordered and nondisordered gamers [67]. Moreover, due to the robust psychometric properties reported across 15 language versions, the IGDS9-SF can also be used in worldwide epidemiological studies examining IGD [40]. Given its brevity and multilanguage capability, the IGDS9-SF can be used in collecting valid and reliable data concerning IGD symptoms in a practical and efficient way that minimizes survey fatigue and can also be used for cross-cultural comparisons in different countries to help advance our understanding of IGD.

Conclusion

This systematic review summarized and reviewed evidence from various populations regarding the IGDS9-SF with regard to its structural validity, internal consistency, cross-cultural validity/measurement invariance, reliability, measurement error, criterion validity, convergent validity, and discriminative or known-group validity. Overall, there was strong evidence demonstrating that the IGDS9-SF has good internal consistency and excellent criterion validity for wide-ranging populations. Responsiveness and reliability as two important psychometric properties of the IGDS9-SF require more evidence because few studies have evaluated them. Regarding the psychometric evidence of different language versions, it was found that all versions of the IGDS9-SF present strong psychometric properties, except for the concurrent validity of the traditional Chinese (Hong Kong and Taiwan), simplified Chinese, Polish, and Persian versions. As such, the psychometric properties of the IGDS9-SF are robust, and the findings regarding its cross-cultural psychometric features in different language versions have been further clarified. Therefore, the IGDS9-SF can be used widely within clinical and research settings.

Acknowledgments

The authors acknowledge financial support from the Ministry of Science and Technology, Taiwan (MOST 110-2410-H-006-115), the 2021 Southeast and South Asia and Taiwan Universities Joint Research Scheme (NCKU 31), and the Higher Education Sprout Project, Ministry of Education to the Headquarters of University Advancement at National Cheng Kung University (NCKU). MDG and HMP are the developers of the IGDS9-SF.

Authors' Contributions

LYJP, TYJC, SWTM, LYN, YLEW, C-YL, and C-WC designed the study and wrote the protocol. SWTM and YLEW conducted literature searches and screened the eligibility. TYJC, LYN, and LYJP were involved in the discussion of eligibility. TYJC, SWTM, LYN, LYJP, and YLEW provided summaries of previous research studies. LYJP, TYJC, SWTM, LYN, YLEW, C-YL, and C-WC wrote the first draft of the manuscript. C-YL and AHP registered the systematic review. AHP, MDG, and HMP critically reviewed the manuscript. HMP helped in drafting and revising the final version of the manuscript. All authors have approved the final manuscript.

Conflicts of Interest

None declared. The Internet Gaming Disorder Scale–Short-Form is not a commercial instrument. Therefore, both MDG and HMP have no conflicts of interest.

Multimedia Appendix 1
Database search.
[DOCX File , 18 KB - jmir_v23i10e26821_app1.docx]

Multimedia Appendix 2
Results of studies on measurement properties.
[DOCX File , 63 KB - jmir_v23i10e26821_app2.docx]

References

1. Lee C, Tsai M, Liang Y, Strong C, Lin C. Adolescent internet use and social integration: analysis from a longitudinal cohort survey. Journal of Adolescent Health 2016 Feb;58(2):S73-S74. [doi: 10.1016/j.jadohealth.2015.10.159]
2. Tsai M, Strong C, Chen W, Lee C, Lin C. Longitudinal impacts of pubertal timing and weight status on adolescent internet use: analysis from a cohort study of Taiwanese youths. PLoS One 2018 May 24;13(5):e0197860 [FREE Full text] [doi: 10.1371/journal.pone.0197860] [Medline: 29795649]
3. Agbaria Q. Internet addiction and aggression: the mediating roles of self-control and positive affect. Int J Ment Health Addiction 2020 Jan 14;19(4):1227-1242 [FREE Full text] [doi: 10.1007/s11469-019-00220-z]
4. Gu M. A longitudinal study of daily hassles, internet expectancy, self-control, and problematic internet use in Chinese adolescents: a moderated mediation model. Personality and Individual Differences 2020 Jan;152:109571 [FREE Full text] [doi: 10.1016/j.paid.2019.109571]

5. Pontes HM, Taylor M, Stavropoulos V. Beyond "Facebook addiction": the role of cognitive-related factors and psychiatric distress in social networking site addiction. Cyberpsychol Behav Soc Netw 2018 Apr;21(4):240-247. [doi: 10.1089/cyber.2017.0609] [Medline: 29589972]
6. Mandryk RL, Frommel J, Armstrong A, Johnson D. How passion for playing World of Warcraft predicts in-game social capital, loneliness, and wellbeing. Front Psychol 2020 Sep 15;11:2165 [FREE Full text] [doi: 10.3389/fpsyg.2020.02165] [Medline: 33071843]
7. Stevens MW, Dorstyn D, Delfabbro PH, King DL. Global prevalence of gaming disorder: a systematic review and meta-analysis. Aust N Z J Psychiatry 2021 Jun 07;55(6):553-568. [doi: 10.1177/0004867420962851] [Medline: 33028074]
8. Wong HY, Mo HY, Potenza MN, Chan MNM, Lau WM, Chui TK, et al. Relationships between severity of internet gaming disorder, severity of problematic social media use, sleep quality and psychological distress. Int J Environ Res Public Health 2020 Mar 13;17(6):1879 [FREE Full text] [doi: 10.3390/ijerph17061879] [Medline: 32183188]
9. Griffiths M, Király O, Pontes H, Demetrovics Z. An overview of problematic gaming. In: Aboujaoude E, Starcevic V, editors. Mental Health in the Digital Age. Grave Dangers, Great Promise. Oxford: Oxford University Press; 2015:A-45.
10. Király O, Sleczka P, Pontes HM, Urbán R, Griffiths MD, Demetrovics Z. Validation of the Ten-Item Internet Gaming Disorder Test (IGDT-10) and evaluation of the nine DSM-5 internet gaming disorder criteria. Addict Behav 2017 Jan;64:253-260. [doi: 10.1016/j.addbeh.2015.11.005] [Medline: 26632194]
11. Jo YS, Bhang SY, Choi JS, Lee HK, Lee SY, Kweon Y. Clinical Characteristics of diagnosis for internet gaming disorder: comparison of DSM-5 IGD and ICD-11 GD diagnosis. J Clin Med 2019 Jun 28;8(7):945 [FREE Full text] [doi: 10.3390/jcm8070945] [Medline: 31261841]
12. Young KS. Internet addiction: the emergence of a new clinical disorder. CyberPsychol Behav 1998 Jan;1(3):237-244. [doi: 10.1089/cpb.1998.1.237]
13. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). Arlington, VA: American Psychiatric Publication; 2013.
14. Chen JH, Zhang MX, Ko C, Tong KK, Yu SM, Sou EKL, et al. The development of a screening tool for Chinese disordered gamers: the Chinese Internet Gaming Disorder Checklist (C-IGDC). Int J Environ Res Public Health 2020 May 14;17(10):3412 [FREE Full text] [doi: 10.3390/ijerph17103412] [Medline: 32422914]
15. Jo S, Yim HW, Lee H, Lee HC, Choi J, Baek K. The Internet Game Use-Elicited Symptom Screen proved to be a valid tool for adolescents aged 10-19 years. Acta Paediatr 2018 Mar 27;107(3):511-516. [doi: 10.1111/apa.14087] [Medline: 28940637]
16. Pan Y, Chiu Y, Lin Y. Development of the Problematic Mobile Gaming Questionnaire and prevalence of mobile gaming addiction among adolescents in Taiwan. Cyberpsychol Behav Soc Netw 2019 Oct 01;22(10):662-669. [doi: 10.1089/cyber.2019.0085] [Medline: 31613156]
17. Pontes HM. Current practices in the clinical and psychometric assessment of internet gaming disorder in the era of the DSM-5: a mini review of existing assessment tools. Ment Health Addict Res 2016;1(1):18-19. [doi: 10.15761/mhar.1000105]
18. Pontes HM, Király O, Demetrovics Z, Griffiths MD. The conceptualisation and measurement of DSM-5 internet gaming disorder: the development of the IGD-20 test. PLoS One 2014 Oct 14;9(10):e110137 [FREE Full text] [doi: 10.1371/journal.pone.0110137] [Medline: 25313515]
19. Pontes HM, Griffiths MD. Measuring DSM-5 internet gaming disorder: development and validation of a short psychometric scale. Computers in Human Behavior 2015 Apr;45:137-143. [doi: 10.1016/j.chb.2014.12.006]
20. Lemmens JS, Valkenburg PM, Gentile DA. The Internet Gaming Disorder Scale. Psychol Assess 2015 Jun;27(2):567-582. [doi: 10.1037/pas0000062] [Medline: 25558970]
21. King DL, Chamberlain SR, Carragher N, Billieux J, Stein D, Mueller K, et al. Screening and assessment tools for gaming disorder: a comprehensive systematic review. Clin Psychol Rev 2020 Apr;77:101831 [FREE Full text] [doi: 10.1016/j.cpr.2020.101831] [Medline: 32143109]
22. Leung H, Pakpour AH, Strong C, Lin Y, Tsai M, Griffiths MD, et al. Measurement invariance across young adults from Hong Kong and Taiwan among three internet-related addiction scales: Bergen Social Media Addiction Scale (BSMAS), Smartphone Application-Based Addiction Scale (SABAS), and Internet Gaming Disorder Scale-Short Form (IGDS-SF9) (Study Part A). Addict Behav 2020 Feb;101:105969. [doi: 10.1016/j.addbeh.2019.04.027] [Medline: 31078344]
23. Yam C, Pakpour AH, Griffiths MD, Yau W, Lo CM, Ng JMT, et al. Psychometric testing of three Chinese online-related addictive behavior instruments among Hong Kong University students. Psychiatr Q 2019 Mar 16;90(1):117-128. [doi: 10.1007/s11126-018-9610-7] [Medline: 30328020]
24. Chen I, Strong C, Lin Y, Tsai M, Leung H, Lin C, et al. Time invariance of three ultra-brief internet-related instruments: Smartphone Application-Based Addiction Scale (SABAS), Bergen Social Media Addiction Scale (BSMAS), and the nine-item Internet Gaming Disorder Scale-Short Form (IGDS-SF9) (study part B). Addict Behav 2020 Feb;101:105960. [doi: 10.1016/j.addbeh.2019.04.018] [Medline: 31072648]
25. Chen I, Ahorsu DK, Pakpour AH, Griffiths MD, Lin C, Chen C. Psychometric properties of three simplified Chinese online-related addictive behavior instruments among mainland Chinese primary school students. Front Psychiatry 2020 Sep 3;11:875 [FREE Full text] [doi: 10.3389/fpsyt.2020.00875] [Medline: 33101070]

26. de Palo V, Monacis L, Sinatra M, Griffiths MD, Pontes H, Petro M, et al. Measurement invariance of the nine-item Internet Gaming Disorder Scale (IGDS9-Sf) across Albania, USA, UK, and Italy. Int J Ment Health Addiction 2018 Apr 30;17(4):935-946. [doi: 10.1007/s11469-018-9925-5]
27. Pontes HM, Griffiths MD. Portuguese validation of the Internet Gaming Disorder Scale-Short-Form. Cyberpsychol Behav Soc Netw 2016 Apr;19(4):288-293. [doi: 10.1089/cyber.2015.0605] [Medline: 26974853]
28. Severo RB, Barbosa APPN, Fouchy DRC, Coelho FMDC, Pinheiro RT, de Figueiredo VLM, et al. Development and psychometric validation of Internet Gaming Disorder Scale-Short-Form (IGDS9-SF) in a Brazilian sample. Addict Behav 2020 Apr;103:106191. [doi: 10.1016/j.addbeh.2019.106191] [Medline: 31887719]
29. Pontes HM, Macur M, Griffiths MD. Internet gaming disorder among Slovenian primary schoolchildren: findings from a nationally representative sample of adolescents. J Behav Addict 2016 Jun;5(2):304-310 [FREE Full text] [doi: 10.1556/2006.5.2016.042] [Medline: 27363464]
30. Wu T, Lin C, Årestedt K, Griffiths MD, Broström A, Pakpour AH. Psychometric validation of the Persian nine-item Internet Gaming Disorder Scale-Short Form: does gender and hours spent online gaming affect the interpretations of item descriptions? J Behav Addict 2017 Jun 01;6(2):256-263 [FREE Full text] [doi: 10.1556/2006.6.2017.025] [Medline: 28571474]
31. Schivinski B, Brzozowska-Woś M, Buchanan EM, Griffiths MD, Pontes HM. Psychometric assessment of the Internet Gaming Disorder diagnostic criteria: an item response theory study. Addict Behav Rep 2018 Dec;8:176-184 [FREE Full text] [doi: 10.1016/j.abrep.2018.06.004] [Medline: 30505924]
32. Beranuy M, Machimbarrena JM, Vega-Osés MA, Carbonell X, Griffiths MD, Pontes HM, et al. Spanish validation of the Internet Gaming Disorder Scale-Short Form (IGDS9-SF): prevalence and relationship with online gambling and quality of life. Int J Environ Res Public Health 2020 Feb 28;17(5):1562 [FREE Full text] [doi: 10.3390/ijerph17051562] [Medline: 32121280]
33. Maldonado-Murciano L, M Pontes H, Griffiths MD, Barrios M, Gómez-Benito J, Guilera G. The Spanish version of the Internet Gaming Disorder Scale-Short Form (IGDS9-SF): further examination using item response theory. Int J Environ Res Public Health 2020 Sep 28;17(19):7111 [FREE Full text] [doi: 10.3390/ijerph17197111] [Medline: 32998358]
34. Sánchez-Iglesias I, Bernaldo-de-Quirós M, Labrador FJ, Estupiñá Puig FJ, Labrador M, Fernández-Arias I. Spanish validation and scoring of the Internet Gaming Disorder Scale - Short-Form (IGDS9-SF). Span. J. Psychol 2020 Jun 19;23:1562. [doi: 10.1017/sjp.2020.26]
35. Arıcak O, Dinç M, Yay M, Griffiths M. Adapting the short form of the Internet Gaming Disorder Scale into Turkish: validity and reliability. Addicta 2018;5:615.
36. Montag C, Schivinski B, Sariyska R, Kannen C, Demetrovics Z, Pontes HM. Psychopathological symptoms and gaming motives in disordered gaming-a psychometric comparison between the WHO and APA diagnostic frameworks. J Clin Med 2019 Oct 15;8(10):1691 [FREE Full text] [doi: 10.3390/jcm8101691] [Medline: 31618950]
37. Suchá J, Dolej? M, Pipová H, Maierová E, Cakirpaloglu P. Hraní digitálních her českými adolescenty. Olomouc, Czech Republic: Univerzita Palackého v Olomouci; 2018.
38. T'ng ST, Pau K. Validation of a translated internet gaming disorder scale (short form) and measurement invariance across sex groups in Malaysian samples. Curr Psychol 2020 Feb 22;39(6):1977-1989. [doi: 10.1007/s12144-020-00668-6]
39. Kim B, Ko H. Psychometric properties of the nine-item Korean Internet Gaming Disorder Scale: Short Form. Cyberpsychol Behav Soc Netw 2020 Dec 01;23(12):854-859. [doi: 10.1089/cyber.2020.0227] [Medline: 32960098]
40. Ko C, Yen J, Chen S, Wang P, Chen C, Yen C. Evaluation of the diagnostic criteria of internet gaming disorder in the DSM-5 among young adults in Taiwan. J Psychiatr Res 2014 Jun;53:103-110. [doi: 10.1016/j.jpsychires.2014.02.008] [Medline: 24581573]
41. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
42. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res 2018 May 12;27(5):1147-1157 [FREE Full text] [doi: 10.1007/s11136-018-1798-3] [Medline: 29435801]
43. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. Qual Life Res 2018 May 19;27(5):1171-1179 [FREE Full text] [doi: 10.1007/s11136-017-1765-4] [Medline: 29260445]
44. Lin C, Luh W, Cheng C, Yang A, Su C, Ma H. Measurement equivalence across child self-reports and parent-proxy reports in the Chinese version of the pediatric quality of life inventory version 4.0. Child Psychiatry Hum Dev 2013 Oct 15;44(5):583-590. [doi: 10.1007/s10578-012-0352-8] [Medline: 23242709]
45. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
46. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42. [doi: 10.1016/j.jclinepi.2006.03.012] [Medline: 17161752]

47. Prinsen CAC, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, et al. How to select outcome measurement instruments for outcomes included in a "core outcome set": a practical guideline. Trials 2016 Sep 13;17(1):449 [FREE Full text] [doi: 10.1186/s13063-016-1555-2] [Medline: 27618914]
48. Evren C, Dalbudak E, Topcu M, Kutlu N, Evren B. The psychometric properties of the Turkish version of the Internet Gaming Disorder Scale. Dusunen Adam 2017 Dec 27:316-324. [doi: 10.5350/dajpn2017300405]
49. Gomez R, Stavropoulos V, Beard C, Pontes HM. Item response theory analysis of the recoded Internet Gaming Disorder Scale-Short-Form (IGDS9-SF). Int J Ment Health Addiction 2018 Mar 26;17(4):859-879. [doi: 10.1007/s11469-018-9890-z]
50. Pontes HM, Stavropoulos V, Griffiths MD. Measurement invariance of the Internet Gaming Disorder Scale-Short-Form (IGDS9-SF) between the United States of America, India and the United Kingdom. Psychiatry Res 2017 Nov;257:472-478. [doi: 10.1016/j.psychres.2017.08.013] [Medline: 28837939]
51. Stavropoulos V, Beard C, Griffiths MD, Buleigh T, Gomez R, Pontes HM. Measurement invariance of the Internet Gaming Disorder Scale-Short-Form (IGDS9-SF) between Australia, the USA, and the UK. Int J Ment Health Addict 2018 Jul 24;16(2):377-392 [FREE Full text] [doi: 10.1007/s11469-017-9786-3] [Medline: 29670499]
52. Monacis L, Palo VD, Griffiths MD, Sinatra M. Validation of the Internet Gaming Disorder Scale - Short-Form (IGDS9-SF) in an Italian-speaking sample. J Behav Addict 2016 Dec;5(4):683-690 [FREE Full text] [doi: 10.1556/2006.5.2016.083] [Medline: 27876422]
53. Pontes H. Internet Gaming Disorder Scale-Short-Form (IGDS9-SF). URL: https://www.halleypontes.com/igds9sf [accessed 2020-10-11]
54. Stavropoulos V, Bamford L, Beard C, Gomez R, Griffiths MD. Test-retest measurement invariance of the nine-item Internet Gaming Disorder Scale in two countries: a preliminary longitudinal study. Int J Ment Health Addiction 2019 May 22:1-18. [doi: 10.1007/s11469-019-00099-w]
55. Tao R, Huang X, Wang J, Zhang H, Zhang Y, Li M. Proposed diagnostic criteria for internet addiction. Addiction 2010 Mar;105(3):556-564. [doi: 10.1111/j.1360-0443.2009.02828.x] [Medline: 20403001]
56. Griffiths MD, van Rooij AJ, Kardefelt-Winther D, Starcevic V, Király O, Pallesen S, et al. Working towards an international consensus on criteria for assessing internet gaming disorder: a critical commentary on Petry et al. (2014). Addiction 2016 Jan 16;111(1):167-175 [FREE Full text] [doi: 10.1111/add.13057] [Medline: 26669530]
57. Kardefelt-Winther D. A critical account of DSM-5 criteria for internet gaming disorder. Addiction Research & Theory 2014 Jul 03;23(2):93-98. [doi: 10.3109/16066359.2014.935350]
58. Kardefelt-Winther D. A conceptual and methodological critique of internet addiction research: towards a model of compensatory internet use. Computers in Human Behavior 2014 Feb;31:351-354. [doi: 10.1016/j.chb.2013.10.059]
59. Teng Z, Pontes HM, Nie Q, Xiang G, Griffiths MD, Guo C. Internet gaming disorder and psychosocial well-being: a longitudinal study of older-aged adolescents and emerging adults. Addict Behav 2020 Nov;110:106530. [doi: 10.1016/j.addbeh.2020.106530] [Medline: 32683173]
60. Baker DA, Algorta GP. The relationship between online social networking and depression: a systematic review of quantitative studies. Cyberpsychol Behav Soc Netw 2016 Nov;19(11):638-648. [doi: 10.1089/cyber.2016.0206] [Medline: 27732062]
61. Frye C, Ly L, Murrieta J, Sun L, Cochancela C, Eikey E. Perceived use and effects of social media for 1 to 2.5 generation immigrant college students with depression: results from a mixed methods survey. Lect Notes Comput Sci 2020:130-150. [doi: 10.1007/978-3-030-43687-2_11]
62. Kraut R, Patterson M, Lundmark V, Kiesler S, Mukophadhyay T, Scherlis W. Internet paradox: a social technology that reduces social involvement and psychological well-being? Am Psychol 1998;53(9):1017-1031. [doi: 10.1037/0003-066x.53.9.1017]
63. Dredge R, Chen S. Chinese online gamers versus nongamers: a difference in social media use and associated well‐being and relational outcomes? Psychol Schs 2020 Jul 10;57(9):1457-1474. [doi: 10.1002/pits.22418]
64. Männikkö N, Ruotsalainen H, Miettunen J, Pontes HM, Kääriäinen M. Problematic gaming behaviour and health-related outcomes: a systematic review and meta-analysis. J Health Psychol 2020 Jan 01;25(1):67-81. [doi: 10.1177/1359105317740414] [Medline: 29192524]
65. Walford G, Tucker E, Viswanathan M. The SAGE Handbook of Measurement. Singapore: Sage Publications; 2010.
66. Humphreys G. Sharpening the focus on gaming disorder. Bull World Health Organ 2019 Jun 01;97(6):382-383. [doi: 10.2471/blt.19.020619]
67. Qin L, Cheng L, Hu M, Liu Q, Tong J, Hao W, et al. Clarification of the cut-off score for nine-item Internet Gaming Disorder Scale-Short Form (IGDS9-SF) in a Chinese context. Front Psychiatry 2020 May 25;11:470 [FREE Full text] [doi: 10.3389/fpsyt.2020.00470] [Medline: 32528331]

Abbreviations

APA: American Psychiatric Association
CERM: Mobile Phone-Related Experiences Questionnaire
COSMIN: Consensus-based Standards for the selection of health status Measurement Instrument
CTT: classical test theory

DIF: differential item functioning
DSM-5: *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition*
ESL: Electronic Sports League
ICD-11: *International Classification of Diseases*, 11th revision
IGD: internet gaming disorder
IGDS9-SF: Internet Gaming Disorder Scale–Short-Form
IGDT: Internet Gaming Disorder Test
IT: item-total correlation
MIC: minimal important change
MMO: massively multiplayer online
MOBA: multiplayer online battle arena
OGD-Q: Online Gambling Disorder Questionnaire
PROM: patient-reported outcome measure
RMSEA: root-mean-square error of approximation
SEM: standard error of measurement
SRMR: standardized root-mean-square residual

Edited by R Kukafka, G Eysenbach; submitted 29.12.20; peer-reviewed by J Borghouts, Z Zrubka; comments to author 13.02.21; revised version received 24.02.21; accepted 16.09.21; published 18.10.21.

Please cite as:
Poon LYJ, Tsang HWH, Chan TYJ, Man SWT, Ng LY, Wong YLE, Lin CY, Chien CW, Griffiths MD, Pontes HM, Pakpour AH
Psychometric Properties of the Internet Gaming Disorder Scale–Short-Form (IGDS9-SF): Systematic Review
J Med Internet Res 2021;23(10):e26821
URL: https://www.jmir.org/2021/10/e26821
doi:10.2196/26821
PMID:34661543

Original Paper

Detrimental Effects of Online Pro–Eating Disorder Communities on Weight Loss and Desired Weight: Longitudinal Observational Study

Johannes Feldhege[1], MSc; Markus Moessner[1], PhD; Stephanie Bauer[1], PhD

Centre for Psychotherapy Research, University Hospital Heidelberg, Heidelberg, Germany

Corresponding Author:
Johannes Feldhege, MSc
Centre for Psychotherapy Research
University Hospital Heidelberg
Bergheimer Strasse 54
Heidelberg, 69115
Germany
Phone: 49 622156 ext 7876
Fax: 49 6221567350
Email: johannes.feldhege@med.uni-heidelberg.de

Abstract

Background: Online pro–eating disorder (pro-ED) communities are considered harmful because of their detrimental effects on their users' body dissatisfaction, dieting, and help seeking. To date, it is unknown to which extent participation in pro-ED communities affects users' body weight and desired weight loss.

Objective: This study aims to investigate the changes in the current and desired body weight of users of a pro-ED community (r/proed) on the social media website Reddit over time.

Methods: Data on 1170 users and the unsolicited weight information they shared with the pro-ED community were collected over a period of 15 months. Linear growth models were used to model changes in the users' current and desired BMI over time.

Results: Both current and desired BMI decreased over time, with a predicted rate of 0.087 and 0.015 BMI points per week, respectively. Weight loss was moderated by the users' activity level in the community, with more active users losing more weight. Users with a higher baseline BMI experienced greater weight loss, but even users with a very low baseline weight (BMI <17 kg/m^2) lost weight during their participation. In addition, users decreased their desired weight over time, with many pursuing extremely low, unrealistic weight goals. Changes in the desired weight were moderated by the baseline current BMI and baseline desired BMI. Users with higher desired weight and lower body weight at baseline decreased their desired weight more over time.

Conclusions: This is the first study to demonstrate the detrimental effects of pro-ED communities in a longitudinal study based on a large data set of user-generated online data. The results extend the literature detailing the harmful effects of online pro-ED communities by showing users' weight loss, decreases in desired weight, and that higher activity levels lead to greater weight loss. Users could be driven to pursue very low, unrealistic weight loss goals by images of very thin bodies presented in these communities.

(J Med Internet Res 2021;23(10):e27153) doi:10.2196/27153

KEYWORDS

pro-eating disorder communities; weight loss; body weight; social media; linear growth models; eating disorders; pro-ED; Reddit

Introduction

Background

Eating disorders (EDs) are severe mental disorders. Many individuals with EDs strive to lose weight to achieve a *thin ideal*, that is, an ideally slim body type portrayed as attractive in the media. Low body weight and marked weight loss contribute to the emergence of physical conditions and elevated mortality rates observed in individuals with ED [1-3].

Online pro-ED communities, also called pro-ana and pro-mia communities for anorexia nervosa and bulimia nervosa, respectively, are widespread over the internet and attract a large

number of vulnerable or impaired users. These communities exist on social media platforms [4] and as self-hosted websites and blogs [5]. They have been identified as harmful to their users: participation leads to increases in ED behaviors while discouraging help seeking and recovery, thereby prolonging the duration of the mental disorder [6,7]. This has led to the ban or removal of pro-ED content and communities from social media platforms [8,9] and, in at least one instance, to the passing of a law that fines websites for hosting content that promotes excessive thinness and dieting [10]. Positive impacts of participating in these communities, such as receiving social support from similar others and feeling less alone with a stigmatized condition, have also been noted [11,12]. However, it has been suggested that this support is conditional on following unhealthy group norms and that it reinforces ED thoughts and behaviors [7,13,14].

One of the central mechanisms through which ED behaviors are promoted in these communities is *tips and tricks* for extreme weight loss, which often include behaviors such as the use of laxatives or diet pills, purging, fasting, and extreme dietary or exercise regimes [15-17]. This is potentially harmful, as adolescents report learning about new weight loss methods on pro-ED websites and using them afterward [5,18]. In addition, exposure to pro-ED websites is associated with increases in dieting [19].

But literature on the body weight and weight loss of users of pro-ED communities is scarce. Users show a variety of different body weights, with the majority having a weight in the healthy weight category as defined by the BMI classification of the World Health Organization (BMI: weight in kg/height in m^2; underweight BMI: <18.5 kg/m^2, healthy weight BMI: 18.5-24.99 kg/m^2, overweight BMI: 25.0-29.99 kg/m^2, and obese BMI ≥30 kg/m^2 [20]) as well as sizable groups that have body weights in the underweight, overweight, and obese ranges [5]. Body dissatisfaction and concerns about body shape and weight tend to be high in users of pro-ED websites, which could lead them to pay close attention to their current weight and weight changes [12,19]. Many users of pro-ED communities use apps and web-based tools to track their body weight and food intake [21-23]. Pro-ED communities provide similar functions as weight tracking apps with *check-ins* or *status updates*, allowing users to share personal information, such as their current body weight, height, age, or gender, with the online community [4,12,24].

Individuals with EDs report lower desired weights than individuals without EDs [25]. Desired weights lower than the current weight or lower than the recommended healthy weight are associated with unhealthy behaviors in adolescents [26] and elevated ED severity in individuals with ED [27-29]. Specific features of pro-ED communities could encourage their users to lower their desired weight. They often compare themselves with other users in terms of weight and compete to be as thin as possible [7]. Another common feature of pro-ED communities, *thinspiration* content, for example, images of unrealistically thin female body shapes, is another source for comparison [15]. Exposure to thinspiration and physical appearance comparisons are positively related to ED symptoms in individuals with ED [30].

An individual's desired weight can change over time [31,32]. Frequent changes are associated with a number of strategies to control weight and increases in physical activity [33]. Users could be encouraged to adopt lower and lower desired weights during their participation in pro-ED communities to align themselves more closely with the propagated thin ideal.

The level of user activity in a pro-ED community is positively associated with disordered eating and ED-related impairment [5]. In weight loss communities, higher levels of activity are related to greater weight loss over time [34-36].

Objectives

Although the detrimental effects of pro-ED communities are frequently emphasized, studies demonstrating these effects longitudinally for users of a pro-ED community are lacking. This study aims at investigating the effects by modeling weight loss and changes in desired weight over the course of participation. In addition, although some studies have noted the existence of voluntarily shared weight information in pro-ED communities, the majority of studies used surveys or laboratory experiments to assess data outside of online communities. To our knowledge, this is the first study that uses data from a pro-ED community to calculate BMI and desired BMI and model their changes longitudinally.

In this study, we explored the pro-ED community r/proed on the social media website Reddit [37]. Reddit consists of thousands of communities called subreddits dedicated to topics such as sports, music, movies, news, politics, or videogames as well as physical health and mental disorders. Reddit users can write posts or comment on other users' posts in these communities. The r/proed community, similar to other pro-ED communities, featured tips and tricks such as circumventing minimums for weight and daily intake in a fitness tracker app [23], thinspiration, and discussions about topics such as low-calorie foods or binge eating as well as social support and treatment [17]. Besides being one of the largest and most active communities related to EDs on Reddit [17], r/proed also had one feature that was relevant for this study—users could display a text field called *flair* next to their username on all their posts and comments in r/proed. The flair (the term is used by Reddit users for both the singular and plural forms) contained self-reported height, current body weight, desired weight, age, and gender. It was possible for users to create, change, or update their flair with new information at any time, thereby allowing us to estimate changes in reported weight values over time. The r/proed community was banned on November 15, 2018, by Reddit's administrators for violating the website's content policy, namely, for *posting of content that encourages physical harm* [38].

We hypothesize that participation in r/proed is associated with weight loss. In addition, we hypothesize that users set lower desired weights over time to align themselves more closely with the propagated thin ideal. Furthermore, we hypothesize that the level of activity in the r/proed is associated with the changes in BMI and desired BMI of the users. Highly active users are

expected to lose more weight and decrease their desired BMI compared with less active users.

Methods

Data Collection

Data from the r/proed community and its users were collected over a period of 15 months between August 15, 2017, and November 14, 2018, until its ban. We assessed current and desired BMI using the values for height and current or desired weight at the time they appeared in the users' flair in r/proed. The values in the flair had to be coded manually, as they could not be assigned automatically to one of the weight categories because of ambiguity of the category to which they belonged. In addition, the units for height and weight measurements were omitted from some flair requiring educated guesses to determine which measurement system was used. Three independent raters manually coded a third of the flair texts for height, body weight, and desired weight. All 3 raters also coded a common set of 322 flair texts to assess interrater reliability for height, current weight, and desired weight. The intraclass correlation coefficients (ICCs) were 1, 0.98, and 0.98. The variable representing activity levels was calculated as the average number of posts and comments a user writes in the pro-ED community per week. After rating, all values for height and weight were converted to meters and kilograms, respectively.

As the data accessed through the official Reddit application programming interface for this study were publicly available, no application for ethics approval was considered necessary. We do not report any personally identifying information in this paper to protect the privacy of all users.

Data Analysis

We estimated separate linear growth models for the current and desired weights with individual measurement time points nested in users. The first available values for the current and desired BMIs of each user were treated as their respective baseline values. The variable time was calculated as weeks between the baseline and the respective time points. The current and desired BMIs were person centered around their respective baseline values. Baseline BMI, baseline desired BMI, and activity levels were grand mean centered and standardized. All linear growth models were estimated with maximum likelihood estimation and the Nelder-Mead optimizer using the R package lme4 [39].

We estimated four models for the current BMI by iteratively adding predictors in each model. The first model was an intercept-only model with random effects from users. This model was used to calculate ICCs to determine the amount of variance due to between-person differences. In the second model, a random slope and the fixed effect for time were added. In the third and fourth models, we added interactions of time with standardized baseline BMI and activity levels, respectively.

A similar procedure was adopted for the desired BMI values. The first model was an intercept-only model, whereas in the second model, time was added. The three following models added the interaction between time and the standardized baseline desired BMI, standardized first current BMI, and the activity level variable.

Interaction effects in the linear growth models were illustrated by simulations using the R package merTools. The graphs of the simulation results are provided.

Results

Descriptive Statistics

A total of 16,241 different flair texts were created by 5372 users during the data collection period. Of the 5372 users, 4101 had to be excluded from the analysis because they had only one measurement point for the current and desired weight. We performed sanity checks of the coded data and excluded 13 users because they had unrealistic values for height, current, or desired weight, such as a goal weight of 0 or a current weight BMI of 100, or because their flair showed implausible changes over time, such as a height increase of 5 cm in 1 day. The final data set consisted of 1170 users with 5193 flair texts. The descriptive statistics of the sample are presented in Table 1. At baseline, the average BMI was 22.32 (SD 4.08), whereas the average desired BMI was 18.55 (SD 2.33). Users participated in the community for an average of 25.00 weeks (SD 19.36) during our data collection. Overall, 1.2% (14/1170) users participated from the beginning of data collection to the last week before the community was banned, and 50% (585/1170) users participated for more than 20 weeks.

Table 1. Demographics, participation characteristics, and baseline values for the current and desired weight of users of an eating disorder community on Reddit (n=1170).

	Values
Demographics	
Gender[a], n (%)	
Female	891 (76.15)
Male	63 (5.38)
Nonbinary	25 (2.14)
Transgender	9 (0.77)
Age[b] (years)	
Value, mean (SD)	22.54 (4.68)
Value, median (range)	22 (13.60-59.00)
Participation characteristics	
Number of flair texts	
Value, mean (SD)	4.44 (3.93)
Value, median (range)	3 (2-62)
Time[c]	
Value, mean (SD)	13.27 (14.67)
Value, median (range)	7.59 (0.00-64.68)
Activity level[d]	
Value, mean (SD)	7.68 (29.85)
Value, median (range)	2.75 (0.0-759.32)
Baseline values	
Height (cm)	
Value, mean (SD)	164.91 (7.72)
Value, median (range)	165.10 (142.24-195.58)
Current weight (kg)	
Value, mean (SD)	60.80 (12.42)
Value, median (range)	58.29 (34.47-135.35)
Current BMI (kg/m^2)[e]	
Value, mean (SD)	22.32 (4.08)
Value, median (range)	21.48 (14.20-46.74)
Desired BMI (kg/m^2)[f]	
Value, mean (SD)	18.55 (2.33)
Value, median (range)	18.31 (12.02-42.29)

[a]Overall, 15.56% (182/1170) of users did not report their gender.

[b]Overall, 50.85% (595/1170) of users did not report their age.

[c]Weeks between the first and last flair of a user.

[d]Number of posts and comments of a user per week.

[e]Current BMI = (current weight in kg)/(height in m^2).

[f]Desired BMI = (desired weight in kg)/(height in m^2).

Linear Growth Models for the Current BMI

The ICC for BMI, calculated using the intercept-only model, was 0.38, indicating considerable variance between users. Compared with the intercept-only model, the three subsequent models, each adding one additional predictor, showed marked improvements in the goodness-of-fit measures Akaike information criterion (AIC) and Bayesian information criterion

(BIC). The fourth model had the lowest AIC and BIC values and thus was chosen as the final model (results for all four models are provided in Table 2).

The final model shows that users lost weight over time, with a predicted rate of 0.087 BMI points per week. Users with a higher baseline BMI and higher activity levels experienced greater weight loss. The interaction between time and baseline BMI on weight loss is shown in Figure 1. It shows that even users with low BMIs at baseline (BMIs between 15 and 18.24) lost some weight during their participation in the pro-ED community.

Table 2. Linear growth models for the current BMI of 1170 users of an eating disorder community on Reddit.

	Model 1	Model 2	Model 3	Model 4
Fixed effect estimates (SE)				
Intercept	−0.473 (0.025)	−0.165 (0.013)	−0.169 (0.013)	−0.165 (0.013)
Time	—[a]	−0.074 (0.004)	−0.072 (0.004)	−0.087 (0.005)
Time×baseline current BMI interaction[b]	—	—	−0.044 (0.004)	−0.043 (0.004)
Time×activity interaction[b]	—	—	—	−0.118 (0.020)
Random effects variance				
Intercept	0.512	0.068	0.068	0.069
Residual	0.829	0.250	0.251	0.250
Slope	—	0.011	0.009	0.009
Goodness-of-fit measures				
Akaike information criterion	15,172.918	10,504.541	10,387.654	10,353.948
Bayesian information criterion	15,192.584	10,543.871	10,433.540	10,406.389
Log likelihood	−7583.459	−5246.271	−5186.827	−5168.974

[a]—: not applicable.

[b]Main effects of baseline current BMI and activity were excluded from their models as their inclusion lowered the model fit.

Figure 1. Interaction effects of time and baseline current BMI on current weight BMI change of 1170 users of an eating disorder community on Reddit.

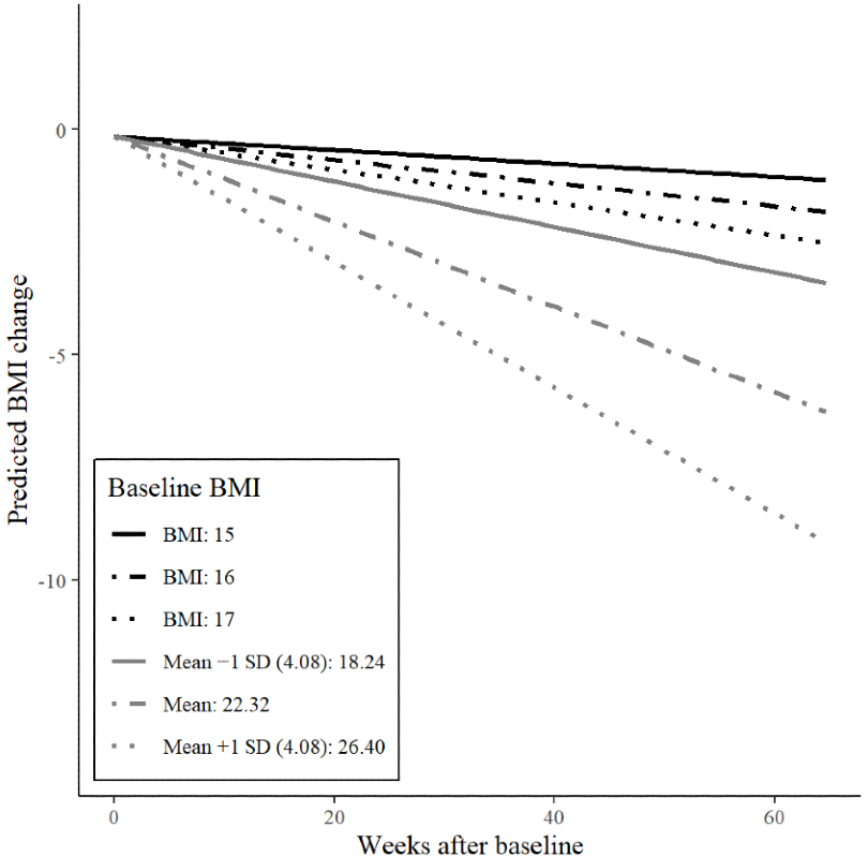
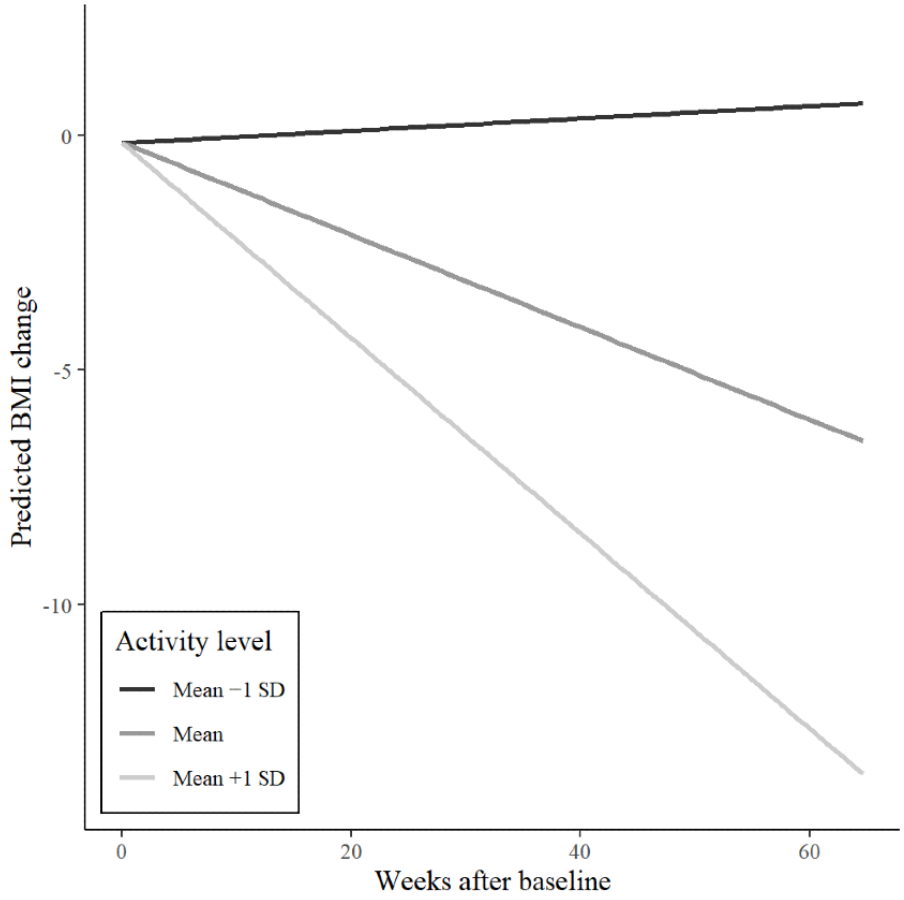


Figure 2 shows the interaction between the activity level and time on weight loss. Although greater activity levels were associated with greater weight loss for users at mean or above mean activity levels, users with low activity levels were predicted to gain weight.

Figure 2. Interaction effects of time and activity level on current weight BMI change of 1170 users of an eating disorder community on Reddit.



Linear Growth Models for the Desired BMI

The intercept-only model for the desired BMI has an ICC of 0.41, which indicates substantial variance between users. The three subsequent models showed marked improvements in the AIC and BIC fit measures. Including the interaction of activity and the time variable (model 5) does not improve the model fit. Therefore, the fourth model was considered as the final model (Table 3).

On average, users' desired weight decreased by 0.014 BMI points per week. Those with average or above average baseline desired BMIs reduced their desired weight, whereas those with below average baseline desired BMI increased their desired BMI over time. Figure 3 illustrates the interactions of time and baseline desired BMI on desired BMI graphically.

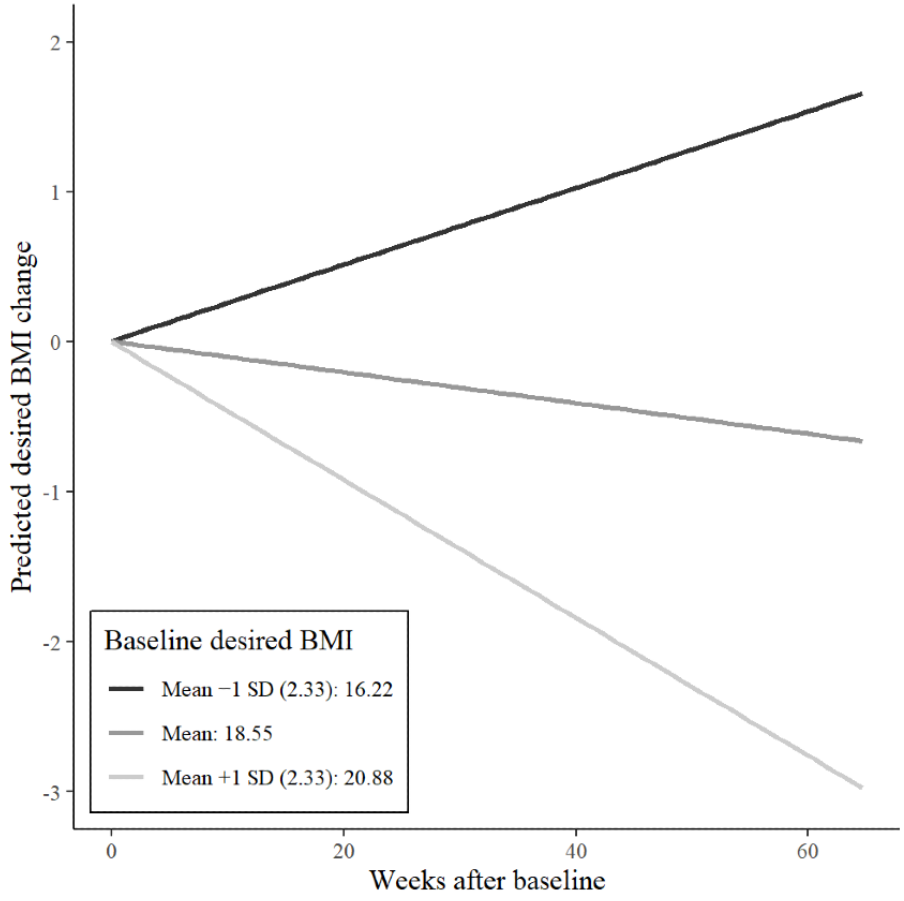
Table 3. Linear growth models for the desired BMI of 1170 users of an eating disorder community on Reddit.

	Model 1	Model 2	Model 3	Model 4	Model 5
Fixed effect estimates (SE)					
Intercept	−0.075 (0.021)	−0.003 (0.01)	0.002 (0.017)	0.002 (0.017)	0.003 (0.017)
Time	—[a]	−0.015 (0.003)	−0.015 (0.003)	−0.014 (0.003)	−0.015 (0.003)
Time×baseline desired BMI interaction[b]	—	—	−0.017 (0.003)	−0.033 (0.003)	−0.033 (0.003)
Time×baseline current BMI interaction[b]	—	—	—	0.026 (0.003)	0.026 (0.003)
Time×activity interaction[b]	—	—	—	—	−0.010 (0.015)
Random effects variance					
Intercept	0.379	0.185	0.187	0.187	0.187
Residual	0.534	0.335	0.336	0.335	0.335
Slope	—	0.004	0.004	0.003	0.003
Goodness-of-fit measures					
Akaike information criterion	12999.600	11627.828	11593.609	11538.863	11540.415
Bayesian information criterion	13019.265	11667.159	11639.494	11591.304	11599.411
Log likelihood	−6496.800	−5807.914	−5789.804	−5761.432	−5761.208

[a]—: not applicable.

[b]Main effects of baseline desired BMI, baseline current BMI, and activity were excluded from their models as their inclusion lowered the model fit.

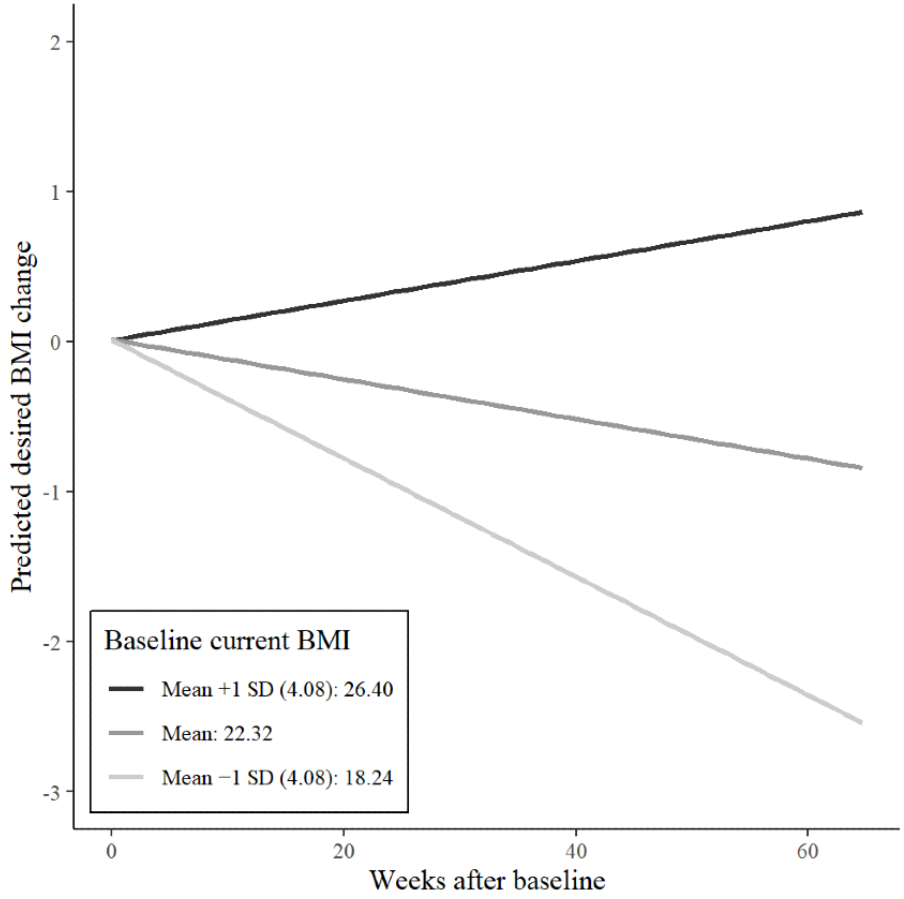
Figure 3. Interaction effects of time and baseline desired BMI on desired weight BMI change of 1170 users of an eating disorder community on Reddit.



The interaction between baseline current BMI and time on desired BMI was positive, indicating that a higher baseline

current BMI was associated with a slower decrease or even an increase in desired BMI over time. Figure 4 illustrates this interaction.

Figure 4. Interaction effects of time and baseline current BMI on desired weight BMI change of 1170 users of an eating disorder community on Reddit.



Discussion

Principal Findings

This study is the first to investigate changes in the current and desired BMI over a period of 15 months using user-generated data from an online pro-ED community. We were able to demonstrate sustained decreases in the BMI of r/proed users over the course of their participation. Although baseline BMI moderated the effect of time, even participants with very low baseline BMIs of 15-17 kg/m^2 still lost weight. The fact that even users with very low baseline BMIs lost weight over the course of their participation is alarming, as this exacerbates the risks associated with low body weight. However, weight loss is also related to clinical complications in individuals with body weight above the underweight category [3]. In addition, the level of activity in the r/proed community was associated with weight loss. The more active users contributed to the pro-ED community, the more weight they lost over time (Figure 2). In addition to evidence that more active users experience greater ED-related impairment [5], our results suggest that they are also at greater risk of unhealthy weight loss over the course of their participation. Baseline BMI as a moderator of weight loss over time is consistent with results from a case study on a weight loss community on Reddit [36]. Those at higher baseline BMIs might have been more motivated to lose weight by a perception of excess weight in a community that puts a high value on the thin ideal [36]. In addition to decreases in BMI, we could also demonstrate decreases in users' desired BMIs over the course of their participation. This suggests that users did not have one fixed weight goal but that ever-decreasing weight goals stayed out of their reach, even when they were losing weight. Baseline BMI moderated changes in the desired BMI (Figure 4). Users with higher baseline BMI increased their desired BMI from initially more unrealistic to healthier values over time. However, users with average or below average baseline BMI decreased their desired BMI over time. This is concerning, as these users are very likely to experience high body weight dissatisfaction although they already had low or underweight body weights.

The interaction of baseline desired BMI and time (Figure 3) demonstrates that the desired BMI of the users aligns over time. In a way, the pro-ED community moves toward a common middle ground of desired weight. Such a concurrence in the level of desired BMI could be because of posts of thinspiration images featuring women of similar, thin body types or because of users comparing and adjusting to the desired weight in the flair of other users.

In summary, the results strongly support the detrimental effects of participating in online pro-ED communities and point to the risks associated with these communities. In view of these risks, potential benefits such as receiving social support and reducing loneliness appear negligible [11].

Users' average desired BMI at baseline was just above the cutoff for underweight and thus at a similar level to that of a sample of women with ED [25] but considerably lower than that of a population sample [32]. In this study, the sample's average baseline desired weight was 20% lower than the average baseline current weight and thus higher than the recommended realistic weight loss goals of 5%-10% below the actual body weight [40]. Users pursuing unrealistic weight loss goals that extend into the underweight range point to the influence of thinspiration and social comparisons on weight goal setting. A lower desired weight in relation to actual body weight has also been linked to greater severity of ED [27,28].

We could not identify the specific processes and factors that caused these effects and yielded weight loss or decrease in the desired BMI. Cross-sectional and experimental studies have shown that exposure to pro-ED communities is linked to increased dieting and body dissatisfaction; however, it is unclear whether these are enduring effects [19]. In addition, online pro-ED communities are considered harmful because they decrease the likelihood of seeking help and lead to increased ED symptoms [7,19,41]. Moreover, encouragement and social reinforcement from other users may foster extreme weight loss techniques [11].

Strengths and Limitations

Our analyses show that a high number of users used features such as flair to self-report their height, current, and desired weight in the pro-ED community and that it is possible to reliably extract meaningful information on ED impairment from online ED communities. Although self-reports of weight and height can be inaccurate, users of pro-ED communities, young women and women with EDs, tend to report their height and weight more accurately [42]. The results of this study apply only to active users in the pro-ED community. It remains unclear whether the findings can be generalized to lurkers who do not actively contribute to the community. In addition, we could not determine whether individuals were active with more than one Reddit account during the period of our study, either using multiple accounts simultaneously or sequentially. Another limitation of the study is that its observational study design does not allow to conclude causality. However, conducting randomized controlled trials to investigate the detrimental effects of potentially harmful online communities longitudinally is not feasible. The users' activity on other social media platforms besides Reddit, such as Facebook or Instagram, was not assessed in this study. These platforms can also have detrimental effects on ED symptomatology [43,44]. However, the moderating effects of the activity level support the validity of the results. Although changes in BMI and desired BMI can be considered valid proxies for ED impairment in this population (especially for the subgroup of users with very low BMIs), they represent only one aspect of ED-related impairment, and more comprehensive assessments of ED impairment would be beneficial.

Conclusions

In November 2018, the administrators of Reddit issued a ban on the r/proed community. Although the subject of this study no longer exists, we consider its results to be an important addition to the literature. They can inform our understanding of other newly created or existing communities with similar processes and effects to those in r/proed. After the ban of r/proed, almost immediately, new communities appeared in its place, with most of them being quickly taken down by Reddit's administrators. One such community, r/EDAnonymous [45], has avoided a ban so far, possibly because of its stricter rules that do not allow weight and height measurements in flair. This is unlikely to prevent or stop detrimental effects, such as weight loss and decreases in desired weight, as we showed in this study, instead rendering these processes invisible to social media admins and researchers. Uncovering these processes was only possible by studying the now-banned r/proed.

In general, it seems unlikely that it will be possible to ban pro-ED communities completely from the internet or even from major social media platforms, such as Pinterest, Instagram, and Tumblr [46]. Therefore, it is of utmost importance to learn more about the detrimental effects of these communities, their social structures, the processes that take place, and the factors mediating the effects of web-based communication on ED symptomatology or mental health in general. Future research should focus on identifying the mechanisms that occur within online communities. Observational studies, such as this study, have to complement cross-sectional and experimental research to gain a deeper understanding of the processes and social structures in these communities and to counteract potentially detrimental effects on their users. The deeper understanding gained from this study and other similar research is also relevant for clinical work. Therapists should discuss social media habits and potentially harmful social media use with their clients. Therapeutic and preventive interventions for EDs, especially those for adolescents and emerging adults, should work toward healthier use of social media, that is, by teaching social media literacy [47].

Acknowledgments

We acknowledge financial support by the Open Access Publishing Fund of Ruprecht-Karls-Universität Heidelberg.

Conflicts of Interest

None declared.

References

1. Arcelus J, Mitchell AJ, Wales J, Nielsen S. Mortality rates in patients with anorexia nervosa and other eating disorders. A meta-analysis of 36 studies. Arch Gen Psychiatry 2011 Jul 04;68(7):724-731. [doi: 10.1001/archgenpsychiatry.2011.74] [Medline: 21727255]
2. Gibson D, Workman C, Mehler PS. Medical complications of anorexia nervosa and bulimia nervosa. Psychiatr Clin North Am 2019 Jun;42(2):263-274. [doi: 10.1016/j.psc.2019.01.009] [Medline: 31046928]
3. Whitelaw M, Lee KJ, Gilbertson H, Sawyer SM. Predictors of complications in anorexia nervosa and atypical anorexia nervosa: degree of underweight or extent and recency of weight loss? J Adolesc Health 2018 Dec;63(6):717-723. [doi: 10.1016/j.jadohealth.2018.08.019] [Medline: 30454732]
4. Juarascio AS, Shoaib A, Timko CA. Pro-eating disorder communities on social networking sites: a content analysis. Eat Disord 2010 Sep;18(5):393-407. [doi: 10.1080/10640266.2010.511918] [Medline: 20865593]
5. Peebles R, Wilson JL, Litt IF, Hardy KK, Lock JD, Mann JR, et al. Disordered eating in a digital age: eating behaviors, health, and quality of life in users of websites with pro-eating disorder content. J Med Internet Res 2012;14(5):e148 [FREE Full text] [doi: 10.2196/jmir.2023] [Medline: 23099628]
6. Harper K, Sperry S, Thompson JK. Viewership of pro-eating disorder websites: association with body image and eating disturbances. Int J Eat Disord 2008 Jan;41(1):92-95. [doi: 10.1002/eat.20408] [Medline: 17634964]
7. Rouleau CR, von Ranson KM. Potential risks of pro-eating disorder websites. Clin Psychol Rev 2011 Jun;31(4):525-531. [doi: 10.1016/j.cpr.2010.12.005] [Medline: 21272967]
8. Hasan H. Instagram bans thinspo content. New York City, NY: Time USA; 2012 Apr 26. URL: https://newsfeed.time.com/2012/04/26/instagram-bans-thinspo-content/ [accessed 2021-07-29]
9. Zafar A. Tumblr bans pro-eating disorder and other self-harm blogs. Time. 2012 Feb 24. URL: https://newsfeed.time.com/2012/02/24/tumblr-bans-pro-eating-disorder-and-other-self-harm-blogs/ [accessed 2021-07-29]
10. Lichfield J. France bans websites promoting anorexia 'cult'. Independent. 2011 Sep 23. URL: https://www.independent.co.uk/news/world/europe/france-bans-websites-promoting-anorexia-cult-809617.html [accessed 2021-07-29]
11. Csipke E, Horne O. Pro-eating disorder websites: users' opinions. Eur Eat Disord Rev 2007 May;15(3):196-206. [doi: 10.1002/erv.789] [Medline: 17676689]
12. Sowles SJ, McLeary M, Optican A, Cahn E, Krauss MJ, Fitzsimmons-Craft EE, et al. A content analysis of an online pro-eating disorder community on Reddit. Body Image 2018 Mar;24:137-144. [doi: 10.1016/j.bodyim.2018.01.001] [Medline: 29414146]
13. Chancellor S, Hu A, De Choudhury M. Norms matter: contrasting social support around behavior change in online weight loss communities. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18: CHI Conference on Human Factors in Computing Systems; April 21 - 26, 2018; Montreal, Canada p. 1-14. [doi: 10.1145/3173574.3174240]
14. Haas SM, Irr ME, Jennings NA, Wagner LM. Communicating thin: A grounded model of Online Negative Enabling Support Groups in the pro-anorexia movement. New Media Soc 2010 Oct 12;13(1):40-57. [doi: 10.1177/1461444810363910]
15. Borzekowski DLG, Schenk S, Wilson JL, Peebles R. e-Ana and e-Mia: A content analysis of pro-eating disorder Web sites. Am J Public Health 2010 Aug;100(8):1526-1534. [doi: 10.2105/AJPH.2009.172700] [Medline: 20558807]
16. Harshbarger JL, Ahlers-Schmidt CR, Mayans L, Mayans D, Hawkins JH. Pro-anorexia websites: what a clinician should know. Int J Eat Disord 2009 May;42(4):367-370. [doi: 10.1002/eat.20608] [Medline: 19040264]
17. Moessner M, Feldhege J, Wolf M, Bauer S. Analyzing big data in social media: text and network analyses of an eating disorder forum. Int J Eat Disord 2018 Jul;51(7):656-667. [doi: 10.1002/eat.22878] [Medline: 29746710]
18. Wilson JL, Peebles R, Hardy KK, Litt IF. Surfing for thinness: a pilot study of pro-eating disorder Web site usage in adolescents with eating disorders. Pediatrics 2006 Dec;118(6):1635-1643. [doi: 10.1542/peds.2006-1133] [Medline: 17142493]
19. Rodgers RF, Lowy AS, Halperin DM, Franko DL. A meta-analysis examining the influence of pro-eating disorder websites on body image and eating pathology. Eur Eat Disord Rev 2016 Jan;24(1):3-8. [doi: 10.1002/erv.2390] [Medline: 26230192]
20. World Health Organization. Obesity : preventing and managing the global epidemic : report of a WHO consultation. Geneva: World Health Organization; 2000.
21. Eikey EV, Reddy MC. "It's definitely been a journey": A qualitative study on how women with eating disorders use weight loss apps. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17: CHI Conference on Human Factors in Computing Systems; May 6 - 11, 2017; Denver Colorado USA p. 642-654. [doi: 10.1145/3025453.3025591]
22. McCaig D, Bhatia S, Elliott MT, Walasek L, Meyer C. Text-mining as a methodology to assess eating disorder-relevant factors: comparing mentions of fitness tracking technology across online communities. Int J Eat Disord 2018 Jul;51(7):647-655. [doi: 10.1002/eat.22882] [Medline: 29734478]
23. McCaig D, Elliott MT, Prnjak K, Walasek L, Meyer C. Engagement with MyFitnessPal in eating disorders: qualitative insights from online forums. Int J Eat Disord 2020 Mar;53(3):404-411. [doi: 10.1002/eat.23205] [Medline: 31762064]
24. Boero N, Pascoe C. Pro-anorexia communities and online interaction: bringing the pro-ana body online. Body Soc 2012 May 24;18(2):27-57. [doi: 10.1177/1357034X12440827]

25. Coker E, Abraham S. Body weight dissatisfaction: a comparison of women with and without eating disorders. Eat Behav 2014 Aug;15(3):453-459. [doi: 10.1016/j.eatbeh.2014.06.014] [Medline: 25064299]
26. Roy M, Gauvin L. Having a personal weight goal that mismatches healthy weight recommendations increases the likelihood of using unhealthy behaviors among a representative population-based sample of adolescents. Eat Behav 2010 Dec;11(4):281-287. [doi: 10.1016/j.eatbeh.2010.08.002] [Medline: 20850064]
27. Boyd HK, Bodell LP, Jennings KM, Graham AK, Crosby RD, Wildes JE. Relationship between desired weight constructs and eating disorder severity following treatment for anorexia nervosa. Int J Eat Disord 2018 Aug;51(8):870-878 [FREE Full text] [doi: 10.1002/eat.22879] [Medline: 29734468]
28. Boyd HK, Kass AE, Accurso EC, Goldschmidt AB, Wildes JE, Le Grange D. Relationship between desired weight and eating disorder pathology in youth. Int J Eat Disord 2017 Aug;50(8):963-969 [FREE Full text] [doi: 10.1002/eat.22720] [Medline: 28429386]
29. Rienecke RD, Ebeling M. Desired weight and treatment outcome among adolescents in a novel family-based partial hospitalization program. Psychiatry Res 2019 Mar;273:149-152. [doi: 10.1016/j.psychres.2019.01.028] [Medline: 30641345]
30. Griffiths S, Castle D, Cunningham M, Murray SB, Bastian B, Barlow FK. How does exposure to thinspiration and fitspiration relate to symptom severity among individuals with eating disorders? Evaluation of a proposed model. Body Image 2018 Dec;27:187-195. [doi: 10.1016/j.bodyim.2018.10.002] [Medline: 30359868]
31. Aars NA, Jacobsen BK. Longitudinal changes in desired body weight compared to changes in body weight: evidence of adaptation to weight gain? BMC Obes 2016;3:40 [FREE Full text] [doi: 10.1186/s40608-016-0120-6] [Medline: 27688891]
32. Maynard LM, Serdula MK, Galuska DA, Gillespie C, Mokdad AH. Secular trends in desired weight of adults. Int J Obes (Lond) 2006 Sep;30(9):1375-1381. [doi: 10.1038/sj.ijo.0803297] [Medline: 16552407]
33. Nothwehr F, Yang J. Goal setting frequency and the use of behavioral strategies related to diet and physical activity. Health Educ Res 2007 Aug;22(4):532-538. [doi: 10.1093/her/cyl117] [Medline: 17032703]
34. Chomutare T, Årsand E, Hartvigsen G. Characterizing development patterns of health-care social networks. Netw Model Anal Health Inform Bioinforma 2013 May 14;2(3):147-157. [doi: 10.1007/s13721-013-0033-y]
35. Hwang KO, Ning J, Trickey AW, Sciamanna CN. Website usage and weight loss in a free commercial online weight loss program: retrospective cohort study. J Med Internet Res 2013;15(1):e11 [FREE Full text] [doi: 10.2196/jmir.2195] [Medline: 23322819]
36. Pappa GL, Cunha TO, Bicalho PV, Ribeiro A, Couto SA, Meira W, et al. Factors associated with weight change in online weight management communities: a case study in the LoseIt Reddit community. J Med Internet Res 2017 Jan 16;19(1):e17 [FREE Full text] [doi: 10.2196/jmir.5816] [Medline: 28093378]
37. Reddit. 2002. URL: https://www.reddit.com [accessed 2021-07-29]
38. proED: banned. Reddit. 2018. URL: https://www.reddit.com/r/proED/ [accessed 2021-07-29]
39. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using. J Stat Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
40. Foster G. Reasonable weights: determinants, definitions and directions. In: Allison DB, Pi-Sunyer FX, editors. Obesity Treatment. NATO ASI Series (Series A: Life Sciences), Vol 278. Boston, MA: Springer; 1995:35-44.
41. Sharpe H, Musiat P, Knapton O, Schmidt U. Pro‐eating disorder websites: facts, fictions and fixes. J of Public Mental Health 2011 Mar 15;10(1):34-44. [doi: 10.1108/17465721111134538]
42. Engstrom JL, Paterson SA, Doherty A, Trabulsi M, Speer KL. Accuracy of self-reported height and weight in women: an integrative review of the literature. J Midwifery Womens Health 2003;48(5):338-345. [doi: 10.1016/s1526-9523(03)00281-2] [Medline: 14526347]
43. Holland G, Tiggemann M. A systematic review of the impact of the use of social networking sites on body image and disordered eating outcomes. Body Image 2016 Jun;17:100-110. [doi: 10.1016/j.bodyim.2016.02.008] [Medline: 26995158]
44. Saiphoo AN, Vahedi Z. A meta-analytic review of the relationship between social media use and body image disturbance. Comp Hum Behav 2019 Dec;101:259-275. [doi: 10.1016/j.chb.2019.07.028]
45. EDAnonymous. Reddit. 2018. URL: http://www.reddit.com/r/EDAnonymous [accessed 2021-07-29]
46. Gerrard Y. Beyond the hashtag: Circumventing content moderation on social media. New Media Soc 2018 May 28;20(12):4492-4411. [doi: 10.1177/1461444818776611]
47. Tamplin NC, McLean SA, Paxton SJ. Social media literacy protects against the negative impact of exposure to appearance ideal social media images in young adult women but not men. Body Image 2018 Sep;26:29-37. [doi: 10.1016/j.bodyim.2018.05.003] [Medline: 29807299]

Abbreviations

AIC: Akaike information criterion
BIC: Bayesian information criterion
ED: eating disorder
ICC: intraclass correlation coefficient

Edited by R Kukafka; submitted 14.01.21; peer-reviewed by J Denison-Day, A Graham, M Wick; comments to author 07.02.21; revised version received 31.03.21; accepted 14.06.21; published 06.10.21.

Please cite as:
Feldhege J, Moessner M, Bauer S
Detrimental Effects of Online Pro–Eating Disorder Communities on Weight Loss and Desired Weight: Longitudinal Observational Study
J Med Internet Res 2021;23(10):e27153
URL: https://www.jmir.org/2021/10/e27153
doi:10.2196/27153
PMID:34612830

Original Paper

Identifying Frequent Health Care Users and Care Consumption Patterns: Process Mining of Emergency Medical Services Data

Laura Maruster[1], PhD; Durk-Jouke van der Zee[1], PhD; Erik Buskens[1,2], MD, PhD

[1]Faculty of Economics and Business, University of Groningen, Groningen, Netherlands

[2]Health Technology Assessment, Department of Epidemiology, University Medical Center Groningen, Groningen, Netherlands

Corresponding Author:
Laura Maruster, PhD
Faculty of Economics and Business
University of Groningen
Nettelbosje 2
Groningen, 9747 AE
Netherlands
Phone: 31 50 3637316
Email: l.maruster@rug.nl

Abstract

Background: Tracing frequent users of health care services is highly relevant to policymakers and clinicians, enabling them to avoid wasting scarce resources. Data collection on frequent users from all possible health care providers may be cumbersome due to patient privacy, competition, incompatible information systems, and the efforts involved.

Objective: This study explored the use of a single key source, emergency medical services (EMS) records, to trace and reveal frequent users' health care consumption patterns.

Methods: A retrospective study was performed analyzing EMS calls from the province of Drenthe in the Netherlands between 2012 and 2017. Process mining was applied to identify the structure of patient routings (ie, their consecutive visits to hospitals, nursing homes, and EMS). Routings are used to identify and quantify frequent users, recognizing frail elderly users as a focal group. The structure of these routes was analyzed at the patient and group levels, aiming to gain insight into regional coordination issues and workload distributions among health care providers.

Results: Frail elderly users aged 70 years or more represented over 50% of frequent users, making 4 or more calls per year. Over the period of observation, their annual number and the number of calls increased from 395 to 628 and 2607 to 3615, respectively. Structural analysis based on process mining revealed two categories of frail elderly users: low-complexity patients who need dialysis, radiation therapy, or hyperbaric medicine, involving a few health care providers, and high-complexity patients for whom routings appear chaotic.

Conclusions: This efficient approach exploits the role of EMS as the unique regional "ferryman," while the combined use of EMS data and process mining allows for the effective and efficient tracing of frequent users' utilization of health care services. The approach informs regional policymakers and clinicians by quantifying and detailing frequent user consumption patterns to support subsequent policy adaptations.

(J Med Internet Res 2021;23(10):e27499) doi:10.2196/27499

KEYWORDS

process mining; frequent users; hospital care; emergency medical services; regional care networks; elderly; Netherlands

Introduction

A large part of regional health care consumption is attributed to "frequent users" (ie, patients who make repeated calls to hospital and nursing health care services) [1]. Although definitions differ, a threshold of 4 to 5 calls or more per year is generally used to classify a patient as a "frequent user" [2,3]. While frequent users represent a minority of emergency department (ED) patients (4.5%-8%), they may account for up to 21%-28% of all ED visits [4]. Frequent users appear heterogeneous as a group; however, they may be clustered into distinct categories relating to their health care needs and the health services provided to them [5].

The high workload and costs incurred by frequent users make them a relevant target group for regional policymakers and clinicians to consider as they attempt to make the best use of scarce resources. For example, due to the various health care needs related to advanced age, the "frail elderly" are known to be frequent users [6-9]. Their frailty is related to their status of being extremely vulnerable to endogenous and exogenous stressors, exposing them to a higher risk of negative health-related outcomes [8]. Importantly, they are often confronted with fragmented health care [6], inappropriate or delayed triage at EDs [10], and incorrect referrals. These observations suggest an inappropriate approach to their health care needs and the potential unnecessary use of health care services. Once their frailness is identified, advanced health care planning may be used to improve their health care continuity [8,11,12]. Therefore, tracing frequent users and their (shared) consumption patterns is a prerequisite for assessing the efficiency and effectiveness of current clinical practice, undertaking appropriate actions to improve it, and evaluating the added value of these actions and related policy changes.

Notably, many frequent users do not make calls to a single health care provider but are network users, making consecutive visits to multiple health care providers. Apart from their preferences, their network use may be explained, for example, by specialized health care needs. Thus, tracing frequent users requires a network-based approach, including all regional health care providers. Unfortunately, collecting data by interviewing health care providers or even extracting data from their local records tends to be cumbersome. Rules on patient privacy, competition among health care providers, incompatibility of information systems, and the effort required may present hurdles that are not easily overcome, adding to data collection costs. Most research designs limit their scope to single or related health care providers, with a primary focus on hospitals [13]. As a result, many frequent users may be overlooked. Moreover, their routings along different health care providers may appear fragmented due to a lack of information on major health care providers outside hospitals, such as nursing homes, which are particularly relevant to the frail elderly.

This article explores an alternative approach to tracing frequent users, relying on emergency medical services (EMS) data (ie, records of ambulance rides containing patient and logistic data). Acting as the "ferrymen" in the regional health care network, EMS theoretically direct subacute or acute patient routings, starting with a time-ordered sequence of ambulance rides. Therefore, their databases potentially offer an efficient means for identifying and tracing frequent users. Importantly, the EMS patient population is likely to accommodate many frequent users [14,15]. Moreover, patients served by EMS tend to require substantial health care resources, as indicated by their need for mobile nursing services and transport. However, the potential of using EMS data to identify frequent users and their consumption patterns has hardly been acknowledged in the literature [16].

This study aimed to show how the analysis of EMS records may contribute to tracing frequent users on a regional scale, especially the frail elderly, and reveal their health care consumption patterns. The novel analysis technique of process mining is instrumental to the study, enabling the automated identification of patient routings (ie, identifying health care providers consulted over time by combining the records related to ambulance rides). Frequent users can be identified by quantifying their number of ambulance rides via process mining. In turn, their consumption patterns are reflected in their routings, specifying health care providers and specialties involved over time. The aggregation of patient routings establishes trends in their annual demand for health care and the associated workload distribution over the network. Process mining has an advantage over other mapping techniques as it uses factual observations retrieved from data rather than man-made process models. Process mining has been successfully used to analyze health care processes, usually in an intrahospital context [17-27]. However, its application to regional health care networks is new.

As relevant background information, we briefly summarize that the Dutch system for chronic healthcare has for a long time been funded through national funds under the General Act Special Care Costs [28]. However, this funding scheme was deemed too inefficient and generous. Accordingly, as of January 1, 2015, a major system change was introduced, accompanied by new legislation (ie, the Long-term Care Act) [29]. The execution and implementation of health care for less severe indications in the home setting were transferred to the municipalities to achieve a better match and a more efficient system. The transfer was expected to result in more elderly people remaining in their homes, thus reducing chronic (elderly) health care costs. As of spring 2021, these expectations, to a large extent, have not materialized or even worsened the situation. Our study reveals problems that have emerged since the system's change, including subsequent nursing home closures.

Using EMS records of the province of Drenthe, the Netherlands, this study shows how ambulance data allows for effective and efficient tracing and quantification of frequent users of health care services on a regional scale, considering frail elderly users as a focal group. The proposed approach builds on the role of EMS as the regional "ferryman," implying the utilization of a single key source for data collection, covering many health care providers at the same time (ie, hospitals, nursing homes, and EMS). Process mining adds to the efficiency of the approach by enabling automated mapping of patient routings (ie, their consecutive visits to health care providers). Furthermore, the structure of patient routes is analyzed at patient and group levels, allowing us to gain insight into regional coordination issues and workload distributions among health care providers, which is helpful to policymakers and clinicians. In particular, we seek evidence for the effects of the Dutch governmental policies described above, including the impact of higher health care need entry requirements for nursing home admissions on regional health care consumption [30-32].

Methods

The Health Care Network of the Province of Drenthe

The province of Drenthe has a population of 491,867, with a population density of 183 inhabitants per square kilometer [33]. Hospital care is provided by 4 hospitals within the province and

by several hospitals located in neighboring provinces. Among the 4 hospitals, 3 offer basic treatment, and 1 hospital has the necessary skills and resources to treat multilevel trauma. The reasons for referral to hospitals outside the province include being close to the patient scene, patient preferences, level of health care, or specialization related to specific treatments. Nursing care is also provided by a few dozen nursing homes, mainly located within the province. EMS is provided by a single operator, relying on a network of 17 bases in 14 towns or villages in Drenthe. Its services include both urgent (A-rides) and planned (B-rides) patient transport to hospitals and planned (B-rides) transport to nursing homes. B-rides are legitimized by patient health care needs prohibiting self-transport or transport by taxi. A-rides assume the presence of staff and equipment, enabling advanced life support (ALS; ie, advanced health care for critical patients). In contrast, B-rides may be offered with either ALS or basic life support, setting less strict medical skillsets and equipment requirements.

Data

Patient data were collected from EMS records of ambulance rides performed between January 1, 2012, and December 31, 2017. The data collected included the ride dates and times, destinations (ie, health care providers), patient age, urgency (A or B), and health care providers' medical specialty as indicated by the patient's health care needs. A unique identification number assigned to each patient identified patients' routings along different health care providers and services by combining their records and organizing them according to ambulance ride dates.

Based on our focus on the frail elderly population, frequent users, and EMS scope of services, 3 categories of health care providers were distinguished: hospitals, nursing homes, and the EMS. The latter was considered a formal health care provider when the treatment provided by the ambulance paramedics on the scene sufficed to address patient health care needs, designated EMS "see and treat" (EMS-S&T). This inclusion may also serve as an indicator of inappropriate or fragmented health care [6].

Process Mining

Like many regions worldwide, the province of Drenthe has increasing numbers of frail elderly people who utilize a significant part of the regional health care system. In terms of their visits to regional hospitals and nursing homes, their consumption patterns remain largely unknown, as regional health care providers do not disclose this type of patient information.

Process mining is a technique combining data science and process management to support the identification and analysis of operational processes (ie, sequential activities undertaken by an organization in satisfying its customers), thereby relying on event logs (ie, recordings of respective activities) stored in a database [34]. Process mining has been successfully used to map health care processes, clarifying how patients are served as a net effect of activities performed by health care providers. So far, most examples of its use in health care are related to an intrahospital context [17-27]. By automatically generating process maps using factual observations retrieved from data, process mining has important advantages over other modeling techniques that rely on manual observations of the actual system or inspection of documents [17,34].

EMS records referring to single rides are anonymized, cleaned by removing empty records (ie, records not relating to patients), and inspected for data accuracy to allow for process mining. If a record lacks information on the health care provider (eg, the destination of a ride), it was marked "unknown." Subsequently, Disco (version 2.2.0; Fluxicon) [35], a tool used to perform the process mining, traced patient routings by combining and ordering (time-wise) patient records referring to unique patients.

The health care consumption of frequent users (ie, patients meeting a threshold of 4 ambulance rides to regional health care providers within a year) is quantified by presenting their numbers and the number of calls, including annual trend figures and the distribution of frequencies. Frail elderly users (patients 70 years of age or more) were considered a subgroup among frequent users. They were analyzed for urgency using the ambulance ride categories A (urgent) and B (planned) as a proxy and for the volume of recognized categories of frequent users among them, specifically patients known to be in frequent need of dialysis, radiation therapy, or hyperbaric medicine (DRH). The "known" patient categories depend on local insight, thereby relying on EMS staff and records. For "unknown" patient categories, such a classification is lacking.

The health care consumption of frail elderly users was assessed at two levels (the patient level and the patient group level) using process mining for structural analysis. At the patient level, patient routings along the various health care providers are identified by the ride destinations, including the specialties being consulted. Patients are distinguished by two categories (low complexity and high complexity) as indicated by their routings. This complexity is considered from the perspective of health care consumption uncertainty. For instance, low-complexity patients are the "known" patients, involving few health care providers, and high-complexity patients are those for whom routings might not be fully understood. Although the term "low complexity" might suggest that these patients have a relatively mild health condition, they are frail and consume many resources and should also be recognized as a relevant subgroup.

Ethics Approval and Consent to Participate

Since the data originally were routinely collected for administrative purposes and completely anonymized, this study does not fall within the scope of the Medical Research Involving Human Subjects Act (Wet Medisch-weteschappelijk Onderzoek 2021) [31]. Accordingly, we obtained a full waiver for using anonymized data from the EMS services from the Medical Ethics Review Board of the University Medical Center Groningen (reference number METc 2018/402).

Results

Quantifying Frequent Users

Overview

Table 1 provides an overview of all patients served by regional health care providers based on EMS records. In total, 126,758 unique patients were identified between 2012 and 2017, involving 212,967 calls for services and omitting 2494 records not linked to patients. Table 1 shows call volumes have increased 25% over the years, from 31,300 to 39,235 calls. Similar changes are observed for frequent users' general health care consumption, including frail elderly users. Frequent users account for approximately 16% of total regional health care consumption, of which more than half is attributable to the frail elderly. The number of frail elderly users increased from 395 to 628 (59%), and their calls rose from 2607 to 3615 (39%). Strikingly, the largest growth in frail elderly users was observed from 2013 to 2015, increasing from 320 to 548 (71%) patients.

Table 1. Overview of patients served by regional health care providers.

Year	All patients, N	All patient calls, N	Frequent users, n	Frequent users, n (%)[a]	Frail elderly, n	Frail elderly, n (%)[b]
2012	22,551	31,300	731	5051 (16)	395	2607 (52)
2013	23,794	32,359	625	4636 (14)	320	2428 (52)
2014	24,355	34,568	844	5681 (16)	446	2792 (49)
2015	25,677	36,742	987	5976 (16)	548	3169 (53)
2016	27,146	38,763	999	6449 (17)	561	3333 (52)
2017	27,671	39,235	1043	6258 (16)	628	3615 (58)
Total	151,194 (126,758)[c]	212,967	5229 (4734)[c]	34,051	2898 (2700)[c]	17,944

[a]Calls made by frequent users as a percentage of calls from all patients.

[b]Calls made by frail elderly users as a percentage of calls from frequent users.

[c]Number of unique patients involved.

Frail Elderly

Details on the health care needs for frail elderly users and their urgency are shown in Table 2, which are categorized by distinguishing their EMS calls according to the medical specialty requested and the urgency of the ride (Tables 1 and 2). For example, among the 395 frail elderly users in 2012 (Table 1), 302 (76%) patients had been transported at least once in an urgent ride (A), while 345 (87%) patients had used at least 1 planned ride (B; Table 2). In addition, among patients transported in planned rides, 75 (19%) patients required dialysis, radiation therapy, or hyperbaric medicine (B-DRH), and 326 (83%) patients had other diverse health care needs and urgencies (B-other). The 2 patient groups overlap due to comorbidity; therefore, the sum of their patient numbers exceeds the overall annual number of frail elderly users. The right-hand side of Table 2 shows the number of calls, indicating the number of rides associated with the groupings mentioned above.

Table 2 reveals the number of urgent (A) calls among frail elderly users more than doubled between 2012 and 2017 (from 825 to 1729, 110%), in contrast to the modest growth in calls for rides planned in advance (B; from 1782 to 1886, 6%). In addition, between 2012 and 2017, the number of frail elderly users requiring either specific, predictable treatment or other treatments (Table 2, see columns B-DRH and B-other) increased from 75 to 104 (39%) and 326 to 487 (49%), respectively. However, annual calls made by DRH patients diminished between 2012 and 2017 (from 728 to 502, –31%), while a considerable growth in the number of calls made had occurred for the remaining group (from 1054 to 1384, 31%).

Table 2. Health care needs and urgency for frail elderly users.

Year	Frail elderly population					Frail elderly calls				
	All, N	Urgency of health care, n		Health care needs, n		All, N	Urgency of health care, n		Health care needs, n	
	A[a]+B[b]	A	B	B-DRH[c]	B-other[d]	A+B	A	B	B-DRH	B-other
2012	395	302	345	75	326	2607	825	1782	728	1054
2013	320	258	271	64	256	2428	664	1764	903	861
2014	446	372	370	74	353	2792	1019	1773	740	1033
2015	548	473	408	83	395	3169	1478	1691	533	1158
2016	561	498	417	99	398	3333	1552	1781	575	1206
2017	628	561	501	104	487	3615	1729	1886	502	1384
Total	2898 (2700)[e]	2464 (2303)[e]	2312 (842)[e]	499 (480)[e]	2,215 (2119)[e]	17,944	7267	10,677	3981	6696

[a]A: urgent transport.

[b]B: planned transport.

[c]B-DRH: patients in need of dialysis, radiation therapy, or hyperbaric medicine making use of EMS planned transport.

[d]B-other: patients with health care needs other than dialysis, radiation therapy, or hyperbaric medicine making use of EMS planned transport.

[e]Number of unique patients involved.

Identifying Regional Health Care Consumption Patterns Among Frail Elderly Users

Patient Level: Low- and High-Complexity Patients

Consumption patterns for frail elderly users are captured by ordered ride lists and process maps (Tables 3 and 4; Figures 1 and 2). Tables 3 and 4 show an excerpt of the routings of 2 frequent users, patient A and patient B, respectively. Patient A needs dialysis and is served by a single hospital (RegHospital 1). Patient A's routing exhibits low complexity, which is clearly shown by the process map in Figure 1. It illustrates how the patient was treated 145 times by RegHospital 1 while living in the nursing home between 2012 and 2017. Arcs in Figure 1 summarize information on the sequence of services consumed; for example, a ride to RegHospital 1 is directly followed by a ride to the nursing home 117 times, and a ride to the nursing home is directly followed by a ride to RegHospital 1 118 times. In some cases, transport to or from the hospital has not been organized by the EMS provider under study. For instance, a ride to RegHospital 1 is followed by another ride to RegHospital 1 27 times. Similarly, a ride to the nursing home is followed by another ride to the nursing home 27 times.

Table 3. Excerpt from the routings of patient A.

Date	Time	Urgency	Destination	Specialty
2012-01-06	16:06:15	B	Nursing home	Other specialties
2012-01-19	15:46:06	B	Nursing home	Internal medicine
2012-01-20	09:30:56	B	RegHospital 1-dialysis	Other specialties
2012-01-23	10:15:32	B	RegHospital 1-dialysis	Other specialties
2012-01-23	15:16:49	B	Nursing home	Internal medicine
2012-01-25	09:11:57	B	RegHospital 1-dialysis	Other specialties

Table 4. Excerpt from the routings of patient B.

ID	Date	Time	Urgency	Activity	Specialty
1	2013-09-13	06:50:39	A1	EMS-S&T[a]	OS[b]
2	2013-06-02	02:42:14	A1	Hospital A	OS
3	2013-06-16	13:50:06	A2	Hospital A	OS
4	2013-07-02	00:04:43	A1	Hospital A	P[c]
5	2014-02-01	20:40:56	A2	Hospital A	OS
6	2014-02-01	23:09:41	A2	Unknown	OS
7	2014-11-26	22:26:42	A1	Hospital A	OS
8	2014-12-06	11:22:00	A1	Hospital A	P
9	2014-12-08	12:51:39	A1	Hospital A	Surgery
10	2014-12-08	16:14:38	B	Unknown	OS
11	2014-02-17	11:42:49	A1	Hospital A	P
12	2014-04-02	09:37:52	B	RegHospital 2	OS
13	2015-09-05	08:43:23	A1	RegHospital 4	General
14	2015-09-05	11:04:53	B	Unknown	Unknown
15	2015-09-23	01:51:15	A2	Hospital D	General
16	2015-10-22	07:57:29	A2	Hospital D	General
17	2015-02-06	09:56:39	A2	Hospital A	P
18	2015-02-06	11:56:25	A2	Nursing home	P
19	2015-02-12	11:05:34	A1	Hospital A	P
20	2015-02-12	12:47:12	B	Nursing home	P
21	2015-02-26	08:57:59	B	Hospital A	OS
22	2015-02-26	10:47:32	B	Nursing home	OS
23	2015-01-07	08:16:58	A1	Hospital A	OS
24	2015-01-07	10:02:35	A1	Unknown	OS
25	2015-01-07	11:48:10	A1	RegHospital 1	Surgery
26	2015-03-15	05:22:40	A1	EMS-S&T	OS
27	2015-01-11	16:09:12	A2	Hospital A	OS
28	2015-01-11	17:46:08	B	RegHospital 1	OS
29	2015-04-30	14:39:22	A2	Hospital A	P
30	2015-04-30	16:02:48	B	Unknown	P
31	2015-01-14	08:55:44	B	Hospital A	P
32	2015-06-23	16:47:33	A2	Hospital A	P
33	2015-06-23	19:00:12	B	Unknown	P
34	2015-06-28	10:43:34	A1	Hospital A	General
35	2015-06-28	21:01:32	A1	RegHospital 1	P
36	2015-07-10	11:45:13	A1	Hospital D	General
37	2015-07-12	12:09:24	A1	RegHospital 1	IM[d]
38	2015-07-12	14:21:53	B	Unknown	IM
39	2015-01-21	21:39:06	A2	Hospital A	P
40	2016-01-14	08:58:03	A2	Hospital D	P
41	2016-01-02	22:08:36	A1	Hospital D	P

[a]EMS-S&T: emergency medical services "see and treat."

[b]OS: other specialties.

[c]P: pulmonology.

[d]IM: internal medicine.

Figure 1. The process map of low complexity patient A.

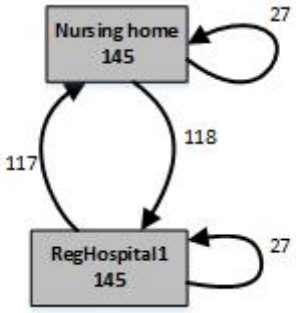
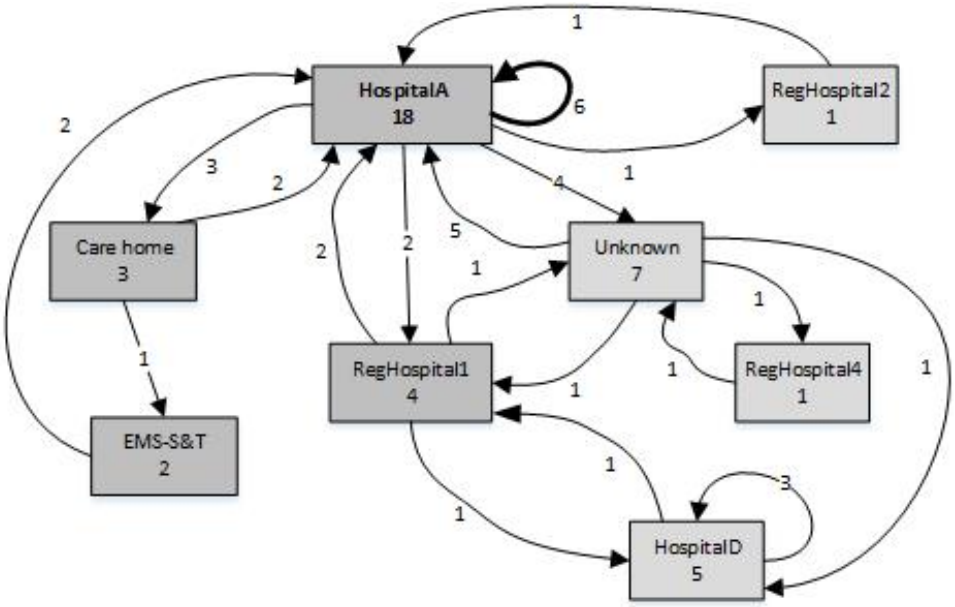


Figure 2. The process map of high complexity patient B. EMS-S&T: emergency medical services "see and treat."



Whereas routings for patient A exhibit low complexity, other patients may have more complex routings, as illustrated by patient B in Table 4 and Figure 2. Patient B is among the very frequent users (between 2012 and 2017), with 41 ambulance rides and diverse health care needs as indicated by the specialties attending to the patient's treatment, thus relying on several health care providers.

Patient Group Level: "Unknown" Patients

Figure 3 and Figure 4 show the workload distribution for frail elderly users who did not belong to a known category for 2012 (326 patients) and 2017 (487 patients; Table 2, see column B-other). Only the health care providers involved in at least 30 treatments and had arcs with frequencies of at least 8 are shown. The number of treatments provided by all health care providers rose considerably between 2012 (Figure 3) and 2017 (Figure 4), although the growth rate is quite different across health care providers. This is paralleled by the higher connectivity among health care providers in 2017, as indicated by the arc frequencies and new arcs (Figure 4, see arcs marked in red). However, developments were not necessarily unidirectional, as connections may disappear over the years (Figure 3, see arcs marked green).

Figure 3. Process maps of 2012 aggregated routings for frail elderly patients (non-dialysis, radiation therapy, or hyperbaric medicine). EMS-S&T: emergency medical services "see and treat."

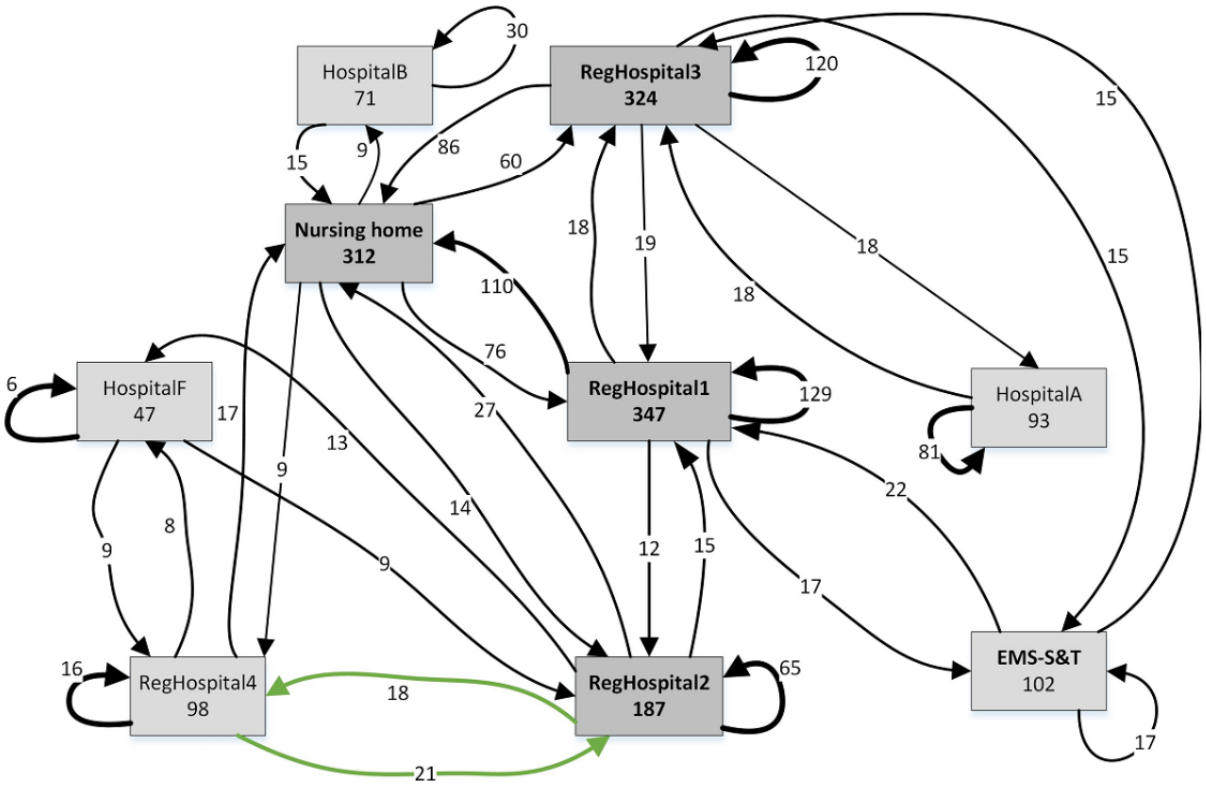
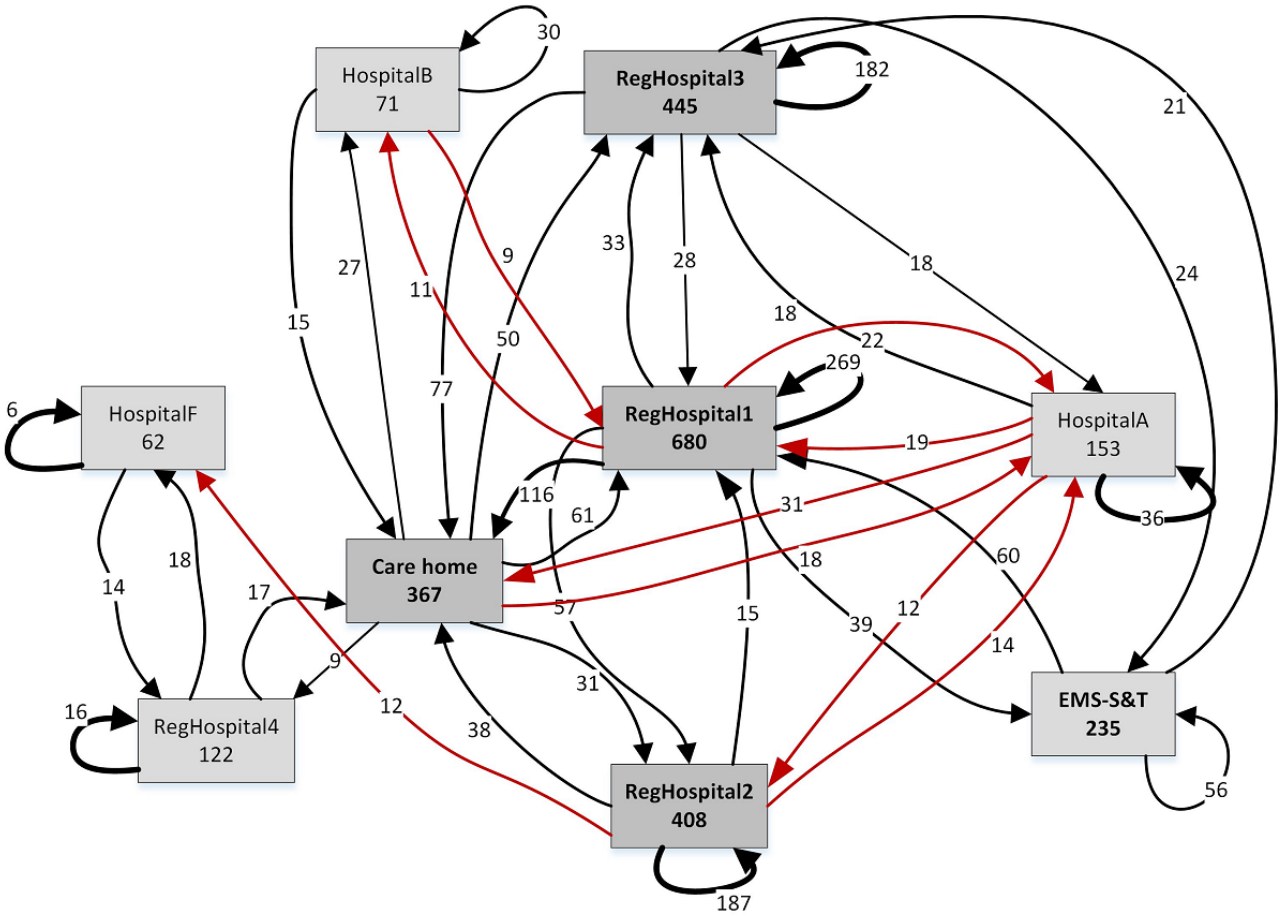


Figure 4. Process maps of 2017 aggregated routings for frail elderly patients (non-dialysis, radiation therapy, or hyperbaric medicine). EMS-S&T: emergency medical services "see and treat."



Discussion

Principal Findings

Frequent users, defined as patients making repeated calls to regional health care services, have a high impact on health care capacities and resource management. Tracing frequent users and their (shared) consumption patterns may be instrumental in regional policymaking. In this study, we combined EMS records and process mining in the Dutch province of Drenthe to trace frequent users and discern different types of users. The approach allowed us to develop and assess patient routings along various regional health care providers by combining their records on ambulance rides. This study demonstrates that this approach can effectively and efficiently trace and quantify frequent users and assess their consumption patterns. Considering frail elderly users as a focal group, the effectiveness of the approach benefits from its broad scope, spanning a large group of health care providers, including nursing homes. Moreover, using EMS records as a single source of data guaranteed the availability of data on all regional hospitals. Notably, data collection and analysis costs were low as the approach relied on a single data source that is routinely collected and the use of automated data analysis by process mining.

Process mining of EMS records confirms the relevance and impact of frail elderly users as a subgroup of frequent users [6-8], representing over 50% of frequent users and meeting a threshold of 4 calls in 1 year. Moreover, the number of frail elderly users and the number of calls they made to health care providers significantly increased during the observation period. Strikingly, the largest growth in frail elderly users was observed in 2014 and 2015, increasing by 100 patients a year, from 320 in 2013 to 548 patients in 2015 (71%). Also, the urgency of their calls increased as indicated by a steep increase in the number of urgent rides in this category (from 664 in 2013 to 1478 in 2015, 123%), substantially exceeding the background annual growth rate of EMS transports of 5.1% over the observation period [30,36].

Our observations parallel structural changes in the Dutch healthcare system of elderly health care (Wet Maatschappelijke Ondersteuning 2015) [37]. Closure of health care homes and stricter health care need entry requirements for nursing homes forced the elderly to continue living independently in their homes for as long as possible [32,38,39]. Although the respective transformations intended to reduce chronic health care costs, they resulted in more hospital admissions and acute situations that were no longer manageable in the home setting, which indeed may be considered counterproductive. We feel the structural change in the trend observed in EMS figures for 2014 and 2015 may indicate an unwanted and unexpected impact of national policy changes. Quantifying the health care consumption of the frail elderly population may be helpful to

policymakers by showing their impact on the system and revealing the urgency to address their needs. Furthermore, understanding the health care consumption of the frail elderly can lead to discussions about residential and other health care homes or other forms of home health care. Our findings reveal a need for further action, such as capacity extensions, tailored home health care services, or advanced health care planning to improve elderly health care and its coordination.

Based on process mining, the structural analysis of patient consumption patterns revealed 2 patient groups: low-complexity "known" patients who require DRH and "unknown" patients often linked to complex routings and the use of several regional health care providers. Importantly, while the known group of frail elderly patients exhibited a growth rate mirroring demographic changes, the unknown group exhibited growth at a much higher pace, implying increased and unpredictable workloads. The workload related to the latter group tends to be increasingly distributed over the several regional health care providers, requiring them to become better connected over time to provide the best health care for the patients they jointly serve. This may be explained by ongoing regional specialization, calling for regional coordination in identifying and addressing patient needs and managing capacities. As a result, process mining further identifies the health care providers involved and the nature of their involvement in terms of specialties.

Of patients with high call frequencies, we evaluated 2 specific cases and noted that some of these patients might be considered low complex. For example, patient A needed frequent dialysis treatments but was mainly served by a single health care provider. Alternatively, patient B was associated with multiple health care providers. The latter patient may benefit from scrutinizing their treatment plan and organizing advanced health care planning if deemed necessary. However, using this method to identify such cases would assume that potential privacy issues are recognized and resolved, which is beyond the scope of this paper.

Policymakers and clinicians may use the results of our analyses to engage in discussions or assess the current standard of care. Our results indicate that frequent users with no clear indication, such as dialysis, are "shopping and hopping," representing an unmet need while utilizing excessive resources. The onus is on general practitioners and nursing home specialists to address this challenge. Advanced care planning and timely and appropriate care at the right location for this category of frequent users might enhance their quality of life while saving scarce resources. Providing sound evidence for the latter would require a different type of study.

Limitations

This study also has limitations. Firstly, only the frail elderly were studied at some depth, using EMS records for only one province in the Netherlands. Nevertheless, while health care consumption patterns are likely to be affected by regional characteristics, it is expected that the success of the proposed approach is not dependent on the latter. Secondly, EMS records only include frequent users who are not capable of self-transport. Thus, frequent users who do not or rarely use EMS will not be traced by the proposed approach. Thirdly, as it is explorative, the paper highlights the potential of the proposed approach in tracing frequent users and enhancing regional policymaking. Ongoing and future research should be directed toward confirming and expanding the method, including comparisons with alternative approaches.

Conclusions

The combined use of EMS data and process mining allows for the effective and efficient tracing of frequent users of health care services. The approach supports regional policymakers and clinicians by quantifying and detailing frequent user consumption patterns to support subsequent policy adaptations.

Acknowledgments

The authors would like to thank Mrs. Harriëtte Holt and Mr. Jaap Hatenboer from the University Medical Center Groningen Ambulancezorg for providing input on the manuscript as domain experts and Coert Schrijver and Roy Huisman for their technical support.

Authors' Contributions

LM, DJZ, and EB contributed to the concept and design of the study and critical revision of the manuscript for important intellectual content. LM and DJZ analyzed and interpreted the data and drafted the manuscript. LM completed the data acquisition and provided statistical expertise.

Conflicts of Interest

None declared.

References

1. Hudon C, Courteau J, Krieg C, Vanasse A. Factors associated with chronic frequent emergency department utilization in a population with diabetes living in metropolitan areas: a population-based retrospective cohort study. BMC Health Serv Res 2017 Aug 04;17(1):525 [FREE Full text] [doi: 10.1186/s12913-017-2453-3] [Medline: 28778161]
2. Urbanoski K, Cheng J, Rehm J, Kurdyak P. Frequent use of emergency departments for mental and substance use disorders. Emerg Med J 2018 Jan 08;35(4):220-225. [doi: 10.1136/emermed-2015-205554]

3. Seguin J, Osmanlliu E, Zhang X, Clavel V, Eisman H, Rodrigues R, et al. Frequent users of the pediatric emergency department. CJEM 2018 May;20(3):401-408. [doi: 10.1017/cem.2017.15] [Medline: 28382879]
4. LaCalle E, Rabin E. Frequent users of emergency departments: the myths, the data, and the policy implications. Ann Emerg Med 2010 Jul;56(1):42-48. [doi: 10.1016/j.annemergmed.2010.01.032] [Medline: 20346540]
5. Cook LJ, Knight S, Junkins EP, Mann NC, Dean JM, Olson LM. Repeat patients to the emergency department in a statewide database. Acad Emerg Med 2004 Mar;11(3):256-263 [FREE Full text] [doi: 10.1111/j.1553-2712.2004.tb02206.x] [Medline: 15001405]
6. Dollard J, Harvey G, Dent E, Trotta L, Williams N, Beilby J, et al. Older People Who Are Frequent Users of Acute Care: A Symptom of Fragmented Care? A Case Series Report on Patients' Pathways of Care. J Frailty Aging 2018;7(3):193-195. [doi: 10.14283/10.14283/jfa.2018.12] [Medline: 30095151]
7. Komenda P, Tangri N, Klajncar E, Eng A, Di Nella M, Hiebert B, et al. Patterns of emergency department utilization by patients on chronic dialysis: A population-based study. PLoS ONE 2018 Apr 17;13(4):e0195323. [doi: 10.1371/journal.pone.0195323]
8. Legramante JM, Morciano L, Lucaroni F, Gilardi F, Caredda E, Pesaresi A, et al. Frequent Use of Emergency Departments by the Elderly Population When Continuing Care Is Not Well Established. PLoS ONE 2016 Dec 14;11(12):e0165939. [doi: 10.1371/journal.pone.0165939]
9. Burns TR. Contributing factors of frequent use of the emergency department: A synthesis. International Emergency Nursing 2017 Nov;35:51-55. [doi: 10.1016/j.ienj.2017.06.001]
10. Rutschmann OT, Chevalley T, Zumwald C, Luthy C, Vermeulen B, Sarasin FP. Pitfalls in the emergency department triage of frail elderly patients without specific complaints. Swiss Med Wkly 2005 Mar 05;135(9-10):145-150. [Medline: 15832233]
11. Goldstein J, McVey J, Ackroyd-Stolarz S. The Role of Emergency Medical Services in Geriatrics: Bridging the Gap between Primary and Acute Care. CJEM 2015 Aug 18;18(1):54-61. [doi: 10.1017/cem.2015.73]
12. Kue R, Ramstrom E, Weisberg S, Restuccia M. Evaluation of an emergency medical services-based social services referral program for elderly patients. Prehosp Emerg Care 2009;13(3):273-279. [doi: 10.1080/10903120802706179] [Medline: 19499461]
13. Chan JS, Tin AS, Chow WL, Tiah L, Tiru M, Lee CE. Frequent attenders at the emergency department: an analysis of characteristics and utilisation trends. Proceedings of Singapore Healthcare 2017 Jun 13;27(1):12-19. [doi: 10.1177/2010105817715271]
14. Edwards MJ, Bassett G, Sinden L, Fothergill RT. Frequent callers to the ambulance service: patient profiling and impact of case management on patient utilisation of the ambulance service. Emerg Med J 2015 May 13;32(5):392-396. [doi: 10.1136/emermed-2013-203496] [Medline: 25312857]
15. Scott J, Strickland AP, Warner K, Dawson P. Describing and predicting frequent callers to an ambulance service: analysis of 1 year call data. Emerg Med J 2014 May;31(5):408-414. [doi: 10.1136/emermed-2012-202146] [Medline: 23413152]
16. Brown JF, Raven MC, Tangherlini NL, Kennedy Hall M. Frequent Users of 9-1-1 Emergency Medical Services: Sign of Success or Symptom of Impending Failure? Prehosp Emerg Care 2018 Aug 21:1-3. [doi: 10.1080/10903127.2018.1475531] [Medline: 30130427]
17. Mans RS, van der Aalst WMP, Vanwersch RJB. Process Mining in Healthcare: Evaluating and Exploiting Operational Healthcare Processes. New York, US: SpringerBriefs in Business Process Management; 2015.
18. Rebuge Á, Ferreira DR. Business process analysis in healthcare environments: A methodology based on process mining. Information Systems 2012 Apr;37(2):99-116. [doi: 10.1016/j.is.2011.01.003]
19. Rovani M, Maggi FM, de Leoni M, van der Aalst WM. Declarative process mining in healthcare. Expert Systems with Applications 2015 Dec;42(23):9236-9251. [doi: 10.1016/j.eswa.2015.07.040]
20. Delias P, Doumpos M, Grigoroudis E, Manolitzas P, Matsatsinis N. Supporting healthcare management decisions via robust clustering of event logs. Knowledge-Based Systems 2015 Aug;84:203-213. [doi: 10.1016/j.knosys.2015.04.012]
21. Pika A, Wynn MT, Budiono S, Ter Hofstede AHM, van der Aalst WMP, Reijers HA. Privacy-Preserving Process Mining in Healthcare. Int J Environ Res Public Health 2020 Mar 02;17(5):1-28 [FREE Full text] [doi: 10.3390/ijerph17051612] [Medline: 32131516]
22. Pereira GB, Santos EAP, Maceno MMC. Process mining project methodology in healthcare: a case study in a tertiary hospital. Netw Model Anal Health Inform Bioinforma 2020 Apr 15;9(28):1-14. [doi: 10.1007/s13721-020-00227-w]
23. Lu F, Li P, Bao Y, Liu C, Zeng Q. Death Risk Prediction of Intensive Care Unit Patients Combined with Treatment Process Mining. J Med Imaging Hlth Inform 2020 Jul 01;10(7):1754-1762. [doi: 10.1166/jmihi.2020.3077]
24. Duma D, Aringhieri R. An ad hoc process mining approach to discover patient paths of an Emergency Department. Flex Serv Manuf J 2018 Dec 7;32(1):6-34. [doi: 10.1007/s10696-018-9330-1]
25. Orellana A, Davila C, Leon I. Detection of Bottlenecks in Hospital Processes from the XAVIA HIS System using Process Mining. IEEE Latin Am. Trans 2019 Jan 21;17(01):18-25. [doi: 10.1109/TLA.2019.8826690] [Medline: 5637250]
26. Ibanez-Sanchez G, Fernandez-Llatas C, Martinez-Millana A, Celda A, Mandingorra J, Aparici-Tortajada L, et al. Toward Value-Based Healthcare through Interactive Process Mining in Emergency Rooms: The Stroke Case. IJERPH 2019 May 20;16(10):1783. [doi: 10.3390/ijerph16101783]

27. Kurniati A, Hall G, Hogg D, Johnson O. Process mining to explore variation in chemotherapy pathways for breast cancer patients. In: British Journal of Cancer.: Springer; 2018 Presented at: NCRI Cancer Conference; November 4-6; Glasgow, Scotland p. 16. [doi: doi.org/10.1038/s41416-018-0299-z]
28. Overheid.nl. [translated from Dutch: General Act Special Medical Expenses (A.W.B.Z.)]. Algemene Wet Bijzondere Ziektekosten. 2014. URL: https://wetten.overheid.nl/BWBR0002614/2014-07-16 [accessed 2021-07-01]
29. Overheid.nl. Wet langdurige zorg. 2021. URL: https://wetten.overheid.nl/BWBR0035917/2021-01-01 [accessed 2021-07-01]
30. Nederlandse_Zorgautoriteit. Monitor acute zorg. 2018. URL: https://puc.overheid.nl/nza/doc/PUC_260889_22/1/ [accessed 2021-07-01]
31. Overheid.nl. Wet medisch-weteschappelijk onderzoek. 2021. URL: https://wetten.overheid.nl/BWBR0009408/2021-07-01
32. Koopmans RT, Pellegrom M, van der Geer ER. The Dutch Move Beyond the Concept of Nursing Home Physician Specialists. J Am Med Dir Assoc 2017 Sep 01;18(9):746-749. [doi: 10.1016/j.jamda.2017.05.013] [Medline: 28668662]
33. Eurostat. Eurostat - Population projections. 2019. URL: https://ec.europa.eu/eurostat/web/population-demography/population-projections [accessed 2021-07-01]
34. van der Aalst W. Process Mining. ACM Trans. Manage. Inf. Syst 2012 Jul;3(2):1-17. [doi: 10.1145/2229156.2229157]
35. Fluxicon. Disco. 2019. URL: https://fluxicon.com/disco/ [accessed 2021-07-01]
36. Rijksoverheid.nl. Kamerbrief over stand van zaken Ambulancezorg en rol in acute zorg. 2019. URL: https://www.rijksoverheid.nl/documenten/kamerstukken/2019/10/01/kamerbrief-over-toegankelijkheid-acute-zorg [accessed 2021-07-01]
37. Overheid.nl. Wet maatschappelijke ondersteuning 2015. 2015. URL: https://wetten.overheid.nl/BWBR0035362/2021-07-01
38. Kroneman M, Boerma W, van den Berg M, Groenewegen P, de Jong J, van Ginneken E. Netherlands: Health System Review. Health Syst Transit 2016 Mar;18(2):1-240 [FREE Full text] [Medline: 27467715]
39. Maarse JAMH, Jeurissen PPP. The policy and politics of the 2015 long-term care reform in the Netherlands. Health Policy 2016 Mar;120(3):241-245. [doi: 10.1016/j.healthpol.2016.01.014] [Medline: 26872702]

Abbreviations

ALS: advanced life support
DRH: dialysis, radiation therapy, or hyperbaric medicine
ED: emergency department
EMS: emergency medical services

Edited by R Kukafka; submitted 27.01.21; peer-reviewed by R van Poelgeest, B Jugdutt; comments to author 20.03.21; revised version received 02.04.21; accepted 21.06.21; published 06.10.21.

Please cite as:
Maruster L, van der Zee DJ, Buskens E
Identifying Frequent Health Care Users and Care Consumption Patterns: Process Mining of Emergency Medical Services Data
J Med Internet Res 2021;23(10):e27499
URL: https://www.jmir.org/2021/10/e27499
doi:10.2196/27499
PMID:34612834

Original Paper

Prediction of Readmission in Geriatric Patients From Clinical Notes: Retrospective Text Mining Study

Kim Huat Goh[1], PhD; Le Wang[2], PhD; Adrian Yong Kwang Yeow[3], PhD; Yew Yoong Ding[4,5], MBBS, MPH, PhD; Lydia Shu Yi Au[6], MBBS, MMed; Hermione Mei Niang Poh[7], BSc (Hons); Ke Li[7], MD; Joannas Jie Lin Yeow[7], BBus; Gamaliel Yu Heng Tan[6,7], MBBS, MMed

[1]Nanyang Business School, Nanyang Technological University, Singapore, Singapore
[2]City University of Hong Kong, Hong Kong, Hong Kong
[3]School of Business, Singapore University of Social Sciences, Singapore, Singapore
[4]Tan Tock Seng Hospital, Singapore, Singapore
[5]Geriatric Education and Research Institute, Singapore, Singapore
[6]Ng Teng Fong General Hospital, Singapore, Singapore
[7]Medical Informatics, National University Health System, Singapore, Singapore

Corresponding Author:
Kim Huat Goh, PhD
Nanyang Business School
Nanyang Technological University
S3-B2A-34, 50 Nanyang Avenue
Singapore, 639798
Singapore
Phone: 65 67904808
Email: akhgoh@ntu.edu.sg

Abstract

Background: Prior literature suggests that psychosocial factors adversely impact health and health care utilization outcomes. However, psychosocial factors are typically not captured by the structured data in electronic medical records (EMRs) but are rather recorded as free text in different types of clinical notes.

Objective: We here propose a text-mining approach to analyze EMRs to identify older adults with key psychosocial factors that predict adverse health care utilization outcomes, measured by 30-day readmission. The psychological factors were appended to the LACE (Length of stay, Acuity of the admission, Comorbidity of the patient, and Emergency department use) Index for Readmission to improve the prediction of readmission risk.

Methods: We performed a retrospective analysis using EMR notes of 43,216 hospitalization encounters in a hospital from January 1, 2017 to February 28, 2019. The mean age of the cohort was 67.51 years (SD 15.87), the mean length of stay was 5.57 days (SD 10.41), and the mean intensive care unit stay was 5% (SD 22%). We employed text-mining techniques to extract psychosocial topics that are representative of these patients and tested the utility of these topics in predicting 30-day hospital readmission beyond the predictive value of the LACE Index for Readmission.

Results: The added text-mined factors improved the area under the receiver operating characteristic curve of the readmission prediction by 8.46% for geriatric patients, 6.99% for the general hospital population, and 6.64% for frequent admitters. Medical social workers and case managers captured more of the psychosocial text topics than physicians.

Conclusions: The results of this study demonstrate the feasibility of extracting psychosocial factors from EMR clinical notes and the value of these notes in improving readmission risk prediction. Psychosocial profiles of patients can be curated and quantified from text mining clinical notes and these profiles can be successfully applied to artificial intelligence models to improve readmission risk prediction.

(J Med Internet Res 2021;23(10):e26486) doi:10.2196/26486

KEYWORDS

geriatrics; readmission risk; artificial intelligence; text mining; psychosocial factors

Introduction

Background

Hospital readmission of older adults is a significant challenge for the individual, caregivers, and health system. For individuals, readmissions can be distressing, may compromise quality of care, and increase the risk of adverse health outcomes. For caregivers, readmission is often burdensome and increases their health care spending. As for health systems, readmissions often cause resource demands and financial costs to escalate [1]. The 30-day readmission rate among patients aged 65 years or older in Singapore has been reported to be 19% [2], which is comparable to the readmission rate of Medicare patients in the United States, most of whom are older adults [3]. Significant risk factors for hospital readmission in adults aged 65 years and older include (a) sociodemographic factors such as higher age, male gender, ethnicity, and poor living conditions; (b) health-related factors such as poor overall condition, comorbidity, functional disability, and recent hospital admissions; and (c) organizational factors such as prolonged length of stay in the index hospitalization and discharge destination [4,5]. These risk factors have been used extensively in predictive models for hospital readmission by health service researchers worldwide [6-10]. Recently, other readmission predictors such as those in the psychosocial domain have begun to receive more attention.

Psychosocial factors can be defined as "the combination and interplay of psychological and social factors that potentially influence health, injury, illness, and disease" [11]. However, a review of the medical literature suggests that different medical specialties have slightly different definitions of psychological factors [11-17]. Based on the various factors identified in earlier studies, we observed that psychosocial factors can be divided into three relevant dimensions: (1) individual psychological well-being, (2) social structures, and (3) resources. Individual psychological well-being factors include psychological conditions such as mood [11,18], attitude [11,19], coping mechanism [11,17], depression [15,16,20], perceived control [13,19], and psychological distress [16,17,21]. Social structures represent the conditions of the environment in which the individual lives, including support structures [11,14,16,17], social relationships [14,18], social norms [19], and family life [22]. Finally, resources represent the means available to the individual, such as financial means, accessibility to health care [13,14], and the health service system [19].

Prior research has shown that these factors—depressive symptoms [23], poor social support, and financial stress—contribute to hospital readmission for specific patient subgroups such as those with chronic obstructive lung disease, chronic kidney disease, and heart failure [24-26]. In general, psychosocial factors could play a significant role in the hospital readmission of older adults and account for a significant proportion of the readmission risk. At the same time, psychosocial factors are indicators of a patient's complex needs that are amenable to tailored care interventions. Such interventions can improve the patient's clinical outcomes and reduce the utilization of health care resources.

Literature Review

There are two conceptual models in the extant literature that link psychosocial factors to hospital readmissions for older adults. The first is Andersen's [27] Behavioral Model of Health Services Use that posits an individual's use of health services as a function of predisposing, enabling, and need factors. Psychosocial factors (ie, individual-level and structural-level variables) can be categorized as the model's predisposing and enabling factors, respectively. The other is Adler and Stewart's [28] Pathways Linking Socioeconomic Status and Health model, which suggests that environmental resources and constraints, as well as psychological influences are mechanisms that lead to health outcomes such as hospital readmission. Individual-level and structural-level psychosocial factors map to the model's psychological and environmental variables.

In contrast to the numerous clinically related risk factors that are stored as structured data in electronic medical records (EMRs), most psychosocial factors are recorded as free text in the patient's clinical notes such as the initial and progress clinical notes of physicians, allied health professionals, case managers, and social workers. Such unstructured textual data in the EMR represent a potentially rich and untapped source of data related to patients' psychosocial factors. The manual extraction of psychosocial keywords from unstructured data is challenging and impractical given the copious and ever-increasing amount of clinical notes recorded in a typical EMR system. As such, there have been systematic efforts by clinicians to capture social and behavioral data, including psychosocial information, as structured data in EMR systems [29]. However, the effectiveness of these efforts in different health care contexts remains unclear. At the same time, other researchers have begun to apply text-mining techniques to efficiently extract and analyze unstructured text data in EMR clinical notes to identify these psychosocial factors.

Text-mining techniques represent a broad range of approaches for analyzing and processing semistructured and unstructured text data to construct structured data. By using powerful algorithms applied to large textual documents such as those typically found in EMR systems, text mining can "turn text into numbers" to be used for further analysis. Topic modeling, which is a specific domain in text mining that examines individual words to identify common topics and concepts, holds significant promise for extracting psychosocial factors from EMR clinical notes.

To date, only a few text-mining studies have set out to identify individuals with psychosocial factors using EMR data [30]. As such, we have limited evidence on the effectiveness of extracting psychosocial information from EMRs for the purpose of secondary health care research or routine clinical care.

Objective

This study proposes a text-mining approach to identify older adults with key psychosocial factors obtained from clinical notes to help predict adverse health and health care utilization outcomes. To validate the efficacy of including psychological factors in the predictive model, we append these psychosocial factors to the commonly used LACE (Length of stay, Acuity

of the admission, Comorbidity of the patient, and Emergency department use) Index for Readmission [31] to improve readmission risk prediction accuracy on an independent, hold-out sample of patients.

Methods

Design

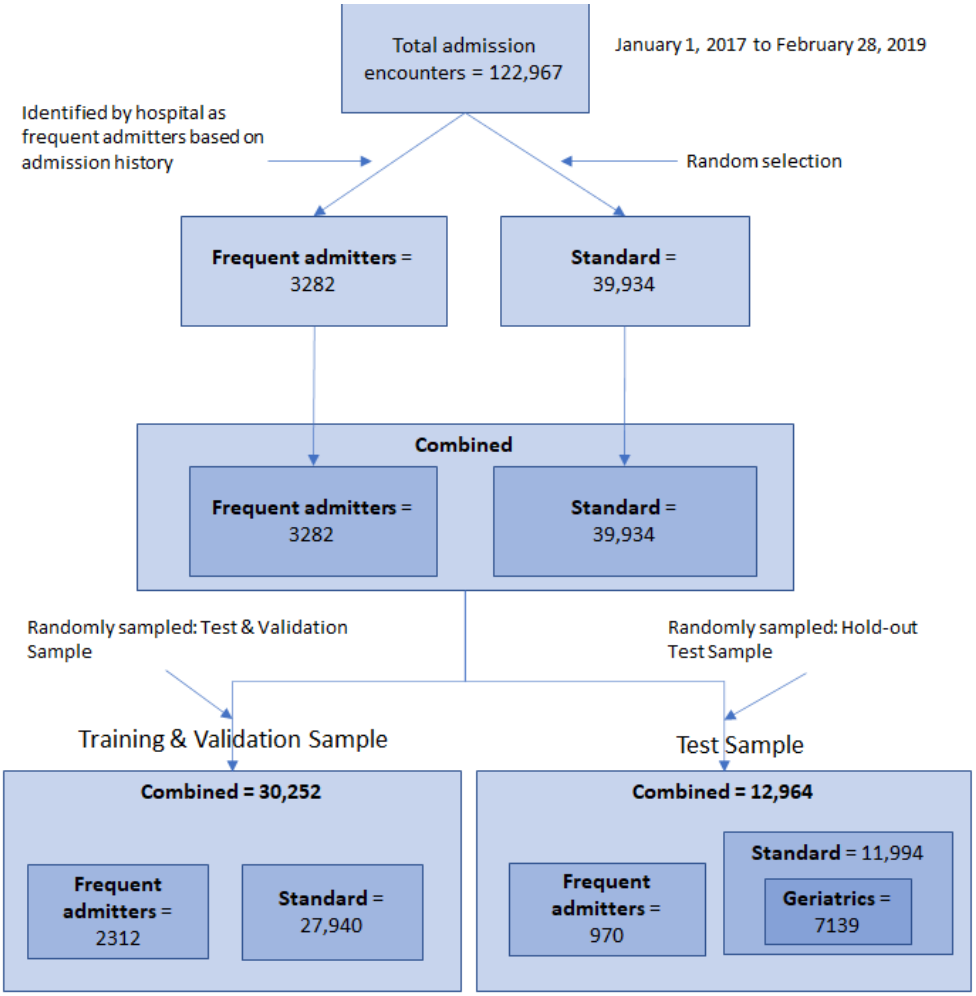
The study was a retrospective analysis of EMR data captured by the EPIC system over a 26-month period from January 1, 2017 to February 28, 2019. Ethical approval was provided by the Domain-Specific Review Board of the National Healthcare Group, Singapore (2018/01072).

Settings and Data Context

The sample consists of 9393 patients with 43,216 admission encounters in a 26-month period from all wards in Ng Teng Fong General Hospital, Singapore. Each clinical record was classified by the role of the author. In this sample, clinical records were authored by physicians, medical social workers, or case managers. Specifically, medical social workers and case managers are assigned to some patients who may require additional social support upon hospital admission. The dataset consists of two cohorts of patients. The first cohort includes 892 patients (3282 admission encounters) identified by the hospital as frequently readmitted patients ("Frequent Admitters" cohort). This cohort consists of patients who (1) are frequently admitted to the hospital despite having their acute medical needs met, (2) have medical conditions that require multidisciplinary care, (3) show signs of caregiver stress, (4) encounter frequent falls (more than two falls in the last 12 months) and require functional management at home, or (5) face medication management issues (eg, noncompliance to their medication regime). The second cohort consists of 9377 randomly selected patients (39,934 admission encounters) admitted to the hospital ("Standard" cohort). The "Standard" cohort consists of patients admitted to the hospital's inpatient wards during the sampling period. The purpose of including the "frequent admitters" cohort was to oversample the frequent readmission cases and facilitate the training of the text-mining algorithm to extract psychosocial topics often associated with readmission risks. This type of oversampling method is commonly applied in health care research to train machine-learning algorithms [32,33]. The combined cohorts were randomly split into a training dataset and a hold-out/test dataset to ensure that both the training and test dataset had similar distributions of patients in the "frequent admitters" and "standard" cohorts. The training dataset comprised 30,252 admission encounters and the hold-out/test dataset comprised 12,964 admission encounters. The unit of analysis is each admission encounter.

We used the 10-fold cross-validation method to train and validate the model with the training dataset. The validated model was subsequently tested against the hold-out/test dataset in four ways. First, we used the "Combined" test dataset, which is the dataset compiled to have a similar distribution as the training dataset, containing proportionally more "frequent admitters" to ensure that the model was tested using a similar distribution of patients as that used to train the model. Second, we used the "Standard" test dataset, which was the sample randomly drawn among patients from the hospital; this sample represents a typical patient that a hospital encounters. We used this dataset to test for the generalizability of the model and to rule out overfitting. Third, we used a "Frequent" test dataset, consisting of "frequent admitters," which was mainly used to test the fit of the model in predicting frequent admitters, as a key concern for many hospitals. Finally, we used a "Geriatrics" test dataset, comprising only geriatric patients (≥65 years old selected from the "standard" cohort test dataset), to test if the model works for the geriatric specialty where it is likely to be deployed. Additional details and procedures of cohort selection are shown in Figure 1.

Figure 1. Sampling methodology. All values represent the admission encounters. For the "frequent admitters" cohort (3282 encounters), there are 892 unique patients and for the "standard" cohort (39,934 encounters), there are 9377 unique patients. The "geriatrics" cohort represents a subsample of patient encounters within the "standard" cohort where the age of the patient at time of encounter is greater or equal to 65 years.



Data Processing and Algorithm Development

For *each* admission encounter, we combined the clinical notes written by authors with similar roles (eg, all notes written by physicians were combined as physician's notes). The notes were combined based on the author's role (physician, medical social worker, and case manager) because each role would potentially document similar issues. Hence, it is more efficient to mine the unstructured clinical notes for each role to identify common or similar topics. We combined the notes for each admission encounter instead of analyzing each note entry as the unit of analysis because a patient's psychosocial conditions are less likely to vary for each admission encounter.

We then applied natural language processing text mining to the clinical notes in the training dataset. We used the latent Dirichlet allocation (LDA) topic modeling algorithm to extract the common topics present in the clinical notes and then numerically weighed each topic's intensity (loadings) in the clinical notes. A vector of lexicographically related words represents each *topic* due to the frequent occurrence of these words in proximity across different notes. A high *loading* value represents the presence of the topic in the clinical note. This process was performed separately for the physician, medical social worker, and case manager notes. A total of 100 topics were extracted from each set of notes based on the clinician's role (ie, physician, medical social worker, case manager).

Two geriatric specialists reviewed and classified these 100 topics into broader themes, specifically dividing them into psychosocial issues or nonpsychosocial-related issues. Additionally, we conducted four interviews with a group of medical social workers and case managers to triangulate if this classification is appropriate. It is important to note that this added classification into broader themes by clinicians is solely to facilitate the reporting and interpretation of results. These broader themes were *not used* in the subsequent development of the readmission risk model, and only the LDA classification loadings were used in the training of the readmission risk model. Further details of the text mining procedure are provided in Multimedia Appendix 1.

We combined the topic's intensity (loadings) for each set of notes with structured predictors of readmission established in the LACE Index for Readmission as predictors for estimating readmission risk. As readmission risk is a function of various factors beyond psychosocial factors, we incorporated the LACE index to take into account some of the factors reported in the literature. The LACE index is a score commonly used to predict a patient's 30-day hospital readmission risk [31]. The index consists of the following variables: (1) the length of stay (L),

(2) the acuity of the current or previous admission (A), (3) comorbidities of the patient as measured by the Charlson Comorbidity Index score (C), and (4) the number of visits to the emergency department in the preceding 6 months (E).

The readmission risk model was fitted using the gradient boosting trees (GBT) algorithm to predict the outcome of readmission within the next 30 days from the discharge date of the current admission. GBT uses an ensemble of multiple trees to generate more accurate prediction models for classification and regression. The algorithm's premise is to build a series of trees, where each tree is trained with the objective to correct the misclassification errors of the previous tree in the series.

We tested the model's predictive accuracy using the four different hold-out test samples described above. To assess the predictive value of the clinical notes, we fitted a LACE baseline readmission model *without* using the topics from the notes. We then compared this baseline model against models that include the physician notes and social notes (ie, medical social worker notes and case manager notes) jointly and separately.

Results

Evaluating the Predictive Value of Psychosocial Information

As expected, we observed that physicians record fewer psychosocial issues than medical social workers and case managers (Table 1). The more detailed distribution of the specific topics extracted is provided in Tables A1-A3 of Multimedia Appendix 1.

The descriptive statistics of the variables used in the readmission risk model for each test cohort are provided in Table 2.

Table 1. Distribution of psychosocial topics (N=100).

Role of author	Proportion of psychosocial topics, n (%)	Proportion of nonpsychosocial topics, n (%)
Physician	25 (25)	75 (75)
Medical social worker	100 (100)	0 (0)
Case manager	88 (88)	12 (12)

Table 2. Descriptive statistics of variables in the LACE (Length of stay, Acuity of the admission, Comorbidity of the patient, and Emergency department use) readmission model (patient encounter level).

Variable	Frequent cohort[a], mean (SD)	Standard cohort[b], mean (SD)	Geriatrics cohort[c], mean (SD)	Combined cohort[d], mean (SD)
Age (years)	72.94 (13.24)	67.07 (15.98)	77.62 (8.09)	67.51 (15.87)
Gender (1: Male, 0: Female)	0.50 (0.50)	0.55 (0.50)	0.50 (0.50)	0.54 (0.50)
Length of stay (days)	6.73 (13.18)	5.47 (10.15)	6.65 (10.81)	5.57 (10.41)
Charlson Comorbidity Index	0.47 (1.39)	0.41 (1.18)	0.49 (1.38)	0.42 (1.20)
Emergency department admission (1: Yes, 0: No)	0.50 (0.50)	0.57 (0.50)	0.58 (0.49)	0.56 (0.50)
Intensive care unit stay (1: Yes, 0: No)	0.03 (0.17)	0.05 (0.22)	0.04 (0.21)	0.05 (0.22)
Emergency department visits in last 6 months	2.86 (3.07)	1.39 (2.77)	1.47 (2.44)	1.50 (2.82)

[a]Patients identified by the hospital as frequent readmission patients.

[b]Sample of a typical hospital patient.

[c]Subset of patients in the "Standard" sample who are 65 years of age or older.

[d]Combination of the "Frequent" and "Standard" samples.

The area under the receiver operating characteristic curve (AUROC) of the LACE baseline predictive model ranged from 0.8288 to 0.8397 (Table 3) for the four different test cohorts (Frequent, Standard, Geriatrics, and Combined). The baseline model only considered common factors identified in the prior literature associated with readmission risks and did not include psychosocial factors extracted from the clinical notes. The receiver operating characteristic curve is a plot representing the diagnostic ability of a binary classifier while varying the discriminatory threshold (ie, the cut-off value to reclassify one state to the other). With varying discriminatory threshold values, the different sets of true positive rate (sensitivity) are plotted against the corresponding false positive rates (1–specificity). Thus, AUROC is a representation of the overall performance of the classifier.

Adding the text-mined notes from the medical social workers and case managers increased the AUROC of the model to 0.8573-0.8707. Further appending the clinical notes from physicians increased the AUROC to 0.8952-0.9100.

Table 3. Results of the readmissions prediction model.

Model	AUROC[a]	Sensitivity	Specificity	PPV[b]	NPV[c]
LACE[d] baseline					
Frequent[e]	0.8288	0.7021	0.7840	0.7466	0.7438
Standard[f]	0.8302	0.7341	0.7606	0.6649	0.8156
Geriatrics[g]	0.8254	0.7479	0.7328	0.6713	0.7994
Combined[h]	0.8397	0.7303	0.7757	0.6696	0.8221
LACE baseline+social[i]					
Frequent	0.8573	0.7598	0.7573	0.7394	0.7767
Standard	0.8621	0.7661	0.7796	0.6922	0.8375
Geriatrics	0.8686	0.7749	0.7832	0.7228	0.8267
Combined	0.8707	0.7763	0.7825	0.6896	0.8490
LACE baseline+physician[j]+social					
Frequent	0.8952	0.8232	0.8136	0.8001	0.8354
Standard	0.9001	0.8224	0.8235	0.7509	0.8776
Geriatrics	0.9100	0.8318	0.8331	0.7843	0.8716
Combined	0.9069	0.8254	0.8318	0.7534	0.8845

[a]AUROC: area under the receiving operating characteristic curve.

[b]PPV: positive predictive value.

[c]NPV: negative predictive value.

[d]LACE: Length of stay, Acuity of the admission, Comorbidity of the patient, and Emergency department use.

[e]Hold-out sample of patients identified by the hospital as frequent readmission patients.

[f]Hold-out sample of a typical hospital patient.

[g]Subset of patients in the "Standard" sample who are 65 years or older.

[h]Combination of "Frequent" and "Standard" hold-out samples.

[i]Social represents the text-mined notes that medical social workers and case managers provided.

[j]Physician represents the text-mined notes provided by physicians.

Comparison Across Patient Profiles

The addition of textual information improved the AUROC of the readmission model. This improvement was particularly more significant for geriatric patients than for other cohorts of patients (Table 4). For geriatric patients, notes from the medical social workers and case managers improved the AUROC by 4.32%. Combining these notes with physician notes further improved the AUROC by 8.46% compared with the baseline LACE readmission model.

Table 4. Improvements of prediction (area under receiver operating characteristic curve) over the baseline LACE (Length of stay, Acuity of the admission, Comorbidity of the patient, and Emergency department use) model for different test cohorts.

Notes	Frequent cohort[a]	Standard cohort[b]	Geriatrics cohort[c]	Combined cohort[d]
Social[e]	2.85%	3.19%	4.32%	3.10%
Social and physician[f]	6.64%	6.99%	8.46%	6.72%

[a]Hold-out sample of patients identified by the hospital as frequent readmission patients.

[b]Hold-out sample of a typical hospital patient.

[c]Subset of patients in the "Standard" sample who are 65 years or older.

[d]Combination of "Frequent" and "Standard" hold-out samples.

[e]The readmission model with clinical notes from the medical social worker and case manager.

[f]The readmission model with clinical notes from the medical social workers and case managers.

Discussion

Principal Findings

The AUROC of our readmission risk model was higher than the typical accuracy of readmission predictive models, ranging from 0.66 to 0.83, as reported in an earlier review of 30 studies [6]. The results also suggest that the readmission predictive algorithm's performance for all four cohorts (frequent admitters, standard, geriatrics, and the combination of frequent and standard groups) are relatively similar. Thus, this model can be applied to geriatric patients as the typical pool of patients who require additional management for readmission risks. Further, when taking into account the psychosocial information captured by nonphysicians (ie, medical social workers and case managers) by adding social topics, the prediction accuracy improved by 0.0285-0.0432. When we added the physicians' textual clinical notes, the AUROC further increased by 0.0362-0.0414 in different cohorts.

Overall, the results show that with the addition of text-mined clinical notes from physicians and other clinicians, the AUROC of readmission prediction improves by 0.0664 to 0.0842, suggesting the added benefits of extracting psychosocial information from textual clinical notes in predicting readmission risk.

This study shows that clinicians could leverage natural language processing to gain more information from the EMR system beyond the traditional structured data commonly used to predict readmission risk. Specifically, this study establishes a proof of concept for the use of text-mining techniques with EMR unstructured free text to identify psychosocial predictors of hospital readmission, particularly among geriatric patients. In doing so, our findings support the viability of the psychosocial approach in potentially reducing readmission rates. Thus, our study represents a T2 translational stage (to patients) of research, paving the way toward the T3 translational stage (to practice). In terms of development along the translational pathway, the next phase will focus on proof of value of embedding text-mining techniques in prediction models used to identify the risk of early readmission among hospitalized patients. The purpose of this phase is to perform a comprehensive geriatric assessment for high-risk patients with the goal of offering tailored care management. By managing patients' specific physical and psychosocial needs, we should observe improvement in the quality of care and a reduction in unnecessary health care utilization. In this way, precious heath care resources can be optimally allocated to patients who will obtain the greatest benefit. This strategy is particularly relevant for older hospitalized patients, who are more likely to have unmet psychosocial needs and for whom our augmented risk prediction model performs the best. To achieve proof of value, future research could use quasiexperimental designs to compare the feasibility and effectiveness of a product that combines text-mined psychosocial factors in a state-of-the-art prediction model with those of a product that only has a prediction model.

Beyond the application of text-mining techniques to the prediction of hospital readmission, this study also presents the broader and extended possibility of using the same technical approach developed for the EMR to identify a set of underdiagnosed clinical conditions in older adults, which will have an important influence on their health and health care utilization outcomes.

Conclusion

Psychosocial profiles of patients can be curated and quantified from text mining clinical notes, and these profiles can be successfully applied to artificial intelligence models to predict readmission risks. The use of text mining improved the accuracy of predicting readmission, and this improved predictive accuracy was higher for geriatric patients than for other patient cohorts.

Acknowledgments

This project is funded by the Social Science Research Council, Singapore (grant number MOE2017-SSRTG-030) and the Ageing Research Institute for Society and Education-Geriatric Education & Research Institute, Singapore (grant number AG2018001).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed methods of text mining approach and Tables A1-A3.
[PDF File (Adobe PDF File), 113 KB - jmir_v23i10e26486_app1.pdf]

References

1. Felix HC, Seaberg B, Bursac Z, Thostenson J, Stewart MK. Why do patients keep coming back? Results of a readmitted patient survey. Soc Work Health Care 2015 Jan 14;54(1):1-15 [FREE Full text] [doi: 10.1080/00981389.2014.966881] [Medline: 25588093]
2. Lim E, Matthew N, Mok W, Chowdhury S, Lee D. Using hospital readmission rates to track the quality of care in public hospitals in Singapore. BMC Health Serv Res 2011 Oct 19;11(S1):A16. [doi: 10.1186/1472-6963-11-s1-a16]
3. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med 2009 Apr 02;360(14):1418-1428. [doi: 10.1056/NEJMsa0803563] [Medline: 19339721]

4. Pedersen MK, Meyer G, Uhrenfeldt L. Risk factors for acute care hospital readmission in older persons in Western countries. JBI Database Syst Rev Implement Rep 2017;15(2):454-485. [doi: 10.11124/jbisrir-2016-003267]
5. García-Pérez L, Linertová R, Lorenzo-Riera A, Vázquez-Díaz JR, Duque-González B, Sarría-Santamera A. Risk factors for hospital readmissions in elderly patients: a systematic review. QJM 2011 Aug 10;104(8):639-651. [doi: 10.1093/qjmed/hcr070] [Medline: 21558329]
6. Kansagara D, Englander H, Salanitro A, Kagen D, Theobald C, Freeman M, et al. Risk prediction models for hospital readmission: a systematic review. JAMA 2011 Oct 19;306(15):1688-1698 [FREE Full text] [doi: 10.1001/jama.2011.1515] [Medline: 22009101]
7. Futoma J, Morris J, Lucas J. A comparison of models for predicting early hospital readmissions. J Biomed Inform 2015 Aug;56:229-238 [FREE Full text] [doi: 10.1016/j.jbi.2015.05.016] [Medline: 26044081]
8. Zhou H, Della PR, Roberts P, Goh L, Dhaliwal SS. Utility of models to predict 28-day or 30-day unplanned hospital readmissions: an updated systematic review. BMJ Open 2016 Jun 27;6(6):e011060 [FREE Full text] [doi: 10.1136/bmjopen-2016-011060] [Medline: 27354072]
9. Jamei M, Nisnevich A, Wetchler E, Sudat S, Liu E. Predicting all-cause risk of 30-day hospital readmission using artificial neural networks. PLoS One 2017;12(7):e0181173 [FREE Full text] [doi: 10.1371/journal.pone.0181173] [Medline: 28708848]
10. Low LL, Liu N, Wang S, Thumboo J, Ong MEH, Lee KH. Predicting 30-day readmissions in an Asian population: building a predictive model by incorporating markers of hospitalization severity. PLoS One 2016 Dec 9;11(12):e0167413 [FREE Full text] [doi: 10.1371/journal.pone.0167413] [Medline: 27936053]
11. Rosenberger PH, Jokl P, Ickovics J. Psychosocial factors and surgical outcomes: an evidence-based literature review. J Am Acad Orthop Surg 2006 Jul;14(7):397-405. [doi: 10.5435/00124635-200607000-00002] [Medline: 16822887]
12. Bonde JPE. Psychosocial factors at work and risk of depression: a systematic review of the epidemiological evidence. Occup Environ Med 2008 Jul 01;65(7):438-445. [doi: 10.1136/oem.2007.038430] [Medline: 18417557]
13. Kimmel PL. Psychosocial factors in dialysis patients. Kidney Int 2001 Apr;59(4):1599-1613 [FREE Full text] [doi: 10.1046/j.1523-1755.2001.0590041599.x] [Medline: 11260433]
14. Kronborg H, Vaeth M. The influence of psychosocial factors on the duration of breastfeeding. Scand J Public Health 2004 Sep 05;32(3):210-216. [doi: 10.1080/14034940310019218] [Medline: 15204182]
15. Singh-Manoux A. Psychosocial factors and public health. J Epidemiol Community Health 2003 Aug 01;57(8):553-556; discussion 554 [FREE Full text] [doi: 10.1136/jech.57.8.553] [Medline: 12883050]
16. Strike PC, Steptoe A. Psychosocial factors in the development of coronary artery disease. Prog Cardiovasc Dis 2004 Jan;46(4):337-347. [doi: 10.1016/j.pcad.2003.09.001] [Medline: 14961456]
17. Zaza C, Baine N. Cancer pain and psychosocial factors. J Pain Symptom Manag 2002 Nov;24(5):526-542. [doi: 10.1016/s0885-3924(02)00497-9]
18. Lutgendorf SK, Costanzo ES. Psychoneuroimmunology and health psychology: An integrative model. Brain Behav Immun 2003 Aug;17(4):225-232. [doi: 10.1016/s0889-1591(03)00033-3]
19. Bradley EH, McGraw SA, Curry L, Buckser A, King KL, Kasl SV, et al. Expanding the Andersen model: the role of psychosocial factors in long-term care use. Health Serv Res 2002 Oct;37(5):1221-1242 [FREE Full text] [doi: 10.1111/1475-6773.01053] [Medline: 12479494]
20. Bigos SJ, Battié MC, Spengler DM, Fisher LD, Fordyce WE, Hansson TH, et al. A prospective study of work perceptions and psychosocial factors affecting the report of back injury. Spine 1991 Jan;16(1):1-6. [doi: 10.1097/00007632-199101000-00001] [Medline: 1825891]
21. Paarlberg K, Vingerhoets AJ, Passchier J, Dekker GA, Van Geijn HP. Psychosocial factors and pregnancy outcome: A review with emphasis on methodological issues. J Psychosom Res 1995 Jul;39(5):563-595. [doi: 10.1016/0022-3999(95)00018-6]
22. Welin C, Lappas G, Wilhelmsen L. Independent importance of psychosocial factors for prognosis after myocardial infarction. J Intern Med 2000 Jun;247(6):629-639 [FREE Full text] [doi: 10.1046/j.1365-2796.2000.00694.x] [Medline: 10886484]
23. Kartha A, Anthony D, Manasseh CS, Greenwald JL, Chetty VK, Burgess JF, et al. Depression is a risk factor for rehospitalization in medical inpatients. Prim Care Companion J Clin Psychiatry 2007 Aug 15;9(4):256-262 [FREE Full text] [doi: 10.4088/pcc.v09n0401] [Medline: 17934548]
24. Coventry PA, Gemmell I, Todd CJ. Psychosocial risk factors for hospital readmission in COPD patients on early discharge services: a cohort study. BMC Pulm Med 2011 Nov 04;11(1):49 [FREE Full text] [doi: 10.1186/1471-2466-11-49] [Medline: 22054636]
25. Flythe JE, Hilbert J, Kshirsagar AV, Gilet CA. Psychosocial factors and 30-day hospital readmission among individuals receiving maintenance dialysis: a prospective study. Am J Nephrol 2017 Apr 14;45(5):400-408 [FREE Full text] [doi: 10.1159/000470917] [Medline: 28407633]
26. Retrum JH, Boggs J, Hersh A, Wright L, Main DS, Magid DJ, et al. Patient-identified factors related to heart failure readmissions. Circ Cardiovasc Qual Outcomes 2013 Mar;6(2):171-177. [doi: 10.1161/circoutcomes.112.967356]
27. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? J Health Soc Behav 1995 Mar;36(1):1-10. [doi: 10.2307/2137284]

28. Adler N, Stewart J. Health disparities across the lifespan: meaning, methods, and mechanisms. Ann N Y Acad Sci 2010 Feb;1186:5-23. [doi: 10.1111/j.1749-6632.2009.05337.x] [Medline: 20201865]
29. Committee On The Recommended Social And Behavioral Domains And Measures For Electronic Health Records. Capturing social and behavioral domains in electronic health records. Washington, DC: National Academies Press; 2014.
30. Ohno-Machado L. Realizing the full potential of electronic health records: the role of natural language processing. J Am Med Inform Assoc 2011 Sep 01;18(5):539 [FREE Full text] [doi: 10.1136/amiajnl-2011-000501] [Medline: 21846784]
31. van Walraven C, Dhalla IA, Bell C, Etchells E, Stiell IG, Zarnke K, et al. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. CMAJ 2010 Apr 06;182(6):551-557 [FREE Full text] [doi: 10.1503/cmaj.091117] [Medline: 20194559]
32. Carnielli CM, Macedo CCS, De Rossi T, Granato DC, Rivera C, Domingues RR, et al. Combining discovery and targeted proteomics reveals a prognostic signature in oral cancer. Nat Commun 2018 Sep 05;9(1):3598-3617. [doi: 10.1038/s41467-018-05696-2] [Medline: 30185791]
33. Xia B, Zhao D, Wang G, Zhang M, Lv J, Tomoiaga AS, et al. Machine learning uncovers cell identity regulator by histone code. Nat Commun 2020 Jun 01;11(1):2696-2612. [doi: 10.1038/s41467-020-16539-4] [Medline: 32483223]

Abbreviations

AUROC: area under the receiver operating characteristic curve
EMR: electronic medical record
GBT: gradient boosting trees
LACE: Length of stay, Acuity of the admission, Comorbidity of the patient, and Emergency department use
LDA: latent Dirichlet allocation

Edited by R Kukafka; submitted 28.12.20; peer-reviewed by M Perez-Zepeda, S Mooijaart; comments to author 16.03.21; revised version received 30.06.21; accepted 27.07.21; published 19.10.21.

Please cite as:
Goh KH, Wang L, Yeow AYK, Ding YY, Au LSY, Poh HMN, Li K, Yeow JJL, Tan GYH
Prediction of Readmission in Geriatric Patients From Clinical Notes: Retrospective Text Mining Study
J Med Internet Res 2021;23(10):e26486
URL: https://www.jmir.org/2021/10/e26486
doi:10.2196/26486
PMID:34665149

Original Paper

Development, Validation, and Assessment of Clinical Impact of Real-time Alerts to Detect Inpatient As-Needed Opioid Orders With Duplicate Indications: Prospective Study

Elsie Rizk[1], PharmD; Joshua T Swan[1,2], PharmD, MPH

[1]Department of Pharmacy, Houston Methodist, Houston, TX, United States

[2]Department of Surgery, Houston Methodist, Houston, TX, United States

Corresponding Author:
Joshua T Swan, PharmD, MPH
Department of Surgery
Houston Methodist
6550 Fannin St, SM1661
Houston, TX, 77035
United States
Phone: 1 713 441 0165
Email: swan.joshua@gmail.com

Abstract

Background: As-needed (PRN) opioid orders with duplicate indications can lead to medication errors and opioid-related adverse drug events.

Objective: The objective of our study was to build and validate real-time alerts that detect duplicate PRN opioid orders and assist clinicians in optimizing the safety of opioid orders.

Methods: This single-center, prospective study used an iterative, 3-step process to refine alert performance by advancing from small sample evaluations of positive predictive values (PPVs) (step 1) through intensive evaluations of accuracy (step 2) to evaluations of clinical impact (step 3). Validation cohorts were randomly sampled from eligible patients for each step.

Results: During step 1, the PPV was 100% (one-sided, 97.5% CI 70%-100%) for moderate and severe pain alerts. During step 2, duplication of 1 or more PRN opioid orders was identified for 17% (34/201; 95% CI, 12%-23%) of patients during chart review. This bundle of alerts showed 94% sensitivity (95% CI 80%-99%) and 96% specificity (95% CI 92%-98%) for identifying patients who had duplicate PRN opioid orders. During step 3, at least 1 intervention was made to the medication profile for 77% (46/60; 95% CI 64%-87%) of patients, and at least 1 inappropriate duplicate PRN opioid order was discontinued for 53% (32/60; 95% CI 40%-66%) of patients.

Conclusions: The bundle of alerts developed in this study was validated against chart review by a pharmacist and identified patients who benefited from medication safety interventions to optimize PRN opioid orders.

(J Med Internet Res 2021;23(10):e28235) doi:10.2196/28235

KEYWORDS

opioid stewardship; pain; as-needed opioids; duplicate orders; automated alerts

Introduction

Duplicate as-needed (PRN) opioid orders that are indicated for the same pain level can lead to medication errors, opioid-related adverse drug events, and confusion among bedside nurses. Hospital accreditation standards and pain management guidelines recommend constructing mutually exclusive pain levels for PRN opioid orders to establish clear indications and avoid therapeutic duplication of pain scales for PRN indications [1-4]. The proportion of patients having multiple PRN opioid orders with a duplicate indication was among the critical quality indicators established for an Opioid Stewardship Program across a multihospital health system [5]. Therefore, there is a need to develop clinical decision support and automated processes to optimize this high-priority quality indicator.

Epic is an electronic health record (EHR) system with one of the largest market shares among hospitals in the United States; however, the current functionality of Epic does not allow for

triggering drug-drug interactions or best practice alerts based on the PRN indication field specified in medication orders [6]. The purpose of this project was to build and validate real-time alerts within a third-party pharmacovigilance software that detect PRN opioid orders with a duplicate pain indication and provide clinicians with a clinically impactful tool to optimize the safety of PRN opioid orders in the hospital EHR.

Methods

Study Description

This was a prospective, program development study. Several study activities were observational and aimed to detect medication safety events in the EHR. One study activity was interventional where a clinical pharmacist improved patient care using real-time alerts in accordance with a new hospital policy. The hospital's Institutional Review Board approved this study with a waiver of informed consent.

The Houston Methodist Opioid Stewardship Program collaborated with VigiLanz Corporation, a third-party pharmacovigilance software company that receives real-time data from the hospital's EHR and creates condition-based alerts that identify specific patient situations or medical events. For example, if a new order for warfarin was placed in the EHR without any international normalized ratio values in the previous 24 hours, an alert would be generated to notify the pharmacist to initiate appropriate therapy monitoring interventions. Other areas of alerting include drug interactions, lab monitoring, antibiotic culture mismatch, medication dosage adjustments, treatment-related adverse events, and other therapy elements [7].

A set of VigiLanz alerts was created to detect patients with 2 or more PRN opioid orders with duplicate pain category indications. The hospital EHR that was used across the Houston Methodist health system to send data to VigiLanz and inform the alerts was Epic 2018 (Epic Systems Corporation).

Best Practices for Medication Safety

Through interprofessional stakeholder engagement among pharmacists, nurses, physicians, and medication safety specialists, the Houston Methodist Opioid Stewardship Program established best practices for inpatient prescribing of PRN opioid orders to reduce the risk of medication errors. Unless clear parameters are provided, multiple PRN opioid orders for the same pain level are inappropriate and may cause confusion within the care team and lead to duplicate opioid administration. Therefore, the Opioid Stewardship Program created a schema to categorize multiple PRN opioid orders for the same pain level as appropriate or inappropriate using the following medication order attributes: (1) route and formulation, (2) use of the linked order group functionality in Epic, and (3) clear administration instructions. For example, a patient might have a PRN order for oral oxycodone for severe pain along with intravenous morphine PRN for severe pain to be administered as an alternative if the patient cannot tolerate oral intake. When duplicate PRN opioid orders for the same pain level are necessary, clear administration instructions should be provided to the bedside nurse to guide selection of the appropriate medication, and orders may be placed in a linked order group. In Epic, orders that are linked using the "OR" linking logic provide a safety enhancement feature that would notify nurses about the risk of duplicate administration if one of the orders was about to be administered too soon following the administration of another order in the linked group. To empower clinical pharmacists to resolve duplicate PRN opioid orders using their clinical judgment, the Opioid Stewardship Program developed a hospital policy that organized duplications into 5 categories and authorized pharmacists to provide specific interventions for each category, as shown in Table 1.

Table 1. Categorization of multiple as-needed opioid orders as appropriate or inappropriate.

Category	Collected information			Evaluation of data	
	Same route and formulation	Clear administration instructions	Linked orders	Appropriate duplication	Potential medication safety intervention
1	No[a]	Yes	Yes	Yes	No change needed
2	No[a]	No	Yes	No	Clarify administration instructions[b]
3	No[a]	Yes	No	No	May add orders to a linked order group using the "OR" linking logic
4	No[a]	No	No	No	Clarify administration instructions and may link orders[b]
5	Yes	Yes or No	Yes or No	No	Discontinue all but one of the duplicate orders

[a]Two orders with the same route (eg, oral) are allowed as long as the orders have different formulations (eg, oral tablet and oral liquid) with clear administration instructions and use the "link" functionality in Epic for enhanced administration safety.

[b]Examples of administration instructions were suggested based on the medication formulation or route as follows: For oral tablets, the instruction was "Give if patient can receive oral tablet medication." For oral solutions, the instruction was "Give if patient cannot receive oral tablet medication but can receive oral solution medication." For intravenous injections, the instruction was "Give if patient is not able to receive oral medication."

Logic to Trigger Real-Time Alerts

Houston Methodist used a Health Level 7 (HL7) message interface to transmit medication order data from the EHR to the third-party pharmacovigilance software in real time. The relevant medication order variables included in this HL7 message were the medication name, route, frequency, dose, administration instructions, start date and time, and PRN indication. The order linking information in Epic was not

included in the HL7 message and was not used by the alerts to categorize multiple PRN opioid orders as appropriate or inappropriate. Investigators developed operational definitions for types of duplications and then created a bundle of alerts to identify each of the operational definitions given in Table 2. Duplicate PRN opioid orders were classified into explicit duplicates that specified pain levels as mild, moderate, or severe and implicit duplicates where at least one of the duplicate orders had unclear PRN instructions that did not specify the pain scale range. Alerts were censored for opioid orders that were verified while a patient was in an operating room or a procedural area due to the temporary nature and unique medical record workflow of these opioid orders. To reduce alert fatigue for end users, each alert was set to create no more than 1 activation every 24 hours for the same patient.

Table 2. Classification of duplicate as-needed opioid orders.

Duplicates	VigiLanz alert name	Example
Explicit duplicates		
2 or more PRN[a] opioid orders for mild pain	Opioids PRN for mild pain	Tramadol 50 mg PO[b] every 8 hours PRN for mild pain (score 1-3) AND acetaminophen-codeine 300-30 mg PO every 4 hours PRN for mild pain (score 1-3)
2 or more PRN opioid orders for moderate pain	Opioids PRN for moderate pain	Hydrocodone-acetaminophen 5-325 mg PO every 6 hours PRN for moderate pain (score 4-6) AND tramadol 50 mg PO every 6 hours PRN for moderate pain (score 4-6)
2 or more PRN opioid orders for severe pain	Opioids PRN for severe pain	Hydromorphone 1 mg IV[c] every 4 hours PRN for severe pain (score 7-10) AND fentanyl 25 mcg IV every 3 hours PRN for severe pain (score 7-10)
Implicit duplicates		
2 or more PRN opioid orders where at least one order has unclear PRN instructions[d]	Opioids PRN for pain	Hydrocodone-acetaminophen 10-325 mg PO every 6 hours PRN for pain AND morphine 15 mg PO every 6 hours PRN for severe pain (score 7-10)
An opioid PCA[e] order and another IV PRN opioid ordered separately[f]	Opioid PRN PCA	Hydromorphone IV PCA AND morphine 2 mg IV every 4 hours PRN for severe pain (score 7-10)

[a]PRN: as-needed.

[b]PO: oral.

[c]IV: intravenous.

[d]The PRN instructions did not specify the pain scale range.

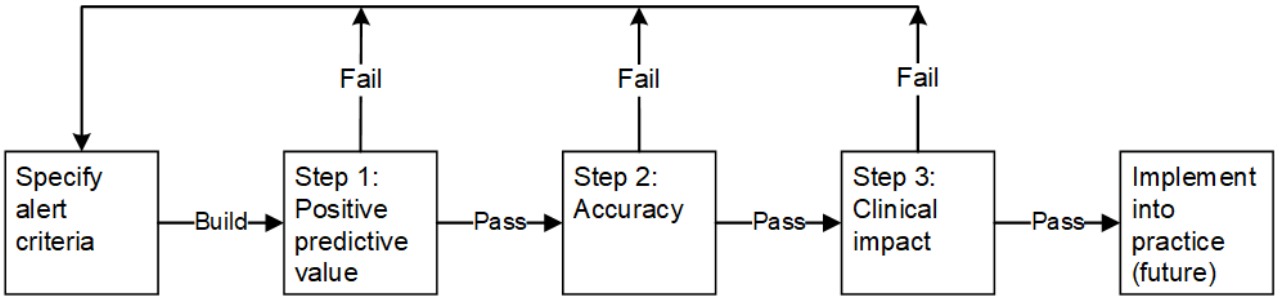
[e]PCA: patient-controlled analgesia.

[f]This includes orders for any pain level (mild, moderate, or severe).

3-Step Process for Specifying, Refining, and Validating Real-Time Alerts

The investigators designed and implemented a 3-step process to develop a bundle of real-time alerts, validate their accuracy, and evaluate their clinical impact when used by pharmacists to identify duplicate PRN opioid orders, as indicated in Figure 1. This process was designed to quickly identify and resolve errors in alert performance by advancing from small sample, positive predictive value (PPV) evaluations (step 1) to intensive accuracy evaluations (step 2) and clinical impact evaluations (step 3). As medical record alerts commonly generate false positives that contribute to alert fatigue, our process measured and optimized the PPV as the first step [8]. After each step, the investigators would determine if the bundle of alerts needed further configuration (fail) or could be advanced to the next step of testing (pass). Results for the final bundle of alerts that passed all 3 steps are reported in this paper.

Figure 1. Alert development process. This 3-step process efficiently pilot-tested, revised, and validated bundles of alerts to detect duplicate as-needed opioid orders.



Step 1: Evaluation of PPV

On September 4, 2018, 10 active alerts for duplicate PRN opioid orders for each of the 2 most common types of alerts (moderate and severe pain) were randomly sampled from alerts generated from August 31, 2018 to September 4, 2018. Alerts were validated against active medication orders on the patients' profiles by a pharmacist. Alerts identifying duplicate orders that were confirmed by the pharmacist's chart review were categorized as true positive, and alerts that were not confirmed were categorized as false positive. The PPV was calculated by dividing the number of true positive alerts by the total number of alerts. The sample size of 10 was chosen to quickly identify and resolve errors in alert performance.

Step 2: Evaluation of Accuracy

A cross-sectional analysis of patients hospitalized on or after August 31, 2018, was conducted over 5 consecutive weekdays (September 6, 2018 until September 12, 2018). Patients admitted to acute care floors or intensive care units were randomly sampled once per day, and a patient could not be sampled more than once. Operating rooms, postanesthesia care units, catheterization labs, and other procedural areas were excluded from sampling. A pharmacist and a pharmacy intern conducted chart review to collect active PRN opioid names, doses, routes, frequencies, and PRN indications for each included patient. Alert-specific data were collected, including alert names, triggering opioid orders, and start and end dates of the triggering orders. As a patient may have more than one set of duplicate PRN opioid orders or more than one alert, this analysis was conducted at the patient level to match the workflow of medication profile review by clinicians.

It is recommended that medical record alerts be evaluated like diagnostic tests by calculating the performance characteristics of accuracy, sensitivity, specificity, PPV, and negative predictive value (NPV) [8]. Duplicate PRN opioid orders identified through this chart review served as the reference standard for this validation of duplicate opioid alerts to identify true and false positives and negatives. Accuracy was the number of patients correctly classified by the alerts (true positives plus true negatives) divided by the total number of patients evaluated. Sensitivity was the proportion of patients with duplicate PRN opioid orders who had an alert. Specificity was the proportion of patients without duplicate PRN opioid orders who did not have an alert. PPV was the proportion of patients with an alert who had duplicate PRN opioid orders. NPV was the proportion of patients with no alerts who did not have duplicate PRN opioid orders.

Step 3: Evaluation of Clinical Impact and Face Validity

Step 3 included 60 patients identified from a random sample of patients with an alert on a specific calendar day (or the previous calendar day) over 6 nonconsecutive days in September and October 2018, and a patient could not be sampled more than once. A pharmacist conducted medication profile review in the EHR to optimize the PRN opioid orders in accordance with the algorithm described in Table 1 through the following actions: (1) discontinuing one or more duplicate PRN opioid orders, (2) clarifying order administration instructions, (3) contacting the prescriber or the bedside nurse, (4) changing the pain level indication, and (5) assigning PRN opioid orders to linked order groups in the medical record to prevent duplicate administration of orders with different routes or formulations. Whenever PRN opioid orders were discontinued, the pharmacist documented the rationale for discontinuation (orders with older start dates, no recent administration, documented adverse events, or no associated order set). For each patient, the pharmacist recorded a start time when the alerts and patient's medical record were opened to evaluate active PRN opioid orders. The pharmacist then recorded an end time after completing all necessary actions and documenting interventions in the pharmacovigilance software. The time in minutes between each start and end time was calculated. To estimate the time effort for future implementation of this program into routine practice for clinical pharmacists and other health care professionals, linear regression was conducted to estimate the time needed by the pharmacist to conduct chart review and resolve duplicate PRN opioid orders based on the extent of actions taken.

Results

Step 1: Evaluation of PPV

During the study period for step 1, 9% (10/114) of moderate pain alerts and 14% (10/71) of severe pain alerts were randomly sampled and evaluated. The PPV was 100% (one-sided, 97.5% CI 70%-100%) for moderate alerts and 100% (one-sided, 97.5% CI 70%-100%) for severe alerts. The investigators interpreted this as success and advanced to step 2.

Step 2: Evaluation of Accuracy

During the study period for step 2, 30% (201/662) of eligible patients were randomly sampled and had 241 active PRN opioid orders, as observed in Table 3. Chart review identified 1 or more

PRN opioid order duplications for 17% (34/201; 95% CI, 12%-23%) of patients. Of these 34 patients, duplication was identified for the moderate pain scale in 12 (35%), severe pain scale in 9 (26%), both moderate and severe pain scales in 7 (21%), and unclear PRN instructions that did not specify the pain scale range in 6 (18%). This bundle of alerts showed high sensitivity (94%) and specificity (96%) for identifying patients who had duplicate PRN opioid orders, as indicated in Tables 4 and 5. Investigators interpreted this as success and advanced to step 3. On September 18, 2018 (after step 2 was completed), 2 implicit alerts were added to the bundle to identify duplication between orders with unclear PRN indications ("breakthrough pain" or "any pain") and any other PRN opioid order.

Table 3. Prevalence of active as-needed opioid orders in step 2 (N=241).

PRN[a] indication pain category	Pain score	Frequency, n (%)
Mild	1-3	3 (1)
Moderate	4-6	123 (51)
Severe	7-10	93 (39)
Mild and moderate	1-6	2 (1)
Moderate and severe	4-10	6 (2)
Patient-controlled analgesia	Not specified	5 (2)
Unspecified pain category	Not specified	9 (4)

[a]PRN: as-needed.

Table 4. Comparison of chart review reference standard with the bundle of alerts in step 2 (N=201).

Alert fired	Reference standard		
	Active duplicate orders	No active duplicate orders	Total
Alert	32 (true positive)	7 (false positive)	39
No alert	2 (false negative)	160 (true negative)	162
Total	34	167	201

Table 5. Performance characteristics of the bundle of alerts in step 2 (N=201).

Performance characteristics	Value (%)	95% CI (%)
Sensitivity	94	80-99
Specificity	96	92-98
PPV[a]	82	67-93
NPV[b]	99	96-100
Accuracy	96	92-98

[a]PPV: positive predictive value.

[b]NPV: negative predictive value.

Step 3: Evaluation of Clinical Impact and Face Validity

During the study period of 6 nonconsecutive days for step 3, 12% of the eligible patients (60/481) were randomly sampled, who accounted for 12% of the unique eligible alerts (79/678). A pharmacist reviewed the charts of all the patients. At least 1 intervention was made to the medication profile for 77% (46/60; 95% CI 64%-87%) of patients, which was interpreted by investigators as having a meaningful clinical impact and face validity. The most common actions taken for these 60 patients were discontinuing inappropriate duplicate PRN opioid orders (32, 53%), linking PRN opioid orders (21, 35%), and clarifying administration instructions (19, 32%), as shown in Figure 2. Using linear regression estimates, the average time needed by the pharmacist to assess alerts, resolve issues, and document interventions was 5 minutes (95% CI 2-9 minutes) if no action was needed, 7 minutes (95% CI 5-10 minutes) if 1 or more orders were discontinued, 14 minutes (95% CI 11-16 minutes) if advanced modifications were performed for at least 1 order (linking orders, clarifying administration instructions, or changing the PRN indication), and 21 minutes (95% CI 16-26 minutes) if a provider or nurse was contacted with or without other order modifications, as shown in Figure 3. The investigators interpreted this as success and advanced the bundle of alerts for future implementation into the workflow of hospital pharmacists.

Figure 2. Clinical impact assessment in step 3 (60 patients). The categories are not mutually exclusive.

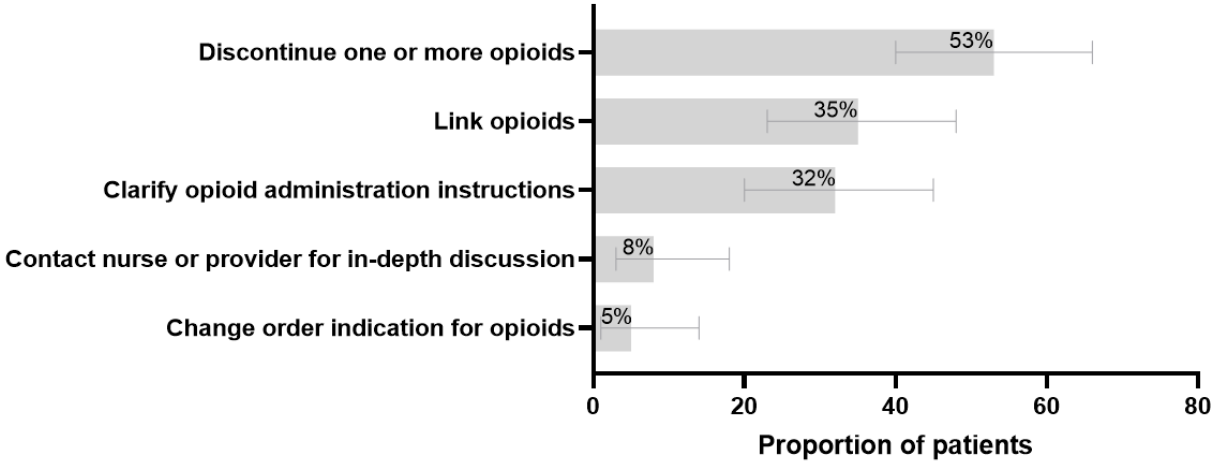
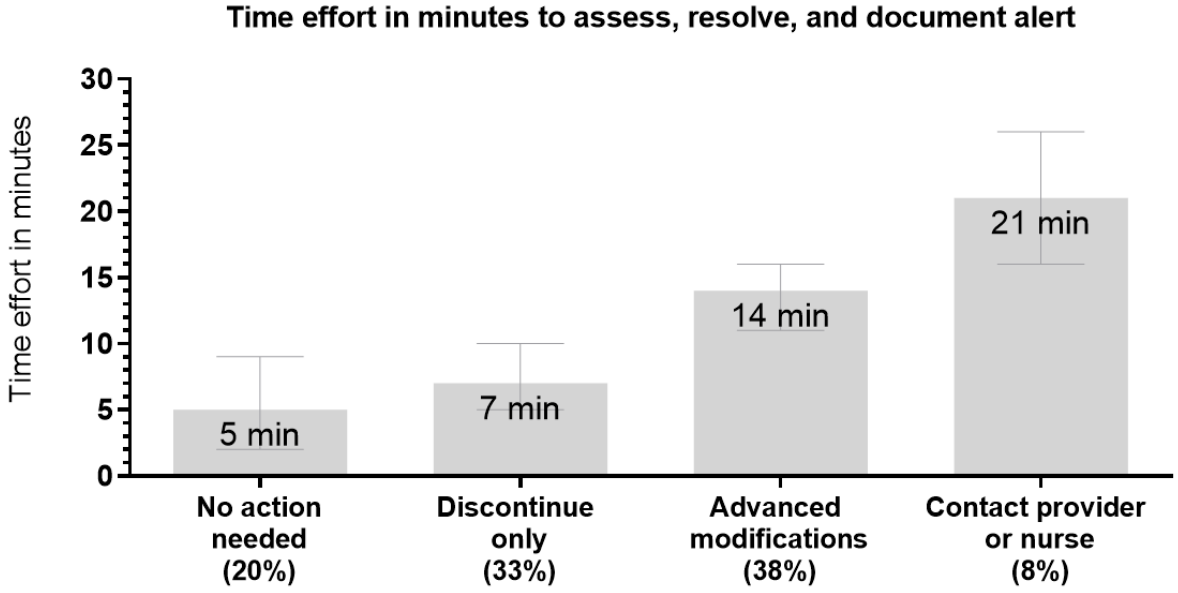


Figure 3. Time effort needed by the pharmacist to assess and resolve duplicate as-needed opioid alerts and document interventions in step 3 (60 patients). The average time effort and 95% CI were calculated by conducting linear regression using the following mutually exclusive categories of actions taken for each patient: (1) no action needed, (2) 1 or more orders discontinued, (3) advanced modifications performed for at least 1order (linking orders, clarifying administration instructions, or changing the PRN (as-needed) indication, and (4) provider or nurse contacted with or without other order modifications. These time estimates also include the time for study data collection, which may require slightly more time (1 or 2 minutes per patient) than routine documentation of clinical interventions by hospital pharmacists.



Discussion

Principal Findings

Decreasing the proportion of hospitalized patients who have PRN opioid orders with duplicate pain level indications was identified as a top priority for opioid stewardship in hospitals and health care systems [5]. This study evaluated the accuracy, validity, and clinical impact of a set of real-time alerts to identify and address duplicate PRN opioid orders. In this 3-step process, step 1 allowed rapid assessment and revision of alerts, step 2 validated a bundle of alerts (sensitivity of 94% and specificity of 96%), and step 3 determined that a clinical pharmacist could make at least 1 intervention for 77% of patients with an alert. In a health system where pharmacists were authorized to optimize PRN opioid orders per hospital policy, the pharmacist spent 7 minutes or less per patient for half of the patients evaluated and resolved most alerts within 14 minutes. This 3-step process could be extended beyond pain intensity scales to develop a bundle of alerts that can identify duplicate analgesic PRN orders based on a patient's functional status or duplicate PRN orders for other clinical indications, such as nausea or vomiting and gut motility.

Limitations

These alerts were validated using a single EHR system (Epic) and would need additional evaluation in other EHR systems. The clinical impact observed in step 3 was evaluated at a single academic medical center, which may not represent the potential

clinical impact at other hospitals with different clinical workflows and patient populations.

Although the linking orders feature provides additional safeguards, the current process for linking 2 or more active orders in Epic requires discontinuation of all relevant orders and substitution with new linked orders. As this workflow may increase the risk of medication errors, multiple hospital committees decided that linking of active orders could be optional if clear administration instructions were provided. When this study was conducted, linked orders were not commonly used, and data feeds from the EHR to the pharmacovigilance software did not include information on order linking. However, if linked orders are used more commonly in the future, this lack of information in the data feed could increase the frequency of false positive alerts, which are estimated to require a pharmacist review time of 5 minutes.

Considerations for Clinical Implementation

The developed bundle of alerts can be used to track each occurrence of duplicate orders (event-based) or identify patient charts that need medication review (patient-based). If used to track each occurrence, these alerts need to be generated every time a duplication occurs (eg, every pair of duplicate orders). However, this tracking approach can generate multiple alerts for each patient daily and cause alert fatigue when identifying patients for medication review. One strategy to prevent alert fatigue is to apply censoring logic to the bundle of alerts so that the bundle will only generate 1 alert in a 24-hour period for each patient, regardless of how many duplicate orders were signed during the 24-hour period. When identifying patients for medication review, alerts can be communicated to pharmacists via automated emails, pages, or an alert queue (internet-based web page).

Conclusions

The bundle of alerts developed in this study was validated against chart review by a pharmacist and identified patients who benefited from medication safety interventions to optimize PRN opioid orders. These alerts provide a real-time automated screening process that replaces intensive chart review needed to identify patients, thus allowing health care team members to spend more time optimizing orders that are unclear and potentially not safe. Our algorithm, which matches categories of inappropriate duplicate PRN opioid orders to potential medication safety interventions, can be used to develop policies that expand the scope of practice for clinicians at other institutions. Further research is needed to evaluate the impact of implementing these alerts into quality surveillance or clinical workflow on the frequency of hospitalized patients who have duplicate PRN opioid orders, which is an important medication safety issue.

Acknowledgments

The authors acknowledge Anh Thu Tran (PharmD) and Lan Bui (PharmD, MPH) for assisting with data collection. VigiLanz Corporation provided research funding to Houston Methodist for a Clinical Pharmacy Fellow in Outcomes Research position that was used to support this research and other research studies. The funding agency was not involved in designing the study, collecting, analyzing, and interpreting data, or writing the manuscript.

Authors' Contributions

ER and JTS conceived the study design and analyzed the data. JTS obtained research funding. ER developed data collection tools, collected data, and drafted the manuscript. ER and JTS revised the manuscript and approved the final version of the manuscript.

Conflicts of Interest

JTS’s employer has received research funding on his behalf from VigiLanz, Heron, Pacira, and Pfizer. ER declares no conflicts of interest.

References

1. DNV GL Healthcare. National integrated accreditation for healthcare organizations (NIAHO®)-Accreditation Requirements, Interpretive Guidelines and Surveyor Guidance - Revision 18. 2018 Feb 05. URL: https://cha.com/wp-content/uploads/2018/05/2018-DNV-NIAHO-Guidelines.pdf [accessed 2021-10-20]
2. Joint Commission. Medication administration - incorporating patient preference into medication administration practices. 2018. URL: https://www.jointcommission.org/standards/standard-faqs/hospital-and-hospital-clinics/medication-management-mm/000002058/ [accessed 2018-08-08]
3. Washington State Agency Medical Directors' Group. Interagency guideline on prescribing opioids for pain. Olympia, WA; 2015. URL: https://agencymeddirectors.wa.gov/Files/2015AMDGOpioidGuideline.pdf [accessed 2021-10-20]
4. Anderson WG, Liao S, Childers JW, De Pinto M, Gordon DB, Pirrello R, et al. Improving Pain Management for Hospitalized Medical Patients, A Society of Hospital Medicine Implementation Guide. 2015. URL: http://tools.hospitalmedicine.org/resource_rooms/imp_guides/Pain_Management/PainMgmt_Final3.4.15.pdf [accessed 2021-10-20]
5. Rizk E, Swan JT, Cheon O, Colavecchia AC, Bui LN, Kash BA, et al. Quality indicators to measure the effect of opioid stewardship interventions in hospital and emergency department settings. Am J Health Syst Pharm 2019 Feb;76(4):225-235. [doi: 10.1093/ajhp/zxy042] [Medline: 30715186]

6. Holmgren AJ, Adler-Milstein J, McCullough J. Are all certified EHRs created equal? assessing the relationship between EHR vendor and hospital meaningful use performance. J Am Med Inform Assoc 2018 Jun;25(6):654-660 [FREE Full text] [doi: 10.1093/jamia/ocx135] [Medline: 29186508]
7. The VigiLanz Platform.: VigiLanz Corporation URL: https://vigilanzcorp.com/the-vigilanz-platform/#/ [accessed 2020-07-18]
8. McGreevey JD, Mallozzi CP, Perkins RM, Shelov E, Schreiber R. Reducing alert burden in electronic health records: state of the art recommendations from four health systems. Appl Clin Inform 2020 Jan;11(1):1-12. [doi: 10.1055/s-0039-3402715] [Medline: 31893559]

Abbreviations

EHR: electronic health record
HL7: Health level seven
NPV: negative predictive value
PPV: positive predictive value
PRN: as-needed

Edited by R Kukafka; submitted 25.02.21; peer-reviewed by D Oyler, N van Helmond; comments to author 30.03.21; revised version received 05.05.21; accepted 12.08.21; published 25.10.21.

Please cite as:
Rizk E, Swan JT
Development, Validation, and Assessment of Clinical Impact of Real-time Alerts to Detect Inpatient As-Needed Opioid Orders With Duplicate Indications: Prospective Study
J Med Internet Res 2021;23(10):e28235
URL: https://www.jmir.org/2021/10/e28235
doi:10.2196/28235
PMID:34694235

Original Paper

Developing a RadLex-Based Named Entity Recognition Tool for Mining Textual Radiology Reports: Development and Performance Evaluation Study

Shintaro Tsuji[1,2], PhD; Andrew Wen[1], MS; Naoki Takahashi[3], MD, PhD; Hongjian Zhang[2], MS; Katsuhiko Ogasawara[2], MBA, PhD; Gouqian Jiang[1], MD, PhD

[1]Department of Health Sciences Research, Department of Radiology, Rochester, MN, United States

[2]Graduate School of Health Sciences, Hokkaido University, Sapporo, Japan

[3]Department of Radiology, Mayo Clinic, Rochester, MN, United States

Corresponding Author:
Gouqian Jiang, MD, PhD
Department of Health Sciences Research
Department of Radiology
200 First Street, SW
Rochester, MN
United States
Phone: 1 507 266 1327
Fax: 1 507 284 1516
Email: Jiang.Guoqian@mayo.edu

Abstract

Background: Named entity recognition (NER) plays an important role in extracting the features of descriptions such as the name and location of a disease for mining free-text radiology reports. However, the performance of existing NER tools is limited because the number of entities that can be extracted depends on the dictionary lookup. In particular, the recognition of compound terms is very complicated because of the variety of patterns.

Objective: The aim of this study is to develop and evaluate an NER tool concerned with compound terms using RadLex for mining free-text radiology reports.

Methods: We leveraged the clinical Text Analysis and Knowledge Extraction System (cTAKES) to develop customized pipelines using both RadLex and SentiWordNet (a general purpose dictionary). We manually annotated 400 radiology reports for compound terms in noun phrases and used them as the gold standard for performance evaluation (precision, recall, and F-measure). In addition, we created a compound terms–enhanced dictionary (CtED) by analyzing false negatives and false positives and applied it to another 100 radiology reports for validation. We also evaluated the stem terms of compound terms by defining two measures: occurrence ratio (OR) and matching ratio (MR).

Results: The F-measure of cTAKES+RadLex+general purpose dictionary was 30.9% (precision 73.3% and recall 19.6%) and that of the combined CtED was 63.1% (precision 82.8% and recall 51%). The OR indicated that the stem terms of *effusion*, *node*, *tube*, and *disease* were used frequently, but it still lacks capturing compound terms. The MR showed that 71.85% (9411/13,098) of the stem terms matched with that of the ontologies, and RadLex improved approximately 22% of the MR from the cTAKES default dictionary. The OR and MR revealed that the characteristics of stem terms would have the potential to help generate synonymous phrases using the ontologies.

Conclusions: We developed a RadLex-based customized pipeline for parsing radiology reports and demonstrated that CtED and stem term analysis has the potential to improve dictionary-based NER performance with regard to expanding vocabularies.

(J Med Internet Res 2021;23(10):e25378) doi:10.2196/25378

KEYWORDS

named entity recognition (NER); natural language processing (NLP); RadLex; ontology; stem term

Introduction

Background

The widespread adoption of electronic medical record (EMR) systems in recent years has increasingly brought opportunities to research communities regarding the secondary use of EMR data such as medical images and clinical notes [1] to support clinical and translational research. It is expected that real-world data will contribute to generating medical evidence, optimizing the use of medical resources, and creating high-quality diagnostic or treatment guidelines [2,3]. To establish effective retrieval and extraction of such data stored in the EMR, standard codes are usually used to describe patient records and make them computable and interpretable. For example, the *International Classification of Diseases* is a standard code system used to classify diseases or diagnoses for medical records [4]. The *International Classification of Diseases* can be used to identify classified disease names from medical records in information-retrieval applications. In addition, a standard code system can also be used to extract features from medical texts such as pathology reports, radiology reports, and family history reports. For example, SNOMED-CT (Systematized Nomenclature of Medicine-Clinical Terms) is a standard terminology in the field of medical care [5], which is often used as a resource for the automatic named entity recognition (NER) of medical texts [6]. Moreover, SNOMED-CT is formalized as an ontology, which has a hierarchical structure of terms and semantic relationships between terms. Such an ontology supports medical reasoning with standard concept definitions and axioms among concepts.

In the field of radiology, a large amount of medical imaging data and diagnostic reporting data is stored in the EMR, which has become an important data source for acquiring knowledge. The use of standard code systems is critical for the effective mining of the data source. RadLex, produced by the *Radiological Society of North America*, is a controlled-standard biomedical ontology that provides codes, conceptual relationships, and procedures of imaging examinations [7]. RadLex was historically developed as indexing teaching files for radiologists, provided by the *American College of Radiology* [8]. Currently, RadLex is widely used to support the creation of templates for generating radiology reports [9], mining radiology reporting data [10], indexing medical images and reports [11], and standardizing examination descriptions [12]. From the perspective of data interoperability in the radiology domain, RadLex is a unique ontology in that it enables semantic parsing of free-text radiology reports by playing a role in integrating identified entities into a higher-level semantic concept such as *anatomical entities*, *clinical findings*, *imaging observation*, and *procedures*.

NER is usually used for preprocessing unstructured data for machine learning research, for example, extracting features from radiology reports [13]. In a previous study on the NER evaluation based on radiologist agreement, it was reported that the F-measure of dictionary-based NER was lower than that of conditional random fields (CRFs) [14,15] and rule-based natural language processing (NLP) [16,17]. However, machine learning–based NER does not provide a relationship between terms, and the reason for the F-measure of dictionary-based NER being lower than that of machine learning–based NER is that it is difficult to identify various patterns of compound terms using standard terminologies or ontologies. For example, in the case of the compound term *right-sided IJ central venous catheter*, all the words in the term except for *catheter* are modifiers. In short, there are several patterns such as *IJcentral venous catheter* and *venous catheter* that can be identified as annotations by radiologists.

Objective

Although an ontology such as RadLex can be leveraged to enhance data interoperability and track relationships and hierarchical structure, we consider that the ontology should also be applied to improve the NER of compound terms in radiology reports. However, few studies have been conducted to evaluate the coverage of RadLex for the NER of compound terms for mining radiology reports. To evaluate and extend the coverage of the lexicon for extracting features from radiology reports, the aim of this study is to develop and assess an NER tool based on RadLex, explore the entities included in RadLex, and subsequently extend the ontology for a higher F-measure on feature extraction by dictionary-based NER.

Methods

RadLex Features

RadLex is a controlled-standard biomedical ontology produced by the *Radiological Society of North America*, which provides unique codes, conceptual mapping based on hierarchal structure, and procedures of imaging examinations [7]. We used and analyzed RadLex version 1.3.4 [18], which includes 46,434 primary terms and 42,831 compound terms.

General Purpose Dictionary

We used a general purpose dictionary (GPD), SentiWordNet [19], to compare RadLex coverage with a general dictionary. SentiWordNet, which is a GPD for sentiment analysis in the context of social network services, provides a negative or positive score of terms. The number of words of parts of speech (POS) is 117,659, including 82,115 distinct nouns, 13,767 verbs, 18,156 adjectives, and 3621 adverbs. The number of compound terms is 48,469.

Clinical Text Analysis Knowledge Extraction System

The clinical Text Analysis Knowledge Extraction System (cTAKES), which is an NLP system for extraction of information from EMR clinical free text, contains an automatic NER tool using a dictionary lookup mechanism [20]. The default dictionary of cTAKES is based on the Unified Medical Language System (UMLS) [21] and provides annotations of diseases or disorders, signs or symptoms, anatomical sites, procedures, and medications. For example, the dictionaries based on SNOMED-CT and RxNORM, which is part of the UMLS, cover the fields of general clinical findings and medications. We investigated the compound terms in each dictionary for the analysis.

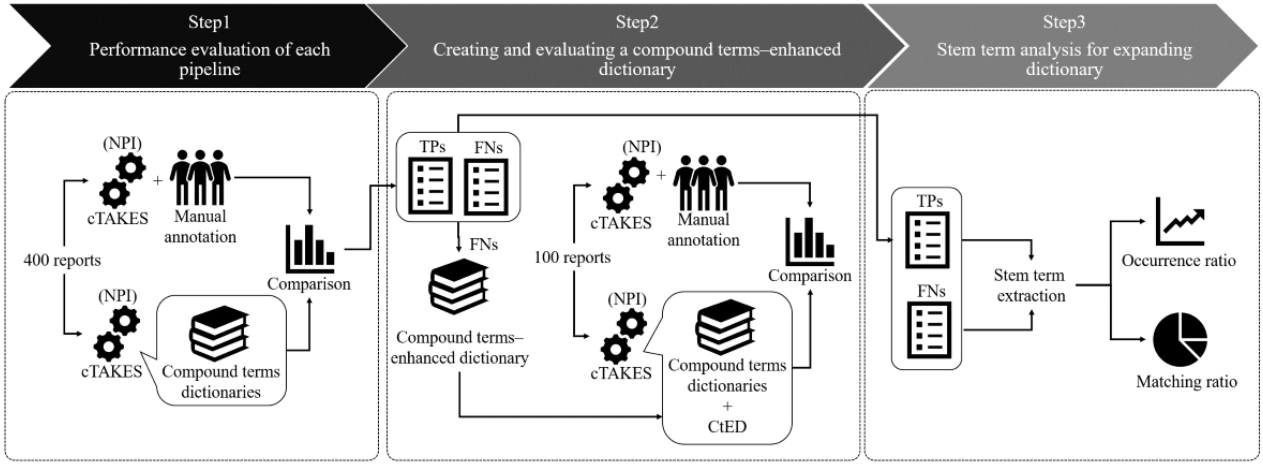
Medical Information Mart for Intensive Care-III

The Medical Information Mart for Intensive Care-III (MIMIC-III) is a free, open database provided by the Massachusetts Institute of Technology Laboratory for Computational Physiology, which includes approximately 60,000 deidentified admissions of patients at the Beth Israel Deaconess Medical Center from 2001 to 2012 [22]. Using PostgreSQL, we queried the note events table of the MIMIC-III database, which includes approximately 520,000 radiology reports.

Procedures

The overall goal of our study is to clarify the coverage of RadLex-based dictionaries with compound terms and to construct and evaluate the NER tools that use the RadLex-based dictionaries for mining free-text radiology reports. First, we customized cTAKES to build the RadLex and GPD dictionaries. As previously mentioned, the default dictionaries of cTAKES provided by the UMLS are SNOMED-CT and RxNORM. Second, we combined these three dictionaries in the following patterns: Default, Default+RadLex, and Default+RadLex+GPD. Third, we removed single terms from each dictionary and evaluated their performance. Finally, we carried out the three processes of analysis (step 1 to step 3) to obtain profiles of the stem terms for improving the performance of NER (Figure 1).

Figure 1. Overview of methods. cTAKES: clinical Text Analysis and Knowledge Extraction System; CtED: compound terms–enhanced dictionary; FN: false negative; NPI: noun phrase identification; TP: true positive.



Creating Annotation Corpus of Radiology Reports

We randomly selected 400 reports of computed tomography (CT), magnetic resonance imaging (MRI), positron emission computed tomography (PET), and radiography (x-ray) from the MIMIC-III database (100 reports for each imaging modality type). These reports were in a free-text format and were categorized into sections; we used the *Findings*, *Interpretations*, and *Impressions* sections, which play a core role in diagnosis. There were 28.9 sentences per report and 179.1 tokens per report. An additional 100 reports (25 reports for each imaging modality type) were randomly selected and used in the validation study for compound terms.

We first conducted stop word removal and exchanged all the characters to the lower case. Next, we leveraged the AggregatePlaintextProcessor of cTAKES to identify noun phrases in the radiology reports so that we could perform a manual annotation for noun phrases. Next, we applied manual reviews to annotate compound terms. The compound terms were also tagged with all conceivable patterns based on the *stem term*. For example, the compound term *right upper lobe* is divided into *right upper lobe* and *upper lobe*. After the annotation, we can obtain two compound terms from *right upper lobe*. In this case, we defined *lobe* as a stem term. We also separated *right upper lobe of lung base* into *right upper lobe* and *lung base*. Thus, we defined the stem term as the modified term of a compound term in this study. These manual annotations were conducted and agreed on by 3 researchers with a background in radiology (n=1) and computer sciences (n=2). Generally, the annotations for compound terms are performed by expert radiologists and are agreed upon through discussion. As some studies have revealed that the annotation patterns of compound terms are different with institutions, we used all the patterns of compound terms as the gold standard.

Developing a RadLex-Based NER Tool

First, we created a customized NER tool using cTAKES, which uses a dictionary lookup–based parser for NER. It extracts terms that can be looked up in the installed dictionary. Some previous studies have attempted to create customized dictionaries (eg, UMLS) [11] for NER, but few studies have investigated NER using RadLex for mining radiology reports [23]. In this study, we built a customized dictionary using RadLex as a domain-specific dictionary and SentiWordNet as a GPD. RadLex can be used to automatically extract technical terms from radiology reports [24], whereas SentiWordNet is usually used for sentiment analysis, which clarifies positive or negative descriptive text on social networking services. Moreover, we created dictionaries for compound terms. The terms of each dictionary were stored in the bar-separated value (BSV) file and located in the dictionary lookup–first directory, which allows the term to be extracted preferentially. cTAKES uses a

SNOMED-CT and RxNORM dictionary by default. Finally, we created a collection of customized dictionaries in the following patterns: Default (SNOMED-CT and RxNORM), Default+RadLex, and Default+RadLex+GPD.

Step 1: Performance Evaluation of Each Pipeline

For each customized pipeline, we evaluated the performance of four different sets of the three dictionary patterns using standard measures (ie, precision, recall, and F-measure). The formulas for the measures are as follows:

Precision = True positives / (True positives + False positives) × 100 **(1)**

Recall = True positives / (True positives + False negatives) × 100 **(2)**

F-measure = 2 × Precision × Recall / (Precision + Recall) **(3)**

Here, true positive (TP) is defined as the number of manual annotations matched with the dictionary phrases, false positive (FP) is defined as the number of dictionary phrases matched with entities other than manual annotations, and false negative (FN) is defined as the number of annotations not matched with the dictionary phrases. We also evaluated the performance of four major imaging modalities: CT, MRI, x-ray, and PET. GATE (General Architecture for Text Engineering) developer version 8.4.1 [25] was used to compute these measures.

Step 2: Creating and Evaluating a Compound Terms–Enhanced Dictionary

We also created a compound terms–enhanced dictionary (CtED) to improve performance (Figure 1). We added these compound terms to the FN category (as identified in the initial evaluation) in the custom dictionaries that were used for parsing 400 radiology reports. At the same time, we removed these compound terms in the FP category from these dictionaries. To validate the performance of the CtED, we carried out NER for another 100 radiology reports (25 reports for each imaging modality type; Figure 1). Finally, we calculated the precision, recall, and F-measure for the performance evaluation.

Step 3: Stem Term Analysis for Expanding Dictionary

To obtain the full benefit of using RadLex, which is an ontology-based tool, we created 2 measures for a stem term. We first defined a measure called the occurrence ratio (OR) to determine the frequency of stem terms in TPs and FNs from step 2. The OR gives priority measures to add compound terms with stem terms into RadLex. For example, if the value of the OR for a stem term in TPs is high, it means that the number of compound terms (containing the stem term) that are correctly identified by the pipeline is high. In contrast, if the value of the OR for the stem term in FNs is high, it means that the number of compound terms (containing the stem term) that are identified as negative by the pipeline is high. Moreover, if a high OR stem with both TP and FN is identified, we can hypothesize that this stem shows that there is a high demand to extract the entity of reports but still lacks the compound terms having the stem. In short, the OR can visualize a profile of the demand and supply of stem term–oriented compound terms in the corpus.

Occurrence ratio (%) = Occurrence of a stem term in TP or FN / Total number of stem terms in TPs or FNs × 100% **(4)**

Second, we defined a measure called the matching ratio (MR) to describe the distribution of stem terms in FNs that are matched with the dictionaries. The MR (%) was calculated using the formula presented below. The MR can guide the basic concept of the RadLex or SNOMED-CT (cTAKES default dictionary) concept that matches the stem terms. For example, if a stem term of *effusion* is found in RadLex, we continue to trace the parent concept until the concept is under the top hierarchy. Finally, we identified the concept of *clinical findings*. The MR provides the criteria for identifying the number of concepts. We used 15 concepts under the RadLex entity (ie, *anatomical entity*, *clinical finding*, *imaging modality*, *imaging observation*, *nonanatomical substance*, *object*, *procedure*, *process step*, *process*, *property*, *RadLex descriptor*, *RadLex nonanatomical set*, *report*, *report content*, and *temporary entity*). Each stem term was tracked using their upper-class ID (RadLex ID). For the cTAKES default dictionary, we used 19 concepts under the top class of SNOMED-CT (RxNORM was excluded because it does not have a hierarchal structure). The class are *Body structure*, *Clinical finding*, *Environment or geographical location*, *Event*, *Observable entity*, *Organism*, *Pharmaceutical/biologic product*, *Physical force*, *Physical object*, *Procedure*, *Qualifier value*, *Record artifact*, *Situation with explicit context*, *SNOMED-CT Model Component*, *Social context*, *Special concept*, *Specimen*, *Staging and scales*, and *Substance*. We manually checked all stem terms based on the criteria of the exact match through the BioPortal site (National Center for Biomedical Ontology) [26].

Matching ratio (%) = Occurrence of a stem term in FN matched with RadLex or SNOMED-CT / Total number of stem terms in FNs × 100% **(5)**

Results

Performance Evaluation of Each Pipeline

The F-measure of the pipeline with the dictionaries Default+RadLex+GPD for compound terms was nearly the same as that of the pipeline with the dictionaries Default+RadLex (31.5% vs 31.4%; Table 1). In step 2—building and evaluating the CtED—the F-measures of the pipeline with the dictionaries Default+RadLex+GPD with and without the CtED were 63.1% and 30.9%, respectively (Table 2).

Table 1. F-measure, precision, and recall of each dictionary (step 1: number of reports=400).

Dictionaries	F-measure, %	Precision, %	Recall, %
Default	27.9	93.4	16.4
Default+RadLex	31.4	94.9	18.8
Default+RadLex+GPD[a]	31.5	93.2	19

[a]GPD: general purpose dictionary.

Table 2. F-measure, precision, and recall of each dictionary (step 2: number of reports=100).

Dictionaries	F-measure, %	Precision, %	Recall, %
Default+RadLex+GPD[a] without enhancement	30.9	73.3	19.6
Default+RadLex+GPD with enhancement	63.1	82.8	51

[a]GPD: general purpose dictionary.

Regarding each imaging modality (Table 3), the F-measure of cTAKES+RadLex+GPD for x-ray was higher (64.3%) than that without enhancement (26.7%). The most frequent stem terms in the FNs were *effusion* (9.1% x-ray), *change* (3.5% CT), *change* (4.1% MRI), and *uptake* (12% PET; Table 4). The number of words in the compound terms in the FPs was mainly 2 (31,774/42,871, 74.12%), 3 (7876/42,871, 18.37%), and 4 (2271/42,871, 5.29%), which is approximately 97.78% (41,921/42,871) of all FNs.

Table 3. F-measure of the compound terms–enhanced dictionary of each modality.

Modality	cTAKES[a]+RadLex+GPD[b] (%)	cTAKES+RadLex+GPD+CtED[c] (%)
Computed tomography	33.5	62.4
MRI[d]	30.7	63.6
PET[e]	30.3	63.4
x-ray	26.7	64.3
All	30.9	63.1

[a]cTAKES: clinical Text Analysis and Knowledge Extraction System.

[b]GPD: general purpose dictionary.

[c]CtED: compound terms–enhanced dictionary.

[d]MRI: magnetic resonance imaging.

[e]PET: positron emission computed tomography.

Table 4. Top five occurrence ratios in each imaging modality.

Modality	Stem	OR[a], n (%)
Computed tomography		
TP[b] (n=1127)		
	lobe	100 (8.87)
	effusion	59 (5.24)
	node	50 (4.44)
	artery	39 (3.46)
	hemorrhage	37 (3.28)
FN[c](n=3532)		
	change	125 (3.54)
	collection	98 (2.77)
	lesion	95 (2.69)
	effusion	94 (2.66)
	evidence	69 (1.95)
MRI[d]		
TP (n=840)		
	artery	146 (17.38)
	lobe	49 (5.83)
	sinus	29 (3.45)
	matter	20 (2.38)
	body	20 (2.38)
FN (n=3732)		
	change	176 (4.72)
	lesion	144 (3.86)
	enhancement	132 (3.54)
	evidence	95 (2.55)
	study	89 (2.38)
PET[e]		
TP (n=1123)		
	node	192 (17.1)
	lobe	102 (9.08)
	gland	69 (6.14)
	nodule	39 (3.47)
	disease	36 (3.21)
FN (n=4708)		
	uptake	567 (12.04)
	node	250 (5.31)
	lesion	180 (3.82)
	avidity	169 (3.59)
	disease	157 (3.33)
x-ray		
TP (n=323)		

Modality	Stem	OR[a], n (%)
	effusion	46 (14.24)
	tube	37 (11.45)
	lobe	27 (8.36)
	edema	18 (5.57)
	lung	17 (5.26)
FN (n=1279)		
	effusion	117 (9.15)
	tube	69 (5.39)
	opacity	67 (5.24)
	pneumothorax	62 (4.85)
	line	57 (4.46)

[a]OR: occurrence ratio.

[b]TP: true positive.

[c]FN: false negative.

[d]MRI: magnetic resonance imaging.

[e]PET: positron emission computed tomography.

In addition, the most frequent FPs that were removed from the cTAKES+RadLex+GPD dictionaries were *related to* (34/239, 14.2%), *abdomen and pelvis* (23/239, 9.6%), and *head and neck* (21/239, 8.8%).

Most Frequent Stem Terms

The ORs of the TPs and FNs in each imaging modality (step 3) are shown in Figure 2. The stem terms of the TPs in the CT reports were more diverse than those in the MRI, PET, and x-ray reports. The FNs in the CT and MRI reports also showed the same trends. The most frequent stem terms in the TPs were *lobe* (100/1127, 8.87% CT), *artery* (146/840, 17.4% MRI), *node* (192/1123, 17.1% PET), and *effusion* (46/323, 14.2% x-ray; Table 4). In contrast, the most frequent stem terms in the FNs were *change* (125/3532, 3.54% CT), *change* (176/3732, 4.72% MRI), *uptake* (567/4708, 12.04% PET), and *effusion* (117/1279, 9.15% x-ray). Table 4 shows that stem terms such as *effusion*, *node*, *tube*, and *disease* had a need in both TPs and FNs.

Figure 2. Occurrence ratio of true positives and false negatives in each imaging modality. CT: computed tomography; FN: false negative; MRI: magnetic resonance imaging; PET: positron emission computed tomography; TP: true positive.

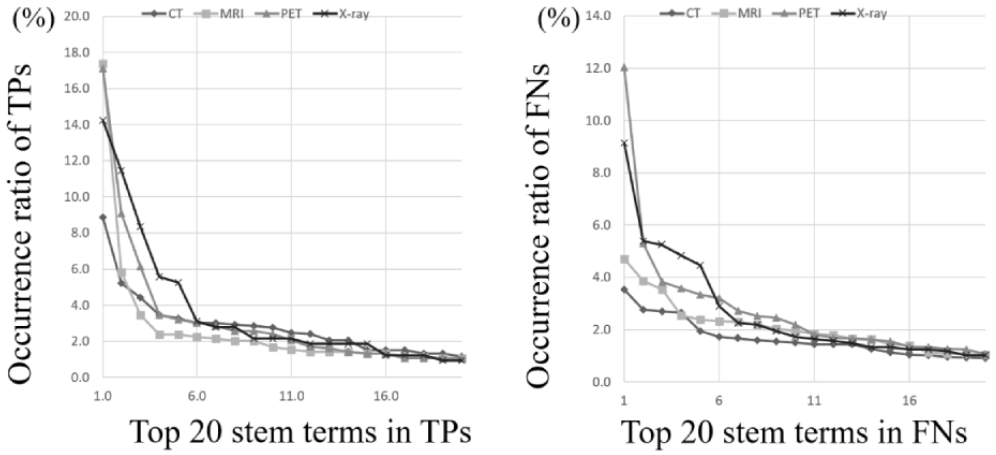


Table 5 illustrates the distribution of the stem terms in FNs that are matched with a RadLex upper concept using the MR. The result of the MR was 71%, which included a connectivity of 47.4% with RadLex and 51.5% with the cTAKES default dictionary (SNOMED-CT and RxNORM). The stem terms that did not match RadLex and the cTAKES default dictionary accounted for 28.15% (3687/13,098). The matched classes in RadLex included clinical finding (1839/13,098, 14.04%), imaging observation (1508/13,098, 11.51%), and process (1000/13,098, 7.63%), and those in the cTAKES default dictionary included Body structure (1428/13,098, 10.9%), Over two category (1265/13,098, 9.66%), and Qualifier value (935/13,098, 7.14%).

Table 5. Classification of stem terms in false negatives based on cTAKES[a], RadLex, and combined dictionary (n=13,098).

Stem terms	Class	Proportion, n (%)
cTAKES default (SNOMED-CT[b])		
	N/A[c]	6349 (48.47)
	Body structure	1428 (10.9)
	Over two categories	1265 (9.66)
	Qualifier value	935 (7.14)
	Clinical finding	878 (6.7)
	SNOMED-CT model component	723 (5.52)
	Procedure	721 (5.5)
	Environment or geographical location	217 (1.66)
	Physical object	206 (1.57)
	Substance	143 (1.09)
	Other	233 (1.78)
RadLex		
	N/A	6893 (52.63)
	Clinical finding	1839 (14.04)
	Imaging observation	1508 (11.51)
	Process	1000 (7.63)
	Anatomical entity	997 (7.61)
	Property	295 (2.25)
	RadLex descriptor	248 (1.89)
	Object	210 (1.6)
	Procedure	91 (0.69)
	Imaging modality	11 (0.08)
	Nonanatomical substance	5 (0.04)
	Report component	1 (0.01)
cTAKES default (SNOMED-CT)+RadLex		
	cTAKES+RadLex	9411 (71.85)
	N/A	3687 (28.15)

[a]cTAKES: clinical Text Analysis and Knowledge Extraction System.

[b]SNOMED-CT: Systematized Nomenclature of Medicine-Clinical Terms.

[c]N/A: not applicable.

Discussion

Overview

In this study, we first constructed RadLex-based NER tools for mining free-text radiology reports and evaluated the coverage of the pipelines (step 1). Second, we built a CtED extracted from the FNs of step 1 to improve performance (step 2). Third, we defined OR and MR to consider the potential of expanding the dictionary using RadLex ontology (step 3).

Performance Evaluation of Each Pipeline (Step 1 and Step 2)

First, the performance of cTAKES+RadLex+GPD was 30.9% (precision 73.3% and recall 19.6%) on its own and 63.1% (precision 82.8% and recall 51%) with the CtED. The CtED for compound terms increased the F-measure by 32.2%, but the F-measure was not obviously changed by the GPD (31.4% vs 31.5%). This indicated that the GPD did not cover the specific compound terms in radiology reports different from the single words. The merit of using RadLex is that we can use the standard vocabularies and relationships such as *Is-A* and *May_cause*. RadLex provides 15 concepts under the top entity,

which can assign labels such as *anatomical entity* and *clinical finding* to each entity.

Our tool using cTAKES was able to customize dictionaries by creating a BSV file, which provides a convenient way to leverage those vocabulary resources that are not covered by the default dictionary. In addition, the BSV file stores IDs that can be used to track the parent concepts for a particular term, which enables the classification or profiling of extracted terms using high-level concept classes defined in a vocabulary.

Stem Term Analysis for Expanding Dictionary (Step 3)

The OR provides profiles of *demand and supply* for stem terms in the corpus. For example, the stem terms of *disease* (PET), *node* (PET), *effusion* (x-ray), and *tube* (x-ray) had a high OR value in both TPs and FNs (Table 4). This means that creating compound terms with high OR–value stem terms in FNs potentially improves precision for capturing entities in each modality's reports compared with the effort of applying the other vast vocabularies in the pipeline. In addition, the features of the FNs also showed that 97.78% (41,921/42,871) of the compound terms consisted of 2-4 words. This fact suggests that NER performance can be effectively improved by identifying *1 to 3 modified words* and *stem term from each imaging modality*. With regard to the MR, RadLex improved 20.33% of the connectivity with stem terms in the FNs compared with the cTAKES default (SNOMED-CT). The contribution of the improvement can provide criteria in terms of whether we should add phrases to RadLex or to SNOMED-CT. Therefore, stem term–related information such as OR and MR would contribute to expanding dictionaries that have ontological structures. This kind of dictionary-based NER would provide ontology-based benefits such as reasoning concepts and using standard codes and vocabularies. Although it is known that CRFs achieve a higher F-measure than dictionary-based approaches, CRFs generate entities that have no hierarchical structure and relationships.

In contrast, our approach is based on an ontology, which enables interoperable processing and data mining of reports. For example, when we identify the term *pleural effusion*, RadLex ontology can guide us to the parent class *effusion* so that we can finally reach the *Clinical findings* tracking upper concepts. RadLex can also provide relationships such as *pleural effusion may cause of vascular cut-off sign*.

Limitations

The limitation of this study is that our pipeline is optimized for identifying short compound terms because we divided compound terms using stop words such as *and*. For example, we set the stop word *and* so that we lead to separate the compound term *abdomen and neck* into *abdomen* and *neck*. This approach has the merit of identifying as possible as the stem term, splitting the long phrase *right pleural effusion and left lung pneumothorax* into *right pleural effusion* and *left lung pneumothorax*. Therefore, in the case of capturing long compound terms, we need to combine short phrases. Generally, noun phrase identification for free-text radiology reports is considered difficult because there are many variants of long compound terms. We believe that our method has the potential to capture long compound terms when applying a combination of single and short compound terms.

Future Work

The annotation tool GATE that we used can identify a partial match with TPs, which means that the types of NER are the same, but the span is not the same. In this study, such partial positives were treated as FNs. We reviewed these uncertainty negatives based on the rule of the stem words and found that 35.4% (90/254) of the partial positives had the potential to change into TPs. This was equivalent to 0.7% of the increased F-measure (cTAKES+RadLex+GPD+CtED). The details of the partial match require further analysis.

The study by Jiang et al [27] demonstrated a state-of-the-art text-mining tool of the Stanford Parser. The study's results showed that POS-based grammatical approaches are efficient in capturing named entities in free-text radiology reports. In future work, we will extract the POS information to define a pattern of the modified words of the compound term.

Lately, Word2Vec technology has been explored for generating synonyms and expanding the radiology-specific dictionary [28,29]. These studies claimed that a machine learning technology such as Word2Vec supports the building of enhanced dictionaries and reduces the annotation cost. We agree with this claim and believe that it is useful to use Word2Vec to calculate vectors of single terms in the noun phrase, creating modifiers for each stem term. In future work, we will generate modified words using this type of machine learning approach. The customized text-mining tool combined with machine learning technology can help further extract features from radiology reports.

Conclusions

In this study, we developed a customized NER tool based on RadLex for the recognition of technical terms. We demonstrated that the CtED and stem term analysis have the potential to improve the performance of the dictionary-based NER with regard to expanding vocabularies.

Conflicts of Interest

None declared.

References

1. Nishimoto N, Terae S, Uesugi M, Ogasawara K, Sakurai T. Development of a medical-text parsing algorithm based on character adjacent probability distribution for Japanese radiology reports. Methods Inf Med 2008;47(6):513-521. [doi: 10.3414/me9127] [Medline: 19057808]
2. Pool F, Goergen S. Quality of the written radiology report: a review of the literature. J Am Coll Radiol 2010 Aug;7(8):634-643. [doi: 10.1016/j.jacr.2010.03.016] [Medline: 20678733]
3. Omary RA, Bettmann MA, Cardella JF, Bakal CW, Schwartzberg MS, Sacks D, Society of Interventional Radiology Standards of Practice Committee. Quality improvement guidelines for the reporting and archiving of interventional radiology procedures. J Vasc Interv Radiol 2003 Sep;14(9 Pt 2):293-295. [doi: 10.1097/01.rvi.0000094601.83406.e1] [Medline: 14514836]
4. International statistical classification of diseases and related health problems (ICD). World Health Organization. URL: https://www.who.int/classifications/classification-of-diseases [accessed 2021-08-19]
5. Markwell D, Sato L, Cheetham E. Representing clinical information using SNOMED Clinical Terms with different structural information models. In: Proceedings of the 3rd international conference on Knowledge Representation in Medicine (KR-MED 2008). 2008 Presented at: 3rd international conference on Knowledge Representation in Medicine (KR-MED 2008); May 31st - June 2nd, 2008; Phoenix, Arizona, USA URL: https://www.semanticscholar.org/paper/Representing-Clinical-Information-using-SNOMED-with-Markwell-Sato/756ae1e77f06c4a93a7b16a342e374424c1cc76c#paper-header
6. Cotik V, Filippo D, Castaño J. An approach for automatic classification of radiology reports in Spanish. Stud Health Technol Inform 2015;216:634-638. [Medline: 26262128]
7. Hong Y, Kahn Jr CE. Content analysis of reporting templates and free-text radiology reports. J Digit Imaging 2013 Oct;26(5):843-849 [FREE Full text] [doi: 10.1007/s10278-013-9597-4] [Medline: 23553231]
8. Langlotz CP. RadLex: a new method for indexing online educational materials. Radiographics 2006;26(6):1595-1597. [doi: 10.1148/rg.266065168] [Medline: 17102038]
9. Oberkampf H, Zillner S, Overton JA, Bauer B, Cavallaro A, Uder M, et al. Semantic representation of reported measurements in radiology. BMC Med Inform Decis Mak 2016 Jan 22;16:5 [FREE Full text] [doi: 10.1186/s12911-016-0248-9] [Medline: 26801764]
10. Segrelles JD, Medina R, Blanquer I, Martí-Bonmatí L. Increasing the efficiency on producing radiology reports for breast cancer diagnosis by means of structured reports. A comparative study. Methods Inf Med 2017 May 18;56(3):248-260. [doi: 10.3414/ME16-01-0091] [Medline: 28220929]
11. Yan Z, Lacson R, Ip I, Valtchinov V, Raja A, Osterbur D, et al. Evaluating terminologies to enable imaging-related decision rule sharing. AMIA Annu Symp Proc 2017 Feb 10;2016:2082-2089 [FREE Full text] [Medline: 28269968]
12. Wang KC, Patel JB, Vyas B, Toland M, Collins B, Vreeman DJ, et al. Use of radiology procedure codes in health care: the need for standardization and structure. Radiographics 2017;37(4):1099-1110 [FREE Full text] [doi: 10.1148/rg.2017160188] [Medline: 28696857]
13. Kundeti S, Vijayananda J, Mujjiga S, Kalyan M. Clinical named entity recognition: challenges and opportunities. In: Proceedings of the IEEE International Conference on Big Data (Big Data). 2016 Presented at: 2016 IEEE International Conference on Big Data (Big Data); Dec 5-8, 2016; Washington, DC, USA. [doi: 10.1109/bigdata.2016.7840814]
14. Pham A, Névéol A, Lavergne T, Yasunaga D, Clément O, Meyer G, et al. Natural language processing of radiology reports for the detection of thromboembolic diseases and clinically relevant incidental findings. BMC Bioinformatics 2014 Aug 07;15(1):266 [FREE Full text] [doi: 10.1186/1471-2105-15-266] [Medline: 25099227]
15. Hassanpour S, Langlotz CP. Information extraction from multi-institutional radiology reports. Artif Intell Med 2016 Jan;66:29-39 [FREE Full text] [doi: 10.1016/j.artmed.2015.09.007] [Medline: 26481140]
16. Spasić I, Zhao B, Jones CB, Button K. KneeTex: an ontology-driven system for information extraction from MRI reports. J Biomed Semantics 2015 Sep 7;6:34 [FREE Full text] [doi: 10.1186/s13326-015-0033-1] [Medline: 26347806]
17. Mowery DL, Chapman BE, Conway M, South BR, Madden E, Keyhani S, et al. Extracting a stroke phenotype risk factor from Veteran Health Administration clinical reports: an information content analysis. J Biomed Semantics 2016 May 10;7:26 [FREE Full text] [doi: 10.1186/s13326-016-0065-1] [Medline: 27175226]
18. Information about RadLex. RSNA Informatics. URL: http://radlex.org/ [accessed 2021-08-19]
19. aesuli/SentiWordNet. GitHub. URL: https://github.com/aesuli/sentiwordnet [accessed 2021-08-19]
20. History. cTAKES. URL: http://ctakes.apache.org/history.html [accessed 2021-08-19]
21. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 2010;17(5):507-513 [FREE Full text] [doi: 10.1136/jamia.2009.001560] [Medline: 20819853]
22. Johnson AE, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 24;3:160035 [FREE Full text] [doi: 10.1038/sdata.2016.35] [Medline: 27219127]
23. Hazen R, Van Esbroeck AP, Mongkolwat P, Channin DS. Automatic extraction of concepts to extend RadLex. J Digit Imaging 2011 Feb;24(1):165-169 [FREE Full text] [doi: 10.1007/s10278-010-9334-1] [Medline: 20838847]

24. Beyer SE, McKee BJ, Regis SM, McKee AB, Flacke S, El Saadawi G, et al. Automatic Lung-RADS™ classification with a natural language processing system. J Thorac Dis 2017 Sep;9(9):3114-3122 [FREE Full text] [doi: 10.21037/jtd.2017.08.13] [Medline: 29221286]
25. Homepage. GATE. URL: https://gate.ac.uk/ [accessed 2021-08-19]
26. Welcome to BioPortal. BioPortal. URL: https://bioportal.bioontology.org/ [accessed 2021-08-19]
27. Jiang M, Huang Y, Fan J, Tang B, Denny J, Xu H. Parsing clinical text: how good are the state-of-the-art parsers? BMC Med Inform Decis Mak 2015;15 Suppl 1(Suppl 1):S2 [FREE Full text] [doi: 10.1186/1472-6947-15-S1-S2] [Medline: 26045009]
28. Maldonado R, Goodwin TR, Skinner MA, Harabagiu SM. Deep learning meets biomedical ontologies: knowledge embeddings for epilepsy. AMIA Annu Symp Proc 2018 Apr 16;2017:1233-1242 [FREE Full text] [Medline: 29854192]
29. Extraction of medical knowledge from clinical reports and chest x-rays using machine learning techniques. University of Alicante. URL: http://rua.ua.es/dspace/handle/10045/102193 [accessed 2021-08-19]

Abbreviations

BSV: bar-separated value
CRF: conditional random field
CT: computed tomography
cTAKES: clinical Text Analysis and Knowledge Extraction System
CtED: compound terms–enhanced dictionary
EMR: electronic medical record
FN: false negative
FP: false positive
GATE: General Architecture for Text Engineering
GPD: general purpose dictionary
MR: matching ratio
MRI: magnetic resonance imaging
NER: named entity recognition
NLP: natural language processing
OR: occurrence ratio
PET: positron emission computed tomography
POS: parts of speech
SNOMED-CT: Systematized Nomenclature of Medicine-Clinical Terms
TP: true positive
UMLS: Unified Medical Language System

Edited by R Kukafka; submitted 10.11.20; peer-reviewed by M Torii, AR Feizi Derakhshi, C Yang, B Alex; comments to author 03.05.21; revised version received 06.07.21; accepted 27.07.21; published 29.10.21.

Please cite as:
Tsuji S, Wen A, Takahashi N, Zhang H, Ogasawara K, Jiang G
Developing a RadLex-Based Named Entity Recognition Tool for Mining Textual Radiology Reports: Development and Performance Evaluation Study
J Med Internet Res 2021;23(10):e25378
URL: https://www.jmir.org/2021/10/e25378
doi:10.2196/25378
PMID:

Original Paper

Factors Influencing the Intention of Actors in Hospitals to Use Indoor Positioning Systems: Reasoned Action Approach

Johannes Wichmann[1,2], BA, MA; Michael Leyer[1,3], Dipl-Volksw, PhD

[1]Chair of Service Operations, Institute of Business Administration, Rostock University, Rostock, Germany

[2]Wismar Business School, Wismar University, Wismar, Germany

[3]Department of Management, Queensland University of Technology, Brisbane, Australia

Corresponding Author:
Johannes Wichmann, BA, MA
Chair of Service Operations
Institute of Business Administration
Rostock University
Ulmenstrasse 69
Rostock, 18057
Germany
Phone: 49 381 498 44
Email: johannes.wichmann@uni-rostock.de

Abstract

Background: Indoor positioning systems (IPS) have become increasingly important for several branches of the economy (eg, in shopping malls) but are relatively new to hospitals and underinvestigated in that context. This research analyzes the intention of actors within a hospital to use an IPS to address this gap.

Objective: To investigate the intentions of hospital visitors and employees (as the main actors in a hospital) to use an IPS in a hospital.

Methods: The reasoned action approach was used, according to which the behavior of an individual is caused by behavioral intentions that are affected by (1) a persuasion that represents the individual's attitude toward the behavior, (2) perceived norms that describe the influence of other individuals, and (3) perceived norms that reflect the possibility of the individual influencing the behavior.

Results: The survey responses of 323 hospital visitors and 304 hospital employees were examined separately using SmartPLS 3.3.3. Bootstrapping procedures with 5000 subsamples were used to test the models (one-tailed test with a significance level of .05). The results show that attitude (β=.536; P<.001; f^2=.381) and perceived norms (β=.236; P<.001; f^2=.087) are predictors of hospital visitors' intention to use an IPS. In addition, attitude (β=.283; P<.001; f^2=.114), perceived norms (β=.301; P<.001; f^2=.126), and perceived behavioral control (β=.178; P=.005; f^2=.062) are predictors of hospital employees' intention to use an IPS.

Conclusions: This study has two major implications: (1) our extended reasoned action approach model, which takes into account spatial abilities and personal innovativeness, is appropriate for determining hospital visitors' and employees' intention to use an IPS; and (2) hospitals should invest in implementing IPS with a focus on (a) navigational services for hospital visitors and (b) asset tracking for hospital employees.

(J Med Internet Res 2021;23(10):e28193) doi:10.2196/28193

KEYWORDS

indoor positioning systems; indoor navigation; indoor localization; hospital; clinic; reasoned action approach; survey; hospital visitors; hospital employees

Introduction

Overview

Hospitals are characterized by high levels of physical movement, with a constant stream of temporary visitors (patients and related visitors), personnel, and mobile technical equipment operating in different locations. While efficiency is a concern, it is also of the utmost importance to ensure high levels of hygiene to avoid contamination and the spread of disease, a necessity highlighted by the COVID-19 pandemic. Consequently, preventing the spread of disease by improving hygiene [1] has been the subject of numerous studies [2,3]. Indoor positioning systems (IPS) can support hospitals' efforts to improve hygiene for visitors and employees in three main ways. First, IPS in hospitals can facilitate wayfinding [4] and support measures against hospital-related infections, such as social distancing [5,6]. Second, IPS can help employees find hospital assets [7] and enable patients to move through different departments [8]. Third, IPS can be used to monitor patients in need of assistance (eg, those with dementia) [9].

Until now, the market penetration for IPS in hospitals has been low because of high implementation costs—roughly US $10200 for approximately 9290 m² [10]. However, as radio-frequency identification tags and Bluetooth beacons have become cheaper, implementing IPS in hospitals is more attractive for hospital management [11]. Commercial implementations of IPS in hospitals in Germany [12] and the United States [13] provide examples of growing interest. Nonetheless, when assessing the costs and benefits, it is important to consider potential user acceptance issues, as high usage rates are necessary to obtain the full benefits of IPS.

Research on the adoption of health care tracking apps has shown the importance of acceptance, notably in the context of COVID-19 [14]. The results highlight the importance of functional and trust-related factors in the use of and intention to use such apps [15]. Some studies have applied model-driven approaches, such as the technology acceptance model [16,17], to different IPS contexts [18]. However, in the hospital context, the only relevant study is that of Anagnostopoulos et al [19], who investigated the IPS needs of employees at Geneva University Hospital.

To investigate the intention of actors in hospitals to use IPS, we adopted the well-established reasoned action approach (RAA) as a causal model to identify relevant influencing factors. The RAA identifies reasons for a specific behavior by considering behavioral, normative, and control beliefs [20]. We surveyed 323 hospital visitors and 304 hospital employees in Germany. We set up a structural equation model (SEM) for both groups that includes factors relevant to the intention to use an IPS.

Our results contribute to understanding which factors influence the intention of actors (ie, hospital visitors and employees) to use systems or applications (ie, IPS) in the health care management context. We show that the RAA, extended to include spatial abilities, can explain the intentions of two major stakeholder groups to use systems in the context of health care management. Hospitals wishing to improve hygiene can apply these insights to encourage IPS usage. This will help tackle a range of issues, from the threat of multiresistant germs to restrictions on hospital visitor numbers during a pandemic. Therefore, we recommend that hospitals invest in the implementation of IPS, taking stakeholder-specific requirements into account.

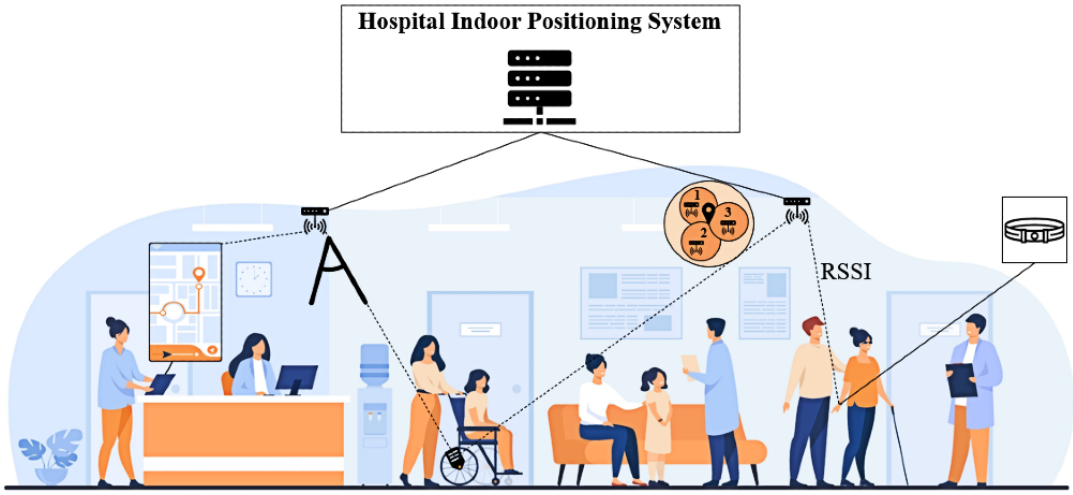
This article is organized as follows: the second and third sections clarify the theoretical background to the research and introduce the hypotheses and research model. The fourth section describes the materials and methods, and the fifth section presents the results, which are discussed in the sixth section. The final section concludes the research, clarifies its implications, and provides an outlook for further investigations.

Theoretical Background

Indoor Navigation/Indoor Localization

An IPS determines the specific position of an individual or an asset [21] using an algorithm that estimates the position of a mobile client. Figure 1 shows how such connections can be established in a hospital setting [22] using a mobile device [23], a tag (ie, attached to a wheelchair), and a wristband [24]. These devices are connected by a set of reference points (ie, routers [25]) within a predefined area [26]. This allows different localization techniques such as Bluetooth or Wi-Fi to be combined with calculation principles to determine specific positions. Frequently used calculation principles are triangulation (represented here by the three circles) and trilateration (represented by the triangle), which use the received signal strength indication of the relevant localization technique [27]. An IPS of this type can be used to track patients in urgent care [8] or to locate insulin pumps [28], ultimately reducing waiting times and redundant activities.

Figure 1. Functional setup of an exemplary positioning system in a hospital [25–28]. RSSI: received signal strength indication.



Research on Indoor Navigation/Indoor Localization in Hospitals

Navigation applications allow the tracking of individuals by connecting localization data with personal data [29]. Research on health care tracking apps has shown the importance of social [14] and behavioral factors [14,15] in relation to usage rates and intention to use. For example, research on COVID-19 apps has established that trust and privacy [15,30,31], as well as voluntary and temporary use, are important factors in acceptance [31]. In addition, a lifestyle that prioritizes hygiene has been identified as a major predictor of using a COVID-19 app, although evidence from Singapore suggests that demographics and situational characteristics are less relevant [32]. Although this previous research has identified factors that may be relevant, it focuses on general app usage. Therefore, we extend it by introducing a specific spatial and organizational context, namely the use of IPS in hospitals.

Model-driven approaches have been adopted in IPS research to account for the navigational requirements of users [18,33]. For example, Arning et al [18] applied the technology acceptance model [16,17] to an IPS that operates using a screen (eg, a smartphone) and a pico-projector. They found strong evidence that disorientation is the most important predictor of screen and projector acceptance. However, their research was limited to young people (ie, university students between the ages of 21-28 years) and may not be generalizable to other age groups. It should also be noted that the technology acceptance model does not include social influences, which are likely to be an important predictor for intention to use an IPS [34].

Within the hospital context, the only relevant study is that of Anagnostopoulos et al [19], who investigated the IPS needs of staff at Geneva University Hospital. They identified five key features of an app: (1) it should show the trajectory toward a destination on a map; (2) it should consider the mobility capabilities of users; (3) it should protect the individual's privacy; (4) it should estimate the position accurately; and (5) it should not require an internet connection to function properly. However, as these results were obtained from a specific case study, they may not be generalizable to users in other contexts.

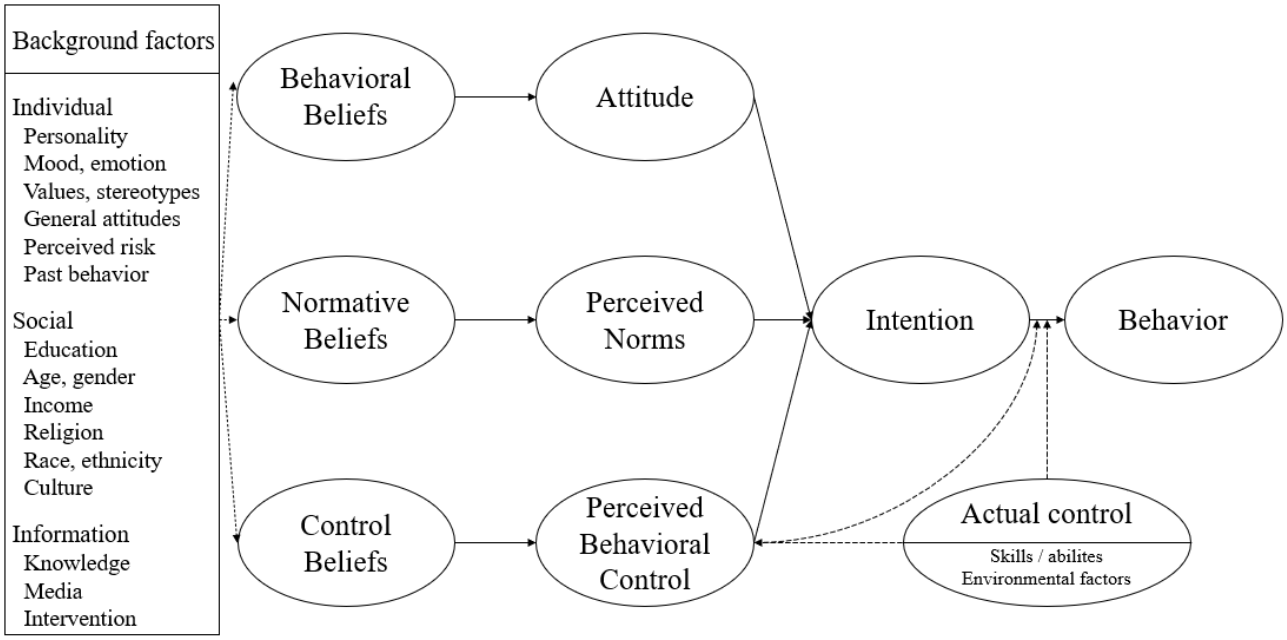
The Reasoned Action Approach

The RAA is a well-established psychological approach based on the theory of reasoned action [35-38], which is widely accepted in psychological studies [39] and is appropriate for ascertaining individual behavior. According to the RAA, individual behavior is caused by behavioral intentions that are rooted in (1) a persuasion that influences the individual's attitude toward the behavior; (2) perceived norms that describe the influence of other individuals; and (3) the opportunity for the individual to affect the behavior, referred to as perceived behavioral control [20]. Figure 2 represents the RAA in greater detail.

An individual's attitude regarding a certain behavior is influenced by his or her beliefs concerning the characteristics and attributes related to the behavior. Thus, an attitude is affected by individual consequences that emerge through assessments of whether or not the behavior is desirable. Therefore, the individual is influenced by whether the behavior is endorsed or opposed by other individuals or groups (those who are most important to her or him in terms of the relevant behavior). The aggregation of motivation and perception assessments for all relevant referent groups is referred to as perceived norms [35,36,38].

Perceived behavioral control determines whether an individual is capable of or directly controls a specific behavior. It is defined by control beliefs that reflect the individual's key personal or situational aspects in relation to the behavior. Ultimately, performing a specific behavior involves the comparison and selection of attitudes, perceived norms, and perceived behavioral controls associated with each of the alternative behaviors in the choice set [40]. Considering these factors together makes it possible to ascertain the likelihood of an individual performing a specific behavior.

Figure 2. The reasoned action approach (RAA) according to Fishbein et al [20].



Hypotheses and Research Model

The RAA is a framework that has to be adjusted to a specific context [20]. In this study, we apply it to the hospital context to predict intention to use an IPS.

First, behavioral beliefs are important for ascertaining the value that an individual perceives in using an IPS. These beliefs cover whether an IPS is perceived as helpful in finding the right location or tracking an object. The positive or negative feelings an individual has toward using an IPS in a hospital (the individual's attitude) are rooted in those beliefs. For the purposes of this study, positive feelings are taken as how the individual feels, as it is the individual who determines whether an IPS is beneficial, satisfactory, relevant, and pleasant to use [41]. The RAA then states that if an individual's attitude toward an IPS is positive, the individual will have a higher intention to use the IPS [20,40]. These considerations lead to the following hypotheses:

- H1: The higher the behavioral beliefs concerning the use of an IPS in a hospital, the more positive an individual's attitude regarding the IPS.
- H2: The more positive an individual's attitude concerning the use of an IPS in a hospital, the higher the intention to use the IPS.

Second, in line with RAA research, we represent the attitudes of other relevant individuals and groups as normative beliefs (subjective norms) [20,42,43]. For hospital visitors, we define family and close friends as relevant social influence groups. For hospital employees, we define immediate colleagues, colleagues in related functional areas, and superiors as relevant influence groups. Normative beliefs generate perceived pressure or motivation, according to whether the individual thinks using an IPS is supported or urged by the reference groups. As implementing an IPS system can be very complex, and the demands on the time and effort of the individual may be high [44], hospital visitors and employees are likely to seek insights from other individuals and groups. In terms of the RAA, the more positive the perception of support from the reference groups, the higher the intention to use an IPS in a hospital. These considerations lead to the following hypotheses:

- H3: The higher the normative beliefs concerning the use of an IPS in a hospital, the more positive an individual's perceived norms regarding the IPS.
- H4: The more positive an individual's perceived norms regarding the use of an IPS in a hospital, the higher the intention to use the IPS.

Third, it is necessary to investigate what facilitates or obstructs an individual's use of an IPS in a hospital. Two of the most critical success factors in relation to information technology projects in hospitals considered are: (1) the complexity of the system and (2) the explanation of how to access it [45]. For the purposes of this investigation, the capability of an individual to use an IPS is dependent on those success factors, which affect whether the individual perceives that she or he controls the new IPS. The individual has to be able to use the IPS under guidance to confirm these control beliefs [46]. Intention to use the system is positively influenced by a higher perceived behavioral control [20,40]. These considerations lead to the following hypotheses:

- H5: The higher the control beliefs concerning an IPS in a hospital, the more positive the perceived behavioral control of an individual regarding the IPS.
- H6: The higher the perceived behavioral control in terms of an IPS in a hospital, the higher the intention to use the IPS.

The navigational skills of the individuals have to be examined to determine confidence in the use of IPS in a hospital (in terms of perceived behavioral control) [47]. Therefore, navigational skills are used here to validate the connection between spatial abilities and intention to use an IPS in a hospital, as well as the connection between spatial abilities and perceived behavioral control. Yao et al [48] determined that spatial abilities are an

important predictor of planning to use a navigational application in outdoor environments. Accordingly, we assume that individuals who are good at navigating through buildings without assistance will be confident about using an IPS in a hospital but will not need to use an IPS urgently. Therefore, we differentiate between hospital visitors and hospital employees. For visitors, we investigate their spatial abilities as a whole, formulating the following hypotheses:

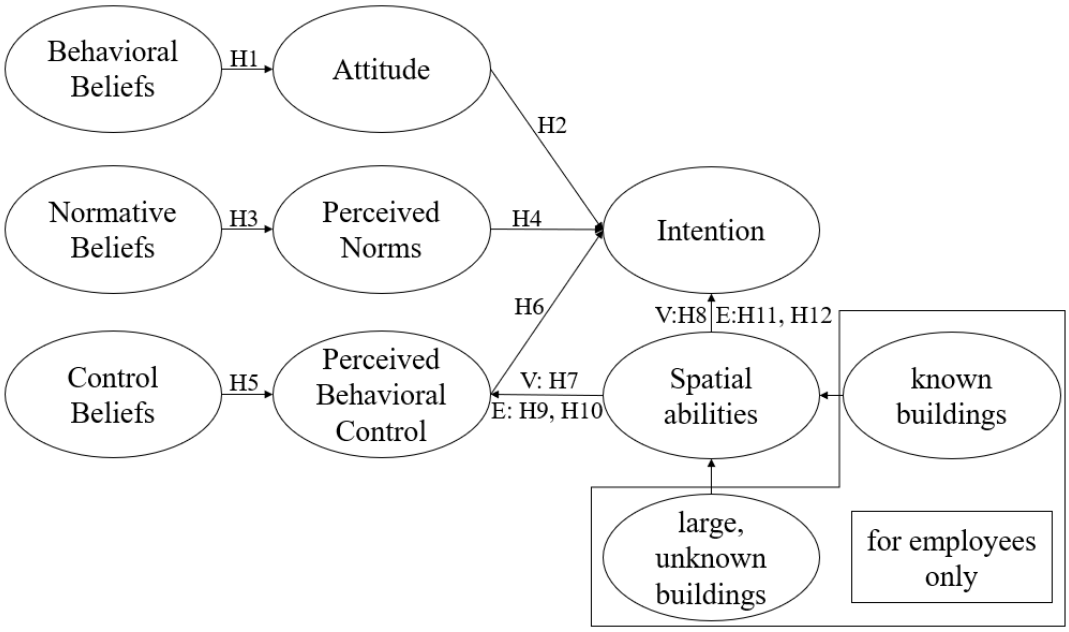
- H7: The higher the spatial abilities, the higher the perceived behavioral control.
- H8: The higher the spatial abilities, the lower the intention to use an IPS in a hospital.

For employees, we investigate their spatial abilities both for buildings that they know (the hospital where they work) and for large unfamiliar buildings, leading to the following hypotheses:

- H9: The higher the spatial abilities for known buildings, the higher the perceived behavioral control.
- H10: The higher the spatial abilities for large unknown buildings, the higher the perceived behavioral control.
- H11: The higher the spatial abilities for known buildings, the lower the intention to use an IPS in a hospital.
- H12: The higher the spatial abilities for unknown buildings, the lower the intention to use an IPS in a hospital.

The research model developed from these hypotheses is shown in Figure 3.

Figure 3. Research model.



Methods

Measures

We used a 7-point Likert scale for each item (from 1 “do not agree at all” to 7 “completely agree”). As Fishbein and Ajzen [20] noted, “it is important to realize that there is no single reasoned action questionnaire. Each investigation requires the construction of a suitable questionnaire.” We, therefore, adjusted the original framework for RAA research to suit the context of IPS in hospitals, creating items that cover relevant behavioral beliefs about time savings and hygiene considerations. In terms of normative beliefs, it is necessary to differentiate between hospital visitors and hospital employees. For visitors, the most relevant normative reference groups are derived from private life, namely family and close friends. For employees, colleagues in the same functional area, colleagues in related functional areas, and superiors are defined as relevant normative groups. In terms of control beliefs, we include ease of access and the necessity of explaining how the IPS functions.

We included several control variables (ie, the size of the hospital and types of buildings, the employees’ work area, how long employees have been working at the hospital, when visitors or patients were present in the hospital, levels of personal innovativeness, and demographic data such as age and gender). The complete questionnaires can be found in A-1 Multimedia Appendix 1.

Participants and Data Collection

The crowdworking platform Clickworker (similar to Amazon MTurk) was used to gather hospital visitors and employees in Germany in April and August 2020. The questionnaires for visitors and employees were separate. We included test questions at the beginning and end of the questionnaire to ensure that the self-reported status was correct. At the beginning of the process, participants also received a text that explained the main function of an IPS. Since the unsupervised online platform paid the participants for their responses, we followed the recommendations of Goodman et al [49] by keeping the questionnaire short and enriching it with attention checks.

Among the hospital visitors, the youngest participant was 18 years of age, and the oldest was 68 years. The mean age was 36.08 years (SD 11.73), with a variance of 137.48 years. A

majority (250/323, 77.4%) were aged between 18 and 44 years, and 22.29% (72/323) were between 45 and 64 years.

For the hospital employees, the mean age was 33.67 years (SD 9.62), with a variance of 92.37 years. We asked the employees to state the main functional area in which they work. The most common area was nursing care (96/304, 31.58%), followed by hospital management (51/304, 16.78%), building services (37/304, 12.17%), diagnosis and therapy (26/304, 8.55%), research, teaching, and training (20/304, 6.58%), emergency medical services (19/304, 6.25%), pastoral care and social services (16/304, 5.26%), supply and waste management (12/304, 3.95%), integrated ambulant care (12/304, 3.95%), kindergarten for employees (11/304, 3.62%), hospice care (3/304, 0.97%), and patient accommodation (1/304, 0.33%).

Validity and Reliability

A partial least squares approach to SEM was used to test the proposed models for hospital visitors and employees. Variance-based SEM is more suitable than covariance-based SEM in cases where the aim is to explain and predict the target construction in structural models or to identify key drivers [50]. Multiple regression analysis, an example of variance-based SEM, develops parameters that "maximize the explained variance of dependent constructs" [50]. We used SmartPLS (version 3.3.3; SmartPLS GmbH) to evaluate our models, estimating our weightings with a path method and determining the significance of the path coefficients using bootstrapping procedures with 5000 samples [50]. We followed the requirements of Hair et al [50] and Hulland [51] by testing (1) internal consistency reliability, (2) indicator reliability, (3) convergent validity, and (4) discriminant validity.

First, composite reliability, used to examine internal consistency, was confirmed for both visitors and employees (A-5 Multimedia Appendix 1). Second, we investigated the reliability of the indicators concerning the reflective variables "attitude," "perceived norms," and "perceived behavioral control" and found the requirements to be fulfilled for both groups (A-2 Multimedia Appendix 1). Third, convergent validity in terms of the reflective variables was confirmed for both groups (A-5 Multimedia Appendix 1). Fourth, the discriminant validity of our measures was investigated using heterotrait-monotrait ratios and confirmed for both groups (A-6 Multimedia Appendix 1). Thus, we conclude that the reliability and validity of the reflective measures are adequate.

The variance inflation factor was used to check for multicollinearity among the indicators for formative belief variables. For both groups, the values were in line with requirements (A-3 Multimedia Appendix 1). The outer weights and loadings used to test the relative and absolute importance of indicators were all significant for both groups (A-4 Multimedia Appendix 1). To check heterogeneity between the indicators, we determined whether the bivariate correlations were higher between an indicator and the variable than between the indicators [52]. Investigation of the results identified no suppressors and no collinear indicators for either group.

We also conducted several tests to ascertain the quality of our structural model. We used the standardized root mean square residual (SRMR) to determine the approximate fit for our composite factor and common factor models [53]. We obtained .075 for the SRMR composite factor model for the visitors and .10 for the SRMR common factor model. For the employees, the values were .55 and .085, respectively. To evaluate the prediction relevance of the models [54], we followed the literature in using blindfolding procedures with an omission distance of 7 [55]. Both tests yielded positive Stone–Geisser Q2 values (A-7 Multimedia Appendix 1), allowing us to conclude that the models have strong overall predictive power [54].

Results

The descriptive statistics and correlations for both our samples are given in Table 1. Note that variable 7 applies to visitors and variables 8 and 9 to employees only.

The results of our analysis concerning the hospital visitors are presented in Figure 4.

For the visitors, strong empirical evidence was found in support of H1 (β=.728; P<.001; f^2=1.153), H3 (β=.767; P<.001; f^2=1.389), and H5 (β=.414; P<.001; f^2=0.179), which indicates that the respective beliefs are relevant antecedents. Furthermore, an increase in R^2 concerning behavioral beliefs resulted in a higher positive attitude, and 60.6% of the variance can be explained by the behavioral beliefs. Regarding the normative beliefs, the explainable variance in perceived norms is similarly strong (63.2%). In contrast, the variance explained by the control beliefs toward perceived behavioral control is comparatively low (23.6%).

Our investigation of H2 (β=.536; P<.001; f^2=.381), H4 (β=.236; P<.001; f^2=.087), and H8 (β=–.089; P=.015; f^2=.019) supported H2 and H4 but not H8. We determined that attitude has a strong influence on intention to use an IPS in a hospital and that perceived norms (as assessments of the intentions of family and close friends) also have an influence. When we consider navigational skills, it is conspicuous that H8 yields a negative value, suggesting that an increase in spatial abilities leads to a lower intention to use IPS in a hospital.

We found that perceived behavioral control is not a predictor of intention to use an IPS (β=.056; P=.129; f^2=.006). H6 is therefore not supported. In contrast, H7 is supported, as spatial abilities are a predictor of perceived behavioral control (β=.137; P<.001; f^2=.023). We used control variables to verify the research model further and found that they had no significant influence, with the exception of personal innovativeness on attitude (β=.114; P=.001; f^2=0.029) and on perceived behavioral control (β=.139; P=.013; f^2=0.020). The results of the research model regarding hospital employees are summarized in Figure 5.

Table 1. Descriptive statistics for the overall sample and correlations among variables for visitors (V) and employees (E).

Variable[a]	Mean (SD)	1	2	3	4	5	6	7	8	9	10
1	V[b]: 33.04 (11.2) E: 29.94 (11.63)	–[c]	V: .51*** E: .70**	V: .67*** E: .71**	V: .77*** E: .74**	V: .53*** E: .63**	V: .38*** E: .51**	V: –.21***	E: –.04	E: .20**	V: .70*** E: .70**
2	V: 23.75 (10.44) E: 27.35 (10.34)		–	V: .52*** E: .688**	V: .56*** E: .63**	V: .79*** E: .74**	V: .21*** E: .54**	V: .03	E: .01	E: .21**	V: .49*** E: .62**
3	V: 30.24 (11.27) E: 29.92 (11.84)			–	V: .67*** E: .67**	V: .58*** E: .65**	V: .28*** E: .50**	V: –.15**	E: –.08	E: .19**	V: .71*** E: .70**
4	V: 5.55 (1.13) E : 5.39 (1.14)				–	V: .59*** E: .66**	V: .36*** E: .43**	V: –.16**	E: .05	E: .11**	V: .74*** E: .65**
5	V: 4.64 (1.30) E: 4.84 (1.23)					–	V: .15** E: .39**	V: –.09	E: .09	E: .07	V: .59*** E: .66**
6	V: 6.04 (0.94)E: 5.39 (1.24)						–	V: .10	E: .03	E: .24**	V: .31*** E: .48**
7	V: 4.26 (1.25)							–	–	–	V: –.19**
8	E: 4.05 (1.50)								–	E: .43**	E: –.10
9	E: 5.30 (1.08)									–	E: .04
10	V: 5.34 (1.54) E: 5.24 (1.47)										–

[a]Number assignment: 1=behavioral beliefs; 2=normative beliefs; 3=control beliefs; 4=attitude; 5=perceived norms; 6=perceived behavioral control; 7=spatial ability; 8=spatial ability large, unknown buildings; 9=spatial ability known buildings; 10=intention.

[b]V: n=323; E: n=304.

[c]Not applicable.

*P<.05; **P<.01; ***P<.001; one-tailed tests.

We found strong empirical evidence for H1 (β=.736; P<.001; f^2=1.038), H3 (β=.719; P<.001; f^2=0.999), and H5 (β=.476; P<.001; f^2=0.244), which again indicates that these beliefs are relevant antecedents. The R^2 results were similar to those for the visitors' model, in that the behavioral and normative beliefs have a strong influence on attitude (56.5%) and perceived norms (58.7%). The influence of the control beliefs on perceived behavioral control (32.9%) is higher than in the visitors' model.

The results support H2 (β=.283; P<.001; f^2=0.114), H4 (β=.301; P<.001; f^2=0.126), H6 (β=.178; P<.001; f^2=0.062), H11 (β=–.023; P=.310; f^2=0.001), and H12 (β=–.140; P<.001; f^2=0.041), although the results for H11 are not significant. Thus, all the reflective variables of the RAA (attitude, perceived norms, and perceived behavioral control) are significant for intention to use. Moreover, in line with H12, positive spatial abilities concerning large unknown buildings negatively influence intention to use. Investigation of H9 (β=.137; P=.014; f^2=0.001) and H10 (β=–.006; P=.460; f^2=0.000) showed that the correlations are not significant. We also established that our control variables, with the exception of gender, had no significant influence on spatial abilities for large unknown buildings (β=.162; P=.003; f^2=0.027), personal innovativeness on intention to use (β=.277; P<.001; f^2=0.168), and the structural unit in which the employees are employed on spatial abilities for known buildings (β=.162; P=.006; f^2=0.023).

Figure 4. Research model results for hospital visitors. *P<.05; **P<.01; ***P<.001; ns: not significant.

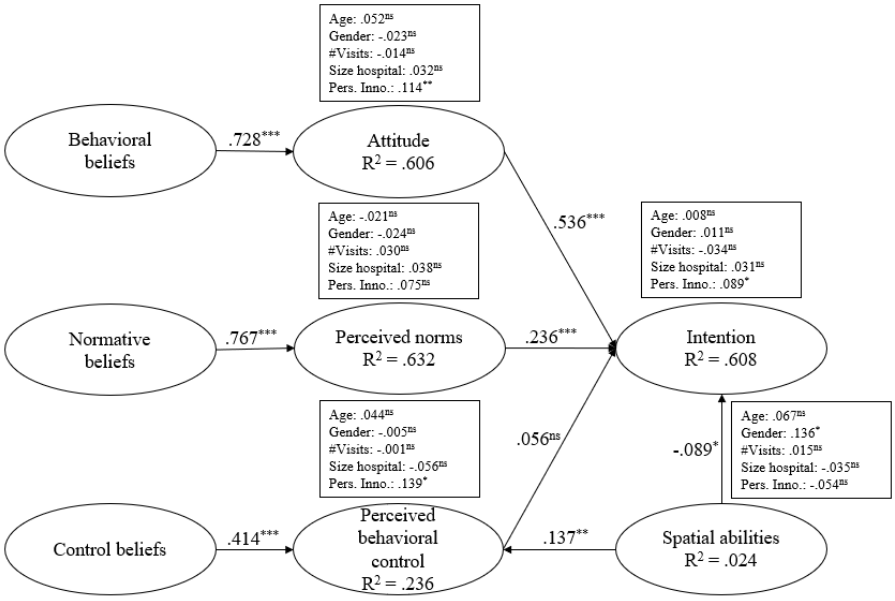
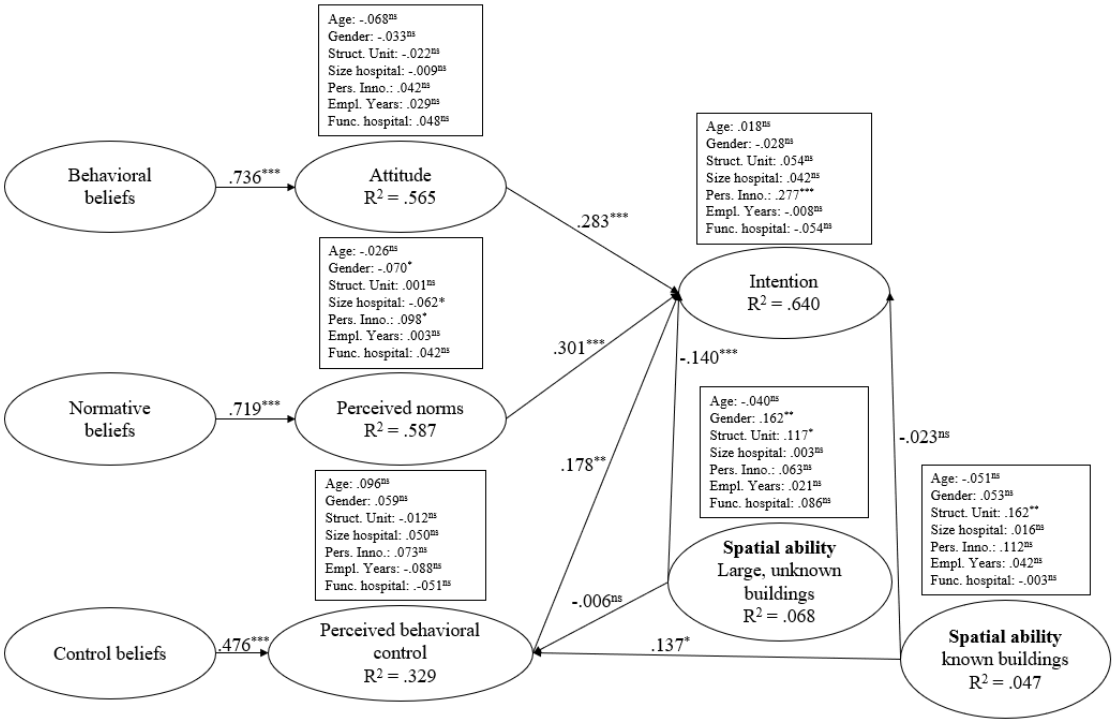


Figure 5. Research model results for hospital employees. *P<.05; **P<.01; ***P<.001; ns: not significant.



As gender and age are important factors in spatial ability [56-61] and personal innovativeness [62,63], we post hoc analyzed our data accordingly. We divided the data set into subgroups for women (visitors: n=109; employees: n=124) and men (visitors: n=211; employees: n=178), for ages 18-33 years (n=162) and 34-68 years (n=161) for visitors (combining both genders), and for ages 18-32 years (n=152) and 33-63 years (n=152) for employees. Table 2 gives the significances for the respective groups.

Table 2. Post hoc analysis by gender and age.

No.	Group	Gender	Age, years	Correlation/(f^2)	β	*P* value
1	Visitors	Women	All	SA[a]>PBC[b]/(0.025)	.226	.008
2	Visitors	Men	All	SA>PBC	.061	.188
3	Visitors	Both	34–68	SA>PBC	.216	.001
4	Visitors	Both	18–33	SA>PBC	.058	.235
5	Visitors	Women	34–68	SA>PBC	.300	.018
6	Visitors	Women	18–33	SA>PBC	.172	.098
7	Visitors	Men	34–68	SA>PBC	.133	.078
8	Visitors	Men	18–33	SA>PBC	–.034	.371
9	Employees	Women	All	SA/KB[c]>PBC/(0.020)	.109	.162
10	Employees	Men	All	SA/KB>PBC	.172	.018
11	Employees	Both	33–63	SA/KB>PBC	.071	.232
12	Employees	Both	18–32	SA/KB>PBC	.177	.025
13	Employees	Women	33–63	SA/KB>PBC	.245	.042
14	Employees	Women	18–32	SA/KB>PBC	.353	.002
15	Employees	Men	33–63	SA/KB>PBC	.169	.046
16	Employees	Men	18–32	SA/KB>PBC	.141	.138
17	Employees	Women	All	SA/LUB[d]>I[e]/(0.041)	–.154	.008
18	Employees	Men	All	SA/LUB>I	–.104	.026
19	Employees	Both	33–63	SA/LUB>I	–.212	.000
20	Employees	Both	18–32	SA/LUB>I	–.076	.086
21	Employees	Women	33–63	SA/LUB>I	–.078	.161
22	Employees	Women	18–32	SA/LUB>I	–.044	.299
23	Employees	Men	33–63	SA/LUB>I	–.161	.019
24	Employees	Men	18–32	SA/LUB>I	–.009	.453
25	Visitors	Women	All	PI[f]>Att[g]/(0.024)	.094	.072
26	Visitors	Men	All	PI>Att	.113	.011
27	Visitors	Both	34–68	PI>Att	.027	.285
28	Visitors	Both	18–33	PI>Att	.200	.000
29	Visitors	Women	34–68	PI>Att	–.060	.239
30	Visitors	Women	18–33	PI>Att	.223	.006
31	Visitors	Men	34–68	PI>Att	.087	.103
32	Visitors	Men	18–33	PI>Att	.119	.043
33	Visitors	Women	All	PI>PBC	.189	.047
34	Visitors	Men	All	PI>PBC	.119	.045
35	Visitors	Both	34–68	PI>PBC	.175	.018
36	Visitors	Both	18–33	PI>PBC	.100	.138
37	Visitors	Women	34–68	PI>PBC	.237	.090
38	Visitors	Women	18–33	PI>PBC	.181	.153
39	Visitors	Men	34–68	PI>PBC	.163	.052
40	Visitors	Men	18–33	PI>PBC	.047	.337
41	Employees	Women	All	PI>I/(0.168)	.248	.001

No.	Group	Gender	Age, years	Correlation/(f^2)	β	*P* value
42	Employees	Men	All	PI>I	.319	.000
43	Employees	Both	33–63	PI>I	.243	.000
44	Employees	Both	18–32	PI>I	.315	.000
45	Employees	Women	33–63	PI>I	.186	.040
46	Employees	Women	18–32	PI>I	.201	.039
47	Employees	Men	33–63	PI>I	.238	.003
48	Employees	Men	18–32	PI>I	.440	.000

[a]SA: spatial ability.

[b]PBC: perceived behavioral control.

[c]KB: known buildings.

[d]LUB: large unknown buildings.

[e]I: intention.

[f]PI: personal innovativeness.

[g]Att: attitude.

Discussion

Principal Findings

In our investigation of intention to use an IPS in a hospital, we identified significant differences between visitors and employees. First, while perceived behavioral control is not significant in determining visitors' intention to use (β=–.056; *P*=.129; f^2=0.016), it is significant for employees (β=.178; *P*=.005; f^2=0.062). Thus, active control over the intention to use an IPS is more relevant for employees than visitors. This might reflect the fact that employees are more experienced than visitors in finding their way around a hospital. Other studies concerning navigational [48] and health care–related [64,65] occupational contexts seem to support this theory.

Second, spatial abilities are significant for perceived behavioral control regarding hospital visitors (β=.137; *P*=.006; f^2=0.023) and known buildings (β=.137; *P*=.014; f^2=0.020). However, they are not significant for large unknown buildings from the viewpoint of hospital employees (β=–.006; *P*=.460; f^2=0.000). Thus, the urgency of using an IPS in a building known to the employee (eg, the hospital where she or he is employed) is lower if the employee's spatial abilities are high, but this is not the case for large unknown buildings. Likewise, spatial abilities are not a predictor of visitors' intention to use an IPS (β=–.089; *P*=.015; f^2=0.016) or the spatial abilities of employees with regards to known buildings (β=–.023; *P*=.310; f^2=0.001). In contrast, spatial abilities are a predictor for employees using an IPS with respect to large unknown buildings (β=–.140; *P*<.001; f^2=0.041), which indicates that employees have an intention to use an IPS if the building is large and unfamiliar.

For visitors, personal innovativeness is not significant for intention to use an IPS (β=.089; *P*=.022; f^2=0.016); however, it is significant for employees (β=.277; *P*<.001; f^2=0.168). This insight aligns with previous research, as personal innovativeness is an important predictor of behavioral intention [56,57].

In current research on spatial abilities, the influence of gender is disputed; research that uses abstract measures, such as mental rotation, indicates that men are better than women at wayfinding [58,59], while research in indoor contexts has identified no major gender differences [64,65]). In this study, for hospital visitors, we found that the older age group (those aged 34-68 years), and especially women, tend to be more realistic about their spatial abilities and their need to use an IPS (see Table 2, numbers 1-8). This suggests that women have greater feelings of uncertainty about wayfinding in a building. However, although women are more likely to use navigation systems [48], actual wayfinding performance does not differ by gender [66].

The findings concerning the impact of the spatial abilities of employees for known buildings on perceived behavioral control align with the findings for visitors. However, it should be noted that the path is also significant for male employees aged 33-63 years (see Table 2, numbers 9-16). These results support the view that physical age and improved experience are positively related, as navigational experience initially increases with age [48], before decreasing in elderly people (an age group not represented in this study) [60,61]. The results in relation to large unknown buildings show that, for both genders and all the age groups under study, higher spatial abilities lead to lower intention to use an IPS in a hospital. However, there is some discrepancy in the results for the different age groups, with a significance for men aged 33-63 years (see Table 2, numbers 17-24), which we ascribe to experience in navigation [48].

For unfamiliar environments, other aspects may be more relevant in determining the urgency of navigational assistance and thus intention to use an IPS, such as the complexity of the environment [67]. In terms of the influence of personal innovativeness on the attitude of hospital visitors, we determined that the path is significant for men and for younger individuals (those aged 18-33 years; see Table 2, numbers 25-32). Concerning personal innovativeness and perceived behavioral control, the path is mainly driven by older participants and is independent of gender (see Table 2, numbers 33-40).

Hence, our results support the consensus in technology adoption research that there is a gender difference. Men's decisions to adopt new technology are driven mainly by their attitude toward

the technology, whereas women's decisions are driven by subjective norms and perceived behavioral control [63]. Concerning the influence of personal innovativeness on attitude and perceived behavioral control, research has determined that attitude toward new technology is more relevant for younger workers, whereas perceived norms and perceived behavioral control are more relevant for older workers [62]. Our findings support these insights by identifying a positive influence of personal innovativeness on intention to use for all genders and age groups (see Table 2, numbers 41-48). An IPS is aimed at individuals who like to explore and experiment with new information technologies, which is a common perception in research on information system adoption and use [68,69].

To clarify the influence of the employees' structural unit on their spatial abilities for known buildings, we post hoc analyzed our data set according to the functional areas in which the individuals are employed. Thus, we distinguished between employees who move through hospital buildings frequently because of their occupation (ie, those in nursing care, building services, and emergency medical services) and those who work mainly in the same place (all the other functional areas represented in our data; see "Data Collection And Participants"). We found that employees who work mainly in the same place are more confident in their spatial abilities in relation to known buildings (β=.194; P=.017, f^2=0.023) than those participants frequently moving (β=.088; P=.172), which we ascribe to the fact that those employees who work mainly in the same place have a lower range of motion in the hospital and have to know a fewer number of floors or buildings, respectively.

Concerning the core model of the RAA, our investigation indicates that attitude and perceived norms are strong predictors of intention to use an IPS in a hospital. For hospital employees, the results are more differentiated; all the reflective variables of the RAA (attitude, perceived norms, and perceived behavioral control) are significant for intention to use, with perceived norms having the strongest influence. Attitude driven by behavioral beliefs is a major predictor of intention to use [20]. Our model indicates that this is the case for hospital visitors and confirms that it is important for hospital employees. In terms of perceived norms, rational choice theorists argue that individual behavior is usually conducted in accordance with self-interest and that we, therefore, accept social norms as limits on those behaviors. In this article, we ensure that social norms do not represent an individual's interest only but that of a larger social system [70]. We established that perceived norms significantly influence intention to use the system for both hospital visitors (with family and close friends as the reference groups) and hospital employees. Moreover, perceived norms are the most important predictors for employees, reflecting the importance of recommendations from immediate colleagues and colleagues working in other functional areas and superiors.

The descriptive statistics for spatial abilities show a mean of 4.18 (SD 1.56) for visitors, and for employers, a mean of 4.05 (SD 1.69) for large unknown buildings and a mean of 5.29 (SD 1.34) for known buildings. These results indicate that employees tend to navigate better through known buildings than through large unknown buildings, although no such tendency is found for visitors. For the influence of the personal innovativeness of employees on their intention to use an IPS, the mean value of 4.86 (SD 1.54) suggests that employees intend to use an IPS if they are personally innovative in terms of new technologies (see A-1 Multimedia Appendix 1). The mean values for all three intention items (on a scale of 1 to 7; visitors: n=323 and employees: n=304) are as follows: intention 1 (visitors: mean 5.35, SD 1.60; employees: mean 5.36, SD 1.53); intention 2 (visitors: mean 5.39, SD 1.57; employees: mean 5.23, SD 1.58); intention 3 (visitors: mean 5.27, SD 1.61; employees: mean 5.15, SD 1.57). Thus, our model indicates that both visitors and employees have a positive intention to use an IPS. Accordingly, we recommend that hospitals pursue IPS implementation.

Conclusions

We analyzed the relevance of IPS in hospitals by considering the perspectives of the main actors, visitors, and employees. The explained variance indicates that intention to use is well predicted and that relevant aspects in the context are covered. This confirms that RAA is an appropriate approach for determining intention to use an IPS in a hospital. Furthermore, our results show that individual attitude and the social norms of relevant reference groups positively impact intention to use an IPS in a hospital. For employees, perceived behavioral control also positively influences intention to use an IPS. These results have many implications for theory, practice, and future research.

Theoretical Implications

Our study design and findings contribute to the literature in several ways. First, we add to the knowledge of how systems or applications, specifically IPS, in the health care management context are accepted by actors in a hospital. Whereas related work regarding general health care tracking apps, including COVID-19-related apps [14], has focused on general use with a broad public interest, we provide insights into a spatially limited organizational context.

Second, we integrate two major stakeholder groups into our analysis: general users, such as patients or visitors, and professional staff. As such, we demonstrate how health care management applications are perceived from a nonexpert perspective, thereby building on previous research, which has generally adopted an expert perspective [15].

Third, we introduce the RAA to analyze intention to use applications in the health care management context, thereby extending the theory conceptually and empirically into a context that considers spatial abilities and personal innovativeness. The high explained variance confirms that the theory is helpful for understanding the reasons for adoption intentions. This increased focus on analyzing the influence of different beliefs from a functional perspective extends other theories that have been applied in the context, such as uncertainty reduction theory [14] and protection motivation theory [71].

Fourth, our extension of the RAA to cover spatial abilities and personal innovativeness contributes to the understanding of gender-related and age-related spatial ability. Hence, we demonstrate that demographics matter and should be considered when analyzing the acceptance of applications in a health care management context.

Practical Implications

From a practical perspective, we recommend that hospitals invest in implementing IPS, as our results show that the potential user intention is high. Furthermore, IPS market research forecasts indicate that low-energy Bluetooth will be one of the most lucrative segments of the IPS market [10], thanks to the low hardware costs and low energy consumption [72]. These forecasts lend support to our recommendation.

However, the IPS design requirements of hospital visitors and employees are different. From our finding that visitors' attitudes and perceived norms are the most important predictors of their intention to use, it follows that the system needs to be simple and self-explanatory. The main focus of the application should be navigation to specific rooms or points of interest. If those services function properly, visitors are likely to recommend the system to reference groups that are important to them (eg, close friends and family), who will then assess and use the system accordingly.

For hospital employees, attitude and perceived norms are also relevant. However, the system needs a different functional focus for employees, whose intentions are determined by perceived behavioral control. Our research model shows that employees that work mainly in the same place are confident in their spatial abilities for known buildings. In other words, they do not need navigational services for specific rooms or points of interest in the hospital building in which they are employed. Asset tracking, in contrast, is more relevant, as this can facilitate daily work and help reduce redundant activity.

Limitations and Future Research

Our study is subject to some limitations that inform future research. First, we used the crowdworking platform Clickworker to gather our participants. This decision partly predetermined the personal innovativeness of our respondents, as individuals who use digital platforms are likely to be more personally innovative than those who respond to a pen and paper survey. Second, our study design involves convenience sampling, albeit with specific criteria for participation. Thus, we cannot claim that our sample is representative, and further research should focus on a defined target population. Third, our participants are from a single country, Germany. Future studies should cover different countries to identify additional relevant factors. Fourth, our research does not consider other settings, such as large hardware stores, that may be relevant to and interact with the hospital context. Therefore, future research should investigate general acceptance of IPS by, for example, determining the likelihood of using an IPS in a hardware store after using it in a hospital.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The Multimedia Appendix contains: the questionnaire (A-1), loadings of reflective variables (A-2), VIF values (A-3), loadings and weights of formative variables (A-4), composite reliability and AVE (A-5), HTMT- (A-6) and Stone-Geisser-values (A-7) of our study.
[DOCX File , 30 KB - jmir_v23i10e28193_app1.docx]

References

1. Aiello AE, Larson EL. What is the evidence for a causal link between hygiene and infections? The Lancet Infectious Diseases 2002 Feb;2(2):103-110. [doi: 10.1016/s1473-3099(02)00184-6]
2. Cerulli Irelli E, Orlando B, Cocchi E, Morano A, Fattapposta F, Di Piero V, et al. The potential impact of enhanced hygienic measures during the COVID-19 outbreak on hospital-acquired infections: A pragmatic study in neurological units. J Neurol Sci 2020 Nov 15;418:117111 [FREE Full text] [doi: 10.1016/j.jns.2020.117111] [Medline: 32892033]
3. Härter M, Bremer D, Scherer M, von dem Knesebeck O, Koch-Gromus U. [Impact of COVID-19-pandemic on Clinical Care, Work Flows and Staff at a University Hospital: Results of an Interview-study at the UKE]. Gesundheitswesen 2020 Sep;82(8-09):676-681 [FREE Full text] [doi: 10.1055/a-1226-6828] [Medline: 32823355]
4. Yoo S, Jung SY, Kim S, Kim E, Lee K, Chung E, et al. A personalized mobile patient guide system for a patient-centered smart hospital: Lessons learned from a usability test and satisfaction survey in a tertiary university hospital. Int J Med Inform 2016 Jul;91:20-30. [doi: 10.1016/j.ijmedinf.2016.04.003] [Medline: 27185506]
5. Fazio M, Buzachis A, Galletta A, Celesti A, Villari M. A proximity-based indoor navigation system tackling the COVID-19 social distancing measures. Piscataway, NJ: IEEE; 2020 Oct 12 Presented at: 2020 IEEE Symposium on Computers and Communications (ISCC); 2020 07 07-10; Rennes, France p. 1-6. [doi: 10.1109/ISCC50000.2020.9219634]
6. Wichmann J, Leyer M. Why do visitors intend to use Indoor Navigation and Indoor Localization Systems in Hospitals? A Quantitative Survey from Germany. In: Proceedings of the 54th Hawaii International Conference on System Sciences. Honolulu, HI. USA: University of Hawai'i at Manoa; 2021 Jan 05 Presented at: Hawaii International Conference on System Sciences 2021; 2021 01 04-08; Shidler College of Business, University of Hawaii at Manoa p. 3694-3703. [doi: 10.24251/HICSS.2021.447]

7. Van Haute T, De Poorter E, Crombez P, Lemic F, Handziski V, Wirström N, et al. Performance analysis of multiple Indoor Positioning Systems in a healthcare environment. Int J Health Geogr 2016 Feb 03;15:7 [FREE Full text] [doi: 10.1186/s12942-016-0034-z] [Medline: 26842830]
8. Konecny J, Prauzek M, Martinek R, Michalek L, Tomis M. Real-time Patient Localization in Urgent Careystem Design and Hardware Perspective. In: 2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom). Piscataway, NJ: IEEE; 2018 Nov 12 Presented at: International Conference on e-Health Networking, Applications and Services (Healthcom); 2018 09 17-20; Ostrava, Czech Republic p. 1-5. [doi: 10.1109/HealthCom.2018.8531110]
9. Hsiao C, Sung Y, Lau S, Chen C, Hsiao F, Chu H, et al. Towards long-term mobility tracking in NTU hospital's elder care center. In: IEEE International Conference on Pervasive ComputingCommunications workshops (PerCom workshops). Piscataway, NJ: IEEE; 2011 May 12 Presented at: IEEE International Conference on Pervasive ComputingCommunications workshops (PerCom workshops); 2011 03 21-25; Seattle, WA, USA p. 649-654. [doi: 10.1109/PERCOMW.2011.5766969]
10. Market Research Report: Indoor Positioning and Indoor Navigation (IPIN) Market 2020. Lanjudkar P. URL: https://www.alliedmarketresearch.com/indoor-positioning-and-indoor-navigation-ipin-market [accessed 2020-09-23]
11. Chen C, Hsieh C, Liao Y, Yin T. Implementation of Wearable Devices and Indoor Positioning System for a Smart Hospital Environment. Piscataway, NJ: IEEE; 2018 Nov 19 Presented at: 2018 International Symposium in Sensing and Instrumentation in IoT Era (ISSI); 2018 09 06-07; Shanghai, China p. 1-5. [doi: 10.1109/ISSI.2018.8538176]
12. Success Story:Helions Klinikum Erfurt. insoft GmbH. URL: https://www.infsoft.com/de/branchen/gesundheit-pharma [accessed 2020-05-19]
13. Technology. Sonitor IPS Holding AS. 2020. URL: https://www.sonitor.com/ultrasound-innovation [accessed 2020-05-19]
14. Oldeweme A, Märtins J, Westmattelmann D, Schewe G. The Role of Transparency, Trust, and Social Influence on Uncertainty Reduction in Times of Pandemics: Empirical Study on the Adoption of COVID-19 Tracing Apps. J Med Internet Res 2021 Feb 08;23(2):e25893 [FREE Full text] [doi: 10.2196/25893] [Medline: 33465036]
15. von Wyl V, Höglinger M, Sieber C, Kaufmann M, Moser A, Serra-Burriel M, et al. Drivers of Acceptance of COVID-19 Proximity Tracing Apps in Switzerland: Panel Survey Analysis. JMIR Public Health Surveill 2021 Jan 06;7(1):e25701 [FREE Full text] [doi: 10.2196/25701] [Medline: 33326411]
16. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
17. Dillon A, Morris M. User acceptance of new information technology - theories and models. Annual Review of Information Science and Technology 1996;31:3-32 [FREE Full text]
18. Arning K, Ziefle M, Li M, Kobbelt L. Insights into user experiencesacceptance of mobile indoor navigation devices. In: Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia. New York, NY: ACM; 2012 Dec 04 Presented at: International Conference on Mobile and Ubiquitous Multimedia; 2012 12 04-06; Ulm, Germany p. 1-10. [doi: 10.1145/2406367.2406417]
19. Anagnostopoulos G, Deriaz M, Gaspoz J, Konstantas D, Guessous I. Navigational needs and requirements of hospital staff: Geneva University hospitals case study. In: 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN). Piscataway, NJ: IEEE; 2017 Nov 23 Presented at: International Conference on Indoor Positioning and Indoor Navigation (IPIN); 2017 09 18-21; Sapporo, Japan. [doi: 10.1109/IPIN.2017.8115958]
20. Fishbein M, Ajzen I. Predictingchanging behavior: The reasoned action approach. New York, NY: Hove: Psychology Press; 2010.
21. Chriki A, Touati H, Snoussi H. SVM-based indoor localization in Wireless Sensor Networks. In: 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC). Piscataway, NJ: IEEE; 2017 Jul 20 Presented at: International Wireless Communications and Mobile Computing Conference (IWCMC); 2017 06 26-30; Valencia, Spain p. 1144-1149. [doi: 10.1109/IWCMC.2017.7986446]
22. pch.vector. Patients and doctors meeting and waiting in clinic hall. hospital interior illustration with reception, person in wheelchair. for visiting doctor office, medical examination, consultation. freepik. URL: https://tinyurl.com/tcukhauz [accessed 2021-02-11]
23. pch.vector. Tiny people using mobile application with map outdoors. freepik. URL: https://www.freepik.com/free-vector/tiny-people-using-mobile-application-with-map-outdoors_9650838.htm#page=1&query=navigation&position=2 [accessed 2021-02-11]
24. smalllikeart. Wristband. flaticon. URL: https://tinyurl.com/2nk29v44 [accessed 2021-02-11]
25. Maksim. Simple router vector icon. Adobe Stock. URL: https://as1.ftcdn.net/v2/jpg/02/41/94/44/500_F_241944442_1im7Vt4pftETwc82ZaAtJJmR37ZCAvZ1.jpg [accessed 2021-02-11]
26. Farid Z, Nordin R, Ismail M. Recent Advances in Wireless Indoor Localization Techniques and System. Journal of Computer Networks and Communications 2013;2013:1-12. [doi: 10.1155/2013/185138]
27. Karimpour N, Karaduman B, Ural A, Challenger M, Dagdeviren O. IoT based Hand Hygiene Compliance Monitoring. In: 2019 International Symposium on Networks, Computers and Communications (ISNCC). InInternational Symposium on Networks, Computers and Communications (ISNCC): IEEE; 2019 Nov 21 Presented at: International Symposium on Networks, Computers and Communications (ISNCC); 2019 06 18-20; Istanbul, Turkey. [doi: 10.1109/ISNCC.2019.8909151]

28. van der Ham MFS, Zlatanova S, Verbree E, Voûte R. Real Time Localization of Assets in Hospitals Using Quuppa Indoor Positioning Technology. ISPRS Ann Photogramm Remote Sens Spatial Inf Sci 2016 Sep 05;IV-4/W1:105-110. [doi: 10.5194/isprs-annals-iv-4-w1-105-2016]
29. Elkhodr M, Mubin O, Iftikhar Z, Masood M, Alsinglawi B, Shahid S, et al. Technology, Privacy, and User Opinions of COVID-19 Mobile Apps for Contact Tracing: Systematic Search and Content Analysis. J Med Internet Res 2021 Feb 09;23(2):e23467. [doi: 10.2196/23467] [Medline: 33493125]
30. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of App-Based Contact Tracing for COVID-19: Cross-Country Survey Study. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
31. Zimmermann BM, Fiske A, Prainsack B, Hangel N, McLennan S, Buyx A. Early Perceptions of COVID-19 Contact Tracing Apps in German-Speaking Countries: Comparative Mixed Methods Study. J Med Internet Res 2021 Feb 08;23(2):e25525 [FREE Full text] [doi: 10.2196/25525] [Medline: 33503000]
32. Saw YE, Tan EY, Liu JS, Liu JC. Predicting Public Uptake of Digital Contact Tracing During the COVID-19 Pandemic: Results From a Nationwide Survey in Singapore. J Med Internet Res 2021 Feb 03;23(2):e24730 [FREE Full text] [doi: 10.2196/24730] [Medline: 33465034]
33. Basiri A, Lohan ES, Moore T, Winstanley A, Peltola P, Hill C, et al. Indoor location based services challenges, requirements and usability of current solutions. Computer Science Review 2017 May;24:1-12. [doi: 10.1016/j.cosrev.2017.03.002]
34. Venkatesh V, Davis FD. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
35. Fishbein M, Ajzen I. Belief, attitude, intention and behavior: An introduction to theory and research. Reading, Mass. Reading, Mass: Addison-Wesley; 1975.
36. Fishbein M. Attitude and the prediction of behavior. In: Fishbein M, editor. Readings in attitude theory and measurement. New York: Wiley; 1967.
37. Ajzen I, Cote NG. Attitudes and the prediction of behavior. In: Prislin R, Crano WD, editors. Attitudes and attitude change. New York, London: Psychology Press; 2008.
38. Ajzen I, Fishbein M. Attitudinal and normative variables as predictors of specific behavior. Journal of Personality and Social Psychology 1973;27(1):41-57. [doi: 10.1037/h0034440]
39. Armitage CJ, Conner M. Efficacy of the Theory of Planned Behaviour: a meta-analytic review. Br J Soc Psychol 2001 Dec;40(Pt 4):471-499. [Medline: 11795063]
40. Sheppard BH, Hartwick J, Warshaw PR. The Theory of Reasoned Action: A Meta-Analysis of Past Research with Recommendations for Modifications and Future Research. J Consum Res 1988 Dec;15(3):325. [doi: 10.1086/209170]
41. Han SH, Kim KJ, Yun MH, Hong SW, Kim J. Identifying mobile phone design features critical to user satisfaction. Hum Factors Man 2004;14(1):15-29. [doi: 10.1002/hfm.10051]
42. Gaither CA, Bagozzi RP, Ascione FJ, Kirking DM. A reasoned action approach to physicians' utilization of drug information sources. Pharm Res 1996 Sep;13(9):1291-1298. [doi: 10.1023/a:1016049311673] [Medline: 8893264]
43. Hüttel S, Leuchten M, Leyer M. The Importance of Social Norm on Adopting Sustainable Digital Fertilisation Methods. Organization and Environment 2020 Jun 01:108602662092907. [doi: 10.1177/1086026620929074]
44. Zafari F, Gkelias A, Leung KK. A Survey of Indoor Localization Systems and Technologies. IEEE Commun Surv Tutorials 2019;21(3):2568-2599. [doi: 10.1109/COMST.2019.2911558]
45. Schönberger M, Čirjevskis A. Successful IT/IS Projects in Healthcare: Evaluation of Critical Success Factors. JEHM 2017 Oct 06;2017:1-18. [doi: 10.5171/2017.956068]
46. Choi W, Rho MJ, Park J, Kim K, Kwon YD, Choi IY. Information system success model for customer relationship management system in health promotion centers. Healthc Inform Res 2013 Jun;19(2):110-120 [FREE Full text] [doi: 10.4258/hir.2013.19.2.110] [Medline: 23882416]
47. O'Neill MJ. Evaluation of a Conceptual Model of Architectural Legibility. Environment and Behavior 2016 Jul 26;23(3):259-284. [doi: 10.1177/0013916591233001]
48. Yao X, Wu Y, Liu H, Zhao X, Bian Y, Qu W. Analysis of Psychological Influences on Navigation Use While Driving Based on Extended Theory of Planned Behavior. Transportation Research Record 2019 May 09;2673(9):480-490. [doi: 10.1177/0361198119845666]
49. Goodman JK, Cryder CE, Cheema A. Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples. J Behav Dec Making 2012 Apr 02;26(3):213-224. [doi: 10.1002/bdm.1753]
50. Hair JF, Ringle CM, Sarstedt M. PLS-SEM: Indeed a Silver Bullet. The Journal of Marketing Theory and Practice 2011 Mar 15;19(2):139-152. [doi: 10.2753/MTP1069-6679190202]
51. Hulland J. Use of partial least squares (PLS) in strategic management research: a review of four recent studies. Strat Mgmt J 1999 Feb;20(2):195-204. [doi: 10.1002/(sici)1097-0266(199902)20:2<195::aid-smj13>3.0.co;2-7]
52. Cenfetelli RT, Bassellier G. Interpretation of Formative Measurement in Information Systems Research. MIS Quarterly 2009;33(4):689. [doi: 10.2307/20650323]
53. Henseler J, Dijkstra TK, Sarstedt M, Ringle CM, Diamantopoulos A, Straub DW, et al. Common Beliefs and Reality About PLS. Organizational Research Methods 2014 Apr 10;17(2):182-209. [doi: 10.1177/1094428114526928]

54. Henseler J, Ringle C, Sinkovics R. The use of partial least squares path modeling in international marketing. In: New challenges to international marketing. Bingley, England: Emerald Group Publishing Limited; 2009.
55. Kortmann S. The Mediating Role of Strategic Orientations on the Relationship between Ambidexterity-Oriented Decisions and Innovative Ambidexterity. J Prod Innov Manag 2014 Jan 24;32(5):666-684. [doi: 10.1111/jpim.12151]
56. Yi MY, Jackson JD, Park JS, Probst JC. Understanding information technology acceptance by individual professionals: Toward an integrative view. Information and Management 2006 Apr;43(3):350-363. [doi: 10.1016/j.im.2005.08.006]
57. Yi MY, Fiedler KD, Park JS. Understanding the Role of Individual Innovativeness in the Acceptance of IT-Based Innovations: Comparative Analyses of Models and Measures. Decision Sciences 2006 Aug;37(3):393-426. [doi: 10.1111/j.1540-5414.2006.00132.x]
58. Malinowski JC. Mental rotation and real-world wayfinding. Percept Mot Skills 2001 Feb;92(1):19-30. [doi: 10.2466/pms.2001.92.1.19] [Medline: 11322586]
59. Halpern D. Sex differences in cognitive abilities. 4. ed. New York, NJ: Psychology Press; 2012.
60. Zijlstra E, Hagedoorn M, Krijnen WP, van der Schans CP, Mobach MP. Route complexity and simulated physical ageing negatively influence wayfinding. Appl Ergon 2016 Sep;56:62-67. [doi: 10.1016/j.apergo.2016.03.009] [Medline: 27184311]
61. Kirasic K. Age differences in adults' spatial abilities, learning environmental layout, and wayfinding behavior. Spatial Cognition and Computation 2000;2(2):117-134. [doi: 10.1023/A:1011445624332]
62. Morris MG, Venkatesh V. Age Differences in Technology Adoption Decisions: Implications for a Changing Work Force. Personnel Psychology 2000 Jun;53(2):375-403. [doi: 10.1111/j.1744-6570.2000.tb00206.x]
63. Venkatesh, Morris, Ackerman. A Longitudinal Field Investigation of Gender Differences in Individual Technology Adoption Decision-Making Processes. Organ Behav Hum Decis Process 2000 Sep;83(1):33-60. [doi: 10.1006/obhd.2000.2896] [Medline: 10973782]
64. Lawton CA, Charleston SI, Zieles AS. Individual- and Gender-Related Differences in Indoor Wayfinding. Environment and Behavior 2016 Jul 26;28(2):204-219. [doi: 10.1177/0013916596282003]
65. Wang C, Chen Y, Zheng S, Liao H. Gender and Age Differences in Using Indoor Maps for Wayfinding in Real Environments. IJGI 2018 Dec 27;8(1):11. [doi: 10.3390/ijgi8010011]
66. Lawton CA. Strategies for Indoor Wayfinding: The Role of Orientation. Journal of Environmental Psychology 1996 Jun;16(2):137-145. [doi: 10.1006/jevp.1996.0011]
67. Maguire EA, Burgess N, O'Keefe J. Human spatial navigation: cognitive maps, sexual dimorphism, and neural substrates. Current Opinion in Neurobiology 1999 Apr;9(2):171-177. [doi: 10.1016/s0959-4388(99)80023-3]
68. Lee C, Shim JP. An exploratory study of radio frequency identification (RFID) adoption in the healthcare industry. European Journal of Information Systems 2017 Dec 19;16(6):712-724. [doi: 10.1057/palgrave.ejis.3000716]
69. Hameed MA, Counsell S, Swift S. A meta-analysis of relationships between organizational characteristics and IT innovation adoption in organizations. Information & Management 2012 Jul;49(5):218-232. [doi: 10.1016/j.im.2012.05.002]
70. Boudon R. Beyond Rational Choice Theory. Annu Rev Sociol 2003 Aug;29(1):1-21. [doi: 10.1146/annurev.soc.29.010202.100213]
71. Kaspar K. Motivations for Social Distancing and App Use as Complementary Measures to Combat the COVID-19 Pandemic: Quantitative Survey Study. J Med Internet Res 2020 Aug 27;22(8):e21613 [FREE Full text] [doi: 10.2196/21613] [Medline: 32759100]
72. Townsend K. Getting started with Bluetooth low energy. Sebastopol, CA: O'Reilly; 2014.

Abbreviations

IPS: indoor positioning system
RAA: reasoned action approach
SEM: structural equation modeling
SRMR: standardized root mean square residual

Edited by R Kukafka; submitted 24.02.21; peer-reviewed by B Alsinglawi, R Bach; comments to author 19.04.21; revised version received 08.06.21; accepted 27.07.21; published 05.10.21.

Please cite as:
Wichmann J, Leyer M
Factors Influencing the Intention of Actors in Hospitals to Use Indoor Positioning Systems: Reasoned Action Approach
J Med Internet Res 2021;23(10):e28193
URL: https://www.jmir.org/2021/10/e28193
doi:10.2196/28193
PMID:34609318

Original Paper

An Augmented Reality Device for Remote Supervision of Ultrasound Examinations in International Exercise Science Projects: Usability Study

Lia Rigamonti[1], MSc; Matteo Secchi[2], BSc; Jimmy B Lawrence[3], MSc; Luca Labianca[4], MD; Bernd Wolfarth[5,6], MD; Harm Peters[7], MD; Klaus Bonaventura[1,8], MD; David Alexander Back[7,9], MD, MBA

[1]University Outpatient Clinic, Sports Medicine & Orthopaedics, Department of Sport and Health Sciences, Center of Excellence "Cognitive Sciences", Faculty of Human Sciences, University of Potsdam, Potsdam, Germany

[2]Italian Association of Hydrotherapists and Newborn Educators (AIIEN), Section of Rome, Rome, Italy

[3]Department of Health and Physical Education, Mercer County Community College, West Windsor, NJ, United States

[4]Department of Orthopaedics and Traumatology, S. Andrea Hospital, Sapienza University of Rome, Rome, Italy

[5]Department of Sports Medicine, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[6]Department of Sports Medicine, Humboldt University Berlin, Berlin, Germany

[7]Dieter Scheffner Center for Medical Education and Educational Research, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[8]Department of Cardiology, Ernst von Bergmann Hospital, Potsdam, Germany

[9]Department of Traumatology and Orthopaedics, Septic and Reconstructive Surgery, Bundeswehr Hospital Berlin, Berlin, Germany

Corresponding Author:
David Alexander Back, MD, MBA
Department of Traumatology and Orthopaedics, Septic and Reconstructive Surgery
Bundeswehr Hospital Berlin
Scharnhorststrasse 13
Berlin, 10115
Germany
Phone: 49 176 40302553
Email: davidback@bundeswehr.org

Abstract

Background: Support for long-distance research and clinical collaborations is in high demand and has increased owing to COVID-19–related restrictions on travel and social contact. New digital approaches are required for remote scientific exchange.

Objective: This study aims to analyze the options of using an augmented reality device for remote supervision of exercise science examinations.

Methods: A mobile ultrasound examination of the diameter and intima-media thickness of the femoral and carotid arteries was remotely supervised using a head-mounted augmented reality device. All participants were provided with a link to a YouTube video of the technique in advance. In part 1, 8 international experts from the fields of engineering and sports science were remotely connected to the study setting. Internet connection speed was noted, and a structured interview was conducted. In part 2, 2 remote supervisors evaluated 8 physicians performing an examination on a healthy human subject. The results were recorded, and an evaluation was conducted using a 25-item questionnaire.

Results: In part 1, the remote experts were connected over a mean distance of 1587 km to the examination site. Overall transmission quality was good (mean upload speed: 28.7 Mbps, mean download speed: 97.3 Mbps, mean ping: 21.6 milliseconds). In the interview, participants indicated that the main potential benefits would be to the fields of education, movement analysis, and supervision. Challenges regarding internet connection stability and previous training with the devices used were reported. In part 2, physicians' examinations showed good interrater correlation (interclass correlation coefficient: 0.84). Participants valued the experienced setting as highly positive.

Conclusions: The study showed the good feasibility of the chosen design and a highly positive attitude of all participants toward this digital approach. Head-mounted augmented reality devices are generally recommended for collaborative research projects with physical examination–based research questions.

(J Med Internet Res 2021;23(10):e28767) doi:10.2196/28767

KEYWORDS

augmented reality; ultrasound; social media; remote; exercise science

Introduction

Background

International collaborations play an important role in addressing research questions in almost every scientific discipline [1]. Exercise science is highly interdisciplinary, without regional boundaries, and closely related to different areas of health care. For example, research on vascular changes in athletes [2] can help understand pathologies and global burdens such as cardiovascular diseases [3]. Although sports techniques are often internationally organized and standardized [4], the number of athletes reached for investigations in a single research institution could sometimes be limited depending on the geographic location and the type of sport [5,6]. Multicenter studies are well established in clinical medicine [7,8] and have been used in clinical exercise science research [9]. It would be advantageous to facilitate collaboration among exercise science researchers worldwide for investigations with common research questions, especially with a small population of athletes.

In addition to the impact of the current COVID-19 pandemic with strict travel and contact restrictions and the usual long distances between collaborating institutions, there is a high demand for new digital solutions to bridge distances and contact barriers [1,10,11]. Different digital media or devices may enhance scientific and professional collaborations and thus the health of patients [12,13]. Various approaches have already been described in this context, with early possibilities such as communication or instruction via email or videos [14]. With the rise of the internet and its increasing capability, social media has become highly sophisticated and widespread in the everyday lives of people worldwide [15]. Consequently, this technology is also used for teaching and research purposes depending on individual features, such as direct interpersonal communication, scientific exchange [12], or instructions on medical examinations or surgical techniques [16].

Despite many promising aspects, many of these options lack real-time personalized participation and direct interaction of remote collaboration with partners and examiners in research settings. New digital modes of interaction may be used, such as augmented reality (AR), which can be defined as “technology that integrates digital information into the user’s real-world environment” [17] and is sometimes also referred to as mixed reality [18]. AR devices have already been practically used in the technical settings of business enterprises [19] and for procedural work in space [20]. However, their use in medicine remains experimental. Potential has been described, especially in undergraduate education [21] and in the context of instructing invasive techniques, such as surgery [22]. In some cases, head-mounted AR devices were used in a clinical setting, where magnetic resonance imaging data were made available directly into the vision of surgeons [23]. For example, in sports science, AR has been used to create volleyball court images to measure athletes’ movements in a laboratory setting [24].

Aim and Hypotheses

This study addressed the feasibility of remote digitally supported exercise science collaborations using two different approaches.

In part 1 of the study, we hypothesized that it is possible to have different experts from various international locations remotely participating in an ultrasound examination setting using an AR device. In addition, their feedback on this method and its potential should be assessed.

In part 2 of the study, we hypothesized that it is possible to have different physicians performing the same ultrasound examination technique while being connected to 2 independent remote supervisors with an AR device.

Methods

Study Design

This project consisted of two parts; the same examination technique was used during both parts to obtain ultrasound-based measurements of the intima-media thickness (IMT) and the diameter of the femoral and carotid arteries. All measurements were obtained using the same healthy male human subject. The procedures or measurements were recorded in advance as videos [25]. For better mobility, a portable ultrasound linear array transducer (Lumify; Philips Healthcare) was chosen and linked to a cell phone (HTC Corp) with the appropriate app. The AR device used was a HoloLens 2 (Microsoft Corp) with Dynamics 365 Remote Access software (Microsoft Corp). Connections between participants were enabled by Microsoft Teams software (Microsoft Corp). Participation in the study was voluntary, with no compensation, and anonymity was ensured. All participants provided informed consent. The ethical committee of the University of Potsdam approved this study (EA1/236/19).

Part 1: Remote Expert Connection and Interviews

In the first approach, different professional experts in engineering and sports science were identified and asked for participation to obtain initial impressions of AR, both from a technical and scientific point of view. None of these participants had previous routine work experience with AR or HoloLens.

All participants received an email in advance with a description of the project design and a link to an examination video uploaded on a private YouTube channel, created specifically for this study (YouTube, LCC). After watching the video in advance, the experts then participated separately in the same setting, and the individual experts themselves were located internationally at their chosen location using their personal electronic devices. The physical location where the ultrasound examination was performed was Berlin, Germany. The standardized examination conditions for the human subject are described in the *Examination Conditions* section. Remote live participation was made possible using a Microsoft Teams

(Microsoft Corp) digital meeting with the examiner wearing the HoloLens (Figure 1).

The experts shared the examiner's view with ultrasound determination of the IMT and diameter of the femoral and carotid arteries. They also presented the options of drawing arrows or lines into the vision of the HoloLens (Figure 2) and optional blending of pictures or videos into the vision display—features that could be used for visual instruction about precise ultrasound techniques (eg, location of the transducer) or measurements (eg, intrainterventional communication about designated landmarks) in the current setting. The internet download speed (Mbps) and latency or ping (milliseconds) of the observing experts were recorded. Thereafter, they were asked 5 interview questions: (1) written or oral, regarding their perception of social media in general; (2) the remote connection mode in particular; (3) the potential for AR in exercise science and sports medicine; (4) other future digital areas with an impact on this field; and (5) challenges for digitalization in work-related settings.

Figure 1. Example of the presented study setting with the investigator wearing the HoloLens 2 (Microsoft Corp) while performing the ultrasound examination with the Lumify transducer (Philips Healthcare).

Figure 2. Examples of possible augmented reality indicators viewed while wearing the HoloLens device. (A) Arrow indicating a position to assist where to measure the intima-media-thickness. (B) Green and red lines drawn to indicate where the intima-media-thickness shall be measured.

Part 2: Examination Setting With Remote Supervision

Before the examination, participating physicians were shown the use of the Lumify transducer along with its mobile app functions and the HoloLens. In addition, they received a link to the unlisted YouTube video, where the examination process and focus data were explained. The participants were asked to wear the HoloLens to measure the diameter and IMT of both the left and right femoral and carotid arteries. In addition, 2 independent remote supervisors were connected with their computer devices by a Microsoft Teams digital meeting to the HoloLens and shared the view of the participating physicians. The performance was rated using the OSAUS (Objective Structured Assessment of Ultrasound Skills) score (maximum 25 points; Multimedia Appendix 1 [26]) and a modified AMIAUD (assessment of mobile imparted arterial ultrasound determination) score (modification: the examiners were not asked to note the results), both recorded by supervisors. This adaptation was made because of the study setting, as the transducer and the mobile phone occupied both examiners' hands (now maximum 25 points; Multimedia Appendix 2 [25]). Each participant received an evaluation sheet with 25 questions (5-point Likert-scale, multiple-choice, or open-ended questions with text answer options). The topics were as follows:

1. Demographic data (gender, sex);
2. Questions about social media and video-based transmission of knowledge (9 questions; six 5-point Likert-scale, 2 multiple-choice questions, 1 open-ended question);
3. Use of the Lumify ultrasound transducer (5 questions; 4 5-point Likert-scale and 1 open-ended question); and
4. Questions about the practicability and acceptance of the HoloLens device (9 questions; seven 5-point Likert-scale and 2 open-ended questions).

Examination Conditions

All examinations were performed on the same 39-year-old healthy male human subject lying in a supine position at a mean room temperature of 23 °C. Electrocardiography was conducted during the examination, and blood pressure was assessed before the examination. The subject was asked to consume the same kind of food and avoid caffeine or alcohol 24 hours before each examination to establish standardized conditions.

Data Analysis

Data were entered in Excel (Microsoft Corporation) and analyzed either descriptively or using SPSS, version 27.0 (IBM Corporation). The interclass correlation coefficient was determined for the interrater correlation of the 2 independent observers of physicians' examinations. Free text-based answers were checked by 2 of the authors for repetitive sequences. The ratio of the left and right arteries was determined. For vessel measurements, mean values and SDs were calculated; for the scores, median values and quartiles (IQR) were calculated.

The expert interviews were analyzed for relevant topics using MaxQDA 2020 qualitative data analysis software (VERBI GmbH). Predominant themes were defined on the basis of subthemes and their particular topics.

Results

Demographic Data

In part 1, the invited sports scientists (2 exercise scientists, 1 physical therapist, and 1 sports physician) were all men and had a mean age of 37 (SD 5.6) years; all engineers (2 automotive engineers, 1 mechanical engineer, and 1 biomedical engineer) were also men and had a mean age of 36 (SD 2.9) years.

In part 2, the 8 participating physicians for the second examination setting (residents from the fields of surgery, orthopedics, anesthesia, and general medicine) were 4 women and 4 men with a mean age of 31 (SD 2.8) years. The remote supervisors were a woman (36 years old) and a man (32 years old).

Part 1: Remote Expert Connection and Interviews

Distance and Internet Connection

The average distance of the remote experts in their different cities in Germany, Italy, England, the Netherlands, Switzerland, and Saudi Arabia to the examination site (Germany) was 1587 (SD 1543) km. The internet connection of all final assessments was stable with a mean upload speed of 28.7 (SD 54.8) Mbps, a download speed of 97.3 (SD 126.3) Mbps, and a ping of 21.6 (SD 11.2) milliseconds.

Interview Questions

Perception of Social Media

The participants were asked about (1) their perception of the technical aspects of obtaining information via the YouTube video, (2) the connection process to the HoloLens via Microsoft Teams, and (3) the internet connection during the examination. The experts stated that the technical aspects worked well overall, and the processes were easy, straightforward, and easily understood. It was positively reported that access to remote sessions was easily manageable from either a regular computer or mobile phone through a standard telecommunication program such as Microsoft Teams. The YouTube video was rated as highly informative and useful. The connection went smoothly with only one minor connection problem reported when logging into Microsoft Teams and once for a perceived low transmitted voice volume by the HoloLens.

Perception of Remote Connection Mode

The general idea of remote digital presence in work- or research-related settings was associated with high approval, also described as an *amazing solution* important for the future; for example, when an expert cannot be physically present at a specific location but is needed. In general, collaboration can be enhanced. Great potential was seen in research teaching and supervision and within business or enterprise. Much of the potential was appreciated because of the current pandemic but also viewed favorably for pre- and postpandemic collaborations across large distances. The use of a relatively common and easy-to-use program may also lower the knowledge barrier required to interact with AR tools. Potential problems of poor internet connection or extended periods to connect were observed. However, compared with lacking alternatives, it is a

fast response to urgent situations that require real-time support and allows for worldwide connection. A remote digital presence can be very handy, convenient, and comfortable for individuals.

Potential for AR in Exercise Science and Sports Medicine

In the fields of sports science and sports medicine, the experts saw the greatest potential for the use of AR to spread knowledge with teaching and supervise from remote locations. Students and many other recipients (such as athletes) could benefit from this form of knowledge or skill transfer. This could better assist those who benefit from a more *hands-on* or interactive approach.

In sports and training, AR can be used for evaluative or observative analyses, for example, of single movements or patterns, and also as remote medical diagnostics.

A future potential was also described regarding connections with artificial intelligence. AR used concurrently with *smart programs* could, for example, analyze shapes, distances, and track movements (with the possibility of combining this with strength or other external sensors). In addition, deep learning algorithms can provide real-time analyses of an athlete's performance to coaches, trainers, scientists, and health and sports workers. Thus, beyond observational analysis, physicians could also perform AR-enhanced examinations combined with live data from circulation and tissues.

Future Digital Areas With an Impact on Exercise Science and Sports Medicine

Participants were asked which areas and tools will affect sports science and sports medicine the most in the future. Most experts mentioned wearables and smart devices (n=56), followed by mobile apps (n=5), telemedicine (n=5), virtual and AR (n=5), education and training (n=4), artificial intelligence and big data (n=3), and social media (n=2). The participants reported several specific aspects of this technology's proposed impact: economic, educational (eg, clinical and practical skills training), increased data collection, and increased outreach with the possibility of reaching more people in remote areas.

Challenges for Digitalization in Work-Related Settings

On the basis of the experts' own experiences, they were asked what the biggest challenges for digitalization in work-related settings will be. The answers provided by the experts are depicted in a sunburst diagram in Figure 3.

Figure 3. Sunburst diagram showing the areas of challenges for digitalization in work-related settings, as indicated by 8 remote experts. DP: data protection; DS: digital solutions; IT: information technology.

Part 2: Examination Setting With Remote Supervision

Gained Examination Data

The mean IMT ratio of the carotid arteries was 1.092 (SD 0.1), and the mean diameter ratio was 1.06 (SD 0.062). The mean IMT ratio of the femoral arteries was 1.032 (SD 0.112), and the mean diameter ratio was 1.001 (SD 0.112). The median of the Objective Structured Assessment of Ultrasound Skills score of the participants was rated at 24.75 (Q1: 23.875; Q3: 25; IQR 1.125) and the mean modified AMIAUD score was 25 (Q1: 24.25; Q3: 25; IQR 0.75), with an interclass correlation coefficient of 0.84 for both raters.

Examiners' Experience Evaluation

During the evaluation of physicians, all stated that they used several social media in their daily lives and had already used videos to acquire professional techniques (mean agreement 4.75/5).

The physician participants considered obtaining video-based knowledge to be practicable in a multitude of capacities. Specifically, the physicians mentioned that it could be used in equipment instruction (n=5), surgical environments (n=4) and emergency rooms (n=2), or further training on examination techniques (n=1).

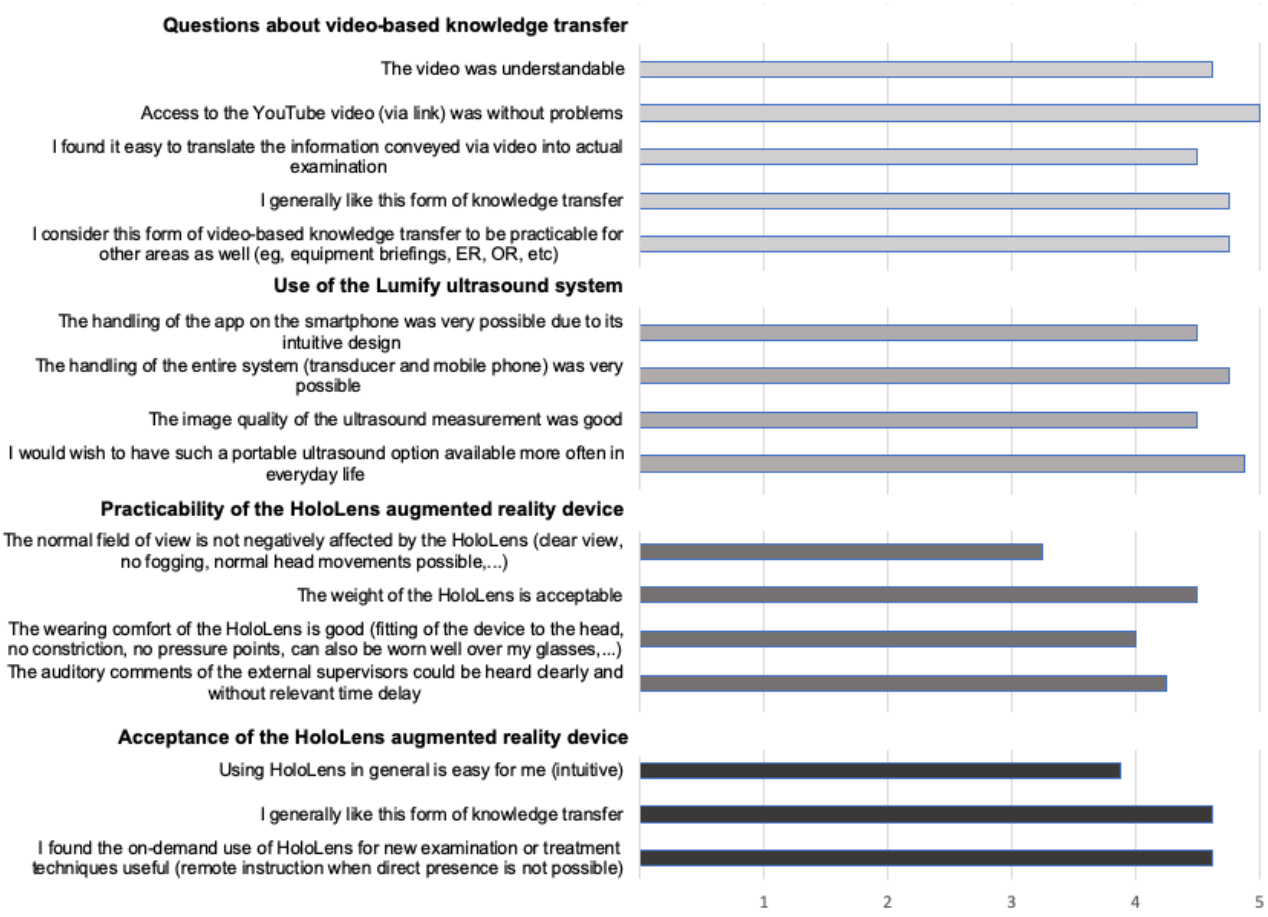
When asked about the advantages or disadvantages of obtaining knowledge via social media, the participants' responses were generally positive. They approved that these resources are constantly available, practical, and easily accessible. This enables learning from home or on the road, regardless of location, and repetition of information is easily possible. However, there are some disadvantages. The physician participants mentioned the lack of opportunity for follow-up

questions and, in some cases, unclear transparency of the scientific quality.

In response to questions about the advantages and disadvantages of using head-mounted AR devices, the participants stated that it would be advantageous for remote experts to collaborate. This could lead to a gain in knowledge with better technical actions. Furthermore, easy handling and fast data transfer were cited. In contrast, the risk of technical malfunction (especially when *relying* on external help) and the need to learn the handling (eg, head position) were mentioned as disadvantages.

The other Likert-scale evaluation results concerning the video-based transmission of knowledge, the use of the app- and mobile phone–based ultrasound technique, and the practicability and acceptance of the AR device are shown in Figure 4.

Figure 4. Answers to questions about the video-based transmission of knowledge, the use of the Lumify ultrasound system, and the practicability and acceptance of the augmented reality device HoloLens (n=8 participants; 5-point Likert scale from 1 fully disagree to 5 fully agree). ER: emergency room; OR: operation room.



Discussion

Principal Findings

Bridging distances between academic institutions and investigators to enhance scientific collaboration has always been highly important for many research projects [1]. Owing to the COVID-19 pandemic, the need to reduce travel and physical contact has reached levels not seen in nearly 100 years [11]. Digital features, such as social media, have already been shown to be sufficient in enhancing scientific exchange and communication [12]. AR technology and its associated devices can positively enhance collaborations across large distances, but the full potential of this technology is not realized, although it is very promising [10].

In this study, remote experts could fulfil the aims and hypotheses and could sufficiently connect to and actively participate in a simulated exercise science examination scenario by collecting the experts' opinions concerning the techniques used. Furthermore, it is possible that different physicians could perform the given ultrasound examination while being remotely supervised by the AR device in a multicenter international collaboration structure.

Both the experts and the examining physicians showed a high preference for the remote setting using social media and procedural connection modalities via Microsoft Teams. Social media had already been established as a firm tool in the scientific community before the COVID-19 pandemic [12]. Examples of this have been shown when learning about movement techniques in sports and exercise science [27], instructions on investigations [25], and data literacy skills [28]. All these features experienced a boost in the last months of the COVID-19 pandemic when the use of digital communication, digital conferencing, and digital remote working conditions became normal [29].

In addition, the participating physicians shared high ratings regarding the use, practicability, and acceptance of the technologies, although they were not remarkably familiar with the mobile ultrasound device. This was accompanied by a good evaluation of the investigation system and its devices. It seems to have potential as a smart digital transmission feature for ultrasound examinations [30].

AR was rated, in general, as highly positive by both experts and examiners with its previously described potential to establish scientific- and work-related contacts between remote participants in real time and across large distances, providing expert support and exchange, thus enabling international collaborations [10]. In scientific study settings, AR could also benefit from independent and objective supervision by integrating different experts internationally into a rater team, as performed in this study.

The participants saw the potential for AR in sports science and sports medicine first in training and education, which can be supported by the authors' experiences, for example, in the field of anatomy [21]. Not explicitly mentioned was that the interactivity of AR might make learning more engaging, interesting, and efficient than pure textbooks [31]. Another interesting suggested approach, especially in the field of sports science, could be to link AR with computed tomography or magnetic resonance imaging data to project these data onto moving people, which could support analyses of movement patterns or pathologies and the development of therapies [32]. In general, AR-based human tracking and its analysis seem to be well suited for sports or exercises that depend on form, movement, and technique.

Remote experts and participating physicians stated that good supervision could be possible with AR. Clinical examples show that supervision via AR is possible in surgical settings [10] and that inexperienced users can achieve similar operative results after AR-supervised training compared with other users with on-site training [33].

This study revealed the relevant technical and organizational challenges to be considered. We showed that, with a working internet connection, well-functioning live connections to examinations using AR devices could easily be possible. A decisive point for future projects would be to determine in advance bandwidths or other internet connection conditions that might pose transmission problems [34].

Remotely coordinating the schedules of subject, investigators, and supervisors proved to be difficult. In larger-scale collaborations, multiple meetings should be made available, with significant time frames between each session and optional alternative dates to accommodate all of those involved.

Despite the potential of the described setting, there is a learning curve with this technique in the current form. Users will benefit most after becoming familiar with the devices. The method used in this study involves simultaneously handling the mobile *touchpad* for ultrasound measurements and using a head-mounted AR device. For the latter, the mere weight of the device must be anticipated, among other issues. For example, the way the virtual program appears in the user´s vision is optimal when looking directly forward through the device. The user must anticipate the radius of the camera of the device in relation to the user´s own field of view. Therefore, excessive eye and head movements are discouraged. This aspect has already been described as relevant in the surgical context [35].

The scores used in this study have already been published before [25,26]. Although the quality of the measurements was not the focus, a high correlation was found among the supervisors. Owing to the high rating scores observed, perhaps the ultrasound task was not very challenging for the participating physicians. In addition, this could suggest that there was less need for supervision and auditory comments. Consequently, this might have had an impact on the examiners' evaluation, although they were asked not only for subjective but also objective answers. It can also be discussed whether AR technology is the only practical choice in this particular setting. Perhaps a head-mounted camera and speaker or microphone could be a sufficient alternative.

However, the chosen study design was intended to serve as an example of the overall practical use of AR devices in an international sports science or medical collaboration setting. Changes in the design according to the scientific requirements and intensified use of AR options will have to be considered for other settings.

Regarding limitations, this study should be viewed as a first step in investigating the feasibility of such technology and techniques because of the small number of participants. In addition, the chosen study design had a stable examination setting with no necessary body movements required by the examiners, thus allowing for good viewing. This could differ in other study designs or real-life scenarios, such as a dynamic setting where numerous and quick body or head movements of an examiner or physician will be necessary (eg, surgical interventions and field research). Furthermore, the long axis of the arteries was used for IMT measurement. However, in future studies, it might be considered that there are fewer influencing factors when performing the measurements on the short axis. Finally, AR options such as showing pictures, videos, and 3D animations (eg, videos in the field of view of an examiner) have not been fully used in this approach.

As there is a lack of larger prospective randomized trials in sports science and medicine that can clearly demonstrate the benefits of AR's practical applications, the aspects mentioned above and the use of a larger cohort should be considered in future studies.

Conclusions

This study showed that the described techniques involving social media video distribution in advance, followed by an examination performed with a head-mounted AR device, can be effectively used in a long-distance international collaboration setting. A stable and sufficient internet connection will always be decisive. To optimize international collaboration, especially during the COVID-19 pandemic, this kind of remote support offers advantages such as bridging distances, shortening travel times, providing real-time interaction, and potentially enhancing objectivity in data collection by including remote experts in the

study setting. Significant challenges would be the time needed to set up such technical settings, user training, and digital coordination of all participants.

Acknowledgments

The authors would like to express their gratitude to all voluntary experts and physicians for their participation in this study who have dedicated their time. The authors also thank the colleagues who helped with the technical support.

Authors' Contributions

This study was conceptualized and designed by LR and DAB. LR, MS, JBL, LL, and DAB participated in data acquisition and analysis. All coauthors participated in the interpretation of the data. LR, JBL, and DAB drafted the manuscript. All authors critically reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
OSAUS (Objective Structured Assessment of Ultrasound Skills) score.
[PNG File , 93 KB - jmir_v23i10e28767_app1.png]

Multimedia Appendix 2
Modified AMIAUD (assessment of mobile imparted arterial ultrasound determination) score.
[PNG File , 116 KB - jmir_v23i10e28767_app2.png]

References

1. Aarons GA, Seijo C, Green AE, Moullin JC, Hasson H, von Thiele Schwarz U, et al. Fostering international collaboration in implementation science and research: a concept mapping exploratory study. BMC Res Notes 2019 Nov 29;12(1):778 [FREE Full text] [doi: 10.1186/s13104-019-4800-4] [Medline: 31783912]
2. Green D, Spence A, Rowley N, Thijssen D, Naylor L. Vascular adaptation in athletes: is there an 'athlete's artery'? Exp Physiol 2012 Mar;97(3):295-304 [FREE Full text] [doi: 10.1113/expphysiol.2011.058826] [Medline: 22179421]
3. Lorenz MW, Markus HS, Bots ML, Rosvall M, Sitzer M. Prediction of clinical cardiovascular events with carotid intima-media thickness: a systematic review and meta-analysis. Circulation 2007 Jan 30;115(4):459-467. [doi: 10.1161/CIRCULATIONAHA.106.628875] [Medline: 17242284]
4. Competition and technical rules. World Athletics. 2019. URL: https://www.worldathletics.org/download/download?filename=0400178d-ce56-43ff-aa9c-6b4433aafcea.pdf&urlslug=Competition [accessed 2021-09-04]
5. Kyparos A, Riganas C, Nikolaidis MG, Sampanis M, Koskolou MD, Grivas GV, et al. The effect of exercise-induced hypoxemia on blood redox status in well-trained rowers. Eur J Appl Physiol 2012 Jun 27;112(6):2073-2083. [doi: 10.1007/s00421-011-2175-x] [Medline: 21947454]
6. Joseph LH, Hancharoenkul B, Sitilertpisan P, Pirunsan U, Paungmali A. Effects of massage as a combination therapy with lumbopelvic stability exercises as compared to standard massage therapy in low back pain: a randomized cross-over study. Int J Ther Massage Bodywork 2018 Dec 07;11(4):16-22. [doi: 10.3822/ijtmb.v11i4.413]
7. Bosiers M, Deloose K, Callaert J, Maene L, Beelen R, Keirse K, et al. BRAVISSIMO: 12-month results from a large scale prospective trial. J Cardiovasc Surg (Torino) 2013 Apr;54(2):235-253. [Medline: 23558659]
8. Frederix I, Hansen D, Coninx K, Vandervoort P, Van Craenenbroeck EM, Vrints C, et al. Telerehab III: a multi-center randomized, controlled trial investigating the long-term effectiveness of a comprehensive cardiac telerehabilitation program--rationale and study design. BMC Cardiovasc Disord 2015 May 07;15(1):29 [FREE Full text] [doi: 10.1186/s12872-015-0021-5] [Medline: 25948479]
9. Kluitenberg B, van Middelkoop M, Smits D, Verhagen E, Hartgens F, Diercks R, et al. The NLstart2run study: incidence and risk factors of running-related injuries in novice runners. Scand J Med Sci Sports 2015 Oct 01;25(5):e515-e523. [doi: 10.1111/sms.12346] [Medline: 25438823]
10. Davis MC, Can DD, Pindrik J, Rocque BG, Johnston JM. Virtual interactive presence in global surgical education: international collaboration through augmented reality. World Neurosurg 2016 Feb;86:103-111 [FREE Full text] [doi: 10.1016/j.wneu.2015.08.053] [Medline: 26342783]
11. Zaer H, Fan W, Orlowski D, Glud AN, Andersen AS, Schneider MB, et al. A perspective of international collaboration through web-based telecommunication-inspired by COVID-19 Crisis. Front Hum Neurosci 2020 Nov 23;14:577465 [FREE Full text] [doi: 10.3389/fnhum.2020.577465] [Medline: 33328931]

12. Van Noorden R. Online collaboration: scientists and the social network. Nature 2014 Aug 14;512(7513):126-129. [doi: 10.1038/512126a] [Medline: 25119221]
13. Gierdowski D. ECAR Study of Undergraduate Students and Information Technology, 2019. Research report. 2019. URL: https://library.educause.edu/-/media/files/library/2019/10/studentstudy2019.pdf?la=en&hash=25FBB396AE482FAC3B765862BA6B197DBC98B42C [accessed 2021-01-06]
14. Wheeler D, Whittlestone K, Smith H, Gupta A, Menon D, East Anglian Peri-Operative Medicine Undergraduate Teaching Forum. A web-based system for teaching, assessment and examination of the undergraduate peri-operative medicine curriculum. Anaesthesia 2003 Nov;58(11):1079-1086 [FREE Full text] [doi: 10.1046/j.1365-2044.2003.03405.x] [Medline: 14616593]
15. Viner RM, Gireesh A, Stiglic N, Hudson LD, Goddings A, Ward JL, et al. Roles of cyberbullying, sleep, and physical activity in mediating the effects of social media use on mental health and wellbeing among young people in England: a secondary analysis of longitudinal data. Lancet Child Adolesc Health 2019 Oct;3(10):685-696. [doi: 10.1016/S2352-4642(19)30186-5] [Medline: 31420213]
16. Pilieci SN, Salim SY, Heffernan DS, Itani KM, Khadaroo RG. A randomized controlled trial of video education versus skill demonstration: which is more effective in teaching sterile surgical technique? Surg Infect (Larchmt) 2018 Apr;19(3):303-312. [doi: 10.1089/sur.2017.231] [Medline: 29406814]
17. Eckert M, Volmerg JS, Friedrich CM. Augmented reality in medicine: systematic and bibliographic review. JMIR Mhealth Uhealth 2019 Apr 26;7(4):e10967 [FREE Full text] [doi: 10.2196/10967] [Medline: 31025950]
18. Kaplan AD, Cruit J, Endsley M, Beers SM, Sawyer BD, Hancock PA. The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: a meta-analysis. Hum Factors 2021 Jun 24;63(4):706-726. [doi: 10.1177/0018720820904229] [Medline: 32091937]
19. Roxo MT, Brito PQ. Augmented reality trends to the field of business and economics: a review of 20 years of research. Asian J Busi Res 2018 Dec 1;8(2):94-117. [doi: 10.14707/ajbr.180051]
20. Braly AM, Nuernberger B, Kim SY. Augmented reality improves procedural work on an international space station science instrument. Hum Factors 2019 Sep 29;61(6):866-878. [doi: 10.1177/0018720818824464] [Medline: 30694084]
21. Uruthiralingam U, Rea P. Augmented and virtual reality in anatomical education - a systematic review. Adv Exp Med Biol 2020;1235:89-101. [doi: 10.1007/978-3-030-37639-0_5] [Medline: 32488637]
22. Kovoor JG, Gupta AK, Gladman MA. Validity and effectiveness of augmented reality in surgical education: a systematic review. Surgery 2021 Jul;170(1):88-98. [doi: 10.1016/j.surg.2021.01.051] [Medline: 33744003]
23. Scherl C, Stratemeier J, Karle C, Rotter N, Hesser J, Huber L, et al. Augmented reality with HoloLens in parotid surgery: how to assess and to improve accuracy. Eur Arch Otorhinolaryngol 2021 Jul 10;278(7):2473-2483. [doi: 10.1007/s00405-020-06351-7] [Medline: 32910225]
24. Adams K, Kiefer A, Panchuk D, Hunter A, MacPherson R, Spratford W. From the field of play to the laboratory: recreating the demands of competition with augmented reality simulated sport. J Sports Sci 2020 Mar 22;38(5):486-493. [doi: 10.1080/02640414.2019.1706872] [Medline: 31865835]
25. Rigamonti L, Kahle P, Peters H, Wolfarth B, Thouet T, Bonaventura K, et al. Instructing ultrasound-guided examination techniques using a social media smartphone app. Int J Sports Med 2021 Apr 19;42(4):365-370. [doi: 10.1055/a-1268-8837] [Medline: 33075835]
26. Tolsgaard MG, Ringsted C, Dreisler E, Klemmensen A, Loft A, Sorensen JL, et al. Reliable and valid assessment of ultrasound operator competence in obstetrics and gynecology. Ultrasound Obstet Gynecol 2014 Apr 12;43(4):437-443 [FREE Full text] [doi: 10.1002/uog.13198] [Medline: 23996613]
27. Bopp T, Vadeboncoeur JD, Stellefson M, Weinsz M. Moving beyond the gym: a content analysis of YouTube as an information resource for physical literacy. Int J Environ Res Public Health 2019 Sep 10;16(18):3335 [FREE Full text] [doi: 10.3390/ijerph16183335] [Medline: 31510001]
28. Alonso Galbán P, Vialart Vidal N. Digital literacy program for the use of social media, aimed at health professionals. Stud Health Technol Inform 2019 Aug 21;264:1895-1896. [doi: 10.3233/SHTI190701] [Medline: 31438395]
29. Beshyah S, Ibrahim W, Hajjaji I, Ben Mami F, Arekat M, Abdelmannan D. Impact of the COVID-19 pandemic on clinical practice, medical education, and research: an international survey. Tunis Med 2020;98(8-9):610-618. [Medline: 33480015]
30. Henry-Tillman R, Kabongo M, Laryea J, Pinder L, Bittenbinder R, Osgood G, et al. The ability to look: management of breast disease in the democratic republic of the Congo using smart ultrasound technology. J Am Coll Surg 2021 Apr;232(4):636-640. [doi: 10.1016/j.jamcollsurg.2020.12.008] [Medline: 33348015]
31. Zahiri M, Nelson CA, Oleynikov D, Siu K. Evaluation of augmented reality feedback in surgical training environment. Surg Innov 2018 Feb 08;25(1):81-87. [doi: 10.1177/1553350617739425] [Medline: 29115184]
32. Watts I, Boulanger P, Kawchuk G. ProjectDR: augmented reality system for displaying medical images directly onto a patient. In: Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology. 2017 Presented at: VRST '17: 23rd ACM Symposium on Virtual Reality Software and Technology; Nov 8-10, 2017; Gothenburg, Sweden p. 1-2. [doi: 10.1145/3139131.3141198]

33. Logishetty K, Western L, Morgan R, Iranpour F, Cobb JP, Auvinet E. Can an augmented reality headset improve accuracy of acetabular cup orientation in simulated THA? A randomized trial. Clin Orthop Relat Res 2019 May;477(5):1190-1199 [FREE Full text] [doi: 10.1097/CORR.0000000000000542] [Medline: 30507832]
34. Leochico CF, Espiritu AI, Ignacio SD, Mojica JA. Challenges to the emergence of telerehabilitation in a developing country: a systematic review. Front Neurol 2020 Sep 8;11:1007 [FREE Full text] [doi: 10.3389/fneur.2020.01007] [Medline: 33013666]
35. Moosburner S, Remde C, Tang P, Queisner M, Haep N, Pratschke J, et al. Real world usability analysis of two augmented reality headsets in visceral surgery. Artif Organs 2019 Jul;43(7):694-698. [doi: 10.1111/aor.13396] [Medline: 30485464]

Abbreviations

AMIAUD: assessment of mobile imparted arterial ultrasound determination
AR: augmented reality
IMT: intima-media thickness
OSAUS: Objective Structured Assessment of Ultrasound Skills

Edited by R Kukafka; submitted 16.03.21; peer-reviewed by P Sun, M Bauer, CI Sartorão Filho; comments to author 18.05.21; revised version received 27.06.21; accepted 27.07.21; published 05.10.21.

Please cite as:
Rigamonti L, Secchi M, Lawrence JB, Labianca L, Wolfarth B, Peters H, Bonaventura K, Back DA
An Augmented Reality Device for Remote Supervision of Ultrasound Examinations in International Exercise Science Projects: Usability Study
J Med Internet Res 2021;23(10):e28767
URL: https://www.jmir.org/2021/10/e28767
doi:10.2196/28767
PMID:34609312

Original Paper

Multiple Sclerosis Progression Discussion Tool Usability and Usefulness in Clinical Practice: Cross-sectional, Web-Based Survey

Tjalf Ziemssen[1], MD; Gavin Giovannoni[2], MBBCHIR, PhD; Enrique Alvarez[3], MD; Virender Bhan[4], MD; Carrie Hersh[5], DO, MS; Olaf Hoffmann[6], MD, PhD; Celia Oreja-Guevara[7], MD; Rene R Robles-Cedeño[8], MD, PhD; Maria Trojano[9], MD; Patrick Vermersch[10], MD, PhD; Pamela Dobay[11], PhD; Mudeer Khwaja[12], BSc; Bianca Stadler[12], PhD; Benedict Rauser[12], PhD; Thomas Hach[12], MD; Daniela Piani-Meier[12], PhD; Jason Burton[13], MD

[1]Center of Clinical Neuroscience, Neurological University Clinic Carl-Gustav Carus, Dresden, Germany

[2]Blizard Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom

[3]University of Colorado School of Medicine, Aurora, CO, United States

[4]University of British Columbia, Vancouver, BC, Canada

[5]Cleveland Clinic Lou Ruvo Center for Brain Health, Las Vegas, NV, United States

[6]Department of Neurology, St Josefs-Krankenhaus, Potsdam, Germany

[7]Department of Neurology, Hospital Clínico San Carlos, Madrid, Spain

[8]Department of Neurology, Girona Neuroimmunology & Multiple Sclerosis Unit, Dr. Josep Trueta University Hospital & Santa Caterina Hospital, Girona, Spain

[9]Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari, Bari, Italy

[10]University of Lille, INSERM U1172, Lille Neuroscience and Cognition, CHU Lille, FHU Precise, Lille, France

[11]Real World Evidence Solutions, IQVIA AG, Basel, Switzerland

[12]Novartis Pharma AG, Basel, Switzerland

[13]Centre for Neuromuscular and Neurological Disorders, Western Australian Neuroscience Research Institute, The University of Western Australia, Perth, Australia

Corresponding Author:
Tjalf Ziemssen, MD
Center of Clinical Neuroscience
Neurological University Clinic Carl-Gustav Carus
Dresden University of Technology
Helmholtzstr. 10
Dresden, 01069
Germany
Phone: 49 3514584465
Email: Ziemssen@web.de

Abstract

Background: A digital tool, Multiple Sclerosis Progression Discussion Tool (MSProDiscuss), was developed to facilitate discussions between health care professionals (HCPs) and patients in evaluating early, subtle signs of multiple sclerosis (MS) disease progression.

Objective: The aim of this study is to report the findings on the usability and usefulness of MSProDiscuss in a real-world clinical setting.

Methods: In this cross-sectional, web-based survey, HCPs across 34 countries completed an *initial* individual questionnaire (comprising 7 questions on comprehensibility, usability, and usefulness after using MSProDiscuss during each patient consultation) and a *final* questionnaire (comprising 13 questions on comprehensibility, usability, usefulness, and integration and adoption into clinical practice to capture the HCPs' overall experience of using the tool). The responses were provided on a 5-point Likert scale. All analyses were descriptive, and no statistical comparisons were made.

Results: In total, 301 HCPs tested the tool in 6974 people with MS, of whom 77% (5370/6974) had relapsing-remitting MS, including those suspected to be transitioning to secondary progressive MS. The time taken to complete MSProDiscuss was

reported to be in the range of 1 to 4 minutes in 97.3% (6786/6974; *initial*) to 98.2% (269/274; *final*) of the cases. In 93.54% (6524/6974; *initial*) to 97.1% (266/274; *final*) of the cases, the HCPs agreed (4 or 5 on the Likert scale) that patients were able to comprehend the questions in the tool. The HCPs were willing to use the tool again in the same patient, 90.47% (6310/6974; *initial*) of the cases. The HCPs reported MSProDiscuss to be useful in discussing MS symptoms and their impact on daily activities (6121/6974, 87.76% *initial* and 252/274, 92% *final*) and cognitive function (5482/6974, 78.61% *initial* and 271/274, 79.2% *final*), as well as in discussing progression in general (6102/6974, 87.49% *initial* and 246/274, 89.8% *final*). While completing the final questionnaire, 94.9% (260/274) of the HCPs agreed that the questions were similar to those asked in regular consultation, and the tool helped to better understand the impact of MS symptoms on daily activities (249/274, 90.9%) and cognitive function (220/274, 80.3%). Overall, 92% (252/274) of the HCPs reported that they would recommend MSProDiscuss to a colleague, and 85.8% (235/274) were willing to integrate it into their clinical practice.

Conclusions: MSProDiscuss is a usable and useful tool to facilitate a physician-patient discussion on MS disease progression in daily clinical practice. Most of the HCPs agreed that the tool is easy to use and were willing to integrate MSProDiscuss into their daily clinical practice.

(J Med Internet Res 2021;23(10):e29558) doi:10.2196/29558

KEYWORDS

multiple sclerosis; relapsing remitting multiple sclerosis; secondary progressive multiple sclerosis; transition; progression; digital; usability

Introduction

Background

Multiple sclerosis (MS) is a chronic debilitating disease of the central nervous system that primarily affects young adults [1]. In most of the patients, the disease evolves as a continuum from the relapsing-remitting phase (known as relapsing-remitting MS [RRMS]) to the secondary progressive phase (known as secondary progressive MS [SPMS]) [2,3]. It is challenging to define this transition from RRMS to SPMS because of the lack of a clear consensus on the diagnostic criteria and the absence of reliable biomarkers of disease progression [4]. This delay in SPMS diagnosis may affect long-term prognosis and treatment decision-making [5]. Previous research has confirmed an unmet need for a tool to facilitate systematic assessment of the early signs of progression to SPMS in routine clinical practice [6].

The Multiple Sclerosis Progression Discussion Tool (MSProDiscuss) is a digital tool for use by health care professionals (HCPs) in clinical practice to raise awareness of the risk of progression from RRMS to SPMS through a structured interaction between HCPs and patients. MSProDiscuss aims to help physicians, in dialog with patients, to evaluate early, subtle signs suggestive of MS disease progression [7]. The tool is based on a set of weighted questions that collect structured information on disease activity (relapses or magnetic resonance imaging activity), symptoms, and impact of the patient's overall symptoms on daily living in the previous 6 months. The tool is completed by the physician during a routine physician-patient interaction. On completion, the tool generates a traffic light output that represents the probability of progression. Green indicates patients who are unlikely to be showing signs of progression, yellow suggests that such signs may be present, and red identifies patients who are very likely showing signs of progression. MSProDiscuss was developed in several phases using a rigorous mixed methods approach. This approach included quantitative analysis of data from a large observational study in patients diagnosed with RRMS and SPMS and qualitative research with MS neurologists and patients with MS. The development of MSProDiscuss involved an iterative feedback process and validation stages [6-8]. The feedback received from patients and physicians was integrated into the next iteration throughout the development phases of MSProDiscuss. The final tool was pilot tested in a separate validation study with clinicians (N=20, from the United States, Germany, and Canada) in a real-world setting, and it demonstrated high sensitivity and specificity to differentiate between patients with RRMS and those with SPMS. The tool also demonstrated evidence of construct validity, suggesting that the items included are relevant in assessing early signs of progression, and the HCPs supported the implementation and usefulness of the tool for clinical practice [8,9]. MSProDiscuss is part of several noninterventional longitudinal studies to further assess changes in the level of progression over the long term [10,11]. MSProDiscuss has been released for use in clinical practice, and the final validated tool can be accessed on the web [12] and on the Neuro-Compass medical education resource website [13].

Objective

We conducted a separate usability testing study to assess the performance of MSProDiscuss on a larger scale, involving HCPs from different geographies and health care systems, with the aims to further test the (1) usability of tool in daily clinical practice, and comprehensibility of items included: (2) usefulness of the tool to assist patient-physician discussion on MS disease progression, (3) feasibility, ease, and willingness of HCPs to integrate MSProDiscuss into their routine clinical practice, and (4) insights gathered from HCPs on areas of improvement on the tool's usefulness. In this paper, we report the findings from physicians on the usability and usefulness of MSProDiscuss while discussing disease progression with patients in a real-world setting.

Methods

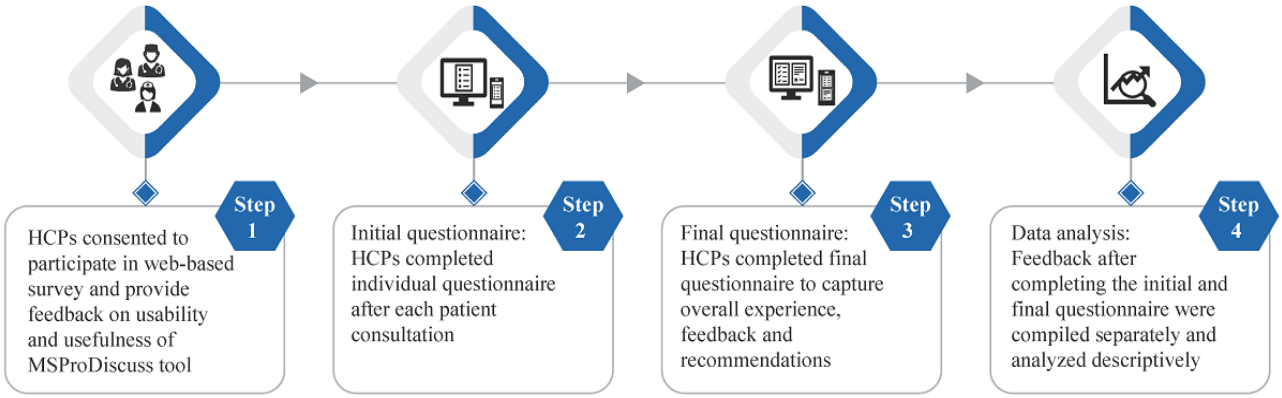
Study Design

This was a multinational, cross-sectional, and noninterventional study surveying HCPs, including MS specialists, general neurologists, or others as indicated by the self-identification choices provided to the participants. The HCPs responsible for the diagnosis, management, and care, or those in charge of symptom evaluation and control, of at least five patients with MS per week in their daily clinical practice were included in the survey. The HCPs were invited by local representatives of the study sponsor, Novartis, to participate in a web-based survey conducted between July 2019 and December 2019 in 34 countries across North America, Europe, Asia, South America, Africa, and Australia. No ethics approval was required because this was a survey. Nevertheless, all respondents were offered detailed information about the survey method, and they provided written consent before participating.

Survey Methodology

The survey was carried out by Real World Evidence Solutions (IQVIA AG). A web-based survey link was sent to the HCPs who consented to participate in the study. The survey consisted of 2 questionnaires to gather feedback on the understanding, usefulness, usability, integration, and adoption of MSProDiscuss in daily clinical practice. The questionnaires were developed based on feedback from HCPs during the tool development phase and according to the principles of implementation science, and inputs provided by the sponsor's medical team and by researchers from Real World Evidence Solutions were included. The initial translations of the questionnaires were provided by IQVIA, which were further evaluated and modified as necessary by the local Novartis medical teams. The HCPs used MSProDiscuss to assess what they felt was a broad range of people with MS, excluding those with clinically isolated syndrome and primary progressive MS. The survey methodology is illustrated in Figure 1.

Figure 1. Steps of the MSProDiscuss usability test survey. HCP: health care professional; MSProDiscuss: Multiple Sclerosis Progression Discussion Tool.



Each HCP was asked to fill two types of questionnaires (Multimedia Appendix 1), with their responses collected on a 5-point Likert scale. The HCPs filled an initial questionnaire after each instance of using MSProDiscuss during a face-to-face individual patient consultation. This initial questionnaire included 7 questions to collect feedback on the time needed to complete MSProDiscuss during a routine clinical visit, comprehensibility of the questions in the tool, and the overall usability and usefulness of MSProDiscuss in facilitating a discussion on disease progression between HCPs and patients in routine practice. Each HCP was expected to fill 10-40 initial questionnaires, depending on the patient population or country size. After completing all individual questionnaires, the HCPs also completed a final questionnaire on their overall assessment of MSProDiscuss, taking into account all previous consultations. The final questionnaire included 13 questions. In addition to those items covered in the initial questionnaire, we gathered feedback on the integration of MSProDiscuss into clinical practice. This two-step process was intended to ensure that the final feedback on usability and usefulness was received only after the HCPs had sufficient experience with the tool itself. The HCPs were also requested to provide general feedback on the features and performance of the tool in free-text fields.

Summary Statistics

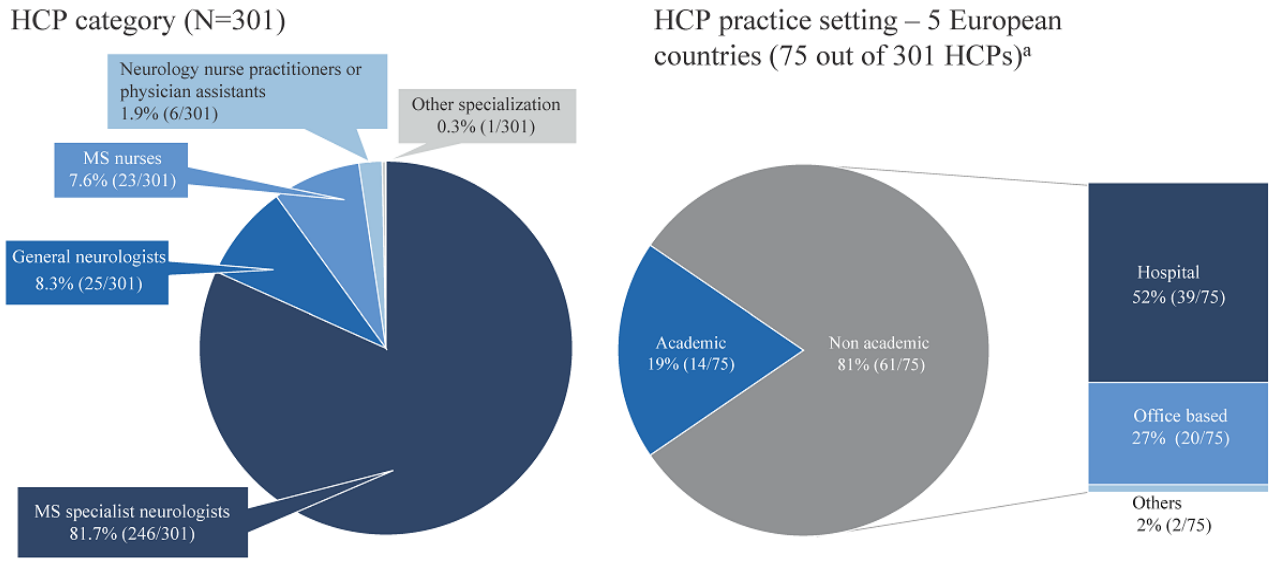
The HCP responses to the 2 usability questionnaires were analyzed separately with regard to the individual and final questionnaires and descriptively on a question-by-question basis and reported as a proportion of the total responses. The differences in the responses to the questionnaires by country were summarized by region and visualized as a heat map. No statistical comparisons were performed. To ensure that the survey sample was representative of the true population, both weighted and unweighted percentages were calculated. The weighted analysis ensured that the results were adjusted to reflect the underlying sample distribution. Therefore, weighting was performed at the country level with regard to the survey sample size. For the individual questionnaire, this refers to the number of patients for whom MSProDiscuss was used, whereas for the final questionnaire, the number of HCPs who participated in the survey was relevant (Multimedia Appendix 2). The feedback provided as free text was analyzed qualitatively by identifying common themes and categorizing them accordingly.

Results

Survey Participant Characteristics

Of the 390 HCPs who were invited, 301 provided feedback on at least one questionnaire. Most of the participants were MS specialists (246/301, 81.7%), but also included general neurologists, MS nurses and nurse practitioners, and physician assistants (Figure 2). In a subanalysis of the participating HCPs from 5 countries in Europe (France, Germany, Italy, Spain, and the United Kingdom, which together represented 24.9% (75/301) of the participating HCPs), 52% (36/75) of the HCPs were from a hospital-based practice setting (Figure 2). Overall, the tool was assessed in different practice settings, ranging from academic hospitals to general hospitals and office-based practices.

Figure 2. The characteristics of the survey participants. France, Germany, Italy, Spain, and the United Kingdom[a]. HCP: health care professional; MS: multiple sclerosis.



Of the 301 HCPs who participated, 232 completed the expected number of questionnaires. During the MSProDiscuss usability test on a total of 6974 patients, the HCPs identified 5370 (77%) patients with RRMS, which also included those who may be transitioning to SPMS. The number of patients per HCP and the total number of individual consultations in the 5 European countries are shown in Table 1. Details from the regions are provided in Multimedia Appendix 3.

Table 1. MSProDiscuss[a] use for the usability test in 5 European countries.

Country	HCPs[b] who used the tool (n=75), n (%)	Patients diagnosed with SPMS[c] at the time of the consultation during usability testing (n=477[d]), n (%)	Patients diagnosed with RRMS[e] at the time of the consultation during usability testing (n=1513), n (%)	Number of times the tool was used in a consultation (n=1990), n (%)
France	9 (12)	5 (1)	11 (0.7)	16 (0.8)
Germany	27 (36)	175 (36.7)	542 (35.8)	717 (36)
Italy	14 (18.7)	167 (35)	357 (23.6)	524 (26.3)
Spain	19 (25.3)	121 (25.4)	583 (38.5)	704 (35.4)
United Kingdom	6 (8)	9 (1.9)	20 (1.3)	29 (1.5)

[a]MSProDiscuss: Multiple Sclerosis Progression Discussion Tool.

[b]HCP: health care professional.

[c]SPMS: secondary progressive multiple sclerosis.

[d]In this cohort of 5 European countries, 23.97% (477/1990) and 76.03% (1513/1990) of the patients were identified as patients with secondary progressive multiple sclerosis and patients with relapsing-remitting multiple sclerosis, respectively.

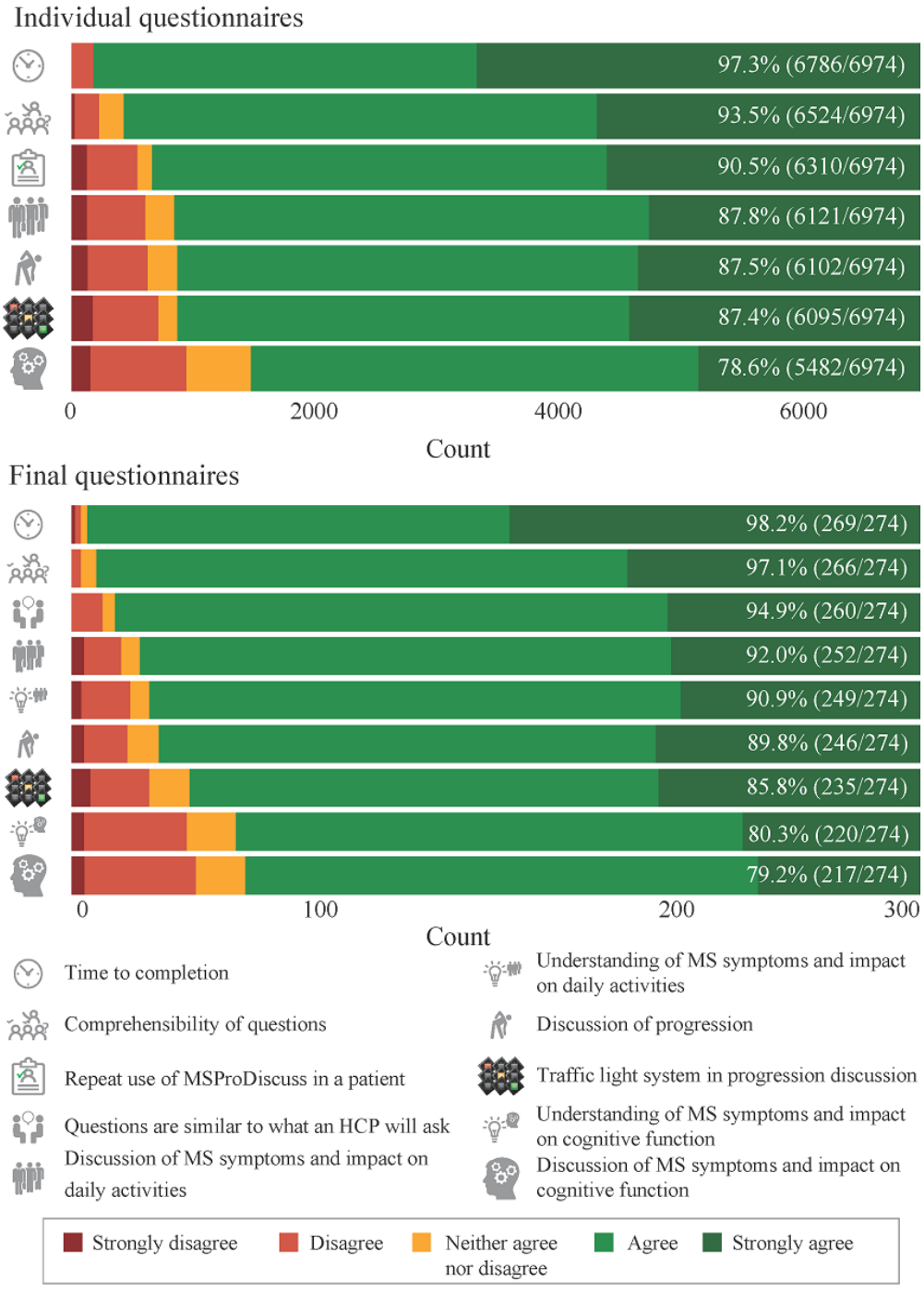
[e]RRMS: relapsing-remitting multiple sclerosis.

Feedback on the Usability and Usefulness of MSProDiscuss

The usability and usefulness of the tool were assessed using both the initial and final questionnaires (Figure 3). The HCPs first completed individual questionnaires after using MSProDiscuss on 6974 people with MS and then completed a final questionnaire (N=274). MSProDiscuss was confirmed to be useful in relation to all the dimensions assessed. Most of the HCPs agreed or strongly agreed (217/274, 79.2%-269/274, 98.2%) that MSProDiscuss is beneficial in their practice (Figure 3). The time taken to complete the tool during a routine

consultation was considered satisfactory (1-4 minutes) in 97.3% (6786/6974) of the initial questionnaires and in 98.2% (269/274) of the final questionnaires. The patients were able to comprehend the questions in the tool in) 93.5% (6524/6974) and 97.1% (266/274) of the cases (individual and final questionnaires, respectively). In 90.5% (6310/6974) of the individual questionnaires, the HCPs were willing to use the tool again with the same patient. MSProDiscuss was also useful in discussing MS symptoms and their impact on daily activities (6121/6974, 87.77% of the individual questionnaires and 252/274, 91.9% of the final questionnaires) and cognitive function (5482/6974, 78.61% for individual and 271/274, 79.2% for final questionnaires), as well as in discussing progression in general (6102/6974, 87.49% and 246/274, 89.8%, respectively). To summarize, there was excellent agreement between the initial and final questionnaires on all items and dimensions of usability and usefulness of the tool. Excellent usability and usefulness were reported from the initial instances, increasing slightly with repeated use.

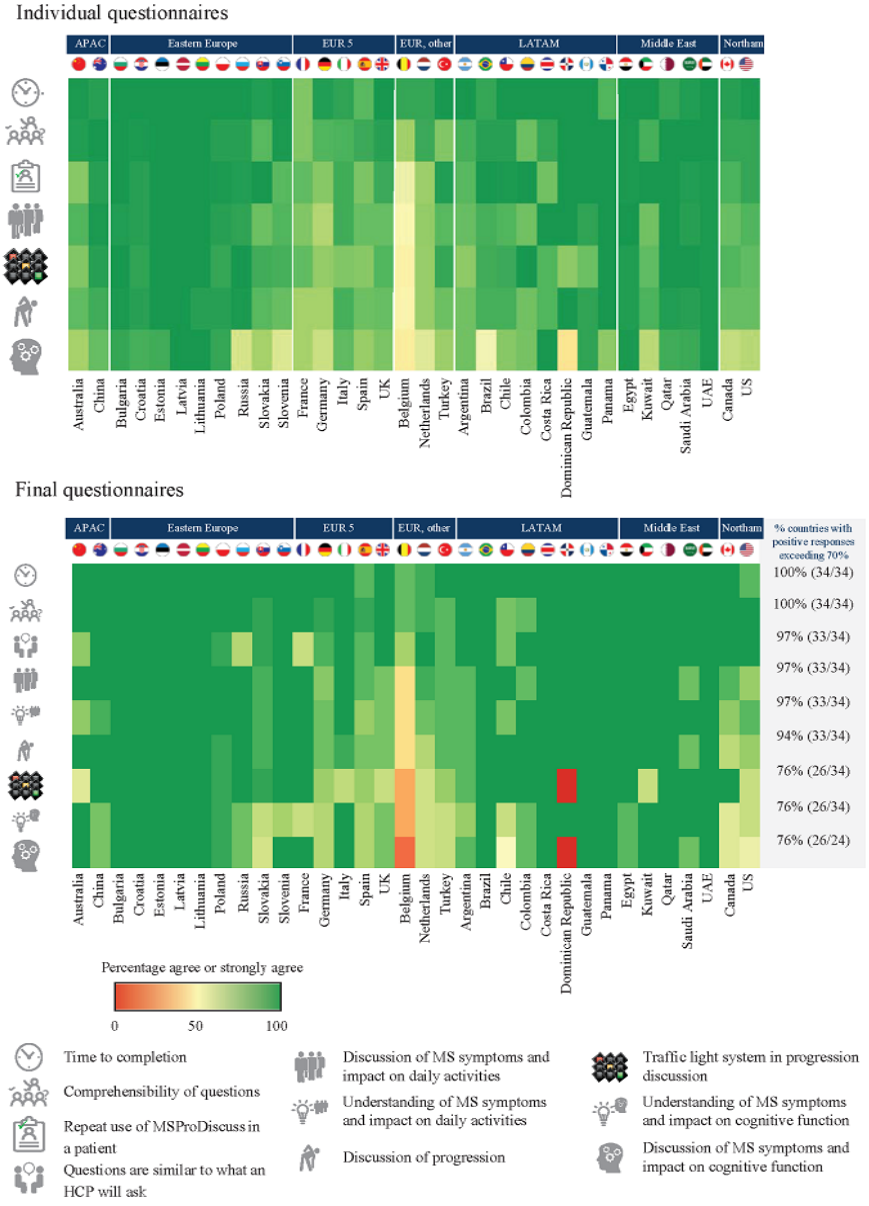
Figure 3. The summary findings from the usability and usefulness testing of the MSProDiscuss individual questionnaire (N=6974) and final questionnaire (N=274). The numbers in the bars reflect the proportion of health care professionals who responded agree or strongly agree for each item. Percentages are based on unweighted results; weighted results were similar. HCP: health care professional; MS: multiple sclerosis; MSProDiscuss: Multiple Sclerosis Progression Discussion Tool.



The final questionnaire assessed additional usability aspects from the perspective of the HCPs. There was general agreement (260/274, 94.9% of the HCPs) that the questions in the tool were similar to those asked by an HCP during a regular consultation. MSProDiscuss was also found to be helpful in understanding the impact of MS symptoms on daily activities (249/274, 90.9% of the HCPs) and cognitive function (220/274, 80.3% of the HCPs). The findings based on the weighted analysis were similar to those based on the unweighted results (Multimedia Appendices 4 and 5). Overall, the tool was confirmed to be highly usable and useful in clinical practice.

On the individual questionnaires, the responses from the individual countries and regions were in line with the overall results, with the exception of Belgium, as shown in Figure 4. Satisfaction with the time taken to complete the tool (30/34, 88% of the countries), comprehensibility of the questions (25/34, 76% of the countries), and equivalence to routine questions asked by an HCP (24/34, 71% of the countries) were the top-scoring dimensions, with unanimous agreement across countries and regions.

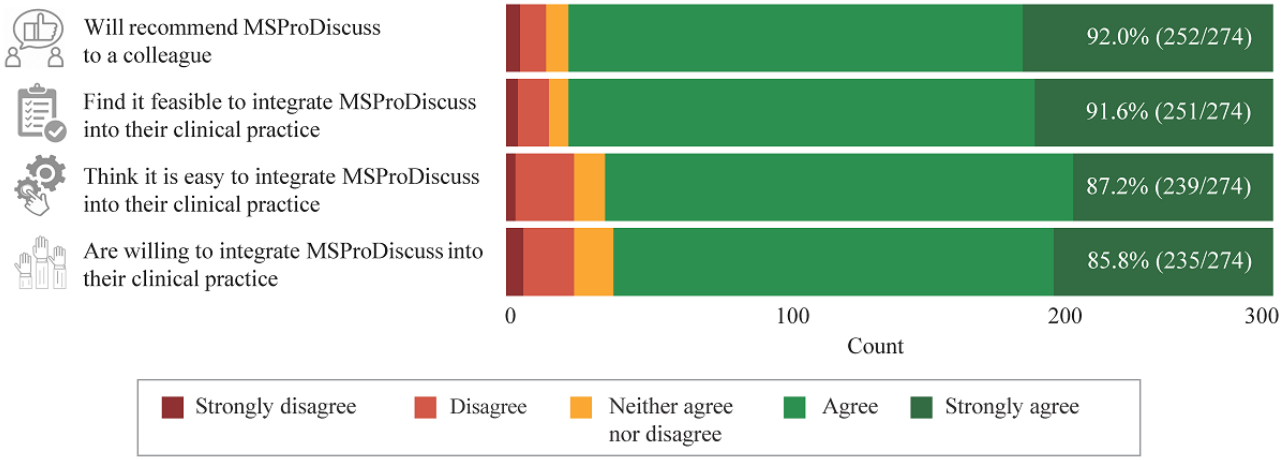
Figure 4. Distribution of responses on the usability and usefulness of MSProDiscuss by country and region: individual questionnaire (N=6974) and final questionnaire (N=274). APAC: Asia Pacific; EUR: Europe; HCP: health care professional; LATAM: Latin America; MS: multiple sclerosis; MSProDiscuss: Multiple Sclerosis Progression Discussion Tool; Northam: North America; UAE: United Arab Emirates.



Feedback on Integration of MSProDiscuss Into Clinical Practice

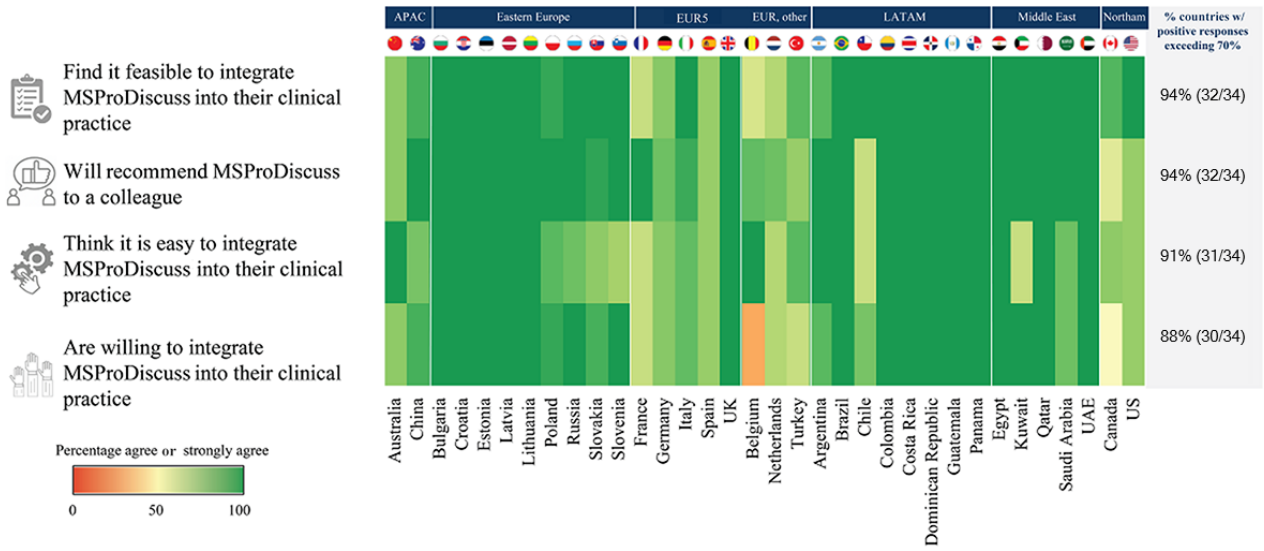
The final questionnaire included items to assess the adaptability of MSProDiscuss and the integration of the tool into routine clinical practice. The responses were very positive. Overall, 91.6% (251/274) of the HCPs thought that adaptability and integration are feasible, 87.2% (239/274) thought that adaptability and integration would be easy, and 85.8% (235/274) were willing to integrate MSProDiscuss into their clinical practice (Figure 5). Finally, 91.9% (252/274) of the HCPs would recommend MSProDiscuss to a colleague.

Figure 5. Integration of MSProDiscuss into clinical practice: summary findings from the final questionnaire (N=274). The numbers in the bars reflect the proportion of health care professionals who responded agree or strongly agree for each item. Percentages are based on unweighted results; weighted results are similar. MSProDiscuss: Multiple Sclerosis Progression Discussion Tool.



The findings at the country and regional levels were reflective of the overall results, with HCPs having high levels of agreement on the ease of implementation of the tool (Figure 6). An average agreement of more than 80% was seen in 76% (26/34) of countries and of more than 90% in 62% (21/34) of countries. In 53% (18/34) of countries, the HCPs unanimously agreed (ie, 100% in all 4 questions) that it is feasible and easy to integrate MSProDiscuss into their clinical practice, that they are willing to do so, and that they would recommend the tool to a colleague. This includes the United Kingdom and approximately half of the participating countries from Eastern Europe, Latin America, and the Middle East (Figure 6). Among the European countries, Belgium was an exception with a lower percentage of HCPs indicating willingness to integrate MSProDiscuss into clinical practice.

Figure 6. Distribution of health care professionals' responses on the final questionnaire (N=274) by country (N=34) and region (N=7). APAC: Asia Pacific; EUR: Europe; LATAM: Latin America; MSProDiscuss: Multiple Sclerosis Progression Discussion Tool; Northam: North America; UAE: United Arab Emirates.



Additional Feedback on MSProDiscuss

Throughout the development of MSProDiscuss, qualitative feedback was gathered, and improvements were implemented iteratively. A meaningful impact of this approach may be seen in the overall positive experience expressed by the HCPs while using MSProDiscuss. Of the 301 HCPs who returned the final questionnaire, 162 made at least one recommendation on improvements, such as including additional variables or expanding the existing variables. Table 2 lists these suggestions by topic and implementation status. Even before the conclusion of the survey on usability, implementation was already ongoing for several recommendations, including (1) integration of MSProDiscuss into electronic health record systems to allow for longitudinal patient follow-up in countries such as the United States, United Kingdom, and Germany; (2) an enhanced user interface for easier navigation; (3) improvement of the sensitivity of the tool for patients with lower Expanded Disability Status Scale (EDSS) scores or confounding symptoms such as fatigue; and, most importantly, (4) development of a patient version of the tool. Some of the other recommendations such as including cognitive or patient-reported outcomes

measures are being considered for future updates of the tool. They need to be carefully weighed against the current high ease of use. Overall, the recommendations were focused on expanding the scope and reach of the tool, rather than improving basic usability, reinforcing the earlier conclusion that the tool is practical and easy to use.

Table 2. Recommendations from health care professionals on improvements.

Response	Number of responses	Details of the recommendation	Recommendation already implemented during usability testing
Expand components included in the tool or include additional questions	162	Add disease duration; add treatment adherence	N/A[a]
Interpretation of the traffic light output	79	Elaborate on the explanation or quantify the output	N/A
Improve or expand on the cognitive assessment	69	Include an interface with cognitive assessment scales	N/A
Include additional variables	57	Add more details on the impact of disease on daily activities, for example, relationships, social, work, sexuality, and emotional state	N/A
Improve tool sensitivity	56	Improve sensitivity of the tool	Improved sensitivity for patients with lower EDSS[b] scores and with confounding symptoms such as fatigue
Include longitudinal follow-up	29	N/A	Included in observational studies for longitudinal monitoring; incorporated in electronic health records
Improve the user interface	29	N/A	User interface was improved on an ongoing basis
Provide a patient version	18	N/A	Development was ongoing in parallel to the usability testing. Meanwhile, the Your MS[c] questionnaire is ready

[a]N/A: not applicable or implemented.

[b]EDSS: Expanded Disability Status Scale.

[c]MS: multiple sclerosis.

Discussion

SPMS Remains a Diagnostic Challenge

The highly variable disease course in individual patients and the lack of consensus on the diagnostic criteria often result in considerable delays in SPMS diagnosis [5,14]. In 2014, the study by Lublin et al [15] reported revised clinical and imaging findings to define clinical phenotypes. However, the study also acknowledged that objective criteria for separating clinical phenotypes are still lacking. There are no clear definitions of the transition from RRMS to SPMS. Until recently, no therapy with proven efficacy was available for a broad range of patients with SPMS. Thus, the diagnosis of SPMS is often delayed to avoid losing treatment options and reimbursement. With the advent of newer treatments for patients with SPMS, tools such as MSProDiscuss could be useful in supporting *real-time* evaluation of early signs of MS progression in routine clinical practice [16]. Tools based on algorithms and nomograms have been developed that use quantitative, data-driven empirical assessments. Although these tools can estimate the future risk of SPMS progression, they cannot easily be translated into current decisions in routine patient management [17-19]. MSProDiscuss provides additional insight through qualitative assessment of disease symptoms and their impact on daily life, thereby including, for the first time, the patient's perspective in the overall shared decision-making. Timely diagnosis will allow for appropriate treatment and better long-term prognosis [14].

Principal Findings

The results of this survey indicate that MSProDiscuss is a useful tool to aid the discussion of disease progression with patients. The findings were consistent between the individual and final questionnaires. Most of the HCPs agreed or strongly agreed that MSProDiscuss was beneficial in their practice. The time taken to complete the tool during routine consultation was considered satisfactory (1-4 minutes). In most instances, the questions in MSProDiscuss were found to be comprehensible by the patients and were similar to those asked by an HCP during a regular consultation. Thus, MSProDiscuss facilitates patient-physician discussion by capturing a structured disease history without imposing an additional time burden on the HCP. Furthermore, these positive results indicate that MSProDiscuss is easy to use and universally helpful, regardless of region, professional background, and practice setting.

More than 90.48% (6310/6974) of instances the HCPs indicated that they would use MSProDiscuss again with the same patients, including both patients with RRMS (4985/6310, 79%) and patients with SPMS (1325/6310, 20.99%). In particular, the HCPs indicated that they would use the tool again with more than 97% (5209/5370) of the patients with RRMS and with 94.01% (1508/1604) of the patients with SPMS. This indicates

a potential role of MSProDiscuss as a complementary disease-monitoring tool for longitudinal follow-up to be added to current empirical measures such as clinical relapses or magnetic resonance imaging lesion counts. Qualitative insight into the symptoms and their impact on patients as captured by MSProDiscuss will help identify early signs of disease progression. In patients with RRMS, the tool might complement clinical assessments and help in treatment decision-making during the earliest stages of SPMS. For patients already diagnosed with SPMS, the tool will be helpful in identifying functional domains that are most affected by progression and in choosing the appropriate modalities for symptom management.

In this survey, 80.3% (220/274) or more of the HCPs indicated that MSProDiscuss is a useful aid in discussing MS symptoms and their impact on daily activities and cognitive function, as well as in discussing progression in general. The fact that 91.9% (252/274) of the HCPs would recommend MSProDiscuss to their colleagues and think that it is feasible to integrate MSProDiscuss into their clinical practice is very encouraging because it suggests that the tool is beneficial to patients as well as HCPs themselves. Key recommendations have already been implemented, including longitudinal follow-up, an enhanced user interface, improved sensitivity at lower EDSS scores, correction for overlapping fatigue symptoms, and creation of a patient-completed version. The suggestion to expand cognitive assessments is a key point; cognitive impairment is an important yet underrecognized sign of disease progression [20,21]. How empirical measures of cognitive performance could be integrated into this tool, without affecting the overall performance of the tool in terms of the time taken and ease of use during regular consultations, is an important goal for future development. Similarly, whether objective patient-reported outcomes measures could be integrated into MSProDiscuss remains to be evaluated.

Along with the traffic light output, considering the patient's responses to the individual questions provides additional insights, supporting the relevance of MSProDiscuss for holistic disease management in individual patients. MSProDiscuss is suitable for longitudinal follow-up and has already been included in large observational studies [22] and integrated into the electronic health records in several countries. This will ensure the systematic recording of a patient's disease evolution and individual patient monitoring over time. The tool will assist in defining a multidisciplinary treatment strategy for individual patients, including physiotherapy, rehabilitation, and relevant symptomatic treatments. As MSProDiscuss covers several functional domains, it provides comprehensive information on the patient's health status, including MS phenotype and symptomatology [23]. Although EDSS assessment is viewed as the gold standard for measuring physical disability, it has been implemented in clinical routine on a limited scale [24-26]. MSProDiscuss will complement EDSS assessment in routine practice by capturing symptoms and their impact between consecutive EDSS assessments.

A patient version of the tool, the Your MS questionnaire, was developed to cover patient-derived information on MS symptoms also contained in MSProDiscuss (Multimedia Appendix 6). The Your MS questionnaire can be completed by the patient in preparation for a clinical visit, potentially with help from a caregiver who can provide additional information [27]. The Your MS questionnaire may further complement patient-physician interactions on disease progression and is expected to not only help reduce the on-call burden on physicians without compromising on the quality of consultation, but also positively involve patients in the management of their own disease. The use of telemedicine is an increasing trend in the management of chronic diseases [28,29]; most recently, the implementation of telemedicine in the management of visits by patients with MS has been accelerated by the COVID-19 pandemic [30,31]. In the context of teleconsultation, the Digital Technologies Web and Social Media Study Group recently suggested a battery for assessing MS disability and relapse, proposing commonly used tools that are suitable before, during, and after a teleconsultation [32]. Being a web-based tool, MSProDiscuss was tested during the COVID-19 pandemic, and it showed promise in assisting with remote visits where the lack of face-to-face interaction can hamper communication. As one of the first tools of its kind, MSProDiscuss is able to assist with deep clinical phenotyping of signs of progression, based on the physician's documented patient history [33]. As a tool with promise, the patient-completed Your MS questionnaire can either be used ahead of traditional face-to-face visits, or it may be integrated into the suggested battery in preparation for a teleconsultation. Although implementation of these digital or web-based tools in the management of MS is a necessity during the pandemic, these tools are unlikely to fully replace face-to-face consultations [30]. In-person evaluation and differential diagnosis of MS symptoms remain vital, and especially in light of technological and psychological limitations, web-based tools will remain complementary to in-person consultations [34].

Overall, the results were consistent at the individual country and region levels, with some minor variations reflective of either individual HCP preference of disease management or general practice guidelines followed in the country. The overall positive feedback from most of the countries and regions suggests that MSProDiscuss is already perceived as a valuable tool. In some countries such as Belgium where the feedback was less encouraging in terms of integration into clinical practice, a root cause analysis could not be performed because of the limitations of the study design. It can be speculated that, with limited treatment options in some geographies, it is likely that there is caution regarding discussing disease progression when no solution can be proposed. In other cases, there may be reluctance to discuss disease progression because of the potential emotional and psychological implications for the patient. It is likely that these factors could have influenced the respondents in our survey. However, this only re-emphasizes the need for clear and transparent discussions regarding the importance of identifying early signs of disease progression, allowing treatments and holistic modalities to be started sooner.

Study Limitations

Gathering feedback on the usability and usefulness of the tool while simultaneously using the tool might have resulted in a potential bias in the HCPs' responses to certain components of the survey, such as integration into practice and time taken for

completion. In general, the questionnaire methodology only highlights trends or attitudes and does not explain the underlying reasons for the responses [35].

Outlook

Overall, the positive findings from our usability and usefulness study of MSProDiscuss are very promising [36]. Although centers with a heavy patient inflow might find it difficult to implement a new tool in their workflow, our results show that HCPs across different practice settings can easily integrate MSProDiscuss into their routine practice. MSProDiscuss does not require extensive data to be collected or curated to assess the level of disease progression, and it is complementary to other approaches based on imaging, neurological examination, or biomarkers. As a valuable disease-monitoring tool, advantages in the long term should outweigh the initial implementation challenges, if any. When used together with Your MS, the patient version, MSProDiscuss can also be a valuable tool in the day-to-day management of people with MS through telemedicine, even during crisis situations. The future impact on long-term disease monitoring and health care resource utilization remains to be evaluated.

Conclusions

The findings from this real-world study suggest that MSProDiscuss is a usable and useful tool to facilitate a physician-patient discussion on disease progression in daily clinical practice. MSProDiscuss facilitates the dialog between the patient and the physician by capturing a structured disease history. Most of the survey participants indicated that MSProDiscuss was beneficial in the discussion of disease progression. Overall, the feedback from the HCPs was very positive regarding the integration of MSProDiscuss into their clinical practice. The tool was used by physicians, MS nurses and nurse practitioners, and pharmacists from very different practice settings and was found to be of value; MSProDiscuss is a tool that is acceptable to all the users involved in the care and management of patients with MS.

Acknowledgments

Funding support was provided by Novartis Pharma AG (Basel, Switzerland). Real World Evidence Solutions, IQVIA Technology and Services, Basel, Switzerland (funded by the research sponsor Novartis), designed and conducted the survey and collected and managed the data. The authors acknowledge all health care professionals who participated in this usability survey. The authors also acknowledge Uma Kundu of Novartis Healthcare Pvt Ltd for medical writing support, which included the literature search, drafting of the article, revising it per author comments, and preparing the manuscript for journal submission, and Suzannah Ryan of Novartis Pharma AG for coordinating author review. The final responsibility for the content lies with the authors.

Conflicts of Interest

TZ has received compensation for consulting from Biogen, Bayer, Celgene, Novartis, Roche, Sanofi, and Teva and for research from Bayer, Biogen, Novartis, Teva, and Sanofi. GG has received compensation for consulting from AbbVie, Actelion, Atara Bio, Biogen, Celgene, Sanofi-Genzyme, Genentech, GlaxoSmithKline, Merck-Serono, Novartis, Roche, and Teva and for research from Biogen, Roche, Merck, Merck-Serono, Novartis, Sanofi-Genzyme, and Takeda. In addition, GG has received personal compensation from Elsevier for serving as an editor on Multiple Sclerosis and Related Disorders. EA has received compensation for consulting from Actelion, Biogen, Celgene, EMD Serono, Genentech, Genzyme, Novartis, Teva, and TG Therapeutics and for research from Biogen, Genentech, Novartis, and Rocky Mountain MS Center. VB has received compensation for consulting from Novartis. CH has received compensation for consulting and research from Novartis, Biogen, and Genentech; for consulting from EMD Serono; and for consulting and serving on a speakers bureau from Genzyme. OH has received compensation for consulting from Biogen, Roche, Merck, Novartis, Sanofi, and Celgene; for non–Continuing Medical Education activities from Alexion, Novartis, Roche, and Sanofi; for research from Novartis, Sanofi, and Biogen; and for travel support from Celgene. COG has received compensation for speaking and/or consultancy from Biogen, Sanofi-Genzyme, Merck, Roche, Teva, and Novartis. RRC has received compensation for consulting from Biogen, Roche, Novartis, Merck, Sanofi, Genzyme, and Teva. MT has received compensation for consulting and serving on a speakers bureau from Biogen, Merck, Roche, and Novartis. PV has received compensation for consulting and/or research as well as registration, travel, and accommodation for meetings from Biogen, Roche, Novartis, Sanofi, Teva, Merck, and Celgene. PD is an employee of IQVIA AG, Basel, Switzerland, which conducted this survey. JB has received compensation for consulting and serving on a speakers bureau from Novartis. MK, BS, BR, TH, and DPM are employees of Novartis.

Multimedia Appendix 1
Usability test questions (responses to each question were provided on a 5-point Likert scale).
[DOCX File , 25 KB - jmir_v23i10e29558_app1.docx]

Multimedia Appendix 2
Illustration showing the principle of weighting.
[DOCX File , 57 KB - jmir_v23i10e29558_app2.docx]

Multimedia Appendix 3
Multiple Sclerosis Progression Discussion Tool use for the usability test by region and patient type.
[DOCX File , 27 KB - jmir_v23i10e29558_app3.docx]

Multimedia Appendix 4
Summary findings from the individual questionnaire: weighted results.
[DOCX File , 127 KB - jmir_v23i10e29558_app4.docx]

Multimedia Appendix 5
Summary findings from the final questionnaire: weighted results.
[DOCX File , 1867 KB - jmir_v23i10e29558_app5.docx]

Multimedia Appendix 6
Screenshots of the Your MS Questionnaire.
[DOCX File , 617 KB - jmir_v23i10e29558_app6.docx]

References

1. Browne P, Chandraratna D, Angood C, Tremlett H, Baker C, Taylor BV, et al. Atlas of multiple sclerosis 2013: a growing global problem with widespread inequity. Neurology 2014 Sep 09;83(11):1022-1024 [FREE Full text] [doi: 10.1212/WNL.0000000000000768] [Medline: 25200713]
2. Kremenchutzky M, Rice GP, Baskerville J, Wingerchuk DM, Ebers GC. The natural history of multiple sclerosis: a geographically based study 9: observations on the progressive phase of the disease. Brain 2006 Mar;129(Pt 3):584-594. [doi: 10.1093/brain/awh721] [Medline: 16401620]
3. Koch M, Kingwell E, Rieckmann P, Tremlett H, UBC MS Clinic Neurologists. The natural history of secondary progressive multiple sclerosis. J Neurol Neurosurg Psychiatry 2010 Sep;81(9):1039-1043. [doi: 10.1136/jnnp.2010.208173] [Medline: 20639385]
4. Inojosa H, Proschmann U, Akgün K, Ziemssen T. A focus on secondary progressive multiple sclerosis (SPMS): challenges in diagnosis and definition. J Neurol 2021 Apr;268(4):1210-1221. [doi: 10.1007/s00415-019-09489-5] [Medline: 31363847]
5. Katz Sand I, Krieger S, Farrell C, Miller AE. Diagnostic uncertainty during the transition to secondary progressive multiple sclerosis. Mult Scler 2014 Oct;20(12):1654-1657. [doi: 10.1177/1352458514521517] [Medline: 24493475]
6. Ziemssen T, Tolley C, Bennett B, Kilgariff S, Jones E, Pike J, et al. A mixed methods approach towards understanding key disease characteristics associated with the progression from RRMS to SPMS: physicians' and patients' views. Mult Scler Relat Disord 2020 Feb;38:101861 [FREE Full text] [doi: 10.1016/j.msard.2019.101861] [Medline: 31865132]
7. Tolley C, Piani-Meier D, Bentley S, Bennett B, Jones E, Pike J, et al. A novel, integrative approach for evaluating progression in multiple sclerosis: development of a scoring algorithm. JMIR Med Inform 2020 Apr 14;8(4):e17592 [FREE Full text] [doi: 10.2196/17592] [Medline: 32286236]
8. Ziemssen T, Piani-Meier D, Bennett B, Johnson C, Tinsley K, Trigg A, et al. A physician-completed digital tool for evaluating disease progression (multiple sclerosis progression discussion tool): validation study. J Med Internet Res 2020 Feb 12;22(2):e16932 [FREE Full text] [doi: 10.2196/16932] [Medline: 32049062]
9. Inojosa H, Akgün K, Haacke K, Ziemssen T. [MSProDiscuss - development of a digital anamnesis tool to identify disease progression in multiple sclerosis]. Fortschr Neurol Psychiatr 2021 Jul;89(7-08):374-381. [doi: 10.1055/a-1397-6851] [Medline: 33723837]
10. Inojosa H, Rauser B, Ettle B, Ziemssen T. The transitional phase of multiple sclerosis: the concept of PANGAEA 2.0 evolution study. Mult Scler Relat Disord 2020 Nov;46:102523 [FREE Full text] [doi: 10.1016/j.msard.2020.102523] [Medline: 32977073]
11. Ziemssen T, Hoffmann O, Klotz L, Schreiber H, Weber MS, Rauser B. Gaining first insights on secondary progressive multiple sclerosis patients treated with Siponimod in clinical routine: protocol of the noninterventional study Amasia. JMIR Res Protoc 2020 Jul 24;9(7):e19598 [FREE Full text] [doi: 10.2196/19598] [Medline: 32499214]
12. MSProDiscuss. URL: https://msprodiscuss.com/ [accessed 2021-08-13]
13. MSProDiscuss: MS Progression Discussion Tool. URL: https://www.neuro-compass.education/en-gb/msprodiscuss/ [accessed 2021-08-13]
14. Giovannoni G, Butzkueven H, Dhib-Jalbut S, Hobart J, Kobelt G, Pepper G, et al. Brain health: time matters in multiple sclerosis. Mult Scler Relat Disord 2016 Sep;9 Suppl 1:S5-48 [FREE Full text] [doi: 10.1016/j.msard.2016.07.003] [Medline: 27640924]
15. Lublin FD, Reingold SC, Cohen JA, Cutter GR, Sørensen PS, Thompson AJ, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. Neurology 2014 Jul 15;83(3):278-286 [FREE Full text] [doi: 10.1212/WNL.0000000000000560] [Medline: 24871874]

16. Inojosa H, Proschmann U, Akgün K, Ziemssen T. Should we use clinical tools to identify disease progression? Front Neurol 2021 Jan 21;11:628542 [FREE Full text] [doi: 10.3389/fneur.2020.628542] [Medline: 33551982]
17. Skoog B, Tedeholm H, Runmarker B, Odén A, Andersen O. Continuous prediction of secondary progression in the individual course of multiple sclerosis. Mult Scler Relat Disord 2014 Sep;3(5):584-592. [doi: 10.1016/j.msard.2014.04.004] [Medline: 26265270]
18. Lorscheider J, Buzzard K, Jokubaitis V, Spelman T, Havrdova E, Horakova D, MSBase Study Group. Defining secondary progressive multiple sclerosis. Brain 2016 Sep;139(Pt 9):2395-2405. [doi: 10.1093/brain/aww173] [Medline: 27401521]
19. Manouchehrinia A, Zhu F, Piani-Meier D, Lange M, Silva DG, Carruthers R, et al. Predicting risk of secondary progression in multiple sclerosis: a nomogram. Mult Scler 2019 Jul;25(8):1102-1112. [doi: 10.1177/1352458518783667] [Medline: 29911467]
20. Moccia M, Lanzillo R, Palladino R, Chang KC, Costabile T, Russo C, et al. Cognitive impairment at diagnosis predicts 10-year multiple sclerosis progression. Mult Scler 2016 Apr;22(5):659-667. [doi: 10.1177/1352458515599075] [Medline: 26362896]
21. Planche V, Gibelin M, Cregut D, Pereira B, Clavelou P. Cognitive impairment in a population-based study of patients with multiple sclerosis: differences between late relapsing-remitting, secondary progressive and primary progressive multiple sclerosis. Eur J Neurol 2016 Feb;23(2):282-289. [doi: 10.1111/ene.12715] [Medline: 25903918]
22. Ziemssen T, Kern R, Cornelissen C. Study design of PANGAEA 2.0, a non-interventional study on RRMS patients to be switched to fingolimod. BMC Neurol 2016 Aug 08;16:129 [FREE Full text] [doi: 10.1186/s12883-016-0648-6] [Medline: 27502119]
23. Ziemssen T, Kern R, Thomas K. Multiple sclerosis: clinical profiling and data collection as prerequisite for personalized medicine approach. BMC Neurol 2016 Aug 02;16(1):124 [FREE Full text] [doi: 10.1186/s12883-016-0639-7] [Medline: 27484848]
24. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 1983 Nov;33(11):1444-1452. [doi: 10.1212/wnl.33.11.1444] [Medline: 6685237]
25. Hobart J, Freeman J, Thompson A. Kurtzke scales revisited: the application of psychometric methods to clinical intuition. Brain 2000 May;123(Pt 5):1027-1040. [doi: 10.1093/brain/123.5.1027] [Medline: 10775547]
26. Inojosa H, Schriefer D, Ziemssen T. Clinical outcome measures in multiple sclerosis: a review. Autoimmun Rev 2020 May;19(5):102512. [doi: 10.1016/j.autrev.2020.102512] [Medline: 32173519]
27. Welcome to your MS questionnaire. Your MS Questionnaire. URL: https://www.yourms.com/#/welcome [accessed 2021-08-13]
28. Bashshur RL, Shannon GW, Smith BR, Alverson DC, Antoniotti N, Barsan WG, et al. The empirical foundations of telemedicine interventions for chronic disease management. Telemed J E Health 2014 Sep;20(9):769-800 [FREE Full text] [doi: 10.1089/tmj.2014.9981] [Medline: 24968105]
29. Yeroushalmi S, Maloni H, Costello K, Wallin MT. Telemedicine and multiple sclerosis: a comprehensive literature review. J Telemed Telecare 2020;26(7-8):400-413. [doi: 10.1177/1357633X19840097] [Medline: 31042118]
30. Scholz M, Haase R, Schriefer D, Voigt I, Ziemssen T. Electronic health interventions in the case of multiple sclerosis: from theory to practice. Brain Sci 2021 Feb 02;11(2):180 [FREE Full text] [doi: 10.3390/brainsci11020180] [Medline: 33540640]
31. Hyland MH, Robb JF. Teleconsultation will replace most face-to-face interactions in the multiple sclerosis clinic - Yes. Mult Scler 2021 Feb;27(2):174-175. [doi: 10.1177/1352458520973640] [Medline: 33325822]
32. Moccia M, Lanzillo R, Brescia Morra V, Bonavita S, Tedeschi G, Leocani L, Digital Technologies WebSocial Media Study Group of the Italian Society of Neurology. Assessing disability and relapses in multiple sclerosis on tele-neurology. Neurol Sci 2020 Jun;41(6):1369-1371 [FREE Full text] [doi: 10.1007/s10072-020-04470-x] [Medline: 32440979]
33. Voigt I, Inojosa H, Dillenseger A, Haase R, Akgün K, Ziemssen T. Digital twins for multiple sclerosis. Front Immunol 2021 May 3;12:669811 [FREE Full text] [doi: 10.3389/fimmu.2021.669811] [Medline: 34012452]
34. Moccia M. Teleconsultation will replace most face-to-face interactions in the multiple sclerosis clinic - No. Mult Scler 2021 Feb;27(2):176-177. [doi: 10.1177/1352458520968815] [Medline: 33325802]
35. Saunders M, Lewis P, Thornhill A. Research Methods for Business Students. Englewood Cliffs, NJ: Prentice Hall; 2007.
36. Sousa VE, Dunn Lopez K. Towards usable e-health. A systematic review of usability questionnaires. Appl Clin Inform 2017 May 10;8(2):470-490 [FREE Full text] [doi: 10.4338/ACI-2016-10-R-0170] [Medline: 28487932]

Abbreviations

EDSS: Expanded Disability Status Scale
HCP: health care professional
MS: multiple sclerosis
MSProDiscuss: Multiple Sclerosis Progression Discussion Tool
RRMS: relapsing-remitting multiple sclerosis
SPMS: secondary progressive multiple sclerosis

Edited by R Kukafka; submitted 13.04.21; peer-reviewed by R Haase, M Moccia; comments to author 10.05.21; revised version received 05.07.21; accepted 27.07.21; published 06.10.21.

Please cite as:
Ziemssen T, Giovannoni G, Alvarez E, Bhan V, Hersh C, Hoffmann O, Oreja-Guevara C, Robles-Cedeño RR, Trojano M, Vermersch P, Dobay P, Khwaja M, Stadler B, Rauser B, Hach T, Piani-Meier D, Burton J
Multiple Sclerosis Progression Discussion Tool Usability and Usefulness in Clinical Practice: Cross-sectional, Web-Based Survey
J Med Internet Res 2021;23(10):e29558
URL: https://www.jmir.org/2021/10/e29558
doi:10.2196/29558
PMID:34612826

Original Paper

Facilitators of and Barriers to Lifestyle Support and eHealth Solutions: Interview Study Among Health Care Professionals Working in Cardiac Care

Talia R Cohen Rodrigues[1], MSc[‡]; David R de Buisonjé[1], MSc; Mike Keesman[1], PhD; Thomas Reijnders[1,2], PhD; Jessica E van der Geer[1], MSc; Veronica R Janssen[1,3], PhD; Roderik A Kraaijenhagen[4,5], PhD; Douwe E Atsma[3], Prof Dr; Andrea W M Evers[1,6,7], Prof Dr

[1]Health, Medical, and Neuropsychology Unit, Faculty of Social and Behavioural Sciences, Leiden University, Leiden, Netherlands

[2]Department of Human-Centered Design, Faculty of Industrial Design Engineering, Technical University of Delft, Delft, Netherlands

[3]Department of Cardiology, Leiden University Medical Center, Leiden, Netherlands

[4]NDDO Institute for Prevention and Early Diagnostics (NIPED), Amsterdam, Netherlands

[5]Vital10, Amsterdam, Netherlands

[6]Department of Psychiatry, Leiden University Medical Center, Leiden, Netherlands

[7]Medical Delta, Leiden-Delft-Erasmus Universities, Delft, Netherlands

[‡]the BENEFIT consortium

Corresponding Author:
Talia R Cohen Rodrigues, MSc
Health, Medical, and Neuropsychology Unit
Faculty of Social and Behavioural Sciences
Leiden University
Wassenaarseweg 52
Leiden, 2333 AK
Netherlands
Phone: 31 71 527 3627
Email: t.r.cohen.rodrigues@fsw.leidenuniv.nl

Abstract

Background: Cardiovascular diseases (CVDs) pose a significant health threat and reduce both people's life expectancy and quality of life. Healthy living is a key component in the effective prevention and treatment of CVD. However, health care professionals (HCPs) experience difficulties in supporting lifestyle changes among their patients. eHealth can provide a solution to these barriers.

Objective: This study aims to provide insights into the factors HCPs find important in the support of patients with CVD in the uptake of and adherence to a healthy lifestyle and the perceived facilitators of and barriers to using eHealth to provide lifestyle support to patients with CVD.

Methods: In-depth interviews were conducted with 16 Dutch HCPs specializing in lifestyle support in cardiac care.

Results: We identified 13 themes, of which the first 12 concerned lifestyle support in general and were related to intervention, patient, or health care. Throughout these themes, the use of eHealth reoccurred as a potential facilitator of or solution to barriers to lifestyle support. Our final theme specifically concerned barriers to the adoption and usability of eHealth.

Conclusions: HCPs do recognize the potential advantages of eHealth while experiencing barriers to using digital tools. Incorporating their needs and values in the development of lifestyle support programs, especially eHealth, could increase their use and lead to a more widespread adoption of eHealth into health care.

(J Med Internet Res 2021;23(10):e25646) doi:10.2196/25646

KEYWORDS

eHealth; digital health; cardiovascular disease; cardiac care; lifestyle change; lifestyle support; intervention; health care professionals; implementation; interview; facilitators; barriers

Introduction

Background

Cardiovascular diseases (CVDs) are the most common cause of death globally [1] and drastically reduce the quality of life [2]. Most CVDs can be prevented and treated by addressing behavioral risk factors such as smoking, poor diet, physical inactivity, and low sleep quality [3,4]. Lifestyle interventions for patients with CVD have been shown to improve risk factors and decrease cardiac readmissions and mortality [5] and are therefore recommended by national and international guidelines on cardiac prevention and rehabilitation [3,6]. Furthermore, a healthy lifestyle has mortality-reducing effects comparable with those of medication intake [7].

Notwithstanding these advantages, health care professionals (HCPs) seem to be hesitant to discuss—let alone prescribe—lifestyle interventions to their patients [8]. For instance, studies among Dutch general practitioners found that they are generally unlikely to mention CVD risk factors, such as lifestyle, during their consultations [9] and only provided advice concerning healthy living in 1 out of 6 consultations to people with hypertension complaints [10]. In addition, only 1 out of 5 primary care physicians indicated that European CVD guidelines concerning lifestyle were being implemented [11]. In line with this, most patients with CVD have an unhealthy lifestyle [12]. A number of barriers have been identified as possible explanations for the low attention paid to lifestyle changes and their respective programs. HCPs have mentioned a low degree of patient motivation, a lack of knowledge about or experience with providing lifestyle advice, insufficient time during consultations, a lack of financial incentives, little external options to refer their patients to, and HCPs' perception that health promotion is ineffective in CVD prevention and treatment [11,13-16]. These barriers could possibly explain why only half of the patients with CVD are actually offered lifestyle modification programs after discharge [17]. To increase the uptake of and adherence to lifestyle interventions among patients with CVD, it is crucial to consider HCPs' needs and barriers.

A potential solution to overcome these barriers could be the increased use of digital tools to provide automated or remote support, which can incorporate interactive web-based components and digital wearables for home measurement, known as eHealth [18]. Recent studies have shown that eHealth can be effective in the prevention and treatment of noncommunicable diseases such as CVD [19,20]. Despite these promising results, the acceptance and successful implementation of digital tools in health care is low because of the barriers that HCPs experience [21-23]. To design digital interventions that have an impact, human and contextual factors should be taken into account, including the needs and values of stakeholders such as HCPs [24,25]. Interview studies uncovering HCPs' views on lifestyle interventions and their own role in health promotion among their patients [15,26-29] and on the use of eHealth in lifestyle support [30-34] have been conducted before. However, to our knowledge, no studies have mapped out the attitudes toward both lifestyle support and eHealth of HCPs specialized in CVD specifically. Importantly, given that the effectiveness and needs related to lifestyle support and eHealth are highly context dependent [24], specific knowledge is needed in the context of cardiac care.

Objective

This study aims to gain insight into the facilitators and barriers that HCPs specialize in cardiac care experience in lifestyle support for the prevention and treatment of CVD and to investigate their views on eHealth tools. We performed in-depth interviews with HCPs to answer two main questions: (1) What factors are important in supporting CVD patients in the uptake of and adherence to a healthy lifestyle? (2) What are the (potential) facilitators of and barriers to eHealth tools in providing lifestyle support to patients with CVD?

Methods

Sample

We interviewed 16 Dutch HCPs (10 women) specializing in supporting patients with CVD and with experience in lifestyle change. To ensure a diverse and representative collection of perspectives, we included professionals with varying backgrounds from multiple institutions located in different parts of the Netherlands (Table 1). Professionals were selected based on eHealth affinity within the department or organization they worked in and asked how they applied eHealth in their own job to verify some level of eHealth experience. In line with these guidelines, 16 interviews would be sufficient for information saturation [35].

Table 1. Organization and professional background of respondents (N=16).

Organization and professional background	Respondents, n (%)
Academic hospital A	
Nurse practitioner working in cardiac rehabilitation	2 (12)
Academic hospital B	
Neurovascular nurse practitioner	1 (6)
Physician assistant specialized in cardiovascular risk factor management	1 (6)
Hospital A	
Physiotherapist working in cardiac rehabilitation	1 (6)
Nurse practitioner working in cardiac rehabilitation	1 (6)
Hospital B	
Physician-researcher working in cardiac rehabilitation	1 (6)
Nurse practitioner working in cardiac rehabilitation	1 (6)
Hospital C	
Neurologist specialized in cardiac rehabilitation	1 (6)
Nurse practitioner working in cardiac rehabilitation	1 (6)
Cardiac rehabilitation center A	
Cardiologist in residence	1 (6)
Lifestyle coach working in cardiac rehabilitation	1 (6)
Cardiac rehabilitation center B	
Physiotherapist working in cardiac rehabilitation	1 (6)
Cardiac rehabilitation center C	
Psychologist specialized in cardiac rehabilitation	1 (6)
General practice center A	
General practitioner specialized in CVD[a] care	1 (6)
Nurse practitioner working in cardiac rehabilitation	1 (6)

[a]CVD: cardiovascular disease.

Procedure

We used convenience sampling and approached organizations within the network of care partners. We asked for professionals within the organization who were most directly involved with lifestyle support of patients with CVD, whom we sent an email with general information about the research goals. After the HCP expressed willingness to cooperate, interview appointments were made by phone and performed at the interviewee's preferred location. Before the start of the interview, information about the research project and the goals and methods of the interview were provided. The interviewee signed the informed consent form, after which the voice recording and interview started. No (financial) compensation was offered to participate.

The interviews were conducted in Dutch, between November 2017 and February 2018, and took 45-90 minutes. One researcher led the interview, whereas another took notes, and roles were alternated between each interview (DRDB and JEVDG). The voice recordings were transcribed and pseudonymized to secure anonymity (Brecht Otto and Pauline van Wolferen). We used a general interview guide approach, as the interviews were based on a semistructured list of questions that allowed for further elaboration based on answers. The questions were divided into six topics (Textbox 1). We asked about both the use of eHealth (digital tools to provide automated or remote support with interactive web-based components) and wearables and sensors (eg, pedometers). Interview topics were defined based on the research questions and assessed whether they would provide answers to these questions. We also included questions about the interviewees' profiles (eg, job description and experience with eHealth). This study only discusses the data from interview topics 1-5, which are relevant to our specific research questions. The data regarding the sixth interview topic (Reward Program to Promote Healthy Living) do not belong to the scope of this study and are used in another publication.

Textbox 1. Interview guide.

Facilitating and impeding factors in the uptake of and adherence to a healthy lifestyle for patients with cardiovascular disease

- What do cardiovascular disease patients need to do in their home environment to achieve sustainable lifestyle change?
- What things that seem to work well for cardiovascular disease patients in changing their lifestyle?
- What impedes cardiovascular disease patients in changing their lifestyle?
- What solutions do cardiovascular disease patients have for these barriers?

Facilitating and impeding factors in providing lifestyle support to patients with cardiovascular disease

- How do you provide lifestyle support to cardiovascular disease patients?
- What works well in providing lifestyle support?
- What impedes providing lifestyle support?
- What solutions do you have for these barriers?

Stakeholders involved in providing lifestyle support to patients with cardiovascular disease

- What do you, as a health care professional, need to better provide lifestyle support to cardiovascular disease patients?
- With whom do you cooperate in providing lifestyle support to cardiovascular disease patients?

Facilitating and impeding factors in using eHealth to provide lifestyle support to patients with cardiovascular disease

- What things go well in your use of eHealth to provide lifestyle support to cardiovascular disease patients?
- What impedes your use of eHealth to provide lifestyle support to cardiovascular disease patients?
- What solutions do you have for these barriers?
- What do you, as a health care professional, need to better make use of eHealth to provide lifestyle support to cardiovascular disease patients?

Facilitating and impeding factors in using wearables and sensors to provide lifestyle support to patients with cardiovascular disease

- To what extent do you use wearables and sensors to provide lifestyle support to cardiovascular disease patients?
- What things go well in your use of wearables and sensors to provide lifestyle support to cardiovascular disease patients?
- What impedes your use of wearables and sensors to provide lifestyle support to cardiovascular disease patients?
- What solutions do you have for these barriers?

Analyses

The transcripts were sorted into meaningful clusters based on a content analysis approach to ensure that insights emerged based on the data [36]. Relevant pieces of data were retrieved from the text and coded and categorized into themes. For each of the transcripts, 2 researchers (DRDB and JEVDG) independently marked quotations in Microsoft Word containing relevant information. These quotations were compared, and a consensus document for each transcript was created. The quotations were transferred to Microsoft Excel, coded in a separate column to allow for interpretation, color coded to indicate whether the quotation was related to eHealth, and subsequently categorized into themes (TRCR). In discussion with a second researcher (MK), a definitive set of 13 themes emerged on which each of the quotations were fit (Textbox 2). Consensus with an independent coder (Magali de Rooy) was reached at once, with an interrater agreement of 74% and sufficient interrater reliability (Krippendorff α=.697), which indicated that the developed list of themes adequately represented the structure of the data. Quotation examples in text were translated into English by 2 researchers (TRCR and DRDB).

Textbox 2. Identified themes after coding.

Intervention-related factors

1. Autonomy

- Factors that concern the extent to which the patient has the freedom to make decisions about lifestyle change for themselves
- Quotations that concern the feeling of control in the process, the amount of self-determination, and insight into one's own health

2. Goal setting

- Factors that are related to setting goals in lifestyle change
- Quotations concerning the quantity, content, and design of these goals

3. Personalization

- Factors that are related to the adjustment of a healthy lifestyle and revalidation program to the needs and wishes of the patient
- Quotations that concern the personal relevance, feasibility, and attractiveness of the revalidation process

Patient-related factors

4. Motivation

- Factors that facilitate or impede the willpower to start or maintain lifestyle change
- Quotations that concern the extent to which patients are willing to work on their lifestyle and their intrinsic and extrinsic motivation

5. Condition of the patient

- Physical, mental, or cognitive impairments that impede the patient in the uptake of and adherence to a healthy lifestyle (eg, pain, depression, stress, addictions, and age)
- Both conditions that already existed and those because of their illness

6. Psychological characteristics

- Characteristics and traits of the patient that facilitate or impede the uptake of and adherence to a healthy lifestyle
- Quotations that concern personality or personal predispositions of the patient (eg, self-efficacy, resistance, and sense of responsibility)

7. Environmental factors

- Factors in the home environment and daily life of the patient that facilitate or impede the uptake of and adherence to a healthy lifestyle
- Quotations that concern the direct surroundings of the patient, which one cannot control (difficult domestic situations, socioeconomic status, and access to health or unhealthy options)

8. Social network

- Factors in the social circle of the patient that facilitate or impede the uptake of and adherence to a healthy lifestyle
- Quotations that concern the role of friends, family, and acquaintances in the patient's lifestyle

Health care–related factors

9. Format of professional support

- Factors that determine the way in which support of the patient is shaped and structured and facilitate or impede the uptake of and adherence to a healthy lifestyle
- Quotations that concern the implementation, frequency, and format of support

10. Relationship with the patient

- Factors that are related to the personal relationship between health care professional and patient
- Quotations that indicate the way in which such a relationship is established and what it should entail

11. Continuity of professional support

- Factors that are related to long-term support of the patient and facilitate or impede maintaining a healthy lifestyle
- Quotations that concern lifestyle change in the long run, outside the health care environment, and continuing the revalidation process by the patient

12. Organization of care

- Practical factors that influence the provided health care, both physical facilities (eg, health care professional's practice) and nontangible influences (eg, regulations, finances) that facilitate or impede lifestyle support
- Quotations that concern the availability of care and the extent to which health care professionals' can do their job and the way they are ought to do

eHealth-related factors

13. Barriers to eHealth

- Factors that are related to the implementation of eHealth (digital tools) in lifestyle support
- Quotations that concern the difficulties in using and implementation of technology and data in the current health care system

Results

Barriers to and Facilitators of Lifestyle Support

Of the 13 identified themes, 12 concerned lifestyle support in general (Textbox 2). The subjects of these themes were related to the intervention, the patient, or health care in general.

Intervention-Related Factors

Autonomy

Nearly all (15/16, 94%) HCPs mentioned that patients need to feel a sense of ownership over their lifestyle change process instead of being just another patient undergoing rehabilitation. One HCP was especially concerned about the lack of choice in cardiac rehabilitation:

> *People are forced to do so many things, they end up in an obligatory trajectory. That is already quite a lot. So I think that can be a barrier.* [Quote 89, HCP 8]

Self-monitoring (eg, heartbeat or weight) and information about both their disease and the benefits of a healthy lifestyle were mentioned by 10 HCPs to be essential for patients to feel a sense of control. This allows them to act independently of their HCPs when they notice irregularities:

> *It is also important for patients that they get more insight themselves...That they can alert us whenever they are training independently and say "my heartrate shows irregularities or is not going up.*" [Quote 73, HCP 1]

However, 3 HCPs mentioned that self-monitoring might have the downside of becoming an obsession, as people could fixate on numbers rather than their own body.

Goal Setting

More than half of the HCPs (9/16, 56%) mentioned the importance of goal setting in a healthy lifestyle. Patients would reach the most success when the number of goals at a given time is limited; when the goals are formulated in a specific, measurable, acceptable, realistic, and timely way; and when the goals are personally relevant for the patient. Accomplishment of these goals provides a rewarding feeling, which increases motivation to continue:

> *I want them to create their own success story...I choose something* [a goal] of which I guess that person will be able to achieve in the upcoming week. And that turns into motivation...[Quote 133, HCP 8]

Personalization

Of all HCPs, 56% (9/16) experienced that a lifestyle intervention will succeed when the provided support is tailored to patients' needs, capabilities, and preferences. For instance, for some, it is more important to work on their eating habits, whereas for others, an increase in physical activity is more relevant. At the same time, HCPs mentioned difficulties in finding out what their patients actually wanted and needed, which made it challenging to individualize the program:

> *And I'd really like to get to know the person on the other side of the table, what kind of information that person would like to receive. I find it hard to know: How would someone like to be motivated.* [Quote 148, HCP 4]

Patient-Related Factors

Motivation

A key theme throughout the interviews (14/16, 88%) was the level of motivation of the patient. Intrinsic motivation was deemed essential to successfully complete—or even start—a lifestyle program. Such intrinsic motivation is not always self-evident because of low awareness about the current and future health impact of an unhealthy lifestyle. About 38% (6/16) of HCPs mentioned how the occurrence of the disease acts as the tipping point for patients to change their lifestyle:

> *People already know that they are unhealthy and that they should make changes. Often you will notice that such a crisis causes them to actually do so.* [Quote 54, HCP 13]

To maintain the level of motivation, after the initial scare from the incident has passed, 56% (9/16) of HCPs mentioned that patients need to see progress of their effort, preferably through tangible results (eg, increased performance durations). Extrinsic motivation, in the form of both material and nonphysical incentives (eg, positive feedback), was mentioned by 44% (7/16) of HCPs to play a role:

> *Rewards are on multiple levels, a reward can also be that you are just being noticed by your significant other, brother, sister, friend. But it can also be a more literal reward, you know, that you buy something for*

yourself. Or that you tell yourself, well done. [Quote 57, HCP 13]

Condition of the Patient

Health-related issues hindering patients from initiating or maintaining a healthy lifestyle were mentioned by 75% (12/16) of HCPs. These issues are physical, cognitive, or mental and are either pre-existing or because of cardiac incidents. For example, reduced mobility in older patients is a physical barrier to physical activities or reaching the clinic. Frequently mentioned mental barriers were depressive symptoms and fear, such as concerns about physical capabilities after a cardiac incident:

Especially people who experience persistent heart complaints, that cause a lot of anxiety, they think: I won't push myself. When I start exercising, I will experience it again. [Quote 177, HCP 4]

Psychological Characteristics

Most HCPs (13/16, 81%) mentioned the role of their patients' personalities either as facilitators or as barriers. Patients need to be disciplined, and most importantly, some level of self-awareness helps to reflect on their own behavior and acknowledge their own role in the process. Patients who come up with excuses for not performing healthy behaviors are most difficult to work with:

But there is also a big group of people who are just very resistant to change, who are mainly externalizing and say: "I can't do anything about it." Or who continuously come up with excuses about why things can't change. Yes, that is the most difficult group to work with. That is also the most unhealthy group. [Quote 215, HCP 15]

In addition to personality characteristics, another frequently mentioned barrier was previously developed bad habits.

Environmental Factors

Factors related to the daily environment of the patients were identified by 56% (9/16) of HCPs. For instance, difficult domestic situations are often given more priority and can therefore reduce the success of a healthy lifestyle initiated in the clinic. Some HCPs (4/16, 25%) explicitly stated that socioeconomic status (eg, language barriers) affects people's lifestyles:

When it comes to handing out flyers as well, I come across situations such as: "I can't read." Not very frequently, but it happens every now and then. [Quote 230, HCP 10]

According to 25% (4/16) of HCPs, government authorities should take responsibility for creating a healthy environment (eg, offering healthy food in hospitals, stricter tobacco and alcohol regulations, or regulating the prices of food) and providing health education.

Social Network

The roles of both close (family and friends) and distant others were mentioned by 50% (8/16) of HCPs. Other people function as social controls or exert some level of group pressure. A sense of cohesion through engaging in healthy activities with others is a great motivator:

...an exemplary role, sociability, a social aspect, controlling aspect, when you are part of a group people will ask about you: How are you doing, where were you? All those kind of things play a role. [Quote 246, HCP 1]

The importance of the social network of patients in providing practical and psychological support was emphasized by 38% (6/16) of HCPs. However, HCPs worried that overly critical family members or friends could also negatively influence the process. A second concern was social norms, as some unhealthy behaviors (such as drinking too much alcohol) are less socially accepted and therefore more difficult for patients to be open about:

Well, the subject is more of a taboo. It is automatically an issue. When you drink too much, you are an alcoholic. Eating too much, well, that happens to all of us. That we are snacking a little too much. [Quote 256, HCP 8]

Health Care–Related Factors

Format of Professional Support

Considering the way support should be provided, 50% (8/16) of the HCPs mentioned the importance of frequency. Through frequent repetition of information, healthy behaviors by the patient, and reminders or feedback, a healthy lifestyle remains a topic of interest. However, the frequency of consultations in current practice is too low to do so. A total of 31% (5/16) of HCPs mentioned that support should be accessible at all times whenever the patient needs it. In addition, to provide tailored support, HCPs need data independent of the patients' self-reports about their progress:

...we have tried to use a logbook, but a pedometer can track the walking process outside. You can respond to the objective information you receive. A logbook is just an estimate, you just have to believe that it's true. [Quote 276, HCP 2]

Most HCPs (10/16, 63%) found education as an important part of the intervention as patients lack knowledge or have misconceptions about their disease and a healthy lifestyle. Therefore, the health care system should play a role in offering trustworthy information, providing patients with concise pieces of information that are easy to understand:

...there is so much information available that they have no idea what to trust..., especially among the older population who have more respect for healthcare professionals, simple advice is really appreciated. [Quote 288, HCP 4]

Relationship With the Patient

More than half of the HCPs (11/16, 69%) mentioned that their relationship with patients has a significant influence on the process. A good relationship helps in understanding the underlying reasons for patients' behavior and motivation and creating a safe environment to share their feelings. Support does

not end at the physical aspect of cardiac rehabilitation but entails mental support as well:

> *They suddenly are obliged to change a lot of things. I try to focus less on things that have to change, but acknowledge how it affects them...Therefrom, they will more easily comply with a lifestyle change in the end.* [Quote 334, HCP 8]

HCPs disagreed about their role as an authority figure. An equal relationship, in which they co-operated with their patients during the revalidation process, was frequently mentioned. However, 19% (3/16) of HCPs recognized that they function as the so-called *big stick* to keep patients on the right track.

Continuity of Professional Support

According to 63% (10/16) of HCPs, long-term support is crucial for maintaining a healthy lifestyle outside the health care environment. When the window of opportunity after a cardiovascular incident disappears, patients are more likely to return to old (unhealthy) habits. However, 44% (7/16) of HCPs mentioned lack of follow-ups or an end evaluation, leaving them with no ability to provide long-term feedback or information about the postrehabilitation success of the lifestyle intervention:

> *...when you want someone to follow through with the lifestyle change, you do have to check whether someone comprehends it and if is able to do so. When you let someone on their own, you will lose that person.* [Quote 353, HCP 4]

Organization of Care

All HCPs (16/16, 100%) mentioned at least 1 factor related to the way health care is organized, varying from physical facilities to nontangible influences. Most HCPs (11/16, 69%) mentioned a lack of financial resources and time. Furthermore, the limited co-operation between HCPs involved in cardiac care and those of other health care disciplines was mentioned as a barrier. There are few options for redirecting patients to another professional, disagreements within the extensive group of care providers, and insufficient information sharing between them:

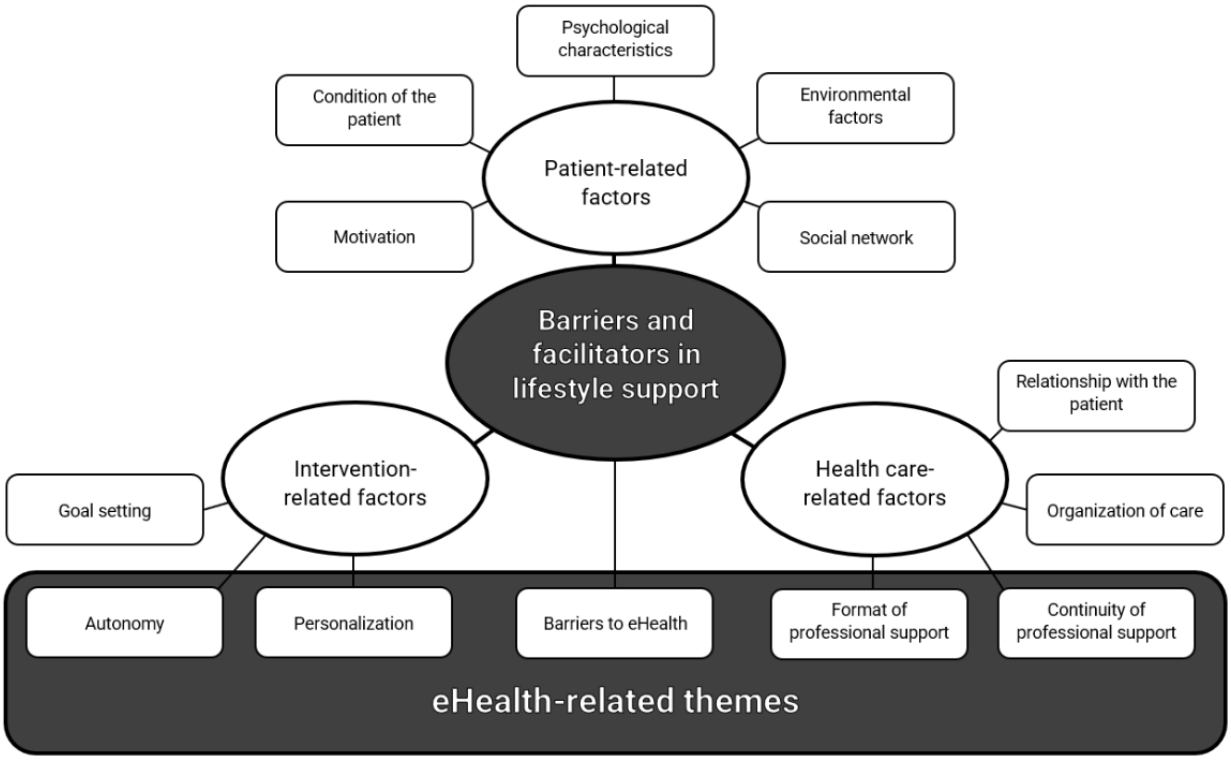
> *The dietician for instance, she keeps her own records, she can't attend the multidisciplinary team meeting due to her planning. I think that is a barrier as well, because we once had a patient who didn't consume enough calories, which caused problems and that person did not feel well...[Quote 393, HCP 6]*

More than half of the HCPs (7/16, 44%) mentioned a lack of attention for lifestyle within cardiac care or health care in general, although this varied greatly between domains (eg, cardiology and neurology). HCPs indicated that they missed regulations and protocols they could follow in providing lifestyle support.

eHealth in Lifestyle Support

Throughout the 12 themes discussed earlier, the use of eHealth reoccurred as a (potential) facilitator or solution to barriers, most prominently within the themes Autonomy, Personalization, Format of support, and Continuity of professional support (Figure 1). Although eHealth facilitators were more strongly related to lifestyle support, the barriers HCPs experienced were rather unrelated to lifestyle themes. To put more emphasis on this, the 13th theme concerning barriers in the implementation of eHealth will be discussed in the final part of this section (*Barriers to eHealth*).

Figure 1. Overview of the identified barriers and facilitators in lifestyle support and their relation to eHealth.



Benefits of eHealth

During analyses, a link between barriers to and facilitators of lifestyle support and potential facilitating benefits of eHealth emerged. The HCPs provided examples of how eHealth could help them.

Autonomy in eHealth

For patients to regain autonomy (theme 1), HCPs indicated a need for education and insight and saw an opportunity for eHealth to provide both. HCPs noticed that by giving patients the opportunity (and therefore the responsibility) of monitoring their own health through a digital tool, they can act whenever necessary (eg, adjusting their diet when they notice a higher blood pressure). According to HCPs, such insights would also enlarge awareness about unhealthy behavior and provide progress feedback to increase their motivation:

> *We offer cardiac patients who we follow-up via eHealth a pedometer, a digital scale that measures body fat percentages, a blood pressure monitor and a device to make an ECG. That provides them with insight into how they are doing.* [Quote 107, HCP 14]

Furthermore, as eHealth can be individually tailored, HCPs indicated that patients can control what information they receive and how they receive it. Patients can also work on their lifestyle at any time and in any way they wish to, increasing the level of self-management.

Personalization in eHealth

Personalization (theme 3) appeared to be key to lifestyle interventions, but HCPs raised the issue of identifying what patients wanted and needed. They thought that eHealth could help them to get more information about patients and their needs, before and during the intervention (eg, through web-based intake questionnaires). HCPs believed this could result in better adjustment of their support and more efficient consultations and lifestyle interventions:

> *In fact, even before someone comes in, you would have to start with: "This is the goal of the consultation." Based on a test or questionnaire, you look at how someone can best be approached: What kind of advice do you prefer? There are probably apps, tools, questionnaires, and other things that can do so.* [Quote 150, HCP 4]

Format of Support in eHealth

Related to the format of support (theme 9), HCPs gave examples of how eHealth, especially its possibilities for remote support, could be beneficial. As digital tools are available at all times and not bound to a physical location, HCPs predicted that it would be much easier for patients to frequently have contact, work on their lifestyle, or receive information. This would also increase the accessibility of support:

> *That is the advantage of eHealth, that it is flexible, 24/7, which is really convenient. I think that a lot of people ruminate at night and would appreciate to write during nighttime. The possibility to do so at that very moment, not only when you meet your coach again. Then it has already faded away.* [Quote 305, HCP 8]

Consistent and automatic digital monitoring would provide more objective data, meaning HCPs would no longer have to rely on single measures during consultations or error-prone self-report measures. Furthermore, eHealth could be used as an educational platform, which HCPs thought they could use to provide patients with reliable and consistent information about their disease and lifestyle.

Continuity of Professional Support in eHealth

Continuity of professional support (theme 11) was mentioned as one of the biggest issues in current cardiac care. Therefore, most HCPs saw prolonged monitoring as a huge advantage of eHealth. Furthermore, this could enable HCPs to provide support in the long run, once patients return to their everyday lives:

> *There are gaps within the healthcare system, which makes it difficult for patients to continue independently. That is where this eCoach steps in. So during cardiac rehabilitation over here, they see the physiotherapist, they see the doctor, they can chat more easily through the portal.* [Quote 359, HCP 6]

Barriers to eHealth

While recognizing these potential advantages of eHealth, HCPs raised some barriers concerning its adoption and usability. The most prominent concern of 63% (10/16) of HCPs was related to the general old age of patients with CVD, as older people are more likely to have little experience with or no interest in technology. Moreover, patients would generally prefer face-to-face contact over digital communication, either during the entire intervention or at least a part of it. One HCP explicitly mentioned the importance of face-to-face intake for a digital intervention to be successful:

> *They tell me: "Oh, I received a mail from online coaching, but I have already so much on my mind, so I just ignored it."...But then they see me and say: "But now I know that it was you, that is nice!" It comes to life for them, in my experience at least.* [Quote 466, HCP 8]

Difficulties with technological tools and devices, such as bugs and slow development of the technology, were mentioned by 31% (5/16) of HCPs. In addition, they mentioned that there was no help desk for patients or HCPs. Furthermore, it was frequently mentioned by 31% (5/16) of HCPs that many current technological aids suffer from a low level of user-friendliness.

Discussion

Principal Findings

This study aimed to gain insights into the facilitators and barriers that HCPs experience in lifestyle support for the prevention and treatment of CVDs and investigate their views on potential eHealth tools. We interviewed 16 HCPs, resulting in 12 themes relevant to lifestyle support, of which four were related to eHealth. The 13th theme was related to eHealth barriers.

Barriers to and Facilitators of Lifestyle Support

First, we aimed to identify the factors that HCPs find important in supporting patients with CVD in the uptake of and adherence to a healthy lifestyle. We found factors related to the intervention, patient, and health care system to help answer this question.

According to the HCPs, a lifestyle intervention should give patients a feeling of autonomy and possibilities for goal setting and allow for personalization. In line with our findings, in interview studies on lifestyle support for patients with diabetes, HCPs indicated that well-formulated goals create realistic patient expectations [15] and that standardized norms should be adjusted to patients' capabilities [26]. Furthermore, HCPs indicated that it is no longer their role to tell patients to change their lifestyle but rather the patient's responsibility [26]. However, although HCPs in this study named autonomy as an additional facilitator within lifestyle change, other studies reported patient responsibility to be a basic necessity because of low patient motivation [26] or even seem unrealistic as patients are not always able to independently start or maintain a healthy lifestyle [14].

With regard to patient factors, motivation to live healthily, the condition of the patient, psychological characteristics, environmental factors, and social networking were mentioned to be of influence within lifestyle support. HCPs working with people with (a high risk of) CVD [9,11,27], patients with diabetes [15,26,28], and chronic diseases in general [29] recognized similar factors, thereby suggesting that these are relevant within different patient populations. However, although HCPs in our study thought that little awareness of the impact of an unhealthy lifestyle on health contributed to a low level of patient motivation for change, a study with primary care HCPs reported that limited knowledge about risks of CVD is only seldom a barrier for engaging in lifestyle modification [14]. This discrepancy in results might be because of methodological differences, as the study by Jallinoja et al [14] included primary care HCPs and not HCPs mostly working in cardiac rehabilitation. Furthermore, primary care HCPs were asked about the relevance of insufficient knowledge to treatment and not lifestyle change per se. In addition, factors reported by patients with CVD themselves are relatively similar to those found in our study [37], which suggests that, at least in part, HCPs are able to recognize what patients need in lifestyle interventions.

Finally, several factors related to health care in general were mentioned, including the format of the provided support, continuity of professional support, the way care is organized, and the relationship between the HCP and patient. A high-quality relationship with the patient was also recognized as a facilitator within lifestyle support in other studies, as it would lead to both more collaborative patients and more motivated HCPs [30,34]. In addition, it would be easier to foster face-to-face encounters [30]. Similar to our results, the lack of time, little governmental responsibility, financial shortcomings, little co-operation between HCPs, and difficulties in referring patients were mentioned as barriers by HCPs involved in the prevention of CVD [11,16,27], type 2 diabetes [15,26], or in (chronic) diseases in general [29,38]. This shows that such barriers are not unique for lifestyle support in CVD rehabilitation, which provides HCPs and researchers with the opportunity to learn from other disciplines and work together to find solutions (eg, eHealth tools).

Other studies reported a lack of skills by HCPs to provide lifestyle support or a feeling that lifestyle interventions are ineffective as a barrier to the provision of lifestyle support [14,15,29,38]. These factors were not mentioned in this study, which might be owing to the nature of our sample that included HCPs who were specifically involved in lifestyle support and therefore might have a bigger skill set for and a more positive attitude toward providing lifestyle support.

Barriers to and Facilitators of eHealth

Second, to determine what the (potential) facilitators of and barriers to eHealth tools would be in providing lifestyle support to patients with CVD, the interviewed HCPs described how eHealth could be applied to strengthen facilitators or solve barriers they encountered in lifestyle support. The statements that HCPs made concerning facilitators of eHealth were related to the intervention-related factors, Autonomy and Personalization. These advantages of eHealth have also been recognized by HCPs in other studies. Macdonald et al [34] reported that HCPs acknowledged that eHealth fosters the *two-way conversation*—a collaborative interaction between patients and HCPs, which explains why eHealth can create well-informed and autonomous patients. As HCPs previously indicated that lifestyle is the responsibility of the patient [11,19], eHealth could offer them tools that foster the patient's autonomy. HCPs from other studies also indicated that eHealth helps them to personalize the program by getting to know their patients' needs through the personal diary within the digital portal [39] and that personalization of an eHealth program is essential to fit the patient's capabilities [40]. Furthermore, meta-analyses have demonstrated a positive relationship between both an autonomy-supportive health care climate and personalization of digital intervention content and successful behavior change [41,42].

With regard to health care–related factors, we found that the Format of professional support and the Continuity of support were important topics related to eHealth. Other studies have reported similar advantages of eHealth. Brandt et al [30] reported HCPs indicated that, because of its format, eHealth provides them with objective and measurable information and that it is not bound to a specific location or moment in time. In addition, some HCPs appreciated being able to follow-up their patients for a longer period, as it can be rewarding and increases their motivation and sense of responsibility to continue providing support [40]. Although we did not find a link between eHealth and the HCP-patient relationship, other studies have reported contradictory findings. Das et al [39] reported that eHealth does not have time constraints, shame, and fear of stigma, which leads to more self-disclosure from patients. However, Brandt et al [30] reported that HCPs indicated it is more challenging to establish an empathic relationship in a digital environment. This contrast might be because HCPs seem positive about tools that are an addition to face-to-face contact

[39], but those that replace face-to-face interactions are perceived as less favorable to build a supportive relationship [30]. Furthermore, although we did not find the advantages of eHealth in the organization of care, other studies did. For example, other studies mentioned additional time by reusing old advice [30], co-operation between HCPs, and accessible alternatives to refer their patients to [32] as advantages of eHealth. Methodological differences related to the different care settings and organizational structures the interviewed HCPs worked in could explain this.

Despite the advantages that were recognized by HCPs from both our and other studies [30,32,34,39], there is a low level of acceptance and implementation of eHealth in health care [23]. HCPs in this study formulated several barriers that could offer an explanation. First, HCPs feel that because patients with CVD are older, they prefer face-to-face contact and have little technological experience; therefore, digital tools would not be suitable for this patient population. HCPs in another study made a distinction between current patients with CVD and future ones, as the latter will have substantially more experience in and affinity to technology [40]. In addition, the eHealth and face-to-face support preferences of patients with CVD vary greatly [43], which raises concerns about uneven eHealth adoption and unequal health benefits [44]. HCPs could possibly contribute to this, as the views and preferences of patients are important in their decision to use eHealth [40]. At the same time, Grünloh et al [45] suggested that some HCPs seem to be unaware of the development of patient skills and knowledge over time. This could mean that once HCPs believe a patient is a technology-averse person, there will be minimal attempts to help the patient become acquainted with eHealth. eHealth acceptance could also be influenced by preference for face-to-face communication of HCPs themselves [30,32], which could be because of concerns regarding the therapeutic alliance with their patients [40]. However, others do not experience this issue, as they use eHealth for information sharing (eg, educational texts) rather than communication purposes (eg, interacting with patients) [31].

Other barriers identified concerned eHealth apps themselves, such as bugs and the slow development of digital tools, the lack of a help desk, and a low level of user-friendliness for both patients and HCPs. Other studies mentioned similar concerns, such as limited innovation, being offered digital tools that were still under development, and digital information that is too difficult to interpret and translate into support for their patients [33,34]. A study on an eHealth tool evaluation showed that, in hindsight, HCPs have specific wishes concerning the utility and design of such tools [32]. If they were included in the development process at an early stage, such barriers could have been prevented [21,24].

In this study, barriers to eHealth were not related to organizational factors. However, HCPs have previously shown concern about the inflexibility of the health care system and indicated that organizational structures and attitudes of HCPs have a major impact on eHealth acceptance and implementation [21,39,40,44]. The lack of financial compensation also played a role according to HCPs from previous studies [21,22]. Therefore, apart from barriers experienced by individual HCPs, overcoming structural obstacles seems necessary for the implementation of eHealth in health care [46]. As many HCPs in our sample were already working with eHealth tools, they might have experienced fewer organizational difficulties and, therefore, did not mention such organizational barriers.

To make eHealth implementation more successful in practice, the results of this study suggest that HCPs do not need to be convinced about the benefits of eHealth but rather that the barriers they experience should be resolved. To overcome these barriers, health policies could play an important role in the provision of support and equipment. This way, HCPs would be able to implement the reported benefits of eHealth in lifestyle support for people with CVD.

Limitations and Future Studies

First, our results were based on the opinions and interpretations of HCPs and not on the actual views of the patients themselves, who might have an alternative view on how eHealth can support them. Future studies could therefore conduct interviews with both HCPs and their patients to compare their views and attitudes toward lifestyle support and the use of eHealth.

Although we intentionally interviewed health care HCPs involved in the lifestyle support of patients with CVD, this specific sample limits the generalizability of our results as our sample has experience with and might be more willing to provide lifestyle support, whereas other HCPs might be less inclined to. It would therefore be interesting to investigate how different levels of experiences with and attitudes toward lifestyle support and eHealth translate into differences in the barriers experienced by HCPs.

Finally, we did not explicitly ask about the association between facilitators of and barriers to lifestyle support and the use of eHealth as a possible solution. Future studies should therefore investigate how eHealth can help overcome barriers related to specific aspects of lifestyle support experienced in various health care settings. These results could provide eHealth developers with a better direction in the development of eHealth interventions.

Conclusions

This study provides insights into Dutch HCPs' views on lifestyle support and eHealth in cardiac care. We identified facilitators and barriers related to intervention-, patient-, and health care–related factors. HCPs in general showed high approval of lifestyle support for patients with CVD and identified the potential benefits of incorporating eHealth. However, the interviews also revealed several barriers that impede HCPs' use of eHealth in lifestyle support. Incorporating their needs and values in the development of lifestyle support programs, especially eHealth, could increase their use and lead to a more widespread adoption of eHealth into health care.

Acknowledgments

This work was supported by the Netherlands Cardiovascular Research Initiative: an initiative with support of the Dutch Heart Foundation, CVON2016-12 BENEFIT, ZonMw (The Netherlands Organization for Health Research and Development), and the members of the BENEFIT consortium. Furthermore, the authors would like to thank Brecht Otto (Bachelor Student Applied Psychology, University of Applied Sciences Leiden) and Pauline van Wolferen (Master Student Health Science, University Twente) for their help with transcribing the interviews and Magali de Rooy (Bachelor Student Psychology, Leiden University) for helping with coding of data.

Authors' Contributions

Conception and design were carried out by DRDB, MK, JEVDG, and AWME; data acquisition was performed by DRDB, MK, and JEVDG; data analysis and interpretation were carried out by TRCR, TR, MK, and AWME; drafting the manuscript was performed by TRCR, TR, and MK; manuscript revision was carried out by TRCR, DRDB, TR, MK, JEVDG, VRJ, RAK, DEA, and AWME. All authors gave final approval and agreed to be accountable for all aspects of the work, ensuring integrity and accuracy.

Conflicts of Interest

RAK is a shareholder in Vital10, CardioVitaal, and the Netherlands Institute for Prevention and E-Health Development (NIPED). The remaining authors have no conflicts to declare

References

1. Cardiovascular diseases (CVDs). World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) [accessed 2019-02-07]
2. De Smedt D, Clays E, Annemans L, Doyle F, Kotseva K, Pająk A, et al. Health related quality of life in coronary patients and its association with their cardiovascular risk profile: results from the EUROASPIRE III survey. Int J Cardiol 2013 Sep 30;168(2):898-903. [doi: 10.1016/j.ijcard.2012.10.053] [Medline: 23201081]
3. Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, ESC Scientific Document Group. 2016 European guidelines on cardiovascular disease prevention in clinical practice: The sixth joint task force on the European society of cardiology and other societies on cardiovascular disease prevention in clinical practice. Eur Heart J 2016 Dec 01;37(29):2315-2381 [FREE Full text] [doi: 10.1093/eurheartj/ehw106] [Medline: 27222591]
4. Wilkins E, Wilson L, Wickramasinghe K, Bhatnagar P, Leal J, Luengo-Fernandez R, et al. European Cardiovascular Disease Statistics 2017. European Heart Network. URL: https://ehnheart.org/cvd-statistics/cvd-statistics-2017.html [accessed 2019-08-13]
5. Janssen V, De Gucht V, Dusseldorp E, Maes S. Lifestyle modification programmes for patients with coronary heart disease: a systematic review and meta-analysis of randomized controlled trials. Eur J Prev Cardiol 2013 Aug;20(4):620-640. [doi: 10.1177/2047487312462824] [Medline: 23022703]
6. Multidisciplinaire Richtlijn Hartrevalidatie 2011. Revalidatiecommissie Nederlandse Vereniging Voor Cardiologie/Nederlandse Hartstichting. URL: https://www.nvvc.nl/Richtlijnen/Multidisciplinaire%20Richtlijn%20Hartrevalidatie%202011%2023052011.pdf [accessed 2020-10-15]
7. Iestra JA, Kromhout D, van der Schouw YT, Grobbee DE, Boshuizen HC, van Staveren WA. Effect size estimates of lifestyle and dietary changes on all-cause mortality in coronary artery disease patients: a systematic review. Circulation 2005 Aug 09;112(6):924-934. [doi: 10.1161/CIRCULATIONAHA.104.503995] [Medline: 16087812]
8. Barnes PM, Schoenborn CA. Trends in adults receiving a recommendation for exercise or other physical activity from a physician or other health professional. NCHS Data Brief 2012 Feb(86):1-8 [FREE Full text] [Medline: 22617014]
9. Noordman J, Verhaak P, van Dulmen S. Discussing patient's lifestyle choices in the consulting room: analysis of GP-patient consultations between 1975 and 2008. BMC Fam Pract 2010 Nov 09;11:87 [FREE Full text] [doi: 10.1186/1471-2296-11-87] [Medline: 21062427]
10. Milder IE, Blokstra A, de Groot J, van Dulmen S, Bemelmans WJ. Lifestyle counseling in hypertension-related visits--analysis of video-taped general practice visits. BMC Fam Pract 2008 Oct 14;9:58 [FREE Full text] [doi: 10.1186/1471-2296-9-58] [Medline: 18854020]
11. Hobbs FD, Erhardt L. Acceptance of guideline recommendations and perceived implementation of coronary heart disease prevention among primary care physicians in five European countries: the Reassessing European Attitudes about Cardiovascular Treatment (REACT) survey. Fam Pract 2002 Dec;19(6):596-604. [doi: 10.1093/fampra/19.6.596] [Medline: 12429661]
12. Kotseva K, De Backer G, De Bacquer D, Rydén L, Hoes A, Grobbee D, EUROASPIRE Investigators*. Lifestyle and impact on cardiovascular risk factor control in coronary patients across 27 countries: Results from the European Society of Cardiology ESC-EORP EUROASPIRE V registry. Eur J Prev Cardiol 2019 May;26(8):824-835. [doi: 10.1177/2047487318825350] [Medline: 30739508]

13. Brotons C, Björkelund C, Bulc M, Ciurana R, Godycki-Cwirko M, Jurgova E, et al. Prevention and health promotion in clinical practice: the views of general practitioners in Europe. Prev Med 2005 May;40(5):595-601. [doi: 10.1016/j.ypmed.2004.07.020] [Medline: 15749144]
14. Jallinoja P, Absetz P, Kuronen R, Nissinen A, Talja M, Uutela A, et al. The dilemma of patient responsibility for lifestyle change: perceptions among primary care physicians and nurses. Scand J Prim Health Care 2007 Dec;25(4):244-249 [FREE Full text] [doi: 10.1080/02813430701691778] [Medline: 17934984]
15. Jansink R, Braspenning J, van der Weijden T, Elwyn G, Grol R. Primary care nurses struggle with lifestyle counseling in diabetes care: a qualitative analysis. BMC Fam Pract 2010 May 25;11:41 [FREE Full text] [doi: 10.1186/1471-2296-11-41] [Medline: 20500841]
16. Riley R, Coghill N, Montgomery A, Feder G, Horwood J. Experiences of patients and healthcare professionals of NHS cardiovascular health checks: a qualitative study. J Public Health (Oxf) 2016 Sep;38(3):543-551 [FREE Full text] [doi: 10.1093/pubmed/fdv121] [Medline: 26408822]
17. De Luca L, Temporelli PL, Lucci D, Gonzini L, Riccio C, Colivicchi F, START Investigators. Current management and treatment of patients with stable coronary artery diseases presenting to cardiologists in different clinical contexts: A prospective, observational, nationwide study. Eur J Prev Cardiol 2018 Jan;25(1):43-53. [doi: 10.1177/2047487317740663] [Medline: 29124952]
18. Barak A, Klein B, Proudfoot JG. Defining internet-supported therapeutic interventions. Ann Behav Med 2009 Aug;38(1):4-17. [doi: 10.1007/s12160-009-9130-7] [Medline: 19787305]
19. Lunde P, Nilsson BB, Bergland A, Kværner KJ, Bye A. The effectiveness of smartphone apps for lifestyle improvement in noncommunicable diseases: systematic review and meta-analyses. J Med Internet Res 2018 May 04;20(5):e162 [FREE Full text] [doi: 10.2196/jmir.9751] [Medline: 29728346]
20. Beishuizen CR, Stephan BC, van Gool WA, Brayne C, Peters RJ, Andrieu S, et al. Web-based interventions targeting cardiovascular risk factors in middle-aged and older people: a systematic review and meta-analysis. J Med Internet Res 2016 Mar 11;18(3):e55 [FREE Full text] [doi: 10.2196/jmir.5218] [Medline: 26968879]
21. Bally EL, Cesuroglu T. Toward integration of mHealth in primary care in the Netherlands: a qualitative analysis of stakeholder perspectives. Front Public Health 2019;7:407 [FREE Full text] [doi: 10.3389/fpubh.2019.00407] [Medline: 32010660]
22. Peeters JM, Krijgsman JW, Brabers AE, Jong JD, Friele RD. Use and uptake of eHealth in general practice: a cross-sectional survey and focus group study among health care users and general practitioners. JMIR Med Inform 2016 Apr 06;4(2):e11 [FREE Full text] [doi: 10.2196/medinform.4515] [Medline: 27052805]
23. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
24. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
25. Keesman M, Janssen V, Kemps H, Hollander M, Reimer WS, Gemert-Pijnen LV, BENEFIT Consortium. BENEFIT for all: an ecosystem to facilitate sustained healthy living and reduce the burden of cardiovascular disease. Eur J Prev Cardiol 2019 Apr;26(6):606-608 [FREE Full text] [doi: 10.1177/2047487318816388] [Medline: 30486650]
26. Stuij M. 'Physical activity, that's a tricky subject.' Experiences of health care professionals with physical activity in type 2 diabetes care. BMC Health Serv Res 2018 Apr 23;18(1):297 [FREE Full text] [doi: 10.1186/s12913-018-3102-1] [Medline: 29685135]
27. Tonnon SC, Proper KI, van der Ploeg HP, Westerman MJ, Sijbesma E, van der Beek AJ. A qualitative study of the anticipated barriers and facilitators to the implementation of a lifestyle intervention in the Dutch construction industry. BMC Public Health 2014 Dec 23;14:1317 [FREE Full text] [doi: 10.1186/1471-2458-14-1317] [Medline: 25539630]
28. Raaijmakers LG, Hamers FJ, Martens MK, Bagchus C, de Vries NK, Kremers SP. Perceived facilitators and barriers in diabetes care: a qualitative study among health care professionals in the Netherlands. BMC Fam Pract 2013 Aug 10;14:114 [FREE Full text] [doi: 10.1186/1471-2296-14-114] [Medline: 23937325]
29. Geense WW, van de Glind IM, Visscher TL, van Achterberg T. Barriers, facilitators and attitudes influencing health promotion activities in general practice: an explorative pilot study. BMC Fam Pract 2013 Feb 09;14:20 [FREE Full text] [doi: 10.1186/1471-2296-14-20] [Medline: 23394162]
30. Brandt CJ, Søgaard GI, Clemensen J, Søndergaard J, Nielsen JB. Determinants of successful ehealth coaching for consumer lifestyle changes: qualitative interview study among health care professionals. J Med Internet Res 2018 Jul 05;20(7):e237. [doi: 10.2196/jmir.9791] [Medline: 29980496]
31. Brandt CJ, Søgaard GI, Clemensen J, Sndergaard J, Nielsen JB. General practitioners' perspective on eHealth and lifestyle change: qualitative interview study. JMIR Mhealth Uhealth 2018 Apr 17;6(4):e88 [FREE Full text] [doi: 10.2196/mhealth.8988] [Medline: 29666045]
32. Smith E, Bradbury K, Scott L, Steele M, Little P, Yardley L. Providing online weight management in primary care: a mixed methods process evaluation of healthcare practitioners' experiences of using and supporting patients using POWeR. Implement Sci 2017 Dec 25;12(1):69 [FREE Full text] [doi: 10.1186/s13012-017-0596-6] [Medline: 28545538]

33. Bellicha A, Macé S, Oppert J. Prescribing of electronic activity monitors in cardiometabolic diseases: qualitative interview-based study. J Med Internet Res 2017 Sep 23;19(9):e328 [FREE Full text] [doi: 10.2196/jmir.8107] [Medline: 28947415]
34. Macdonald GG, Townsend AF, Adam P, Li LC, Kerr S, McDonald M, et al. eHealth technologies, multimorbidity, and the office visit: qualitative interview study on the perspectives of physicians and nurses. J Med Internet Res 2018 Dec 26;20(1):e31 [FREE Full text] [doi: 10.2196/jmir.8983] [Medline: 29374004]
35. Guest G, Bunce A, Johnson L. How many interviews are enough?: an experiment with data saturation and variability. Field Methods 2006 Feb 01;18(1):59-82. [doi: 10.1177/1525822X05279903]
36. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
37. Murray J, Craigs CL, Hill KM, Honey S, House A. A systematic review of patient reported factors associated with uptake and completion of cardiovascular lifestyle behaviour change. BMC Cardiovasc Disord 2012 Dec 08;12:120 [FREE Full text] [doi: 10.1186/1471-2261-12-120] [Medline: 23216627]
38. Brotons C, Björkelund C, Bulc M, Ciurana R, Godycki-Cwirko M, Jurgova E, et al. Prevention and health promotion in clinical practice: the views of general practitioners in Europe. Prev Med 2005 May;40(5):595-601. [doi: 10.1016/j.ypmed.2004.07.020] [Medline: 15749144]
39. Das A, Faxvaag A, Svanæs D. The impact of an ehealth portal on health care professionals' interaction with patients: qualitative study. J Med Internet Res 2015 Nov 24;17(11):e267 [FREE Full text] [doi: 10.2196/jmir.4950] [Medline: 26601678]
40. Walsh DM, Moran K, Cornelissen V, Buys R, Cornelis N, Woods C. Electronic health physical activity behavior change intervention to self-manage cardiovascular disease: qualitative exploration of patient and health professional requirements. J Med Internet Res 2018 May 08;20(5):e163 [FREE Full text] [doi: 10.2196/jmir.9181] [Medline: 29739740]
41. Ng JY, Ntoumanis N, Thøgersen-Ntoumani C, Deci EL, Ryan RM, Duda JL, et al. Self-determination theory applied to health contexts: a meta-analysis. Perspect Psychol Sci 2012 Jul;7(4):325-340. [doi: 10.1177/1745691612447309] [Medline: 26168470]
42. Lustria ML, Noar SM, Cortese J, Van Stee SK, Glueckauf RL, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
43. Anttila M, Kivistö H, Piirainen A, Kokko K, Malinen A, Pekkonen M, et al. Cardiac rehabilitees' technology experiences before remote rehabilitation: qualitative study using a grounded theory approach. J Med Internet Res 2019 Feb 07;21(2):e10985 [FREE Full text] [doi: 10.2196/10985] [Medline: 30730298]
44. Lingg M, Lütschg V. Health system stakeholders' perspective on the role of mobile health and its adoption in the swiss health system: qualitative study. JMIR Mhealth Uhealth 2020 May 11;8(5):e17315 [FREE Full text] [doi: 10.2196/17315] [Medline: 32391802]
45. Grünloh C, Myreteg G, Cajander A, Rexhepi H. "Why do they need to check me?" Patient participation through eHealth and the doctor-patient relationship: qualitative study. J Med Internet Res 2018 Jan 15;20(1):e11 [FREE Full text] [doi: 10.2196/jmir.8444] [Medline: 29335237]
46. King G, O'Donnell C, Boddy D, Smith F, Heaney D, Mair FS. Boundaries and e-health implementation in health and social care. BMC Med Inform Decis Mak 2012 Sep 07;12:100 [FREE Full text] [doi: 10.1186/1472-6947-12-100] [Medline: 22958223]

Abbreviations

CVD: cardiovascular disease
HCP: health care professional

Edited by R Kukafka; submitted 10.11.20; peer-reviewed by R Cruz Martínez, M Lingg; comments to author 12.12.20; revised version received 21.01.21; accepted 10.08.21; published 15.10.21.

Please cite as:
Cohen Rodrigues TR, de Buisonjé DR, Keesman M, Reijnders T, van der Geer JE, Janssen VR, Kraaijenhagen RA, Atsma DE, Evers AWM
Facilitators of and Barriers to Lifestyle Support and eHealth Solutions: Interview Study Among Health Care Professionals Working in Cardiac Care
J Med Internet Res 2021;23(10):e25646
URL: https://www.jmir.org/2021/10/e25646
doi:10.2196/25646
PMID:34652280

Original Paper

An Audio Personal Health Library of Clinic Visit Recordings for Patients and Their Caregivers (HealthPAL): User-Centered Design Approach

Paul J Barr[1,2], MSc, PhD; William Haslett[2], PhD; Michelle D Dannenberg[1,2], MPH; Lisa Oh[3], BA; Glyn Elwyn[1], MSc, MD, PhD; Saeed Hassanpour[4,5], PhD; Kyra L Bonasia[6], PhD; James C Finora[6], BS; Jesse A Schoonmaker[1,2], MPH, MD; W Moraa Onsando[1,2], MPH, MD; James Ryan[7], DO; Martha L Bruce[8], MPH, PhD; Amar K Das[4], MD, PhD; Roger Arend[9], BS; Sheryl Piper[9]; Craig H Ganoe[4], MS

[1]The Dartmouth Institute for Health Policy & Clinical Practice, Geisel School of Medicine, Dartmouth College, Lebanon, NH, United States
[2]The Center for Technology and Behavioral Health, Geisel School of Medicine, Dartmouth College, Lebanon, NH, United States
[3]Department of Computer Science, Dartmouth College, Hanover, NH, United States
[4]Department of Biomedical Data Science, Dartmouth College, Hanover, NH, United States
[5]Department of Epidemiology, Dartmouth College, Hanover, NH, United States
[6]Geisel School of Medicine, Dartmouth College, Hanover, NH, United States
[7]Ryan Family Practice, Ludington, MI, United States
[8]Department of Psychiatry, Geisel School of Medicine, Dartmouth College, Hanover, NH, United States
[9]Patient Partner, Lebanon, NH, United States

Corresponding Author:
Paul J Barr, MSc, PhD
The Dartmouth Institute for Health Policy & Clinical Practice
Geisel School of Medicine
Dartmouth College
One Medical Center Drive
Lebanon, NH, 03756
United States
Phone: 1 603 653 0863
Email: paul.j.barr@dartmouth.edu

Abstract

Background: Providing digital recordings of clinic visits to patients has emerged as a strategy to promote patient and family engagement in care. With advances in natural language processing, an opportunity exists to maximize the value of visit recordings for patients by automatically tagging key visit information (eg, medications, tests, and imaging) and linkages to trustworthy web-based resources curated in an audio-based personal health library.

Objective: This study aims to report on the user-centered development of HealthPAL, an audio personal health library.

Methods: Our user-centered design and usability evaluation approach incorporated iterative rounds of video-recorded sessions from 2016 to 2019. We recruited participants from a range of community settings to represent older patient and caregiver perspectives. In the first round, we used paper prototypes and focused on feature envisionment. We moved to low-fidelity and high-fidelity versions of the HealthPAL in later rounds, which focused on functionality and use; all sessions included a debriefing interview. Participants listened to a deidentified, standardized primary care visit recording before completing a series of tasks (eg, finding where a medication was discussed in the recording). In the final round, we recorded the patients' primary care clinic visits for use in the session. Findings from each round informed the agile software development process. Task completion and critical incidents were recorded in each round, and the System Usability Scale was completed by participants using the digital prototype in later rounds.

Results: We completed 5 rounds of usability sessions with 40 participants, of whom 25 (63%) were women with a median age of 68 years (range 23-89). Feedback from sessions resulted in color-coding and highlighting of information tags, a more prominent play button, clearer structure to move between one's own recordings and others' recordings, the ability to filter recording content by the topic discussed and descriptions, 10-second forward and rewind controls, and a help link and search bar. Perceived usability

increased over the rounds, with a median System Usability Scale of 78.2 (range 20-100) in the final round. Participants were overwhelmingly positive about the concept of accessing a curated audio recording of a clinic visit. Some participants reported concerns about privacy and the computer-based skills necessary to access recordings.

Conclusions: To our knowledge, HealthPAL is the first patient-centered app designed to allow patients and their caregivers to access easy-to-navigate recordings of clinic visits, with key concepts tagged and hyperlinks to further information provided. The HealthPAL user interface has been rigorously co-designed with older adult patients and their caregivers and is now ready for further field testing. The successful development and use of HealthPAL may help improve the ability of patients to manage their own care, especially older adult patients who have to navigate complex treatment plans.

(J Med Internet Res 2021;23(10):e25512) doi:10.2196/25512

KEYWORDS

patient-centered care; health communication; audiovisual aids; user-centered design; software; natural language processing; patients; caregivers

Introduction

Background

Higher recall of medical information is associated with improved disease management, treatment adherence, and higher patient satisfaction [1,2]. Recall, however, is often low, with 40%-80% of medical information from a clinical visit being forgotten immediately by patients [3-8]. Although the poor recall of medical information is pervasive, it is most acute among older adults. As people age, they process information more slowly and have reduced working memory [9,10]. Older patients experience more challenges in recalling drug information, treatment recommendations, appointments, and disease information [11], especially those with multimorbidity [1,6,8,12-14] who report the *endless struggle* of managing their conditions [15,16]. Poor recall also impacts caregivers [17]. In a recent national survey of caregivers in the United States, 84% of respondents wanted more information on caregiving topics [17].

The last decade has seen significant efforts to increase patient access to medical information, especially clinic visit information. Mandated initially by the Health Information Technology for Economic and Clinical Health Act's meaningful use standards, clinics across the United States now offer patients an after-visit summary (AVS) [18]. The AVS is a summary of the clinic visit generated from the electronic medical record, printed during visits, or available via the patient portal and includes diagnoses, medications, allergies, clinician visited, and clinician comments. Although intended as a means of promoting self-management, there have been concerns about the AVS, including accuracy of medication lists, layout, and use of medical terminology on the AVS [19]. This is particularly challenging for patients who often report low health literacy and struggle with exclusively text-based information [2,3,20-22]. AVS can also represent a significant burden on clinician workload [7,23]. These factors have resulted in low AVS use [24]. An adjunct to the AVS may exist in recordings of the clinic visit.

Patients rely on verbal communication with their doctor [8] and some are now audio recording clinic visits to capture this valuable information [25-27]. In response, a small but growing number of clinics across the United States are beginning to offer patients recordings of clinic visits. Systematic reviews found that access to recordings leads to increased patient and family engagement, understanding, and recalling visit information; reduced anxiety; increased satisfaction; improved treatment adherence; reduced patients' clinic phone calls; and reduced decisional regret [25,28-33].

The absence of a safe and secure recording system is a barrier to the broader uptake of clinic recordings [27]. A recording provides all the visit details, yet navigating recordings is a challenge, as the benefit "depends on picking out...the crucial points..." of the visit [27]. Unstructured visit information increases the risk of overwhelming patients [27,34].

Electronic personal health libraries (PHLs) may be a solution, as they allow patients to manage, maintain, and organize health information on the web [34,35]. PHLs can range from medical records that patients can access tethered to a health system to stand-alone platforms where patients collect and manage their own data. PHLs are becoming more advanced through the application of data science methods such as natural language processing (NLP) [36]. These methods can identify patterns in unstructured data and classify text based on its meaning. Such NLP methods have been used to predict hospital readmissions [37], future radiology utilization [38], and medical conditions in clinical decision support systems [39]. In PHLs, data science methods have been used to automatically provide tailored information via guided searches for disease and self-care information [36]. Despite the availability of these methods, existing PHLs are yet to facilitate the integration of clinic recordings [40].

To address this gap, we planned to develop an audio PHL to facilitate the acquisition, organization, and management of clinic audio recordings—HealthPAL (personal audio library). On the basis of a review of patients' information-seeking behavior and needs [41], the basic features of HealthPAL include (1) identifying, organizing, and tagging elements of the clinic visit audio recording deemed important to patients; for example, parts of the visit recording where medications are mentioned would be automatically highlighted for patients; (2) a search function, allowing end users to search for information from their visit; and (3) linkage of key medical terms from the clinic visit audio recordings to trustworthy, layperson resources such as MEDLINE Plus, which can be retrieved, organized, edited, and shared by patients. For example, a hyperlink to learn more

about the medication mentioned would be available to the patient. In the system's background, a transcript of the medical visit is automatically generated using speech-to-text software. However, because of concerns of inaccurate speech-to-text potentially providing incorrect written medical information, we chose not to expose full transcript text in our design.

Our user-centered design work falls in the history of design studies around *meeting browsers* [42]—software multimedia browsers of meeting recordings and associated meeting artifacts—where HealthPAL is the first to consider design and usability for the specialized context of patients meeting with their primary care provider. *Classroom 2000* [43] initiated the modern genre of meeting browsers, which focused on capturing a *recording* of a meeting or classroom lecture and its context, often focusing on live audio and video and linking to slides of a PowerPoint-like presentation, with some allowing users to add their own annotations or notes. Early evaluation work with a meeting browser found that such a system can allow users to more accurately answer questions about a meeting [44], whereas features that allow the user to focus on key phrases allowed them to answer questions about the audio content more quickly [45]. We hypothesize similar uses of annotated audio recordings in health care.

Objective

This study reports on the development of the HealthPAL user interface (UI). Results from our data science models are reported elsewhere. By adopting a user-centered design framework, we engaged end users throughout the iterative development of HealthPAL [40,41]. We followed the usability specification and evaluation framework developed by Yen and Bakken [46], which consisting of 5 stages: (1) system requirements, (2) system component development, (3) usability evaluation in controlled settings, (4) pilot field testing, and (5) large-scale field testing. We report on steps 1-3 in this paper. We hypothesized that through user-centered design, HealthPAL would be highly usable with high end user satisfaction.

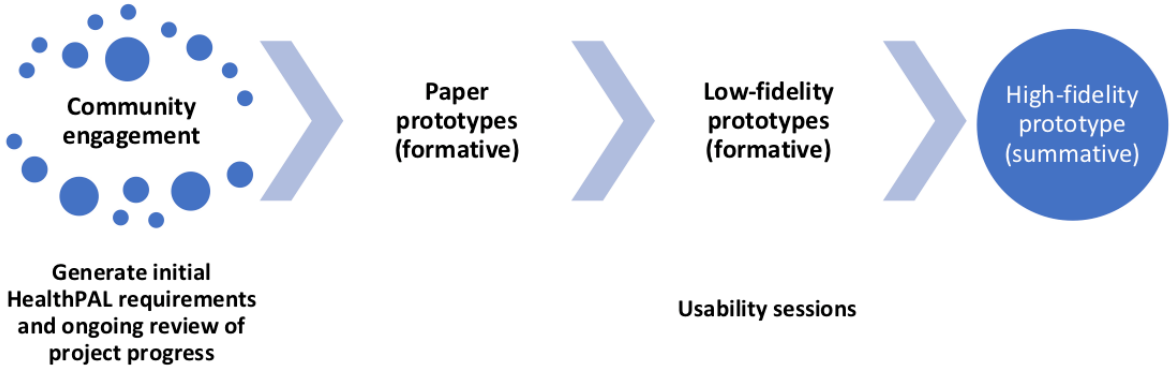
Methods

Study Design: Overview

Our approach incorporated UI development conducted through 5 iterative rounds of *usability sessions*. During the sessions, participants were asked to play the role of a patient or caregiver and complete a series of typical tasks within HealthPAL. The UI was iteratively refined in each round, with additional tasks added to assess the newly added features. We began with the paper prototype and formative sessions (rounds 1-3) in which participants worked with paper and low-fidelity software prototype designs before moving to the summative software sessions (rounds 4 and 5) in which they interacted with higher fidelity prototypes of the software (Figure 1). The sessions were structured so that participants would first listen to a fictitious clinic visit recording, and in the final summative software session (round 5), patients used their actual clinic visit recordings. Participants were presented with a set of typical user tasks to complete in the system being tested during their session.

Before each new iterative round of user testing, the research team completed heuristic evaluations and cognitive walkthroughs with HealthPAL to mitigate common usability problems before working with participants.

Figure 1. Overview of the user-centered design stages of HealthPAL.



Settings

Participants were recruited from the Upper Valley of New Hampshire and Vermont between 2016 and 2019. Individuals were recruited from multiple settings, including public libraries, the Dartmouth-Hitchcock Medical Center (DHMC) simulation and human-computer interaction lab, Dartmouth-Hitchcock (D-H) Heater Road Primary Care, D-H Patient and Family Advisors group, the D-H Aging Resource Center, and a local senior living community. This study was approved by the Committee for the Protection of Human Subjects, Dartmouth College and the D-H Health Human Research Protection Program (Committee for the Protection of Human Subjects Study #30397, 30531; D-HH HRPP 00030531).

Participants

Participants included individuals who represented the views of patients and caregivers. All participants were 18 years or older, able to communicate in English, and able to provide informed consent. Individuals with serious mental illness, self-reported significant uncorrectable hearing or visual impairments, or significant cognitive impairment (score of 4 or less on a 6-item screener) [47] were excluded from the study. Caregiver participants were individuals who self-identified as having previously cared for a family member or loved one. In the final round of user testing, we also recruited primary care clinicians to record the clinic visits of patients to be used during the usability evaluation sessions. Consented clinicians identified

patients who met the eligibility criteria (18 years or older with 2 chronic health conditions) and who were facing a treatment decision or were discussing a diagnosis or medication; patients visiting solely for a procedure, such as blood draw, were excluded.

Our project initially focused on the general public, but additional funding received from the National Library of Medicine during the study allowed us to expand the proposed features and focus on older adults who account for the greatest use of health care and may benefit the most from the proposed system. Therefore, we oversampled older adults (≥65 years) and focused on this group in later rounds. As we moved to later rounds of user testing, it was also necessary for participants to have internet access at home to review the recording system before the usability session. A US $25-$30 honorarium was provided to participants. We targeted 5 individuals per round, a sample size that is considered adequate to detect up to 80% of usability issues [48,49]. All participants provided written informed consent. Participants from previous rounds could not participate in later rounds to reduce the potential impact of learning effects inflating usability evaluations.

Recruitment

In the initial paper prototype and formative sessions (rounds 1 and 2), we recruited participants from the DHMC Patient and Family Advisors group and from public spaces at a local library. Participants were approached by a research team member; informed about the project; and if interested, they were taken to a private space, they provided consent, and they began the usability session. For the final paper prototype and formative session (round 3), we focused on older adults, and recruitment moved to the Aging Resource Center and an older adult living community. Participants were sent physical mail and an email to inform them about the project. The screening was conducted by telephone with interested participants to determine eligibility, and participants were met on the day of the session to complete informed consent before beginning the usability session.

In the summative software sessions (rounds 4 and 5), we specifically targeted individuals from both *patient* and *caregiver* stakeholder groups separately. Caregivers were recruited from the Aging Resource Center using the procedures described above. Patients were recruited from a local primary care clinic, identified by participating clinicians, and received a screening telephone call from a research team member. Eligible patients were asked to come to the clinic 30 minutes earlier than their appointment to complete the informed consent process; once they provided consent, their clinic visits were recorded using the software. Participants were then asked to meet with the research team within a week to complete the usability session.

Community Engagement

In addition to the 5 rounds of user testing, 2 *Lunch and Listen* workshops were conducted with patients and family volunteers from DHMC, Lebanon, New Hampshire. These workshops were codeveloped and led by our study team's patient partners (RA and SP) to discuss key system concepts with community members. Responses from these workshops informed the initial layout, features, and desired functionality of the HealthPAL system that was tested during the usability sessions.

Usability Sessions

Overview

All sessions began with a description of the usability session, and participants were asked to *think aloud* [50,51] as they completed the tasks with the software. Paper prototype sessions were video-recorded along with participant and facilitator audio, and in later rounds where software prototypes were used by participants, screen video and audio were recorded to capture participants' use of the prototype and their verbal feedback. Upon completion of the round-specific tasks, participants completed a semistructured interview about the system and desired functionality. The session facilitator in the room made written field notes related to participants' interactions with the prototype.

UI Design

In all sessions, the UI primarily consisted of 2 pages. The first page was an interface allowing the user to choose which recordings the user is viewing (eg, choosing between their own visit recordings and the visit recordings of someone who they act as a caregiver for). The second page offered basic playback features (eg, play and pause, rewind, skip forward or back 10 seconds, and mute) for a visit recording. In addition, on that page, individual audio segments of the visit recording were *tagged* for 8 classes of information: diagnosis, follow-up, medication, patient education, recommendation, signs, symptoms and problems, test and imaging, and treatment options. The user could click on a segment to start playback at that point. New features were added to each round based on the user feedback.

Fictitious Clinic Visit Recordings

Fictitious primary care visit recordings were created and used throughout the user testing sessions, except in the final round of patient sessions. We created 2 fictitious characters, Chris Hill, a 58-year-old male patient, and his 81-year-old mother Linda, both of whom met with a fictitious primary care clinician named Dr Adams. The fictitious clinic visit recordings for both the characters were produced by rerecording 2 real primary care clinic visits, stripped of patient identifiers (a description of the recording is provided in Textbox 1).

Textbox 1. Fictitious clinic recordings.

Chris Hill

- Chris's recording was a clinic visit of approximately 7 minutes in which Chris and Dr Adams discuss Chris's allergies, and Dr Adams suggests increasing his dosage of Flonase. They also talk about Chris's stomach pain, and Dr Adams suggests increasing his dosage of Omeprazole. Chris talks about his emergency room visit from when he was having bad stomach pain, and Dr Adams reviews some test results from that visit.

Linda Hill

- Linda's recording was a clinic visit of approximately 11 minutes in which Linda and Dr Adams discuss the pain in her foot and back. Linda tells Dr Adams that she has stopped taking the nerve pain controller Gabapentin because of the side effects, but Dr Adams suggests Linda to try it again. They also discuss Linda's arthritis and the ganglion cyst on her hand. Dr Adams informs her that the cysts can be surgically removed if they bother her and also suggests that she visits a pain clinic for overall pain control.

Paper Prototyping and Formative Usability Sessions (Rounds 1-3)

Overview

The initial round began with a paper prototype before moving onto the low-fidelity prototypes of the software in rounds 2 and 3. During the paper prototype sessions, the facilitator adopted a *Wizard of Oz* technique, where they played back the appropriate recorded patient clinic visit audio in response to the participant's interactions with the prototypes [52]. Participants in these sessions (rounds 1-3) were asked to complete tasks in both patient and caregiver roles.

Patient Role

Participants were given the role of Chris. They began the round by listening to the entire fictitious audio recording for Chris before being presented with the paper prototype and before being asked to complete the tasks (Textbox 2). This was designed to replicate the experience of a patient who was present during the visit.

Textbox 2. Example of a role-based scenario and task for participants to complete in the prototypes.

Role

- In this study, you will play the role of *Chris*. Chris Hill is a 58-year-old male patient of Dr Adams, who last saw this doctor on June 20, 2017—1 week ago. Chris's mother, Linda, who is 81 years old, also sees Dr Adams.

Scenario

- Chris's doctor creates audio recordings of patients' visits and provides a web-based software app for patients to listen to their own visit recordings after they leave from each visit. Patients can also grant permission for other family members or caregivers to listen to their visit recordings. Chris also has access to listen to his mother's clinic visits, so that he can stay up to date on her health. Chris knows that his mother visited Dr Adams on June 14, 2017—her most recent visit. Chris also knows his mother and Dr Adams discussed possible surgery for the pain she has in her hand. They also discussed her restarting a medication for her foot pain that Linda had previously stopped taking. Later you will be asked to find and listen to important audio segments of her most recent visit in the website prototype.

Task

- Find where Dr Adams talks about considering surgery for Linda's hand pain (ganglion cyst) in her most recent visit. When you believe you are finished with the task, say "I found it" aloud.

Caregiver Role

Once participants completed tasks with Chris's recording, they were asked to find Linda's recordings in the system and asked to complete a series of similar tasks. They did not listen to Linda's recording before the tasks, replicating the experience of a caregiver who was not present at a visit.

Summative Software Usability Sessions (Rounds 4 and 5)

Starting from round 4, we recruited individuals who identified as either caregivers or patients and asked them to play only that role. All participants completed the same tasks, which required finding and listening to important parts of Linda's recording and new features, including finding and using hyperlinks to additional information on a medical concept. In the final round of user testing, round 5, patients reviewed a recording of their primary care visit with a high-fidelity version of HealthPAL with all desired features; caregivers continued to review fictitious recordings. In parallel with UI development, we are developing NLP models to automatically annotate the classes of information in recordings. However, at this stage of development, we used human transcription and manual tagging of recordings by 2 clinically trained annotators (JAS and WMO). Patients were emailed a link to log in to the system to access their recording within 3 days of their appointment. Participants were then asked to return for an in-person session to demonstrate how they used the system and complete a set of specific tasks within the system.

Data Collection and Analysis

Overview

We collected data on participant age, sex, race, ethnicity, and educational attainment for all rounds of usability testing. The System Usability Scale (SUS) [53] was administered from round

3 onward, as we moved to a web-based version of the software. Usability evaluation metrics included the SUS, critical incidents [54], and task completion ratios. We also gathered data on participant suggestions and views—general attitudes participants had toward the system.

Task Completion

For each task given to participants, we recorded whether the participant completed the task on their own, completed the task with help from the facilitator, or did not complete the task.

Critical Incidents

Recordings were coded for key critical incidents related to usability and interface design. These incidents included whenever a participant verbally or implicitly made known that they were struggling with an action, whether it was controlling the audio or finding the correct segment, regardless of task completion. Critical incidents also included whenever a participant took an action that deviated from the expected path, such as when participants clicked on a hyperlink instead of playing an audio segment.

Participant Suggestions

We recorded any feedback that participants provided about improving the website. This feedback included suggestions such as how the participants would like a feature to work or what the participants were expecting to see on a certain page.

Participant Views

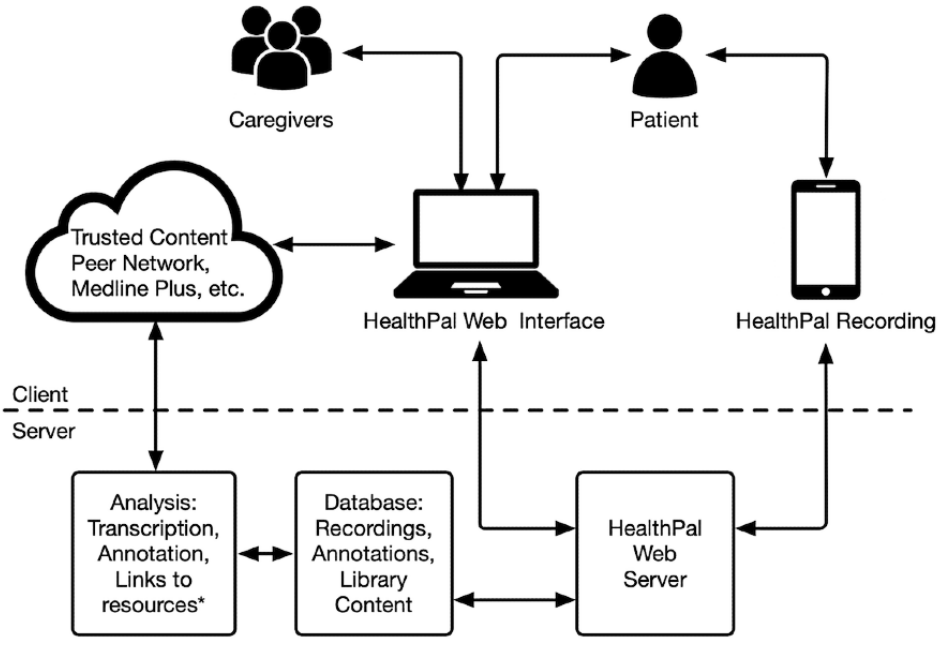
We recorded any comments that participants made regarding their general attitudes toward the system. These comments included positive remarks, such as how the system could be beneficial for keeping track of important health information, and concerns, such as the security and privacy of their health information.

Descriptive statistics of median and range were used for continuous data, and proportions and ranges were used for categorical data. SUS scores were calculated on a scale of 0-100, with a score of >68 indicating above-average satisfaction with the usability of a system [55-57]. Descriptive summary statistics of the above-mentioned usability metrics were guided by the TURF (Task, User, Representation, and Function) framework [58], with a focus on task completion and system usability. We used summative content analysis to identify key issues and suggestions from the usability sessions. Transcripts from usability sessions were reviewed by 2 coders (LO and CHG). Commonly identified suggestions, views, and critical incidents were coded and grouped into the main themes.

Software Architecture

We built the software prototype for this work as a web app, using the mature and widely adopted Ruby on Rails app development framework. The app is hosted on Dartmouth's secure server infrastructure, and all client-server communications are encrypted using an HTTPS protocol. In addition, the app software communicates with a PostgreSQL relational database, hosted on Dartmouth's infrastructure, and app data were encrypted at rest. Using this architecture, we were able to securely deliver the app to a range of devices, supporting both audio recording and playback while avoiding storing audio or other protected health information on users' devices. This architecture also allowed us to rapidly implement design changes that were made based on the results of each round of user testing (Figure 2). By using the Git version control system and an automated app deployment pipeline, we were able to track every deployment of the app code, thus producing a history of precisely which version of the software was running at any given time. The app prototype is available as open-source software under the MIT license. The source code can be obtained at the website [59].

Figure 2. HealthPAL system architecture.



Results

Participant Characteristics

We completed usability sessions with 40 participants (including 10 self-identified caregivers) from October 2017 to May 2019. Participants were predominantly White non-Hispanic (39/40, 98%), with a median age of 68 years (range 23-89 years). There were 63% (25/40) female participants, and most participants (37/40, 93%) had some college education or higher (Table 1). The evolution of the UI during rounds of user testing is shown in Figures 3-7.

Table 1. Participant demographics (N=40).

Participant characteristics	Round 1 (n=8)	Round 2 (n=6)	Round 3 (n=5)	Round 4		Round 5	
				CG[a] (n=5)	P[b] (n=6)	CG (n=5)	P (n=5)
Age (years), median (range)	54 (23-80)	48 (26-81)	80 (68-89)	70 (52-83)	71 (66-88)	72 (59-77)	62 (30-67)
Females, n (%)	3 (38)	3 (50)	3 (60)	5 (100)	3 (50)	4 (80)	4 (80)
Race[c], n (%)							
Hispanic White	0 (0)	0 (0)	1 (20)	0 (0)	0 (0)	0 (0)	0 (0)
Non-Hispanic White	8 (100)	6 (100)	4 (80)	5 (100)	6 (100)	5 (100)	5 (100)
Black or African American	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (20)
Education, n (%)							
High school or no degree	0 (0)	1 (17)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
High school graduate	1 (13)	1 (17)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Some college	1 (13)	1 (17)	0 (0)	1 (20)	0 (0)	0 (0)	1 (20)
College degree (bachelors or associate)	2 (25)	2 (33)	3 (60)	3 (60)	2 (33)	2 (40)	4 (80)
Masters, doctorate, or professional school	4 (50)	1 (17)	2 (40)	1 (20)	4 (67)	3 (60)	0 (0)

[a]CG: caregiver role.

[b]P: patient role.

[c]More than 1 response allowed.

Figure 3. Initial paper prototypes of HealthPAL (round 1)—interface for finding a patient's recording (left) and playing back the recording (center and right).

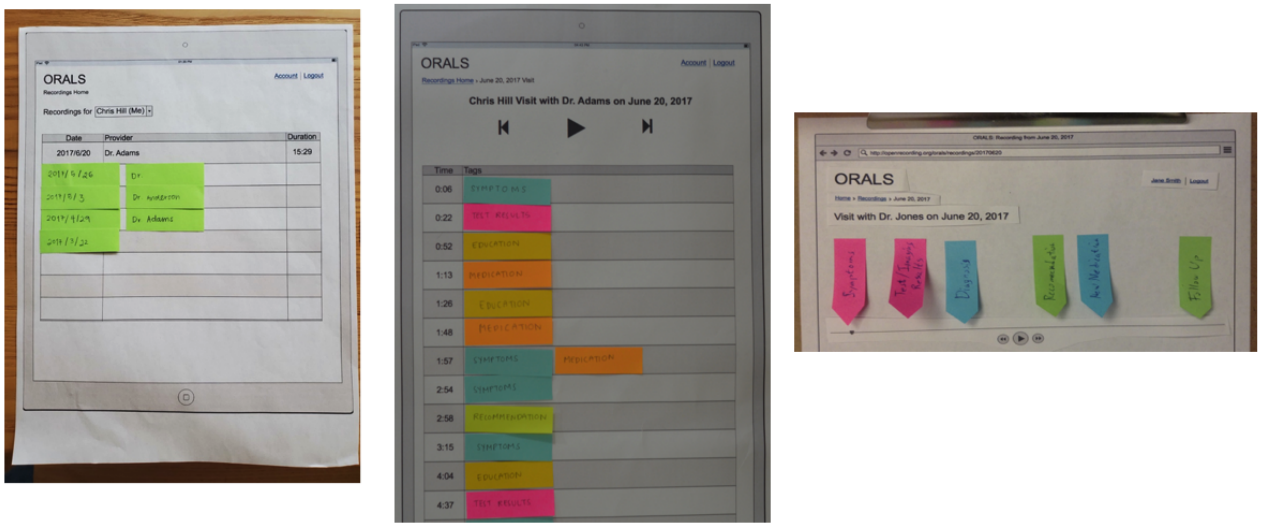


Figure 4. Initial software prototype (round 2)—interface for finding a patient's recording (top) and playing back the recording (bottom).

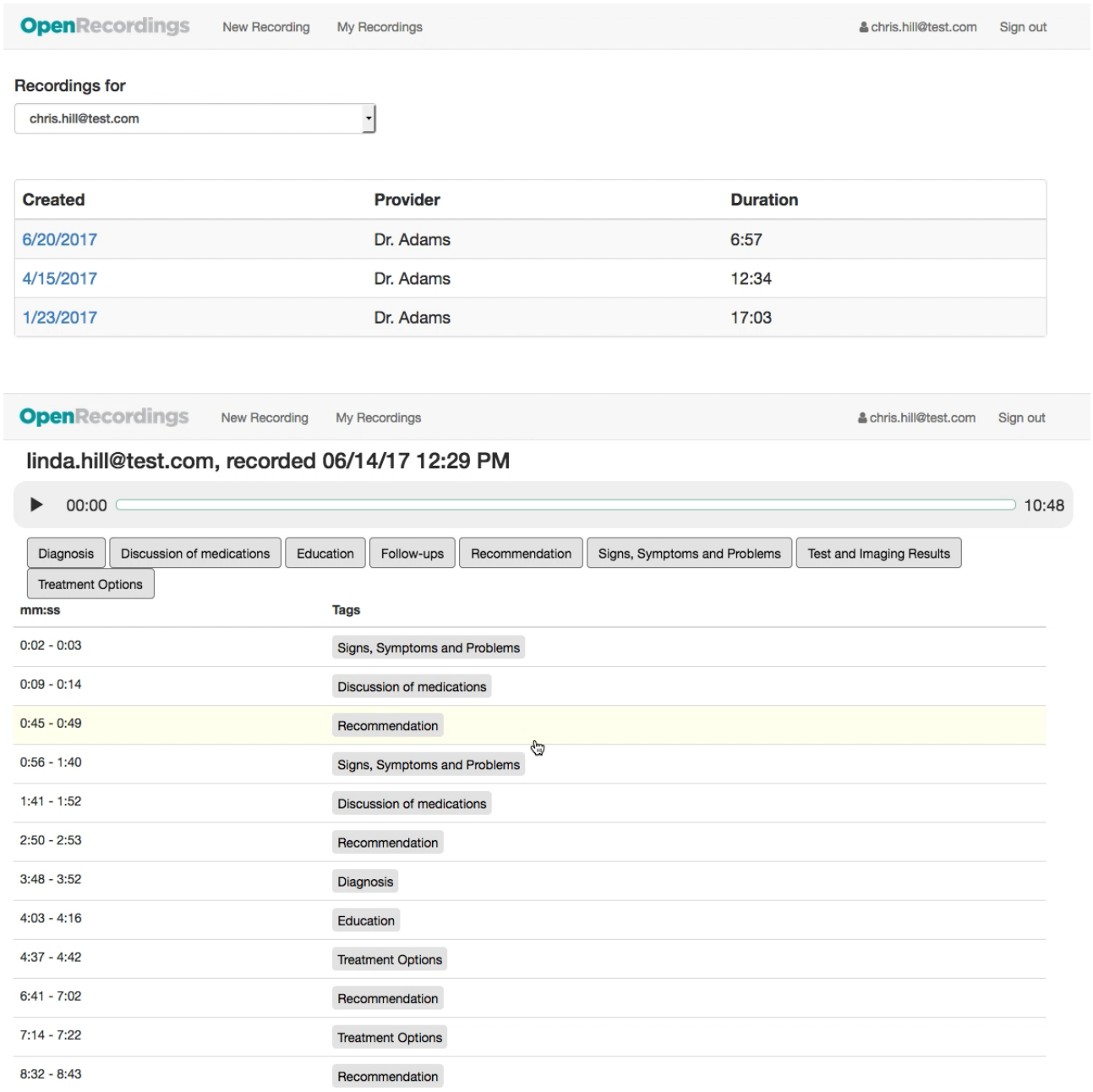


Figure 5. Updated software prototype (round 3)—interface for finding a patient's recording (top) and playing back the recording (bottom).

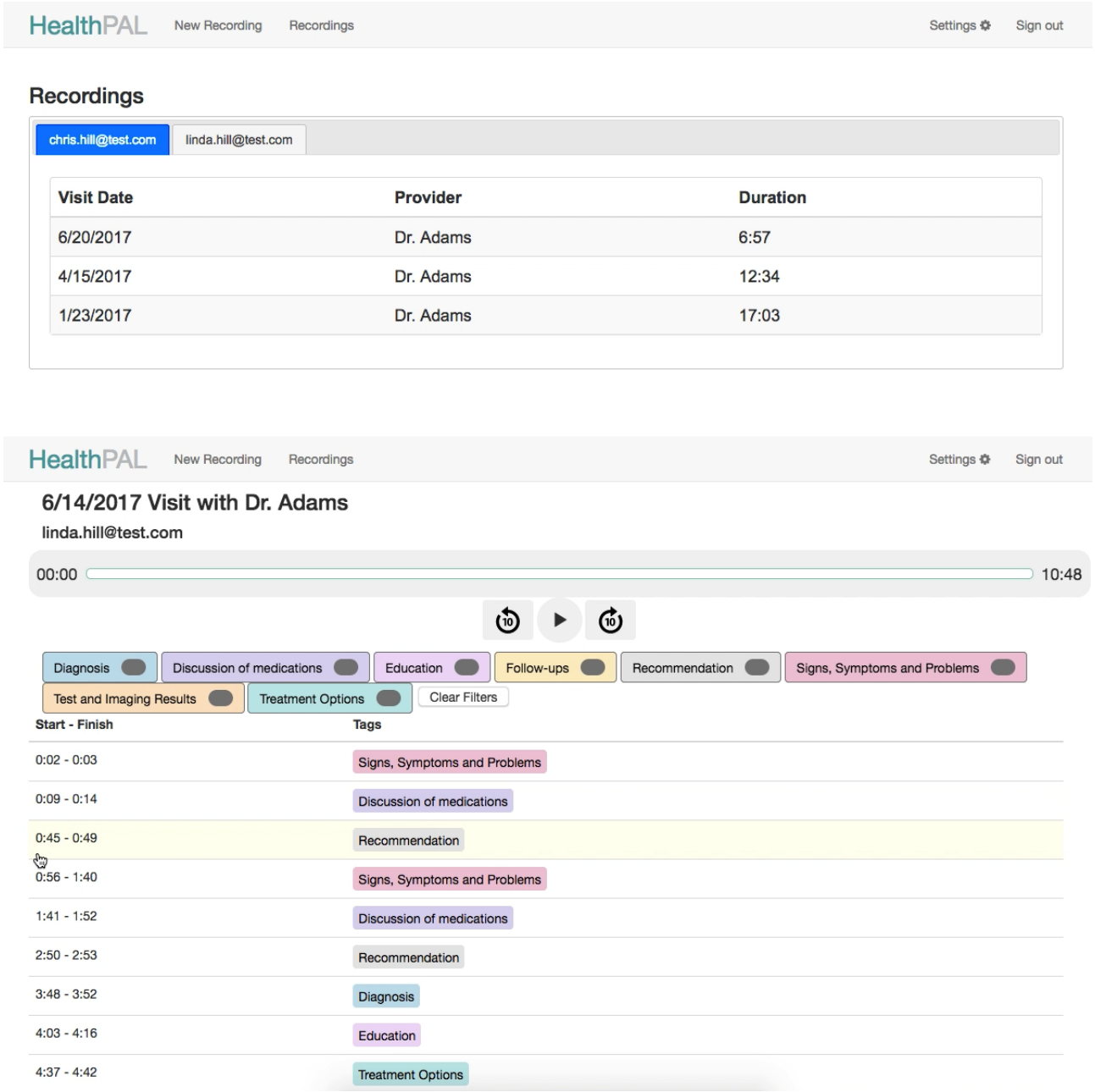


Figure 6. Updated software prototype (round 4)—interface for finding a patient's recording (top) and playing back the recording (bottom).

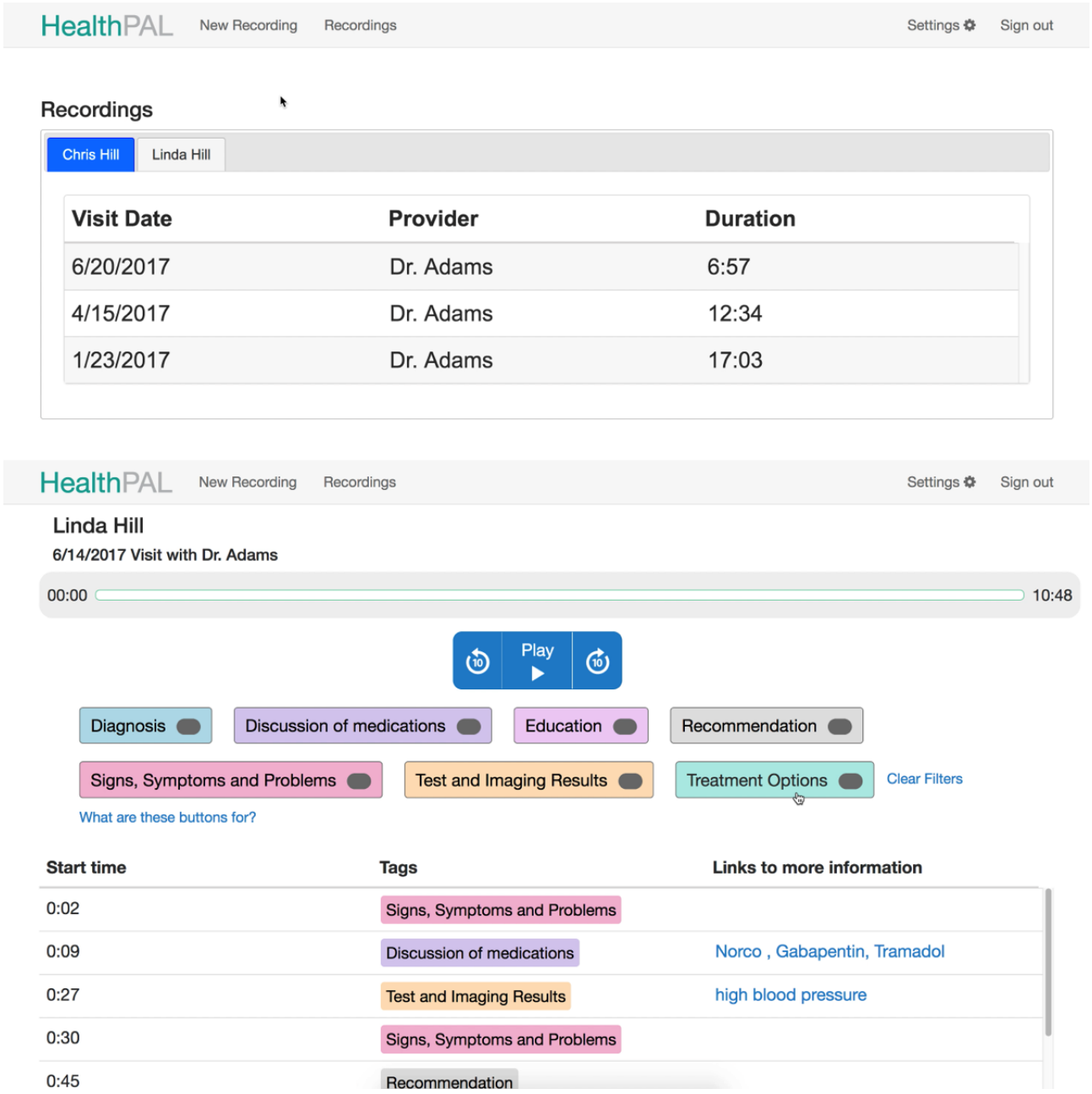
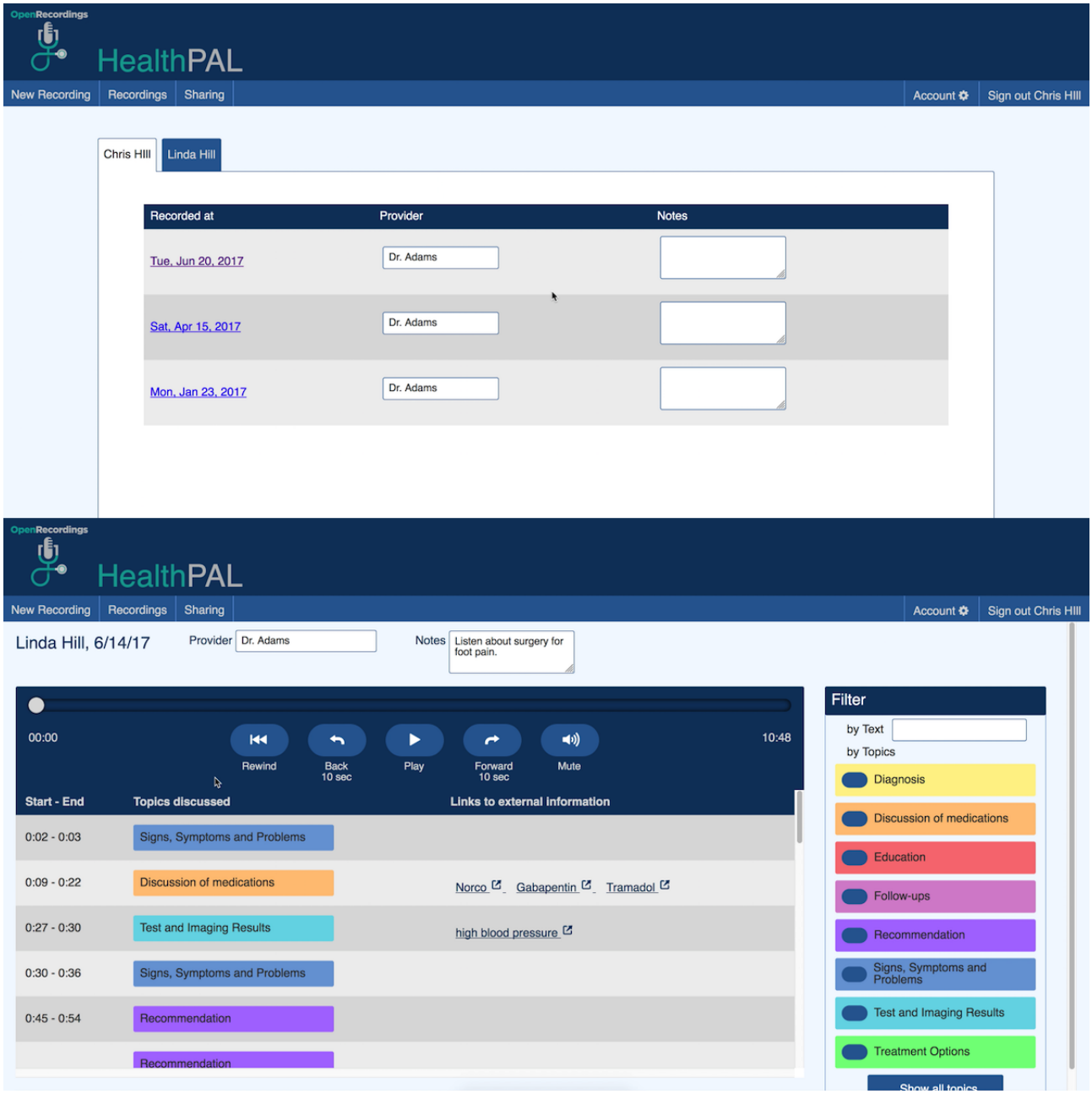


Figure 7. High-fidelity version of the software prototype (round 5)—interface for finding a patient's recording (top) and playing back the recording (bottom).



Evaluation Outcomes

Task Completion and System Usability

Task completion followed an inverted *U*-shaped distribution, with high completion rates in the early and later rounds and a drop in performance in between (Table 2). However, as new features were added, some tasks required further assistance (ie, find and play where medications are discussed on the recording), whereas the initial introduction of hyperlinks to further information resulted in poor task completion for these tasks. By round 5, the majority of tasks were completed without assistance. SUS assessments were introduced in round 3; the median SUS score improved across the rounds: round 3, 40 (range 38-68); round 4, 73 (range 35-100); and round 5, 78 (range 20-100). Scores for caregivers were lower than those for patients in both rounds 4 and 5.

Table 2. Task completion ratios and system usability (N=40).

Evaluation outcomes	R[a]1 (n=8)	R2 (n=6)	R3 (n=5)	R4 (n=11)		R5 (n=10)	
				CG[b] (n=5)	P[c] (n=6)	CG (n=5)	P (n=5)
Task (completed task alone:completed task with help:did not complete task)							
Find visit recording task	4:3:1	3:2:1	2:3:0	3:2:0	5:1:0	4:1:0	4:1:0
Find or play health issue task	8:0:0	5:1:0	3:1:1	2:2:1	4:2:0	2:3:0	4:1:0
Find or play medication task	8:0:0	5:1:0	0:3:2	1:4:0	6:0:0	5:0:0	4:0:1
Find more about health issue task	N/A[d]	N/A	N/A	2:1:2	1:0:5	5:0:0	3:0:2
Find more about medication task	N/A	N/A	N/A	3:1:1	3:0:3	5:0:0	5:0:0
Task completion, n (%)							
Total completed task (alone or with help)	23 (96)	17 (94)	12 (80)	21 (84)	22 (73)	25 (100)	22 (88)
Completed task with help	3 (13)	4 (22)	7 (47)	10 (40)	3 (10)	4 (16)	2 (8)
SUS[e] score (0-100), median (range)	N/A	N/A	40[f] (38-68)	73 (35-100)	75 (55-100)	65 (20-90)	95 (43-100)

[a]R: round.

[b]CG: caregiver role.

[c]P: patient role.

[d]N/A: not applicable.

[e]SUS: System Usability Scale; not administered in the first 2 rounds.

[f]Missing data for 1 person.

Critical Incidents

A summary of critical incidents (challenges that prevented participants from completing a task independently) is presented in Table 3. The greatest number of challenges was observed in round 4, where hyperlinks were added. This resulted in confusion in finding sections of audio recordings to play using tags and using associated hyperlinks to find further information. When asked to find additional information about a medical term in the visit recording (with our intent being they use the hyperlink we provided in HealthPAL), some participants simply opened a new tab in the browser and conducted a web search; we counted these responses as not completing the task. Further refinements to the UI resulted in a lower proportion of critical incidents by round 5.

Table 3. Summary of key critical incidents that occurred during user testing (N=40).

Critical incidents[a]	R[b]1 (n=8), n (%)	R2 (n=6), n (%)	R3 (n=5), n (%)	R4 (n=11), n (%)		R5 (n=10), n (%)	
				CG[c] (n=5)	P[d] (n=6)	CG (n=5)	P (n=5)
Issues switching to or from another user's recordings	6 (75)	3 (50)	4 (80)	1 (20)	—[e]	1 (20)	—
Issues understanding which user the recordings belonged to	3 (38)	2 (33)	4 (80)	2 (40)	1 (17)	1 (20)	—
Issues navigating through the site	2 (25)	1 (17)	0 (0)	4 (80)	2 (33)	1 (20)	1 (20)
Issues finding the appropriate segment	7 (88)	4 (67)	4 (80)	2 (40)	2 (33)	1 (20)	0 (0)
Issues controlling or traversing the audio	5 (63)	0 (0)	5 (100)	5 (100)	5 (83)	3 (60)	3 (60)
Issues using the filters by topic	—	3 (50)	2 (40)	4 (80)	2 (33)	2 (40)	2 (40)
Issues using hyperlinks to find additional information	—	—	—	5 (100)	4 (67)	3 (60)	2 (40)
Issues using the filters by text	—	—	—	—	—	3 (60)	1 (20)

[a]Proportion of sessions with a critical incident.

[b]R: round.

[c]CG: caregiver role.

[d]P: patient role.

[e]Feature not available or not part of testing.

Suggestions

Participants made several suggestions regarding improvements and functionality of the system (Table 4). These suggestions were grouped into common themes across the rounds and were used to improve the UI and functionality. By round 5, few additional suggestions emerged.

Table 4. Proportion of key suggestions given during user testing by category (N=40).

Suggestion	R[a]1 (n=8), n (%)	R2 (n=6), n (%)	R3 (n=5), n (%)	R4 (n=11), n (%)		R5 (n=10), n (%)	
				CG[b] (n=5)	P[c] (n=6)	CG (n=5)	P (n=5)
Suggestions for making segments easier to find within a recording	7 (88)	4 (67)	2 (40)	2 (40)	2 (33)	2 (40)	1 (20)
Suggestions for making specific visits easier to find	5 (63)	2 (33)	3 (60)	0 (0)	0 (0)	1 (20)	1 (20)
Suggestions to make switching between recordings of different users more intuitive (ie, switching from personal to loved one's recording)	4 (50)	3 (50)	3 (60)	0 (0)	—[d]	1 (20)	—
Suggestions for supplementary text to include with recordings	3 (38)	3 (50)	2 (40)	3 (60)	0 (0)	1 (20)	2 (40)
Suggestions to make audio controls more intuitive	2 (25)	2 (33)	2 (40)	1 (20)	2 (33)	1 (20)	0 (0)
Suggestions to improve the filter by topic feature	—	1 (17)	0 (0)	1 (20)	0 (0)	3 (60)	0 (0)
Suggestions to improve the hyperlink feature	—	—	—	1 (20)	1 (17)	2 (40)	0 (0)
Suggestions to improve the filter by text feature	—	—	—	—	—	4 (80)	1 (20)
Suggestions for additional features	2 (25)	3 (50)	3 (60)	2 (40)	3 (50)	2 (40)	4 (80)

[a]R: round.

[b]CG: caregiver role.

[c]P: patient role.

[d]Feature not available or not part of testing.

Participant Views of an Audio PHL

Participants were overwhelmingly positive about the proposed system. Participants identified several benefits of having a visit recording. These included better recall of information for patients and the use of recordings as a historical artifact that could be revisited. For example, a participant said, "Because I had breast cancer 25 years ago and I'd like to go back and say, 'What was that now? What did they say?'" [r3p05]. Communication of clinic visit information to caregivers was considered a significant benefit, as a caregiver whose mother has cognitive issues stated, "it's cumbersome to try and get it from the doctor, so to have that in a place where you can go and access it [would be helpful]" [r4p04]. The added benefit of recording in comparison with written summaries was also mentioned. A participant considered written summaries as *minimal*, whereas another mentioned the ability "to listen to how the doctor said something and how much emphasis he or she was giving..." [r4p09] to clarify instructions; this reflects the added value recordings bring, including the information communicated through voice inflections. The use of hyperlinks was identified as an important feature as "you're getting the information a whole lot quicker than going and sifting through what Google results come up" [r6p03].

Some concerns were also raised. These included the need to keep the UI as user-friendly as possible and not assume "computer capability" [r3p01]. Concerns of information privacy were also raised, "You made this easy for an outside user to access some elements of it – you wouldn't want to let them in the whole thing" [r3p03].

Even when recordings would be shared with family members, patients reported the need for some caution: "I would want to be very careful about inviting relatives to get into the act" [r3p03]. It was mentioned that enabling features that allow partial sharing of recording may reduce this concern, which is not currently a function available in the system. Finally, there was concern from participants about the clinician giving permission for recording:

> *I can't imagine him agreeing to it, my parents' physician...I don't know, it could be used, in an odd way, against the doctor.* [r5p05]

Key Changes to HealthPAL Across Rounds

Textbox 3 outlines key changes made to the system and functionality added in response to user feedback, including color coding and highlighting of information tags, adding more prominent play and pause buttons, creating a clearer structure for switching between user accounts, adding tag filtering and descriptions, adding a 10-second forward and rewind control, a help link, and a search bar. Following our final round of user testing, we made minor modifications to clarify the elements of the interface that should be *clicked* for playback versus hyperlinks to new external information that was approved by our patient partners (RA and SP). Finally, a combination of user feedback and an attempt to optimize our approach to

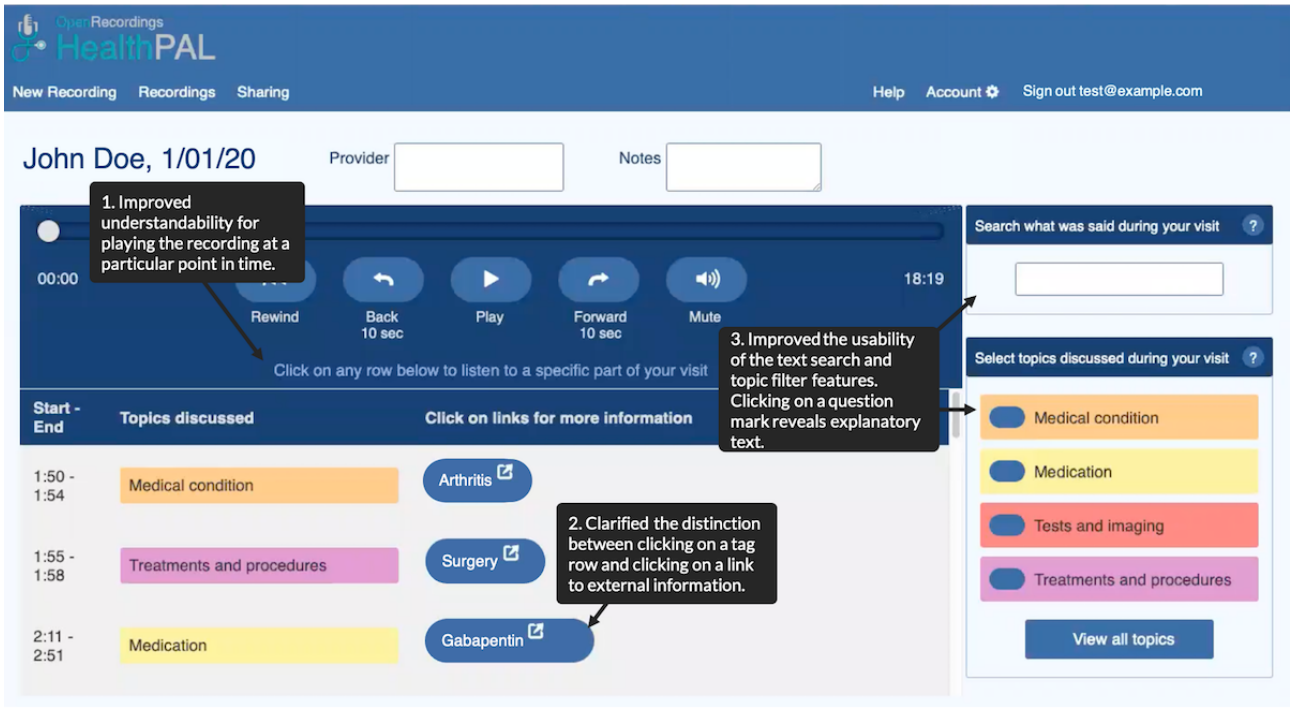
annotations, we reduced the information classes to medication, medical condition, test and imaging, and treatment and procedures.

Textbox 3. Summary of key changes to HealthPAL user interface.

Prototype Description and Updates

- Round 1
  - Initial paper prototype with audio controlled by the facilitator (Figure 3)
  - User quotes
    - "If there's a way to get a finer level of detail [in the topic tags], that would probably be helpful." [r1p03]
    - "[I would like] some sort of indication of where I am in the recording." [r1p05]
  - Round 2
    - Initial software prototype based on feedback from round 1. Included clearer buttons to simplify navigation between recordings, highlighting of audio segments as a place marker, and topic filters to make audio segments easier to find (Figure 4)
    - User quotes
      - "[The topic tags] need to be different colors." [r2p02]
      - "I didn't even look – I looked up and saw the recordings [but not the account name they belonged to]." [r2p05]
- Round 3
  - Updated prototype to include color-coded topic tags, new affordances for selected filters, ahead and back 10-second controls, more prominent play or pause button, clear filters button, and more prominent display of all account names the user has access to (Figure 5)
  - User quotes
    - "I was wondering how I could [turn the audio] off." [r3p05]
    - "Should you have some way to alert this [segment] is [about] the hand, the foot, blood pressure..." [r3p02]
- Round 4
  - Updated prototype to make playback controls more prominent; added external links to MEDLINE Plus for medications, diagnosis, and test results topics; and added help link and popover dialog for filter controls (Figure 6)
  - User quotes
    - "Why is there so much? We have the same things [in the filters] as [in the list of audio segments]...Can we make it so [the filters are] clearly going to help [find audio segments]?" [r4p11]
    - "Put a search in to specifically search for [the desired topic]." [r4p04]
- Round 5
  - Updated prototype to move audio playback closer to top of the window, moved filters to a shopping-like sidebar, added text search or filter, added editable notes field to recordings, renamed the clear filters button, added a mute button, and added a button to go back to the beginning of recording (Figure 7)
  - User quotes
    - "It's a little confusing whether [the text search] is part of [the topic filters]." [r5p03]
    - "I did not see 'Links to external information' so I clicked on [the hyperlink]." [r5p04]
- Final user interface
  - Improved understandability for playing a particular audio segment, clarified the distinction between clicking on a tag-row and clicking on an external information link, improved the usability of the text search and topic filter features, and reduced the number of information classes (Figure 8)

Figure 8. Final version of the software playback user interface.



Discussion

Principal Findings

Following 5 rounds of user design sessions, we iteratively developed a highly usable interface that enables end users to seamlessly interact with desired functions, including playback and sharing of recordings, identifying key segments of the recording, and linking to trustworthy web-based resources. When presented with tasks to find specific topics in a visit recording, participants readily chose to use features (eg, tags, filters, and text search) that helped them more quickly find and play the audio related to that topic by simply listening to the entire recording. Participants were overwhelmingly positive about the concept of accessing a curated audio recording of a clinic visit; however, some participants reported concerns about privacy and the ability of participants to use a computer-based system to access recordings. Although patient partners felt that our final edits addressed confusion about the use of hyperlinks and playback in HealthPAL, further usability testing in less controlled settings is needed.

Comparison With Previous Work

Previous studies of audio or video recording in the health care context have focused primarily on providing a hard copy of a recording to a patient, for example, a CD, cassette, or digital recorder [27]. In more recent developments, commercial apps have emerged that allow the recording, sharing, and tagging of audio recording; however, user-centered design of the app is rarely reported, and concerns about the collection and sharing of patient data have been raised [60].

An exception is *SecondEars*, a recording app developed by Lipson-Smith et al [61] for use by patients receiving cancer treatment. Similar desired features in a recording platform were identified using the MoSCoW (Must Have, Should Have, Could Have, and Won't Have) method in their study. Interestingly, although the SecondEars app focuses on providing a simple recording of oncology clinic visits, patients noted that the ability to link notes to a particular section (ie, minute and second) of an audio recording would be desirable—a unique feature in HealthPAL. The HealthPAL design and our evaluation align with this previous work in that our iterative design incorporated many of these effective meeting browser features: a compressed view of the recording, showing key terms with segment boundaries, and text search of the transcript. Our methodology took learning effects into consideration in our focus on usability for first-time use of the system by including unique patients in each round. In addition, we increased the validity of our findings by asking participants to adopt roles that were reflective of real-world use, that is, participants playing the patient role listened to the recording before using HealthPAL, as patients would be part of the clinic visit in the real world.

It is likely that the inverted *U*-shaped distributions of performance were the result of a younger age demographic and use of paper prototyping in earlier rounds, in addition to the introduction of newer features in the low-fidelity prototype. Although participants became comfortable with most features, some features such as hyperlinks, filters, and the advanced search caused some confusion in the final round of testing. These challenges may be explained by a lack of familiarity with the modern UI design [62], especially in the absence of explicit feedback on actions. Previous usability studies have also reported that, although older adults understand hyperlinks, they can become disoriented when trying to use them [63], and it is unclear which elements of the display can be clicked. It is recommended that hyperlinks appear *touch* interactive [64]. In future iterations of HealthPAL, we will take these additional insights to further improve the usability of the system for older adults before evaluation.

Participants' comments regarding the potential of HealthPAL to improve recall and understanding are supported by previous reviews, which found that sharing recordings can lead to such improvements [28]. However, previous research fails to determine the impact of sharing audio recordings on the ability of patients to manage their own care or the added value of annotated visit recordings. We plan to explore this knowledge gap through a pilot trial of our system.

Limitations

Our sample was predominately White and college educated, reflecting the demographics of the region where our study took place. Further work is needed with individuals from more diverse ethnic and racial groups and from those with lower educational attainment levels. Our project was conducted in controlled settings, where participants were asked to think aloud and received assistance, if needed, with tasks. Although this is important at this stage of user design, it does not reflect the user experience in naturalistic settings. We plan to conduct further field testing in less controlled settings, where clinic visits will be recorded, annotated using our machine learning models, and used at home by patients. During this phase of testing, we will gather information on implementation factors and include clinician feedback. We will also obtain feedback from caregivers using actual clinic recordings of their loved ones' visit. Some participants may have been unclear on the task instructions related to finding additional information (opening tabs outside of the UI), which may have resulted in the higher reported critical incidents in round 5 relative to our other usability metrics, that is, high SUS scores and task completion rate. In addition, our definition of *critical incident* was broad, including any change from the anticipated task path, not only those that resulted in task failure. Finally, we used a laptop computer for all usability testing sessions, but we hope to create a mobile adaptive UI, which will require further testing. Given the simplistic design concept, we believe that the interface can be quickly adapted to a mobile interface.

Conclusions

Sharing visit recordings with patients is an emerging strategy for improving the transparency and communication of visit information. We have developed a highly usable audio PHL, HealthPAL, designed to allow patients and their caregivers to access easy-to-navigate recordings of clinic visits, with key concepts tagged and hyperlinks provided to further information. The interface has been rigorously co-designed with older adult patients and their caregivers and is now ready for further field testing. Our design work has identified and evaluated key features: a tag-based visualization for finer-grained playback of the visit recording coupled with tag-based filtering and text search on audio segments of the recording, which we believe will inform future design of such systems. The successful development and use of HealthPAL may help improve the ability of patients to manage their own care, especially older adult patients who have to navigate complex treatment plans.

Acknowledgments

The research reported in this publication was supported by the National Library of Medicine of the National Institutes of Health under award number R01 LM012815 and the Gordon and Betty Moore Foundation under award number GBMF-4952. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the Gordon and Betty Moore Foundation.

Conflicts of Interest

PJB, WH, MDD, LO, KLB, JCF, SH, AKD, JAS, WMO, MLB, CHG, RA, and SP have no conflicts of interest to declare. GE has edited and published books that provide royalties on sales by publishers, including *Shared Decision Making* (Oxford University Press) and *Groups* (Radcliffe Press). GE's academic interests are focused on shared decision-making and coproduction. He owns a copyright in measures of shared decision-making and care integration, namely, collaboRATE, integRATE (measure of care integration), consideRATE (patient experience of care in serious illness), coopeRATE (measure of goal setting), incorpoRATE (clinician attitude to shared decision-making), Observer OPTION-5, and Observer OPTION-12 (observer measures of shared decision-making). In the past, he provided consultancy for organizations, including (1) Emmi Solutions LLC, which developed patient decision support tools; (2) the National Quality Forum on the certification of decision support tools; (3) Washington State Health Department on the certification of decision support tools; and (4) SciMentum LLC, Amsterdam (workshops for shared decision-making). He is the founder and director of &think LLC, which owns the registered trademark for Option Grids patient decision aids, and the founder and director of SHARPNETWORK LLC, a provider of training for shared decision-making. He provides advice in the domain of shared decision-making and patient decision aids to (1) Access Community Health Network, Chicago (adviser to Federally Qualified Medical Centers), (2) EBSCO Health (consultant), (3) Bind On-Demand Health Insurance (consultant), 4) PatientWisdom Inc (adviser), and (5) abridge AI Inc (chief clinical research scientist).

References

1. Bayliss EA, Ellis JL, Steiner JF. Barriers to self-management and quality-of-life outcomes in seniors with multimorbidities. Ann Fam Med 2007 Sep;5(5):395-402 [FREE Full text] [doi: 10.1370/afm.722] [Medline: 17893380]
2. Henry PC, Caldwell M. Self-empowerment and consumption. Eur J Mark 2006 Sep;40(9/10):1031-1048. [doi: 10.1108/03090560610680998]

3. Watson P, McKinstry B. A systematic review of interventions to improve recall of medical advice in healthcare consultations. J R Soc Med 2009 Jun;102(6):235-243 [FREE Full text] [doi: 10.1258/jrsm.2009.090013] [Medline: 19531618]
4. Ley P. Memory for medical information. Br J Soc Clin Psychol 1979 Jun;18(2):245-255. [doi: 10.1111/j.2044-8260.1979.tb00333.x] [Medline: 454984]
5. Kessels RP. Patients' memory for medical information. J R Soc Med 2003 May 01;96(5):219-222 [FREE Full text] [doi: 10.1258/jrsm.96.5.219] [Medline: 12724430]
6. Jansen J, Butow PN, van Weert JC, van Dulmen S, Devine RJ, Heeren TJ, et al. Does age really matter? Recall of information presented to newly referred patients with cancer. J Clin Oncol 2008 Nov 20;26(33):5450-5457. [doi: 10.1200/jco.2007.15.2322]
7. Butow P, Brindle E, McConnell D, Boakes R, Tattersall M. Information booklets about cancer: factors influencing patient satisfaction and utilization. Patient Educ Couns 1998 Feb;33(2):129-141. [doi: 10.1016/s0738-3991(97)00071-2] [Medline: 9732653]
8. McCarthy D, Waite K, Curtis L, Engel K, Baker D, Wolf M. What did the doctor say? Health literacy and recall of medical instructions. Med Care 2012 Apr;50(4):277-282 [FREE Full text] [doi: 10.1097/MLR.0b013e318241e8e1] [Medline: 22411440]
9. Grady CL, Craik FI. Changes in memory processing with age. Curr Opin Neurobiol 2000 Apr;10(2):224-231. [doi: 10.1016/s0959-4388(00)00073-8] [Medline: 10753795]
10. Bopp KL, Verhaeghen P. Aging and verbal memory span: a meta-analysis. J Gerontol B Psychol Sci Soc Sci 2005 Sep;60(5):223-233. [doi: 10.1093/geronb/60.5.p223] [Medline: 16131616]
11. Brown SC, Park DC. Roles of age and familiarity in learning health information. Educ Gerontol 2010 Nov 10;28(8):695-710. [doi: 10.1080/036012702900997 23]
12. Jones G, Tabassum V, Zarow GJ, Ala TA. The inability of older adults to recall their drugs and medical conditions. Drugs Aging 2015 Apr 1;32(4):329-336. [doi: 10.1007/s40266-015-0255-z] [Medline: 25829296]
13. Wallace E, Salisbury C, Guthrie B, Lewis C, Fahey T, Smith SM. Managing patients with multimorbidity in primary care. Br Med J 2015 Jan 20;350:h176. [doi: 10.1136/bmj.h176] [Medline: 25646760]
14. Moffat K, Mercer SW. Challenges of managing people with multimorbidity in today's healthcare systems. BMC Fam Pract 2015 Oct 14;16(1):129 [FREE Full text] [doi: 10.1186/s12875-015-0344-4] [Medline: 26462820]
15. May C, Montori VM, Mair FS. We need minimally disruptive medicine. Br Med J 2009 Aug 11;339(12):b2803 [FREE Full text] [doi: 10.1136/bmj.b2803] [Medline: 19671932]
16. O'Brien R, Wyke S, Guthrie B, Watt G, Mercer S. An 'endless struggle': a qualitative study of general practitioners' and practice nurses' experiences of managing multimorbidity in socio-economically deprived areas of Scotland. Chronic Illn 2011 Mar 25;7(1):45-59. [doi: 10.1177/1742395310382461] [Medline: 20974642]
17. Hounsell C, Jed JW, Seals CL. Caregiving in the U.S. AARP Research Report. 2015. URL: http://www.aarp.org/content/dam/aarp/ppi/2015/caregiving-in-the-united-states-2015-report-revised.pdf [accessed 2017-04-07]
18. Hummel J, Evans P, Roger CM, Bearden T, McBride S, Deering P, et al. Providing clinical summaries to patients after each office visit: a technical guide. HealthIT.gov. Seattle, Washington: Qualis Health; 2012. URL: http://www.hitecla.org/sites/default/files/Clinical%20Summaries%20Tech%20Guide.pdf [accessed 2021-06-14]
19. Federman A, Sanchez-Munoz A, Jandorf L, Salmon C, Wolf M, Kannry J. Patient and clinician perspectives on the outpatient after-visit summary: a qualitative study to inform improvements in visit summary design. J Am Med Inform Assoc 2017 Apr 01;24(e1):61-68 [FREE Full text] [doi: 10.1093/jamia/ocw106] [Medline: 27497793]
20. Greene J, Hibbard JH. Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes. J Gen Intern Med 2012 May 30;27(5):520-526 [FREE Full text] [doi: 10.1007/s11606-011-1931-2] [Medline: 22127797]
21. Kutner M, Greenburg E, Jin Y, Paulsen C. The health literacy of America's adults: results from the 2003 National Assessment of Adult Literacy. U.S. Department of Education - NCES-483. 2006. URL: https://nces.ed.gov/pubs2006/2006483.pdf [accessed 2021-06-14]
22. Irizarry T, DeVito DA, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
23. Gaston C, Mitchell G. Information giving and decision-making in patients with advanced cancer: a systematic review. Soc Sci Med 2005 Nov;61(10):2252-2264 [FREE Full text] [doi: 10.1016/j.socscimed.2005.04.015] [Medline: 15922501]
24. Pavlik V, Brown AE, Nash S, Gossey JT. Association of patient recall, satisfaction, and adherence to content of an electronic health record (EHR)-generated after visit summary: a randomized clinical trial. J Am Board Fam Med 2014 Mar 07;27(2):209-218 [FREE Full text] [doi: 10.3122/jabfm.2014.02.130137] [Medline: 24610183]
25. Tsulukidze M, Durand M, Barr PJ, Mead T, Elwyn G. Providing recording of clinical consultation to patients - a highly valued but underutilized intervention: a scoping review. Patient Educ Couns 2014 Jun;95(3):297-304. [doi: 10.1016/j.pec.2014.02.007] [Medline: 24630697]
26. Tsulukidze M, Grande SW, Thompson R, Rudd K, Elwyn G. Patients covertly recording clinical encounters: threat or opportunity? A qualitative analysis of online texts. PLoS One 2015 Apr 2;10(5):e0125824 [FREE Full text] [doi: 10.1371/journal.pone.0125824] [Medline: 25933002]

27. Elwyn G, Barr PJ, Grande SW. Patients recording clinical encounters: a path to empowerment? Assessment by mixed methods. BMJ Open 2015 Aug 11;5(8):e008566 [FREE Full text] [doi: 10.1136/bmjopen-2015-008566] [Medline: 26264274]
28. Good DW, Delaney H, Laird A, Hacking B, Stewart GD, McNeill SA. Consultation audio-recording reduces long-term decision regret after prostate cancer treatment: a non-randomised comparative cohort study. Surgeon 2016 Dec;14(6):308-314. [doi: 10.1016/j.surge.2014.10.006] [Medline: 25636362]
29. Ford S, Fallowfield L, Hall A, Lewis S. The influence of audiotapes on patient participation in the cancer consultation. Eur J Cancer 1995 Dec;31(13-14):2264-2269. [doi: 10.1016/0959-8049(95)00336-3]
30. Scott JT, Entwistle VA, Sowden AJ, Watt I. Giving tape recordings or written summaries of consultations to people with cancer: a systematic review. Health Expect 2001 Sep;4(3):162-169 [FREE Full text] [doi: 10.1046/j.1369-6513.2001.00127.x] [Medline: 11493322]
31. McClement SE, Hack TF. Audio-taping the oncology treatment consultation: a literature review. Patient Educ Couns 1999 Mar;36(3):229-238. [doi: 10.1016/s0738-3991(98)00095-0] [Medline: 14528558]
32. Krackow KA, Buyea CM. Use of audiotapes for patient education, medical record documentation, and informed consent in lower extremity reconstruction. Orthopedics 2001 Jul;24(7):683-685. [doi: 10.3928/0147-7447-20010701-22]
33. Santo A, Laizner AM, Shohet L. Exploring the value of audiotapes for health literacy: a systematic review. Patient Educ Couns 2005 Sep;58(3):235-243. [doi: 10.1016/j.pec.2004.07.001] [Medline: 16054796]
34. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006 Mar;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]
35. Price M, Bellwood P, Kitson N, Davies I, Weber J, Lau F. Conditions potentially sensitive to a personal health record (PHR) intervention, a systematic review. BMC Med Inform Decis Mak 2015;15:32 [FREE Full text] [doi: 10.1186/s12911-015-0159-1] [Medline: 25927384]
36. Luo G, Tang C, Thomas SB. Intelligent personal health record: experience and open issues. J Med Syst 2012 Aug;36(4):2111-2128. [doi: 10.1007/s10916-011-9674-5] [Medline: 21399913]
37. Bayati M, Braverman M, Gillam M, Mack KM, Ruiz G, Smith MS, et al. Data-driven decisions for reducing readmissions for heart failure: general methodology and case study. PLoS One 2014 Oct 8;9(10):e109264 [FREE Full text] [doi: 10.1371/journal.pone.0109264] [Medline: 25295524]
38. Hassanpour S, Langlotz CP. Predicting high imaging utilization based on initial radiology reports: a feasibility study of machine learning. Acad Radiol 2016 Jan;23(1):84-89. [doi: 10.1016/j.acra.2015.09.014] [Medline: 26521688]
39. Shin H, Markey MK. A machine learning perspective on the development of clinical decision support systems utilizing mass spectra of blood samples. J Biomed Inform 2006 Apr;39(2):227-248 [FREE Full text] [doi: 10.1016/j.jbi.2005.04.002] [Medline: 15963765]
40. Roehrs A, da Costa CA, da Rosa RR, de Oliveira KS. Personal Health Records: A systematic literature review. J Med Internet Res 2017 Jan 06;19(1):e13 [FREE Full text] [doi: 10.2196/jmir.5876] [Medline: 28062391]
41. Clarke MA, Moore JL, Steege LM, Koopman RJ, Belden JL, Canfield SM, et al. Health information needs, sources, and barriers of primary care patients to achieve patient-centered care: a literature review. Health Informatics J 2016 Dec;22(4):992-1016 [FREE Full text] [doi: 10.1177/1460458215602939] [Medline: 26377952]
42. Whittaker S, Tucker S, Lalanne D. Meeting browsers and meeting assistants. In: Popescu-Belis A, Bourlard H, Carletta J, Renals S, editors. Multimodal Signal Processing: Human Interactions in Meetings. Cambridge, United Kingdom: Cambridge University Press; 2012:204-217.
43. Abowd GD. Classroom 2000: An experiment with the instrumentation of a living educational environment. IBM Syst J 1999;38(4):508-530. [doi: 10.1147/sj.384.0508]
44. Wellner P, Flynn M, Tucker S, Whittaker S. A meeting browser evaluation test. In: Extended Abstracts on Human Factors in Computing Systems. 2005 Presented at: CHI05: CHI 2005 Conference on Human Factors in Computing Systems; April 2-7, 2005; Portland OR USA p. 2021-2024. [doi: 10.1145/1056808.1057082]
45. Basu S, Gupta S, Mahajan M, Nguyen P, Platt J. Scalable summaries of spoken conversations. In: Proceedings of the 13th International Conference on Intelligent User Interfaces. 2008 Presented at: IUI08: 13th International Conference on Intelligent User Interfaces; January 13-16, 2008; Gran Canaria Spain p. 267-275. [doi: 10.1145/1378773.1378809]
46. Yen P, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012 May;19(3):413-422 [FREE Full text] [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
47. Callahan CM, Unverzagt FW, Hui SL, Perkins AJ, Hendrie HC. Six-item screener to identify cognitive impairment among potential subjects for clinical research. Med Care 2002 Sep;40(9):771-781. [doi: 10.1097/00005650-200209000-00007] [Medline: 12218768]
48. Nielsen J, Molich R. Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1990 Presented at: CHI90: Conference on Human Factors in Computing; April 1-5, 1990; Seattle Washington USA p. 249-256. [doi: 10.1145/97243.97281]
49. Lewis JR. Sample sizes for usability studies: additional considerations. Hum Factors 1994 Jun;36(2):368-378. [doi: 10.1177/001872089403600215] [Medline: 8070799]

50. van den Haak M, De Jong M, Schellens P. Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue. Behav Inf Technol 2003 Sep;22(5):339-351. [doi: 10.1080/0044929031000]
51. Lewis C. Using the "Thinking-Aloud" method in cognitive interface design. IBM TJ Watson Research Center. 1982. URL: https://www.worldcat.org/title/using-the-thinking-aloud-method-in-cognitive-interface-design/oclc/38214134 [accessed 2021-06-14]
52. Kelley JF. An iterative design methodology for user-friendly natural language office information applications. ACM Trans Inf Syst 1984 Jan;2(1):26-41. [doi: 10.1145/357417.357420]
53. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Stud 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
54. Flanagan JC. The critical incident technique. Psychol Bull 1954 Jul;51(4):327-358. [doi: 10.1037/h0061470] [Medline: 13177800]
55. Sauro J. Measuring usability with the System Usability Scale (SUS). MeasuringU. 2011. URL: https://measuringu.com/sus/ [accessed 2020-10-01]
56. Brooke J. SUS: A 'Quick and Dirty' usability scale. In: Usability Evaluation In Industry. Boca Raton, Florida, United States: CRC Press; 1996:1-6.
57. Brooke J. SUS: A retrospective. J Usability Stud 2013;8(2):29-40. [doi: 10.5555/2817912.2817913]
58. Zhang J, Walji MF. TURF: Toward a unified framework of EHR usability. J Biomed Inform 2011 Dec;44(6):1056-1067 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.005] [Medline: 21867774]
59. HealthPAL. GitHub, Inc. URL: https://github.com/openrecordings/healthpal [accessed 2021-09-01]
60. Elwyn G, Barr PJ, Piper S. Digital clinical encounters. Br Med J 2018 May 14;361:k2061. [doi: 10.1136/bmj.k2061] [Medline: 29760149]
61. Lipson-Smith R, White F, White A, Serong L, Cooper G, Price-Bell G, et al. Co-design of a consultation audio-recording mobile app for people with cancer: the secondears app. JMIR Form Res 2019 Mar 12;3(1):e11111 [FREE Full text] [doi: 10.2196/11111] [Medline: 30860487]
62. Dodd C, Athauda R, Adam M. Designing user interfaces for the elderly: a systematic literature review. In: Proceedings of the Australasian Conference on Information Systems (ACIS 2017). 2017 Presented at: ACIS 2017: Australasian Conference on Information Systems; December 4-6, 2017; Hobart, Tas p. A.
63. Ziefle M, Schroeder U, Strenk J, Michel T. How younger and older adults master the usage of hyperlinks in small screen devices. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2007 Presented at: CHI07: CHI Conference on Human Factors in Computing Systems; April 28-May 3, 2007; San Jose California USA p. 307-316. [doi: 10.1145/1240624.1240676]
64. McLaughlin A, Pak R. Designing Displays for Older Adults. Boca Raton, Florida, United States: CRC Press; 2010:1-211.

Abbreviations

AVS: after-visit summary
D-H: Dartmouth-Hitchcock
DHMC: Dartmouth-Hitchcock Medical Center
MoSCoW: Must Have, Should Have, Could Have, and Won't Have
NLP: natural language processing
PHL: personal health library
SUS: System Usability Scale
TURF: Task, User, Representation, and Function
UI: user interface

Edited by R Kukafka; submitted 06.11.20; peer-reviewed by M Bestek, A Hyatt; comments to author 19.12.20; revised version received 01.03.21; accepted 27.04.21; published 22.10.21.

Please cite as:
Barr PJ, Haslett W, Dannenberg MD, Oh L, Elwyn G, Hassanpour S, Bonasia KL, Finora JC, Schoonmaker JA, Onsando WM, Ryan J, Bruce ML, Das AK, Arend R, Piper S, Ganoe CH
An Audio Personal Health Library of Clinic Visit Recordings for Patients and Their Caregivers (HealthPAL): User-Centered Design Approach
J Med Internet Res 2021;23(10):e25512
URL: https://www.jmir.org/2021/10/e25512
doi:10.2196/25512
PMID:34677131

Original Paper

Engaging Institutional Stakeholders to Develop and Implement Guidelines for Recruiting Participants in Research Studies Using Social Media: Mixed Methods, Multi-Phase Process

Elizabeth Flood-Grady[1,2], PhD; Lauren B Solberg[3], JD, MTS; Claire Baralt[4], MS; Meghan Meyer[2], MBA; Jeff Stevens[5], MAMC; Janice L Krieger[1,2], PhD

[1]STEM Translational Communication Center, College of Journalism and Communications, University of Florida, Gainesville, FL, United States

[2]Clinical Translational Science Institute, University of Florida, Gainesville, FL, United States

[3]Department of Community Health and Family Medicine, University of Florida, Gainesville, FL, United States

[4]University of Florida Health, University of Florida, Gainesville, FL, United States

[5]University of Florida Health Web Services, Communications & Information Technology, University of Florida, Gainesville, FL, United States

Corresponding Author:
Elizabeth Flood-Grady, PhD
STEM Translational Communication Center
College of Journalism and Communications
University of Florida
PO Box 118400
Gainesville, FL, 32611
United States
Phone: 1 352 273 0240
Email: efloodgrady@ufl.edu

Abstract

Background: Limited regulatory guidance surrounding the use of social media channels for participant recruitment is an interdisciplinary challenge. Establishing stakeholder-informed procedures is essential for ethical and effective use of social media for participant recruitment.

Objective: This study aims to provide replicable procedures for developing and implementing guidelines for using social media to recruit participants in research studies.

Methods: Social media use cases at the university were used to identify institutional stakeholders for the initiative. After establishing workflow procedures, a scoping review of web-based materials about recruitment and research on the internet and social media from 19 peer institutions and 2 federal agencies was conducted to inform the structure of the policies and procedures. End users (investigators and study coordinators; N=14) also provided feedback on the policies and procedures and implementation.

Results: Representatives (n=7) from 5 institutional offices and 15 subject-matter experts from 5 areas were identified as stakeholders in the development of policies and procedures. Peers with web-based materials (n=16) identified in the scoping review revealed 4 themes that served as a basis for developing our policies and procedures. End user feedback further informed the policies and procedures and implementation. A centrally managed social media account for communicating with participants and hosting advertising campaigns on social media was also established and, when combined with the policies and procedures, resulted in 39 advertising campaigns, and 2846 participants were enrolled in health and clinical research studies.

Conclusions: Our policies and procedures allow research teams to harness the potential of social media to increase study recruitment and participation; the transparent, stakeholder-informed process can be replicated by institutional administrators to establish policies and procedures that meet the interests and needs of their research community.

(J Med Internet Res 2021;23(10):e23312) doi:10.2196/23312

KEYWORDS

social media; research recruitment; stakeholder engagement; health communication

Introduction

Background

Social media hold great potential for recruiting prospective participants (ie, identifying, reaching, and advertising studies) into research studies (eg, social and behavioral research, observational studies, and clinical trials) [1]. These channels enable the dissemination of rich information, such as pictures, videos, and links to news articles, which plays an important role in attracting attention and engaging large audiences in dialog about health [2,3]. A recent study of Clinical Translational Science Association web-based strategies for communicating about clinical research participation found that among Clinical Translational Science Associations that hosted separate participant-facing websites, half of those sites linked to social media accounts [4], presumably to direct community members to additional information about research.

Research teams also engage in social media directly for study recruitment. For instance, researchers used Twitter to recruit mothers to participate in a web-based study by tweeting their study invitation and asking users to retweet that information [5]. Teams also used Facebook advertising campaigns to cost-effectively recruit women into human papillomavirus vaccine effectiveness studies [6], screen and recruit adult smokers and heavy-drinking smokers into treatment research [7], and enroll rural adults into health care intervention development studies [8]. Social media are also cost-effective and successful channels for recruiting certain hard-to-reach and underrepresented populations in clinical research [8-10].

Despite the potential of social media as a recruitment tool, there is limited federal guidance for those interested in using social media for research [11], for study recruitment [12], or for institutional review boards (IRBs) charged with reviewing protocols. The federal regulations governing the review and conduct of research with human subjects that went into effect in 2018 [13,14] do not specifically reference the use of social media in research. Federal agency recommendations on the use of social media tools to conduct research remain minimal [11,15]. For instance, in 2013, the Secretary's Advisory Committee on Human Research Protection published considerations and recommendations on using the internet to conduct human subjects research, and a National Institutes of Health page with guidance regarding social media tools was last updated in 2016 [16,17]. Institutions are thus left to develop policies and procedures that are consistent with federal and state laws on the ethical use of social media in research, and more specifically, the use of social media for research recruitment.

Developing policies and procedures for recruiting research participants via social media that are consistent with applicable law and ethical practice, respect potential participants and their privacy, and are accessible to end users and those enforcing its implementation is a challenge. In an era where cybersecurity, web-based privacy, and terms and conditions of social media are routinely in the news, institutions and their researchers are right to approach this process with caution. Engaging institutional stakeholders involved in research recruitment is a critical step in establishing informed social media recruitment policies and procedures. Stakeholder engagement (SE) relies on multiple methods to identify agreement and disagreement among individuals (stakeholders) affected by decision-making and determine the potential reasons for these differences [18]. Incorporating stakeholder perspectives to develop policies and procedures promotes transparency, increases the quality and trustworthiness of the information, and helps facilitate the implementation of policies and procedures as intended [18-20], ultimately benefiting patients [19]. SE is particularly beneficial when the individuals involved face a similar problem, acknowledge the issue exists, and organize to fix it [21]. In other words, engaging institutional stakeholders involved and affected by social media research and recruitment decision-making should ensure that the process, implementation, and resulting policies and procedures are informed, transparent, and credible.

Objective

Our goal is to provide replicable procedures for developing and implementing guidelines, which for the purpose of clarity, we refer to as policies and procedures, for using social media to recruit participants into research studies. This process and the resulting policies and procedures can be used by researchers and administrators to develop and implement policies and procedures that should be tailored to the laws, policies, and practices applicable to other institutions.

Methods

Overview

During the fall of 2016, the institution identified the need for a more coordinated approach to address privacy, information security, and other questions pertaining to IRB submissions that included increasingly sophisticated use cases for social media. The social media initiative was established in response to this need. The initiative was led by the institution's Clinical and Translational Science Institute (CTSI). No aspect of this paper represents an official institutional position on social media research and recruitment or the development of policies and procedures. Below, we describe our methods for completing the 6 phases of the initiative.

Phase 1: Identify Key Stakeholders

SE is the process of engaging individuals, groups, or members of an organization who are (actual) or may be (potential) affected by decision-making or who can influence implementation [18]. Therefore, the initial step was to identify individuals at the institution who were involved in and affected by social media recruitment and research decision-making. Modeled after the CTSI Scientific Advisory Committee at the university, the Directors of Research Services and Strategy and Planning for the institution's CTSI (author) used protocols requesting the use of social media in recruitment and research that were under review by the IRB as use cases to identify institutional stakeholders. Use case protocols reflected interest in using an application through Facebook to conduct an intervention through the social media site, contracting a third-party vendor to facilitate social media advertisements for study recruitment, and requesting to use social media channels to recruit participants into research studies.

Phase 2: Establish Workflow Procedures

Next, the Directors of Research Services and Strategy and Planning for the institution's CTSI convened stakeholders to establish workflow procedures for developing timely social media policies and procedures. Stakeholder concerns, interests, and workflow needs surrounding social media research and recruitment at the institution were discussed and documented. Communication about the policies and procedures workflow that occurred via email (eg, questions and responses) between the initial (phase 1) and follow-up meetings were documented and included as data. Stakeholder input was synthesized and organized into workflow categories.

Phase 3: Complete Benchmarking and Scoping Review of Existing Peer Guidance

After establishing workflow procedures, a scoping review of existing web-based policies and procedures surrounding social media research and recruitment at our institution and across several peer institutions and federal agencies was completed. Scoping reviews identify available types and sources of evidence and key concepts in a particular research area, determine the extent, range, and nature of existing research activities, summarize and disseminate research findings, and identify existing research gaps across the literature and policies [22]. The goal of this scoping review was to summarize the key elements of existing social media recruitment policies and procedures, map consistencies or inconsistencies and gaps across available policies and procedures, and identify use cases to serve as a baseline for developing policies and procedures at our institution. Benchmarking with peer institutions provided an opportunity to incorporate the perspectives of peer stakeholders with experience and knowledge in this area to shape the direction and scope of future social media endeavors.

As a large, publicly funded, medical research institution, we engaged in a web search to identify materials from 19 peer medical research institutions and 2 federal agencies (N=21). Guidelines, recommendations, including templates and best practices, and policies identified on the 21 websites and corresponding landing pages that addressed the *internet* or *social media* or *social networking sites* as *venues* or *channels* or *platforms* for conducting research or participant recruitment in their scope were considered relevant and included as data. Thematic analysis techniques were used to identify, organize, and refine themes from the data [23].

Phase 4: Develop, Review, and Refine Social Media Recruitment Policies and Procedures

The themes identified in the scoping review were used to draft the policies and procedures. Policies and procedures have also been developed in alignment with existing institutional social media policies. The proposed policies and procedures underwent a rigorous yearlong review and refinement by stakeholders identified in phase 1 received feedback from potential users.

Phase 5: Seek Feedback from End Users

This phase was modeled after the focus or working group method, wherein end users provided feedback on the proposed policies and procedures [22]. The intended audience and primary end users of the policies and procedures were investigators (n=6) and study coordinators (n=8) at the institution who participated in separate, small group discussions where they read a comprehensive draft of the policies and procedures and provided feedback on the content, structure, and implementation. Predetermined questions were used to encourage open-ended discussions about the policies and procedures and to obtain information about stakeholder preferences and opinions in a relaxed, nonthreatening environment [22].

Phase 6: Policies and Procedures Approval and Implementation

In December 2017, the proposed policies and procedures were submitted to senior university administration for potential endorsement and guidance on which institutional officials would need to approve the policies and procedures before implementation. A process for implementing and evaluating policies and procedures at the university was also included in the proposal.

Results

Overview

Early in the initiative, it became clear that the most common social media use cases encountered by the IRB were requests to use social media channels for research recruitment. *Therefore, the focus of the initiative was refined to develop policies and procedures to use social media to recruit research participants.* The results of the six phases of the initiative are presented below.

Phase 1: Identify Key Stakeholders

Using social media to recruit research participants poses a number of ethical, legal, and practical implications for an institution. Seven representatives from 5 institutional offices were identified as stakeholders. In total, 15 subject-matter experts—those with established credibility and expertise in areas salient to the policies and procedures [22,24] were also identified to participate in the initiative. Stakeholders were ultimately divided into two groups: the Social Media in Research Committee and the Health Communication Social Media in Research Task Force. Table 1 includes office or title, institutional representatives, and subject-matter experts who participated in the initiative and a rationale for their inclusion. Stakeholders are further labeled by their participation as either committee or task force group members. The Director of Strategy and Planning and Director of Research Services for the CTSI co-led the initiative, participated as members of the task force, and were included as subject-matter experts (n=15).

Table 1. Stakeholders who participated in the initiative.

Office or title	Group	Rationale for inclusion
IRB[a]		
• Chair of Health Sciences, Medical Research • Chair and Vice Chair of Social, Behavioral, Educational Research	Committee	• By federal regulation, federal guidance, and institutional policy, the IRB reviews recruitment materials for accuracy and to ensure the content presented is not coercive [25].
Privacy office		
• Director of Privacy	Committee	• Representatives who oversee privacy understand the dynamic, technical aspects of social media platforms and applicable laws (eg, Health Insurance Portability and Accountability Act) pertaining to the privacy and security of an individual's information, whose data might be viewed or accessed in the process of advertising via social media.
• Information Security Office: Information Security Manager	Committee	• Those who oversee security and information technology also understand the dynamic, technical aspects of social media platforms and applicable laws (eg, Health Insurance Portability and Accountability Act) pertaining to the privacy and security of an individual's information, whose data might be viewed or accessed in the process of advertising via social media.
• General Counsel's Office: Senior University Counsel for Health Affairs	Committee	• Representatives from general counsel understand the dynamics surrounding compliance and conduct with the site's terms of use or potential infringements on social media users' rights to free speech by removing comments on posted advertisements. Concerns about compliance and conduct represented intersecting stakeholder interests across general counsel, privacy, and security.
Office of Research		
• Director of Research Operations and Services	Committee	• Representatives from research administration leadership ensured that the scope and resulting policies and procedures were generally consistent with the university's mission and goals.
Communications professionals		
• Associate Director of Communications • Communication Specialist; Assistant Manager for Web Services	Task force	• Professionals in marketing and strategic communications, and those with social media expertise ensured policies and procedures dovetailed with existing guidance (eg, approval and governance of communications channels) and best practices for external communications representing the institution (eg, adherence to institutional brand standards). • Communications professionals also anticipated the technical and practical how-to aspects of implementation, and advised on unique considerations, such as how audiences historically interacted with institutional social media.
Health Communication scientists	Task force	• Research scientists with expertise in communication and health and science translation ensured the policies and procedures were accessible, understandable, and usable by intended audiences.
Bioethics and legal experts	Task force	• Experts in medical professionalism and clinical and research ethics ensured policies and procedures addressed the ethical and regulatory considerations when using social media in research.
Recruitment and community engagement specialists	Task force	• Individuals directly involved in study development and day-to-day recruitment activities informed the clarity and implementation of the policies and procedures.
Regulatory navigators	Task force	• Individuals directly involved in study development and day-to-day recruitment activities informed the clarity and implementation of the policies and procedures.
Investigators and research coordinator end users	Task force	• Individuals directly involved in study development and day-to-day recruitment activities informed the clarity and implementation of the policies and procedures.

[a]IRB: institutional review board.

Phase 2: Establish Workflow Procedures

Committee and task force members agreed on the importance of establishing a central process for developing timely policies and procedures. Stakeholders met and established workflow procedures, which were organized into four categories of objectives: (1) establish a project plan, (2) define the scope of planned policies and procedures, (3) address privacy concerns, and (4) contribute to research. Table 2 includes descriptions of the four workflow categories and strategies identified to achieve the objectives.

Table 2. Objectives, strategies, and descriptions of workflow procedures.

Objective	Description	Strategies identified for achieving objectives
Establish a project plan	Identify relevant tasks associated with developing and implementing policies and procedures and assign responsibilities to stakeholders	• Identify information needed to make decisions on social media protocols pending approval (IRBa) • Establish target timeline for developing the policies and procedures (Office of Research) • Complete peer benchmarking to identify available types and sources of web-based policies and procedures from peer research institutions (task force) • Identify Taskforce personnel responsible for drafting policies and procedures and establish a coordinated process for receiving committee feedback were also identified as relevant tasks • Identify the person or office responsible for approving and implementing social media policies and procedures at the institution (committee) • Develop social media recruitment templates for submitting IRB protocols and developing theoretically derived advertisements [8], and a risk matrix with examples of high and low-risk social media recruitment activities and the level of review required for each social media recruitment activity (task force and committee) • Establish a cost structure for services (task force)
Define scope of planned policies and procedures	Identify subject matter included in the policies and procedures and policy and procedure classification	• Determine if policies and procedures will address social media as a channel for advertising studies and recruiting prospective participants exclusively, or if it will also include social media as a channel for hosting studies, interventions, or data collection • Decide if policies and procedures will serve as a formal policy, process, or best practice recommendations
Address privacy concerns	Ensure policies and procedures respect prospective participants' rights to privacy and respond to potential issues of privacy and security for the institution (eg, the university, health system, and institutional researchers)	• Address social media users' reasonable expectations of privacy • Address potential concerns about the collection of sensitive and private information on the internet (eg, what is considered Protected Health Information and de facto private and what is private on the internet) • Identify steps for responding and managing the exchange of potentially sensitive information on social media • Assess institutional risks and liabilities that may result from developing and implementing social media recruitment policies and procedures • Identify any legal or ethical obligations to explain what happens when users click on social media recruitment links in advertisements • Identify how to address the changing terms of use and privacy policies on social media sites and ensure teams comply with applicable policies
Contribute to research	Enhance public participation in research and disseminate policy and procedure recommendations and resulting research findings	• Increase interest and understanding about research participation and generate recruitment leads to increase participation in institutional studies • Identify and track metrics for evaluating social media efficacy across individual and multisite studies • Develop evidence-based practices relating to recruitment through social media and publishing academic manuscripts on the policies and procedures process and campaign results

[a]IRB: institutional review board.

Phase 3: Complete Benchmarking and Scoping Review of Existing Peer Guidance

Members of the Taskforce completed peer benchmarking through a scoping review. At the time the scoping review was conducted, most institutions (16/21, 76.2%) had policies or guidelines web-based surrounding the use of the internet and social media to recruit research participants or to conduct research and collect research data. Four themes emerged across existing peer policies and guidelines regarding recruitment and

research on the internet and social media, as identified inductively by the task force. These themes address (1) compliance with platform terms of use; (2) social media as a tool for participant recruitment; (3) participant privacy, confidentiality, and data security; and (4) social media as a venue for research. Textbox 1 presents themes and descriptions identified in the scoping review.

Textbox 1. Themes and descriptions identified in the web-based scoping review of peer policies and procedures.

Theme name and description

- Compliance with platform terms of use: focused on the importance of complying with existing site terms of agreement, terms of use, university policies, and adhering to applicable laws when engaging social media for recruitment and research. Understanding and complying with site terms were described as a joint responsibility between investigators and institutional review boards (IRBs).
- Social media as tool for participant recruitment: described recruitment as direct and indirect communication with prospective participants to advertise a study, including posts or paid study advertisements and 2-way communication via researcher-initiated social media accounts to potential research participants through direct messages. Few policies and guidance included steps associated with purchasing and placing study recruitment advertisements. Developing and reviewing recruitment materials was positioned as joint responsibility of investigators and IRBs.
- Participant privacy, confidentiality, and data security: addressed the need to explain privacy and data security processes (eg, how the data are transmitted and maintained on the web), to address the potential risks to using social media, and to acknowledge the limits to confidentiality while emphasizing steps in place to uphold it. Privacy, confidentiality, and data security were positioned, primarily, as responsibilities of investigators or research teams. Teams were also tasked with explaining this information to the IRB in their protocols and to potential research participants (if interested).
- Social media as venue for research: described social media as a channel for hosting research (ie, the location where the observation or intervention will take place), including directly initiating contact with participants and in some cases, studying information about participants that is available without direct contact.

Phase 4: Develop, Review, and Refine Social Media Recruitment Policies and Procedures

Members of the task force drafted the policies and procedures based on the stated goals and needs of the committee. The committee provided feedback and input during the drafting process and ultimately determined when the policies and procedures could be presented to senior university administration for review and implementation. The three main themes (sections) in our social media recruitment policies and procedures include (1) compliance with platform terms of use (Table 2); (2) participant privacy, confidentiality, and data security (Table 2); and (3) procedures and considerations for using social media to recruit participants.

Theme 3, procedures and considerations for using social media to recruit participants, described the roles and responsibilities of institutional offices and research teams interested in using social media for recruitment; the permissible types, strategies, and considerations for recruiting participants on social media; and mandatory information to include in the social media management plan. Textbox 2 includes a description of the subcategories included in the procedures and considerations theme. The complete policies and procedures can be found in Multimedia Appendix 1 [14,16,26-30].

Textbox 2. Descriptions of social media recruitment procedures and considerations addressed in policies and procedures.

Section 3 subcategories in policies and procedures

- Review of social media recruitment content: the role of the institutional review board (IRB) in reviewing and approving social media recruitment materials (eg, the IRB will review the content of social media recruitment materials in accordance with existing IRB guidelines for traditional media recruitment, such as flyers and news advertisements) as well as the process for research teams submitting recruitment materials to social media sites for approval and posting study materials on social media was described.
- Permissible types of social media recruiting: this section outlined specific options for interacting with prospective participants on social media for recruitment, including static and interactive recruitment materials, recruiting participants via public and private groups, and considerations for private messaging. The responsibilities of research teams interested in recruiting via public and private groups and two-way communication between research team members and prospective participant on social media (eg, private messaging on Facebook and Twitter direct messages) were also included in this section).
- Hosting social media recruitment campaigns: this section described the social media account options for teams to use to facilitate recruitment advertising campaigns on social media, restrictions surrounding the use of personal accounts to purchase and place initial recruitment materials for studies, and parameters for sharing study recruitment materials posted through official university-approved accounts to personal pages and accounts. Details surrounding use of the official *UF Studies* Facebook page, an official account managed by the CTSI Recruitment Center was also included.
- Screening participants and tracking recruitment: this section addressed the procedures for screening prospective participants recruited through social media (for instance, all screening and data collection must occur off social media, such as via phone call or via a secure, university-approved platform, eg, REDCap [Research Electronic Data Capture]), incorporating a waiver of consent for survey studies, and if applicable, a description of the research team's plan for storing and using any identifiable data collected.
- Developing a social media management plan: the mandatory information required in the social media management plan as part of the IRB protocol's recruitment plan and that plans must be approved by the IRB before starting recruitment were described in this section.

Phase 5: Seek Feedback From End Users

During this phase, end users provided general feedback on the usability of policies and procedures for recruiting prospective participants, and the potential for policies and procedures to inform their current recruitment and research practices. End users provided feedback on the scope, clarity, and implementation of the proposed policies and procedures. End users suggested incorporating a definition of personal health information with regard to recruiting on social media. The Director of Privacy provided the Health Insurance Portability and Accountability Act of 1996 definition of individually identifiable information that was provided to the Health System Communications Office in the past for inclusion in the policies and procedures. To increase clarity, end users asked for a list of university-approved resources as they pertain to recruiting on social media (eg, REDCap [Research Electronic Data Capture] for screening) and for additional information regarding sharing recruitment materials from personal accounts. The need for transparent workflow processes and protocol review procedures for teams submitted to study protocols was discussed as part of the implementation. This feedback, along with suggestions for improving readability (eg, removing acronyms, jargon) were incorporated into the policies and procedures and used to inform its implementation.

Phase 6: Policies and Procedures Approval and Implementation

Policies and Procedures Approval and Implementation Proposed Process

The integrated risk management group at the institution endorsed the policies and procedures and proposed a process for implementation. Policies and procedures were hosted on the Office of Research webpage and the IRB website and the CTSI Recruitment Center's Research Resources webpage linked to the policies and procedures and related resources. The IRB enforced guidelines with support from the Committee and the CTSI Recruitment Center. The CTSI Recruitment Center activated the workgroup and committee structure to facilitate ancillary review of complex recruitment strategies and other social media use cases that fall outside the guidelines and provide consults and services for investigators regarding social media recruitment. The CTSI Recruitment Center also evaluated the 1-year pilot of the policies and procedures and the *UF Studies* Facebook page to assess the effectiveness and resources required for long-term maintenance.

Official UF Studies Facebook Page

The *UF Studies* Facebook page, an official account managed by the CTSI Recruitment Center, was established as channel through which teams were permitted to post recruitment materials. The committee and task force collaborated to develop, secure approval for, and launch the *UF Studies* Facebook page, an official account managed the CTSI Recruitment Center. The page served as a centrally managed resource that was made available for both education and recruitment purposes. Community members and prospective participants were the primary audience, and page objectives included educating the community about research and disseminating information about health and science, including research frequently asked questions, results from institutional studies, opportunities to participate in studies, or join established institution-affiliated research registries.

Researchers interested in using social media for study recruitment were the secondary audience. The policies and procedures provided an option for investigators to work with the CTSI Recruitment Center to advertise studies through this page using a streamlined process or follow the policies and procedures for advertising through other official social media accounts approved by the institution. Social media recruitment templates for submitting IRB protocols, developing theoretically derived advertisements, and tracking recruitment metrics and a risk matrix to determine high- and low-risk social media recruitment activities were developed as part of a toolkit for investigators using this central resource. The full policies and procedures, including a risk matrix, is included in Multimedia Appendix 1. Social media recruitment templates are included in Multimedia Appendix 2 and elsewhere [8].

Results of Policies and Procedures and UF Studies Facebook Page Implementation

To date, the CTSI Recruitment Center has managed 39 Facebook advertising campaigns for research teams at the institution. A total of 2846 participants were enrolled in health and clinical research studies as a direct result of implementing the social media recruitment policies and procedures and *UF Studies* Facebook pages.

Discussion

Principal Findings and Implications

We developed policies and procedures surrounding the recruitment of research participants on social media because our institutional use cases overwhelmingly reflected requests to use social media as a tool for research recruitment. The SE framework was used to identify and engage representatives from key institutional offices and subject-matter experts to facilitate the development and implementation of our policies and procedures. The results include a replicable process for establishing and implementing policies and procedures, an official channel for research teams to use for study recruitment and communicating with prospective participants about research studies, and enrollment of over 2800 participants into health and clinical research studies.

The SE framework was beneficial for identifying and understanding the needs, interests, and perspectives of diverse stakeholders and influencing multiple interconnected systems at our institution to develop policies and procedures. This theoretical process enabled the realization of similar interests, needs, and concerns among stakeholders. Stakeholders were united by their commitment to protect the rights of participants and to establish and maintain ethical research practices at the university, while being responsive to the needs and interests of research teams and advances in technology. Collaboration among stakeholders to develop key considerations for social media recruitment policies and procedures is an important step in establishing trust.

The central channel for running study recruitment campaigns (ie, placing, hosting, purchasing, and tracking), social media recruiting templates for teams and IRB personnel, and prioritizing research education and dissemination of health and science information to the local community through the central channel reflect the practical outcomes produced through the initiative. The central channel benefits research teams who are interested in using social media for recruitment but who may lack the resources needed to develop and manage study recruitment campaigns effectively. The theoretically driven templates for advertising studies ensure that recruitment messages resonate with intended populations [8], and the investigator templates facilitate the timely submission and review of IRB protocols, benefiting researchers and those charged with reviewing social media protocols. Finally, because institutions use social media to communicate with the public about health topics [31] and social media can generate interest and awareness in research [32], including educating the community about research participation as a top priority could increase research literacy and understanding of the research process, ultimately leading to more informed research participants.

Finally, it is important to remain vigilant to changes in social media terms of use and agreement through a continuous review of the social media channel procedures teams used for study recruitment. Indeed, institutions can and should expect researchers to comply with terms of the platform, IRB, and other regulations, and other applicable laws, policies, or guidelines. Developing and maintaining an established communication structure for responding to complex recruitment strategies, social media use cases that fall outside the guidelines, and the changing terms of use of social sites are important for minimizing risk factors and enforcing ethical research recruitment.

Limitations and Future Directions

Our policies and procedures address study recruitment on social media, although the bulk of our resources aim to streamline study recruitment on Facebook. We focused on Facebook because it is the largest and most popular social networking site [33], and developing resources for multiple social channels would be labor- and time-intensive. Before developing resources for other social media channels, institutions should identify the diversity of audiences and reach of other channels, and whether a particular site has broad enough reach to merit, devoting that level of resources. Due to the increased number of third-party services (eg, TrialSpark) available for managing participant recruitment across multiple social media sites, institutions would benefit from establishing policies and procedures for using outside vendors for recruitment.

Conclusions

Our goal was to provide limited parameters and procedures for establishing policies and procedures to recruit participants in research studies using social media. The stakeholder-informed, replicable policies and procedures allow research teams to harness the potential of social media to increase study participation and recruitment, while simultaneously managing the potential risks associated with the ubiquity of these channels. Both theory and practice contributed to the development of policies and procedures, enabling a transparent process for administrators at other institutions to replicate and establish policies and procedures that meets the interests and needs of their research community.

Acknowledgments

The authors wish to acknowledge and thank members of the Social Media in Research Committee and the Health Communication Social Media in Research Taskforce for their helpful feedback on the development of the guidelines, including Andrew Eisman, JD, MBA, Senior University Counsel for Health Affairs and Director of Contracts Unit, General Counsel's Office; Dianne Farb, JD, Assistant Director of Research and Institutional Review Board (IRB)-01 Vice Chair; Ira Fischler, PhD, Chair, University of Florida (UF) IRB-02; Cheryl Granto, Information Security Manager, Information Security, UF Information Technology; Peter Iafrate, PharmD, Chair, UF IRB-01/OneFlorida IRB; Michael Mahoney, Director, Research Operations and Services, UF Office of Research; David Wilkens, Director of Privacy, UF Health Shands Privacy Office; Teresa d'Angelo and Holly Morris, Director of Research Services, UF Clinical and Translational Science Institute (CTSI); Lauren Light, Recruitment Coordinator, UF CTSI; Deaven Freed, Communications Specialist, UF CTSI; Tiffany Pineda, Research Navigator, UF CTSI; Erick Edwing, Shawn Alicea, and Chad Douglas, HealthStreet Communication Specialist; and Vicki Piazza, Director of Operations, HealthStreet. The authors also thank Dr Katie McAllindon and the reviewers who provided feedback on this manuscript.

The work reported in this publication was supported in part by the National Center for Advancing Translational Sciences of the National Institutes of Health under the UF Clinical and Translational Science Award UL1TR001427. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authors' Contributions

All authors contributed to the project design and implementation (project conception, development of the overall research plan, and project oversight). EFG, LBS, JS, and CB completed peer benchmarking. EFG, LBS, CB, and JLK contributed to the scoping review. EFG and MM engaged end users to provide feedback on the guidelines. EFG, LBS, and JLK drafted the manuscript with support from CB, MM, and JS. EFG finalized the manuscript. All authors have read and approved the final manuscript before submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full policies and procedures with risk matrix.
[PDF File (Adobe PDF File), 123 KB - jmir_v23i10e23312_app1.pdf]

Multimedia Appendix 2
Institutional review board process template: social media management plan.
[PDF File (Adobe PDF File), 116 KB - jmir_v23i10e23312_app2.pdf]

References

1. Topolovec-Vranic J, Natarajan K. The use of social media in recruitment for medical research studies: a scoping review. J Med Internet Res 2016 Nov 07;18(11):e286 [FREE Full text] [doi: 10.2196/jmir.5698] [Medline: 27821383]
2. Strekalova YA, Krieger JL. A Picture Really is Worth a Thousand Words: Public engagement with the national cancer institute on social media. J Cancer Educ 2017 Mar;32(1):155-157 [FREE Full text] [doi: 10.1007/s13187-015-0901-5] [Medline: 26351003]
3. Strekalova YA, Krieger JL. Beyond Words: Amplification of cancer risk communication on social media. J Health Commun 2017 Oct;22(10):849-857. [doi: 10.1080/10810730.2017.1367336] [Medline: 28956723]
4. Flood-Grady E, Paige SR, Karimipour N, Harris PA, Cottler LB, Krieger JL. A content analysis of Clinical and Translational Science Award (CTSA) strategies for communicating about clinical research participation online. J Clin Transl Sci 2017 Dec;1(6):340-351 [FREE Full text] [doi: 10.1017/cts.2018.2] [Medline: 29707256]
5. O'Connor A, Jackson L, Goldsmith L, Skirton H. Can I get a retweet please? Health research recruitment and the Twittersphere. J Adv Nurs 2014 Mar;70(3):599-609. [doi: 10.1111/jan.12222] [Medline: 23909740]
6. Subasinghe AK, Nguyen M, Wark JD, Tabrizi SN, Garland SM. Targeted facebook advertising is a novel and effective method of recruiting participants into a human papillomavirus vaccine effectiveness study. JMIR Res Protoc 2016 Jul 22;5(3):e154 [FREE Full text] [doi: 10.2196/resprot.5679] [Medline: 27450586]
7. Bold KW, Hanrahan TH, O'Malley SS, Fucito LM. Exploring the utility of web-based social media advertising to recruit adult heavy-drinking smokers for treatment. J Med Internet Res 2016 May 18;18(5):e107 [FREE Full text] [doi: 10.2196/jmir.5360] [Medline: 27194456]
8. Flood-Grady E, Hough D, Damiani RE, Gebru NM, Fedele DA, Leeman RF, et al. Communication strategies for designing Facebook advertising campaigns to recruit rural participants to develop healthcare delivery interventions. J Clin Transl Sci 2020 Jan 16;4(5):398-407 [FREE Full text] [doi: 10.1017/cts.2020.5] [Medline: 33244428]
9. Ramo DE, Prochaska JJ. Broad reach and targeted recruitment using Facebook for an online survey of young adult substance use. J Med Internet Res 2012 Mar 23;14(1):e28 [FREE Full text] [doi: 10.2196/jmir.1878] [Medline: 22360969]
10. Martinez O, Wu E, Shultz AZ, Capote J, Rios JL, Sandfort T, et al. Still a hard-to-reach population? Using social media to recruit Latino gay couples for an HIV intervention adaptation study. J Med Internet Res 2014 Apr 24;16(4):e113 [FREE Full text] [doi: 10.2196/jmir.3311] [Medline: 24763130]
11. Solberg L. Regulating human subjects research in the information age: date mining on social networking sites. North Ky Law Rev 2012;39(2):1-32 [FREE Full text]
12. Gelinas L, Pierce R, Winkler S, Cohen IG, Lynch HF, Bierer BE. Using social media as a research recruitment tool: ethical issues and recommendations. Am J Bioeth 2017 Mar;17(3):3-14 [FREE Full text] [doi: 10.1080/15265161.2016.1276644] [Medline: 28207365]
13. Federal Policy for the Protection of Human Subjects ('Common Rule'). Office for Human Research Protections. 2009. URL: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html [accessed 2019-08-20]
14. 45 CFR 46. Office for Human Research Protections. 2016. URL: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html [accessed 2019-08-20]
15. Solberg L. Complying with facebook?S terms of use in academic research: a contractual and ethical perspective on data mining and informed consent. UMKC Law Rev 2014;82:787-800 [FREE Full text]
16. Hills S. Considerations and recommendations concerning internet research and human subjects research regulations, with revisions. HHS.gov. 2013. URL: https://www.hhs.gov/ohrp/sites/default/files/ohrp/sachrp/mtgings/2013%20March%20Mtg/internet_research.pdf [accessed 2021-08-26]
17. Guidance regarding social media tools. National Institutes of Health (NIH). 2016. URL: https://www.nih.gov/health-information/nih-clinical-research-trials-you/guidance-regarding-social-media-tools [accessed 2019-08-21]
18. Lemke AA, Harris-Wai JN. Stakeholder engagement in policy development: challenges and opportunities for human genomics. Genet Med 2015 Dec 12;17(12):949-957 [FREE Full text] [doi: 10.1038/gim.2015.8] [Medline: 25764215]

19. Selby JV, Beal AC, Frank L. The Patient-Centered Outcomes Research Institute (PCORI) national priorities for research and initial research agenda. J Am Med Assoc 2012 Apr 18;307(15):1583-1584. [doi: 10.1001/jama.2012.500] [Medline: 22511682]
20. Sniderman AD, Furberg CD. Why guideline-making requires reform. J Am Med Assoc 2009 Jan 28;301(4):429-431. [doi: 10.1001/jama.2009.15] [Medline: 19176446]
21. Hallahan K. Inactive publics: the forgotten publics in public relations ∗. Public Rel Rev 2000;26(4):499-515. [doi: 10.1016/s0363-8111(00)00061-8]
22. O'Haire C, McPheeters M, Nakamoto E, LaBrant L, Most C, Lee K, et al. Methods for engaging stakeholders to identify and prioritize future research needs. Methods Future Research Needs Report No. 4. (Prepared by the Oregon Evidence-based Practice Center and the Vanderbilt Evidence-based Practice Center under Contract No. 290-2007-10057-I.) AHRQ Publication No. 11-EHC044-EF. Agency for Healthcare Research and Quality (US). Rockville, MD; 2011 Jun. URL: https://www.ncbi.nlm.nih.gov/sites/books/NBK62565/ [accessed 2019-03-11]
23. Braun V, Clarke V. Thematic analysis. In: Cooper H, Camic PM, Long DL, Panter AT, Rindskopf D, Sher KJ, editors. APA Handbook of Research Methods in Psychology, Vol. 2. Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological. Washington, DC: American Psychological Association; 2012:57-71.
24. IAP2's Public Participation Toolbox. International Association for Public Participation (IAP2). 2006. URL: https://icma.org/sites/default/files/305431_IAP2%20Public%20Participation%20Toolbox.pdf [accessed 2021-08-26]
25. Recruiting study subjects. US Food & Drug Administration. 1998. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/recruiting-study-subjects [accessed 2020-11-24]
26. IRB Policies, Guidelines and Guidances. University of Florida. URL: https://irb.ufl.edu/index/irb-policies-guidelines-and-guidances.html [accessed 2017-04-21]
27. Social Media Use, Guidelines and Policy Implications. University of Florida. URL: http://hr.ufl.edu/manager-resources/policies-2/social [accessed 2017-04-21]
28. Social Media Guidelines for UF Health. University of Florida Health. URL: https://webservices.ufhealth.org/policies/social-media-guidelines-for-uf-health/ [accessed 2021-09-15]
29. List of accounts. University of Florida. URL: http://socialmedia.ufl.edu/accounts/ [accessed 2017-06-06]
30. Follow UF Health on Social Media. University of Florida Health. URL: https://ufhealth.org/social [accessed 2021-09-09]
31. Thackeray R, Neiger BL, Smith AK, Van Wagenen SB. Adoption and use of social media among public health departments. BMC Public Health 2012 Mar 26;12:242 [FREE Full text] [doi: 10.1186/1471-2458-12-242] [Medline: 22449137]
32. Yuan P, Bare MG, Johnson MO, Saberi P. Using online social media for recruitment of human immunodeficiency virus-positive participants: a cross-sectional survey. J Med Internet Res 2014 May 01;16(5):e117 [FREE Full text] [doi: 10.2196/jmir.3229] [Medline: 24784982]
33. Demographics of social media users and adoption in the United States. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/ [accessed 2020-11-21]

Abbreviations

CTSI: Clinical and Translational Science Institute
IRB: institutional review board
REDCap: Research Electronic Data Capture
SE: stakeholder engagement

Edited by R Kukafka; submitted 07.08.20; peer-reviewed by DS Morgan, M McLaughlin; comments to author 29.09.20; revised version received 24.11.20; accepted 19.04.21; published 08.10.21.

Please cite as:
Flood-Grady E, Solberg LB, Baralt C, Meyer M, Stevens J, Krieger JL
Engaging Institutional Stakeholders to Develop and Implement Guidelines for Recruiting Participants in Research Studies Using Social Media: Mixed Methods, Multi-Phase Process
J Med Internet Res 2021;23(10):e23312
URL: https://www.jmir.org/2021/10/e23312
doi:10.2196/23312
PMID:34623319

Original Paper

General Practice and Digital Methods to Recruit Stroke Survivors to a Clinical Mobility Study: Comparative Analysis

Katja Reuter[1,2], PhD; Chang Liu[3], PhD; NamQuyen Le[4], MPH; Praveen Angyan[2], MS; James M Finley[3,5], PhD

[1]Department of Public Health & Preventive Medicine, SUNY Upstate Medical University, Syracuse, NY, United States

[2]Southern California Clinical and Translational Science Institute, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, United States

[3]Department of Biomedical Engineering, University of Southern California, Los Angeles, CA, United States

[4]Annenberg School for Communication and Journalism, University of Southern California, Los Angeles, CA, United States

[5]Division of Biokinesiology and Physical Therapy, University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Katja Reuter, PhD
Department of Public Health & Preventive Medicine
SUNY Upstate Medical University
Weiskotten Hall Rm. 2263
766 Irving Avenue
Syracuse, NY, 13210
United States
Phone: 1 315 464 5540
Fax: 1 315 464 1701
Email: reuterk@upstate.edu

Abstract

Background: Participant recruitment remains a barrier to conducting clinical research. The disabling nature of a stroke, which often includes functional and cognitive impairments, and the acute stage of illness at which patients are appropriate for many trials make recruiting patients particularly complex and challenging. In addition, people aged 65 years and older, which includes most stroke survivors, have been identified as a group that is difficult to reach and is commonly underrepresented in health research, particularly clinical trials. Digital media may provide effective tools to support enrollment efforts of stroke survivors in clinical trials.

Objective: The objective of this study was to compare the effectiveness of general practice (traditional) and digital (online) methods of recruiting stroke survivors to a clinical mobility study.

Methods: Recruitment for a clinical mobility study began in July 2018. Eligible study participants included individuals 18 years and older who had a single stroke and were currently ambulatory in the community. General recruiting practice included calling individuals listed in a stroke registry, contacting local physical therapists, and placing study flyers throughout a university campus. Between May 21, 2019, and June 26, 2019, the study was also promoted digitally using the social network Facebook and the search engine marketing tool Google AdWords. The recruitment advertisements (ads) included a link to the study page to which users who clicked were referred. Primary outcomes of interest for both general practice and digital methods included recruitment speed (enrollment rate) and sample characteristics. The data were analyzed using the Lilliefors test, the Welch two-sample t test, and the Mann-Whitney test. Significance was set at P=.05. All statistical analyses were performed in MATLAB 2019b.

Results: Our results indicate that digital recruitment methods can address recruitment challenges regarding stroke survivors. Digital recruitment methods allowed us to enroll study participants at a faster rate (1.8 participants/week) compared to using general practice methods (0.57 participants/week). Our findings also demonstrate that digital and general recruitment practices can achieve an equivalent level of sample representativeness. The characteristics of the enrolled stroke survivors did not differ significantly by age (P=.95) or clinical scores (P=.22; P=.82). Comparing the cost-effectiveness of Facebook and Google, we found that the use of Facebook resulted in a lower cost per click and cost per enrollee per ad.

Conclusions: Digital recruitment can be used to expedite participant recruitment of stroke survivors compared to more traditional recruitment practices, while also achieving equivalent sample representativeness. Both general practice and digital recruitment methods will be important to the successful recruitment of stroke survivors. Future studies could focus on testing the effectiveness of additional general practice and digital media approaches and include robust cost-effectiveness analyses. Examining the

effectiveness of different messaging and visual approaches tailored to culturally diverse and underrepresented target subgroups could provide further data to move toward evidence-based recruitment strategies.

(J Med Internet Res 2021;23(10):e28923) doi:10.2196/28923

KEYWORDS

clinical trial; stroke; falls; digital media; social media; advertising; participant recruitment; Facebook; Google; clinical research; research methods; recruitment practices; enrollment

Introduction

Participant recruitment remains one of the main challenges in conducting clinical research, with many trials failing due to recruiting an inadequate number of study participants [1,2]. This is also true with regard to recruiting people poststroke to clinical research studies [3]. The disabling nature of a stroke, which often includes functional and cognitive impairments, makes recruiting patients particularly complex and challenging [3,4]. There is generally no centralized system through which researchers can contact people poststroke once they have completed rehabilitation.

More than 795,000 people in the United States experience a stroke every year [5], and those who recover are often left with a range of impairments affecting cognitive and motor function [6,7]. Because of the persistence of these impairments and their influence on the patients' quality of life, many research groups are developing a range of novel interventions to enhance poststroke recovery [8-11]. Clinical trials are considered the gold standard for evaluating the effects of interventions on health-related biomedical or behavioral outcomes [12]. However, older adults, defined as 65 years of age and older [13], despite being at the highest risk of stroke, have been identified as difficult to reach and are commonly underrepresented in health research, particularly clinical trials [14-16]. One potential reason for the underrepresentation of older adults in clinical trials is that they may not be aware of ongoing trials for which they may be eligible [17]. Digital media and social media (SM) may provide a cost-effective approach to complementing existing recruitment issues and increase the enrollment of stroke survivors in research.

The social networks Facebook and Twitter are among the most popular platforms used for research recruitment [18,19]. A smaller subset of studies has experimented with search engine marketing on Google [20-24] and other search engines, such as Yahoo and Bing, to recruit to studies [25]. Today, a growing body of literature reports the successful use of digital media to recruit older adults. Langbaum et al [26], for example, referred cognitively healthy adults aged 55-75 years in the United States online into a registry of Alzheimer's disease prevention studies (GeneMatch). Over half of those participants (45,210/75,351, 60%) joined GeneMatch via SM advertisements (ads). Another study reported dwindling participant recruitment and later showed a significant increase in the recruitment of middle-to-older-aged people into a blood pressure randomized controlled trial after implementing a Facebook advertising campaign [27]. Lam and Woo [28] demonstrated the cost-effectiveness of Facebook to recruit elder Chinese-speaking Americans into a health education study. However, there is little evidence of using digital media and advertising for recruiting stroke survivors.

The objective of this case study was to compare the effectiveness of general practice (traditional) and digital methods (online) of recruiting stroke survivors to an under-enrolling clinical study of factors affecting fall risk during walking. In this study, general practices included calling individuals listed in an institutional review board (IRB)–approved stroke registry, contacting local physical therapists, and placing study flyers; digital methods included ads on Facebook and Google search pages. The primary outcomes of interest for both conventional and digital methods included recruitment speed (enrollment rate) and sample representativeness. To support related recruitment efforts by other research teams, we also shared the recruitment strategy and materials.

Methods

Study Overview

The clinical mobility study, which began in July 2018, was designed to identify factors that impact balance and fall risk in people poststroke and determine how manipulating these factors influences self-reported perceptions of walking quality and the ability to recover from experimentally imposed loss of balance. The accrual goal was 40 participants. The study required a single visit to a research lab based on a university medical campus in Los Angeles and involved video-based recording of the participants' walking patterns. After providing informed consent and completing a screening questionnaire to determine eligibility, participants completed a set of clinical assessments. These included the lower-extremity portion of the Fugl-Meyer Assessment, which quantifies motor impairment; the Berg Balance Scale (BBS) [29]; the Functional Gait Assessment [30]; a 10-m walking test; the Activity-Based Confidence Scale [31]; and a fall efficacy and fall history questionnaire if they had experienced at least 1 fall within the past year.

The participants then completed 4 walking trials on a dual-belt treadmill. They first walked on the treadmill for 2 minutes at their self-selected speed. The participants then completed 3 subsequent tests, where they were asked to modify their step lengths to match the lengths of visual targets displayed on a screen in front of the treadmill. In each test, the participants responded to rapid accelerations of the treadmill belts, which acted as perturbations of the participants' walking patterns. The participants were compensated at a rate of $20 per hour for the time that they spent in the lab. The trial was approved by the IRB of the University of Southern California (USC; HS-18-00417-AM002).

To support research participant recruitment, researchers in the lab where the research was conducted relied on a preexisting participant registry and word-of-mouth referrals from local physical therapists to reach out to stroke survivors who might like to participate in new studies. However, this strategy often had a relatively low rate of return. As of March 2019, with 4 months left in the award period, this study was behind its enrollment target of 40 participants, having recruited only 15 participants. Missing this enrollment target would have resulted in an underpowered study and wasted resources. Therefore, the team decided to include digital recruitment methods on Facebook and Google during the latter portion of the study in 2019.

Study Population

Eligible study participants included individuals 18 years and older who had a single stroke and were currently ambulatory in the community. Inclusion criteria included (1) a unilateral brain lesion from a single stroke; (2) paresis confined to one side; (3) ability to walk on a treadmill for 2 minutes; (4) absence of cognitive impairment, as demonstrated by a Mini-Mental State Examination score greater than 24; and (5) ability to provide informed consent.

Recruitment Methods

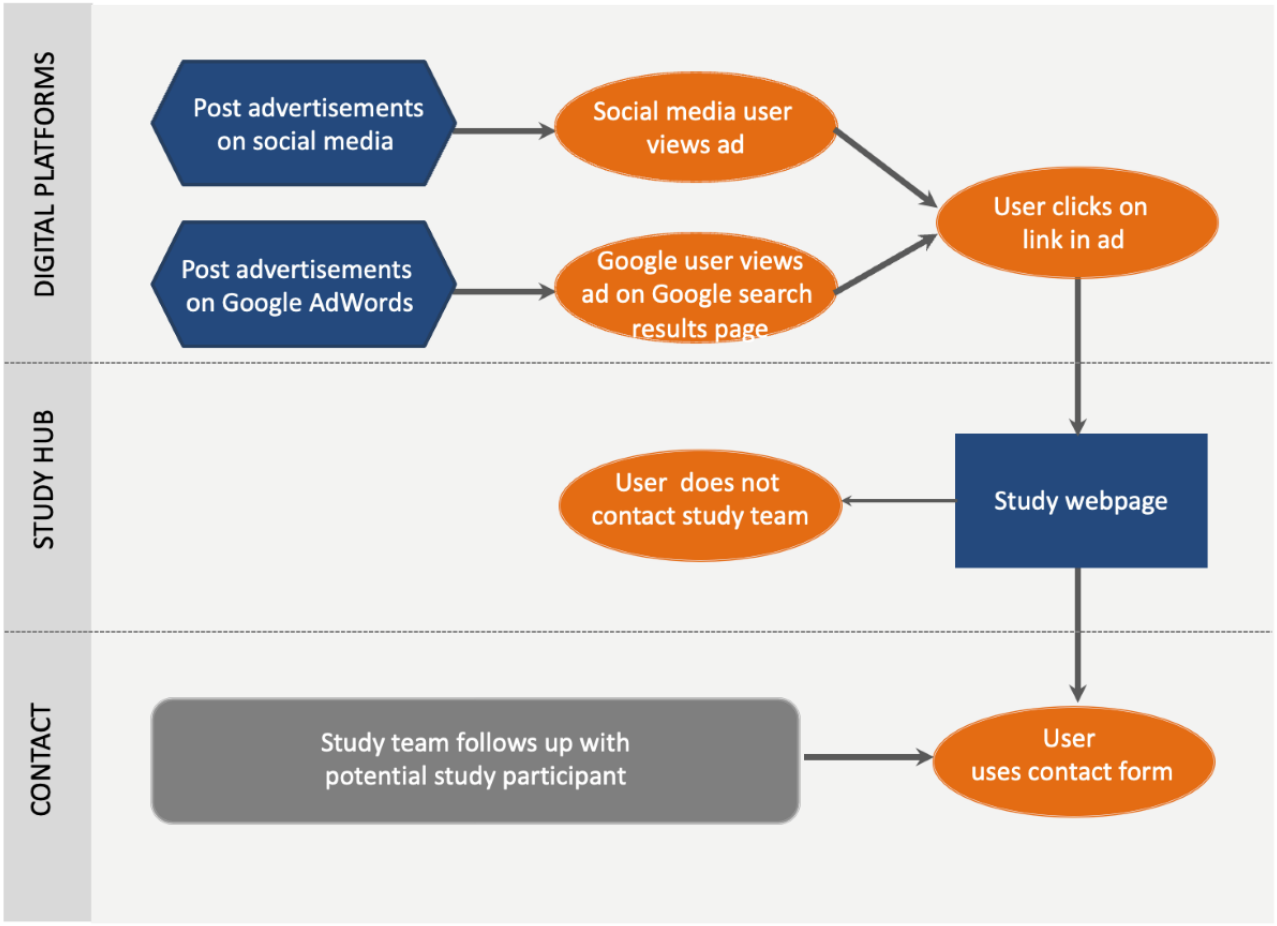
General Practice Recruitment

The standard recruiting practice in the lab was to call individuals listed in an IRB-approved stroke registry or contact local physical therapists affiliated with USC or the Rancho Los Amigos National Rehabilitation Center. Flyers for the study (see Multimedia Appendix 1 for an example) were physically placed throughout USC's Health Sciences Campus.

Digital Recruitment

Between May 21, 2019, and June 26, 2019, the study was promoted digitally using the social network Facebook and the search engine marketing tool Google AdWords. The recruitment ads (Multimedia Appendix 2) were posted on Facebook and Google search results pages. The recruitment ads included a link to the study page (Multimedia Appendix 3) to which users who clicked were referred. On the study page, interested users could use the contact form to get in touch with the study team for further information and screening. Figure 1 outlines the digital recruitment process.

Figure 1. Digital recruitment process for the social network Facebook and Google search results pages.



The text and images of the ads were reviewed and approved by the Health Sciences Campus IRB of USC (HS-18-00417). The text and images (Multimedia Appendix 2) were tailored to people of all ages who had a stroke using a mix of images, such as infographics and photos. We did not target the ads using any other characteristics of potential viewers. The type of ad we used per platform and targeting criteria are listed in Multimedia Appendix 4.

Outcome Measures

The primary outcomes of interest for both conventional and digital methods included recruitment speed (enrollment rate) and sample characteristics. We captured the number of participants enrolled, the time required for enrollment, and participant demographic and clinical characteristics. These measures were used to compute the enrollment rate for each method and to determine differences in the sample representativeness. Secondary outcome measures for our digital recruiting methods included the number of post impressions (ie, the number of times a post was seen without a user necessarily interacting with it), the number of clicks on the link in the message, the number of resulting study page contacts, the number of people screened, the cost per click, and the cost per enrollee.

Data Collection

Recruitment-related data for each ad were collected in three different ways: (1) SM-based engagement data, such as clicks on the ads and impressions (the number of times an ad was displayed), were collected from Facebook and Google; (2) study website engagement data, including contact requests, were collected using Google Analytics; and (3) enrollment data were provided by the study team. Multimedia Appendix 5 details the technical techniques we used to collect the data from the digital platforms.

Analysis

Information on recruitment sources was collapsed into two categories: (1) general practice and (2) digital (ie, Facebook and Google). Digital recruitment was further broken down to compare the recruitment results by digital platform. All statistical analyses were performed in MATLAB 2019b (The Mathworks, Natick, USA). We analyzed the normality of participant characteristics, including age, the Fugl-Meyer Assessment score, and the BBS score, using the Lilliefors test. Normally distributed data were expressed as the mean (SD) of the corresponding mean, and we compared the participant characteristics between groups (ie, recruited via general practice or digital approaches) using the Welch two-sample *t* test. Nonnormally distributed data were expressed as the median (IQR). We used the Mann-Whitney test to compare between groups. Significance was set at the P=.05 level.

Results

Participant Demographics

Overall, the participants (N=40) included 15 (37.5%) women and 25 (62.5%) men, and were aged 29-78 years (mean 59, SD 12 years) (Table 1). This age range is consistent with the current literature, which indicates that stroke is most prevalent in adults aged 60 years and older [32]. Most of the recruited participants were non-Hispanic, that is, Asian (12/40, 30%), followed by one-quarter Hispanic (10/40, 25%), white (9/40, 22.5%), more than one race (9/40, 22.5%), and African American/black (8/40, 20%).

Table 1. Characteristics of enrolled study participants by recruitment method.

Characteristic	Recruited through general practice (n=31), n (%)	Recruited through digital media (n=9), n (%)	Total recruited (N=40), n (%)
Age			
18-29 years	1 (3.2)	0 (0)	1 (2.5)
30-39 years	2 (6.5)	0 (0)	2 (5.0)
40-49 years	3 (9.7)	3 (33.3)	6 (15.0)
50-59 years	9 (29.0)	1 (11.1)	10 (25.0)
60-69 years	11 (35.5)	4 (44.4)	15 (37.5)
70-99 years	5 (16.1)	1 (11.1)	6 (15)
Sex			
Male	21 (64.5)	4 (44.4)	25 (62.5)
Female	10 (35.5)	5 (55.6)	15 (37.5)
Other	0 (0)	0 (0)	0 (0)
Ethnicity/racial background			
African American/Black	6 (19.4)	2 (22.2)	8 (20.0)
American Indian/Alaska Native	0 (0)	0 (0)	0 (0)
Asian/Pacific Islander	7 (22.6)	5 (55.6)	12 (30.0)
Hispanic	9 (29.0)	1 (11.1)	10 (25.0)
Middle Eastern	0 (0)	0 (0)	0 (0)
White	8 (25.8)	1 (11.1)	9 (22.5)
Other	9 (29.0)	0 (0)	9 (22.5)

Results of General Practice Recruitment

Throughout 2018, general practice recruitment proceeded slowly as most of the members of the stroke registry who were contacted either did not return the phone calls or were unable to participate in the study. Between July 1, 2018, and May 20, 2019 (46 weeks), we recruited 26 participants through our standard recruiting procedures. Between May 21, 2019, and July 16, 2019 (8 weeks), we recruited an additional 5 participants through general practice recruiting methods. Of these 31 participants, we recruited 20 participants from the existing registry and 11 through word-of-mouth referrals from local physical therapists and participants. The mean age of those recruited through general practice was 58.7 (SD 11.9) years.

Results of Digital Recruitment

A total of 8 advertised messages were posted (6 on Facebook and 2 on Google) over 5 weeks (between May 21, 2019, and June 26, 2019) (Table 2). Figure 2 shows the recruitment flow diagrams for Facebook and Google. The combined digital recruitment efforts on Facebook and Google resulted in a total of 85 valid referred potential study participants, of which 25 (29.4%) contacted the study team using the web-based contact form. Of these 25, 9 participants met the inclusion criteria, enrolled, and completed the study. The mean age of the participants recruited through digital recruitment was 59.0 (SD 10.7) years, comparable to the mean age of the participants recruited through general practice approaches.

Table 2. Performance of digital recruitment ads.

Ad run dates	Impressions, n	Link clicks, n	Cost ($)	Study page contacts, n	People screened, n	People enrolled, n	Cost per click ($)	Cost per enrollee ($)[a]
Digital platform: Facebook								
5/21/2019-6/26/2019	51,435	580	886.43	15	6	4	1.53	221.60
5/22/2019-6/26/2019	75,459	1033	1604.13	21	10	2	1.55	802.07
5/24/2019-6/26/2019	106,992	2011	1825.5	18	1	0	0.91	—[b]
5/24/2019-6/26/2019	10,852	183	226.46	2	0	0	1.24	—
5/21/2019-6/30/2019	84,798	1118	1523.89	2	3	1	1.36	1523.89
5/24/2019-6/26/2019	33,767	373	537.28	16	3	1	1.44	537.28
Digital platform: Google								
05/13/2019-06/03/2019	9473	291	964.42	3	1	1	3.31	964.42
05/13/2019-06/03/2019	876	23	50.59	2	0	0	2.20	—
Mean	46,706.50	701.50[c]	952.34	9.88	3	1.13	1.36[d]	846.52[e]
Total	373,652	5612	7618.70	79	24	9	N/A[f]	N/A

[a]Values were calculated by dividing the cost by the number of enrollees for each ad. – indicated that no enrollee was recruited through this ad.

[b]Not available.

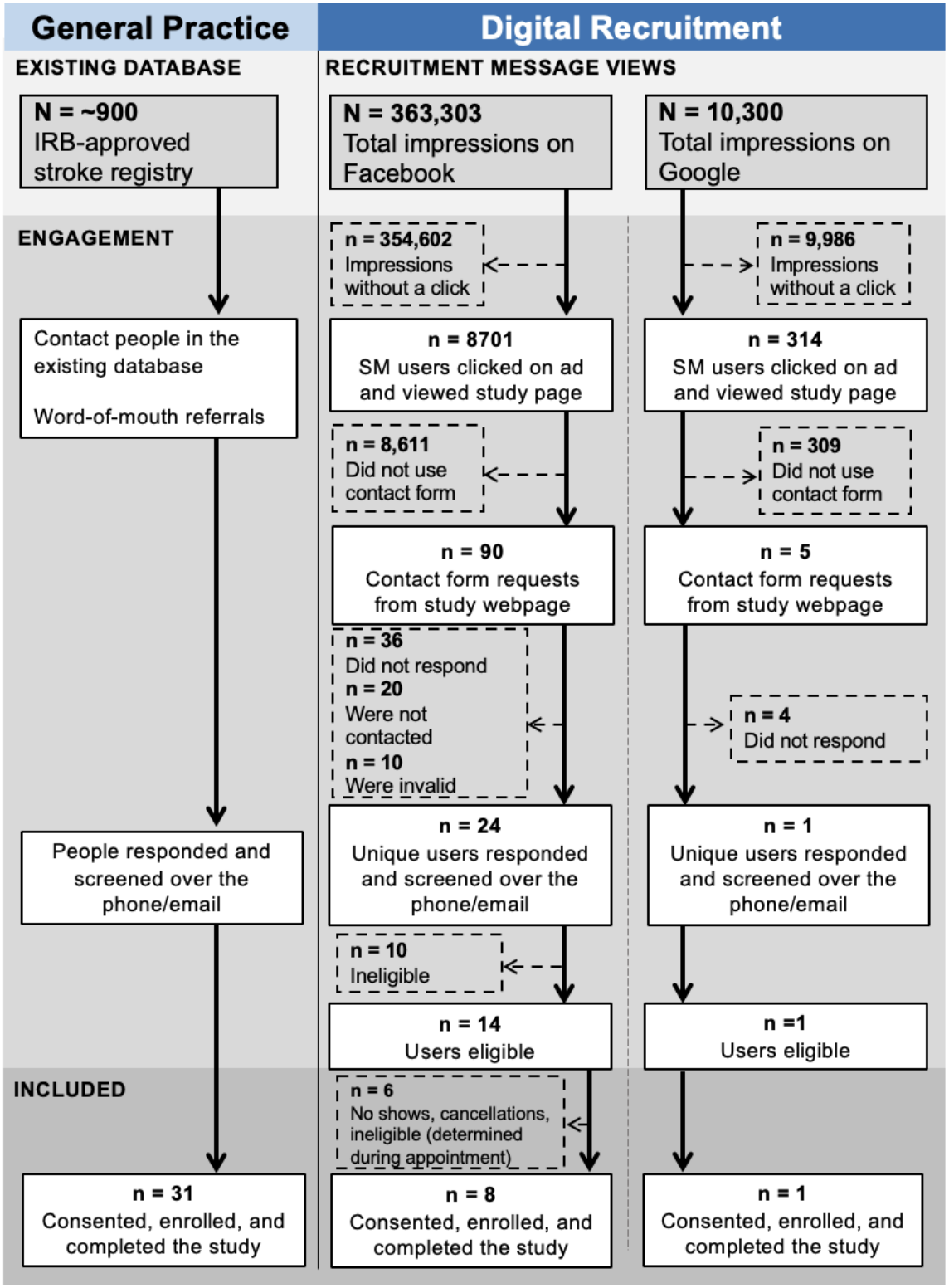
[c]Means for clicks and engagement values were rounded off to the nearest whole number. Unless otherwise indicated, these were calculated by dividing by the total number of ads (n=8).

[d]This value was calculated by dividing the total cost of the ads for both platforms by the total link clicks.

[e]This value was calculated by dividing the total cost of the ads for both platforms by the number of total enrollees.

[f]N/A: not applicable.

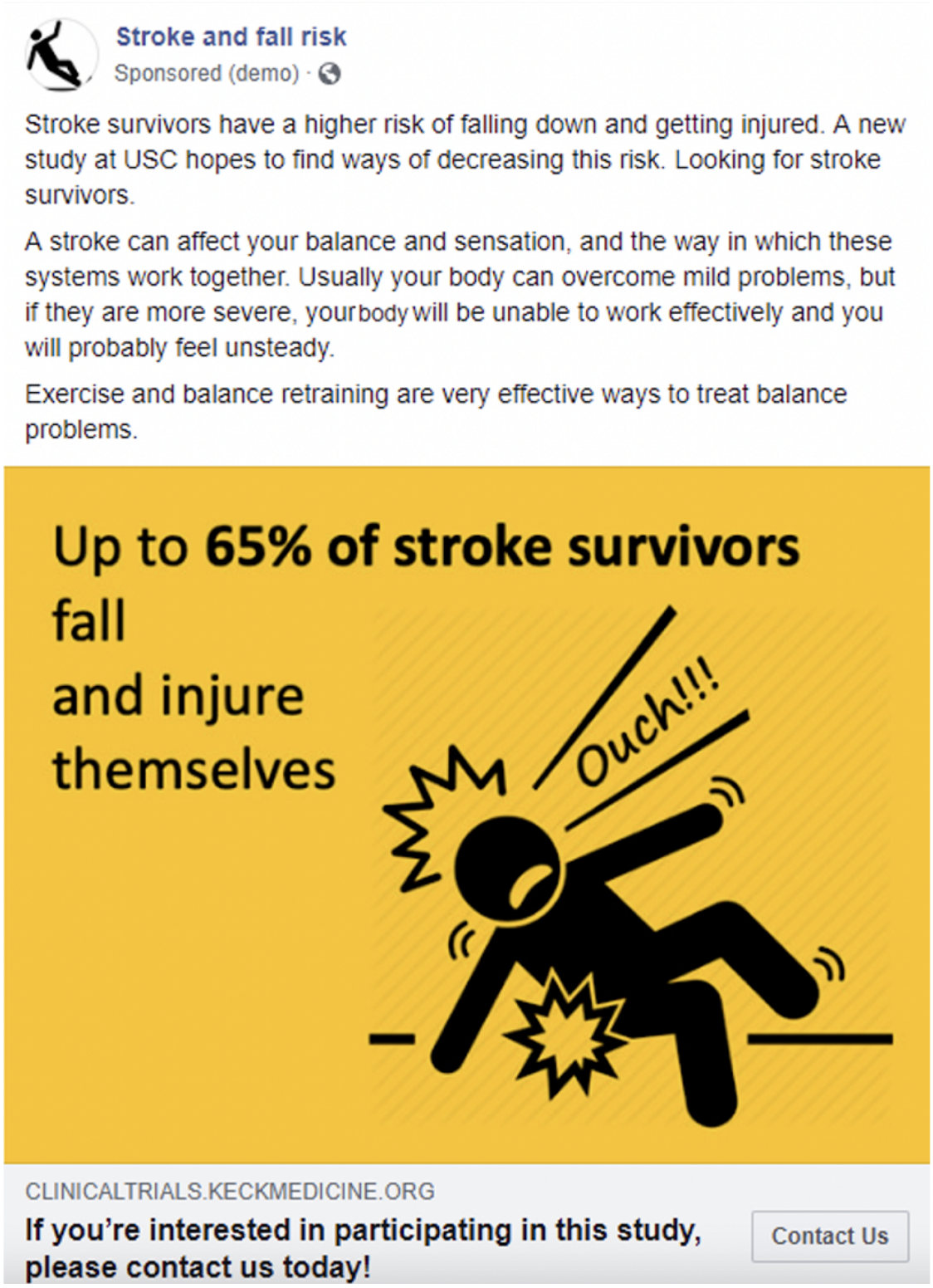
Figure 2. Recruitment flow diagrams for general practice and digital recruitment using Facebook and Google. IRB: institutional review board; SM: social media.



Most Cost-Effective Digital Recruitment Messages

The digital recruitment ads used different types of images, including infographics, images of older couples, and older individuals after a fall. Figure 3 shows the ad message that ran on Facebook, with the lowest cost per enrollee ($221.60) compared to a higher cost per enrollee on Google ($964.42) (Table 2). Multimedia Appendix 2 shows the recruitment ad messages used in this study. We only ran two simultaneous Google ads for about 3 weeks because of the cost of the ads and the available budget.

Figure 3. Example of the advertisement that ran on Facebook, with the lowest cost per enrollee.



Comparison of Recruitment Methods

We found that the rate of enrollment was higher when using digital media (1.8 people/week) than when using general practice approaches (0.57 people/week). We did not find any significant differences in age, motor impairment level measured by Fugl-Meyer scores, and functional balance measured by the BBS (Table 3). Because of the small sample size, we were not able to statistically compare the outcomes related to gender and race/ethnicity for general practice and digital recruitment methods.

Table 3. Comparison of participant characteristics recruited via general practice versus digital approaches.

Characteristics	Recruited through general practice	Recruited through digital media	*P* value
Age (years), mean (SD)	59 (12)	59 (11)	.95
Fugl-Meyer Assessment score, median (IQR)	28 (25-30)	25 (20-29)	.22
BBS[a], median (IQR)	53 (48-54)	52 (49-54)	.82
Enrollment rate (number of people enrolled/week)[b]	0.57 (31/54)	1.8 (9/5)	N/A[c]

[a]BBS: Berg Balance Scale.

[b]The rate of enrollment was calculated as the total number of people enrolled from each recruiting method/week between the first and last enrollment.

[c]N/A: not applicable.

Discussion

Primary Results

Our results indicate that digital recruitment methods can address recruitment challenges regarding stroke survivors. The use of digital approaches allowed us to enroll study participants and close the recruitment gap at a faster rate (1.8 participants/week) compared to using general practice approaches (0.57 participants/week). The effectiveness of SM-based recruitment methods, for example, in comparison to more general (traditional) approaches has been shown by a range of studies [33]. However, comparable data on the effectiveness of recruitment efforts for stroke survivors are currently limited. The characteristics of the enrolled stroke survivors did not differ significantly by age (*P*=.95) or clinical scores (*P*=.22; *P*=.82) between digital and general practice approaches. However, we observed a trend toward recruiting a higher percentage of female participants through digital methods as compared to general practice approaches. Our findings demonstrate that digital and general recruitment practices can achieve similar levels of sample representativeness, with the possibility that digital methods may even result in samples with distributions of sex that are more representative of the general population. There is limited research on the representativeness of participants recruited from digital media and SM. However, although some studies have reported issues related to representativeness and selection bias [34,35], our data support previous findings from other studies that reported the successful use of SM to recruit comparable and representative samples, as described in the systematic review by Whitaker et al [36]. Additionally, Yu et al [37], for example, were able to recruit geographically representative samples of individuals with myeloproliferative neoplasms in the United States for a survey study using multiple recruitment strategies, including Google and Facebook.

Furthermore, comparing the cost-effectiveness of Facebook and Google, we found that the use of Facebook resulted in a lower cost per click and cost per enrollee per ad. The effectiveness (with varying costs) of Facebook as a recruitment mechanism has been reported in a number of studies across disease or health topics and study types [36], such as mental health [38], mobile health studies in psycho-oncology [39], a clinical trial involving healthy elderly [40], and suicide prevention research [35]. Google ads have also been reported as recruitment methods but less frequently [41,42].

Limitations

The external validity of our findings is limited. Study participants were targeted and recruited in LA County. Therefore, it is unclear to what extent the same recruitment and targeting approach will lead to comparable results in other US regions or internationally. It is also important to note that certain populations are less likely to respond to social media ads. These populations include people who are older, live in rural areas, have little or no access to the internet, have a low level of internet literacy, or do not use SM. Additionally, we were not able to conduct a complete cost comparison between digital and general recruitment approaches. At the time of study setup, the study team did not plan to compare the cost-effectiveness of different recruitment methods and did not collect related data, for example, the percentage effort spent by the study coordinator and the principal investigator on recruitment efforts and the number of potential participants contacted and screened with general practice methods. Lastly, it should be noted that the sample size recruited using SM was small in this study, and thus the statistical comparison of differences in age and clinical scores between recruiting methods should be confirmed with larger samples. Future studies should also determine whether there are meaningful differences in sex or race between participants recruited via conventional and digital methods, as these methods may differ in their ability to identify a diverse pool of potential study participants.

Conclusion

Digital recruitment can be used to expedite participant recruitment of stroke survivors compared to general, more traditional recruitment practices, while also achieving equivalent sample representativeness. Both general practice and digital media recruitment methods will be important for the successful recruitment of stroke survivors. The data we provide here demonstrate the potential of digital recruitment methods to aid in meeting the accrual goal without delay. Future studies could focus on testing the effectiveness of different digital platforms and include robust cost-effectiveness analyses. Additionally, examining the effectiveness of different messaging and visual approaches tailored to culturally diverse and underrepresented target subgroups could provide further data to develop evidence-based recruitment strategies.

Acknowledgments

The digital recruitment for this study was supported by the Digital Recruitment Service provided by the Southern California Clinical and Translational Science Institute (SC CTSI) at USC. The SC CTSI is supported by a grant (UL1TR000130) from the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH). The content is solely the responsibility of the authors and does not necessarily represent the official views of the SC CTSI, USC, the NCATS, and the NIH.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Flyer used for general practice recruiting.
[DOCX File , 53 KB - jmir_v23i10e28923_app1.docx]

Multimedia Appendix 2
Recruitment messages used for advertising on Facebook and Google.
[DOCX File , 2617 KB - jmir_v23i10e28923_app2.docx]

Multimedia Appendix 3
Study page to which users who clicked digital ads were referred. ad: advertisement.
[PNG File , 810 KB - jmir_v23i10e28923_app3.png]

Multimedia Appendix 4
Type of ad and targeting criteria used on Facebook and Google. ad: advertisement.
[DOCX File , 15 KB - jmir_v23i10e28923_app4.docx]

Multimedia Appendix 5
Description of technical approaches used to collect the data from the digital platforms.
[DOCX File , 15 KB - jmir_v23i10e28923_app5.docx]

References

1. Reynolds T. Clinical trials: can technology solve the problem of low recruitment? BMJ 2011 Jun 14;342:d3662. [doi: 10.1136/bmj.d3662] [Medline: 21672994]
2. Campbell MK, Snowdon C, Francis D, Elbourne D, McDonald AM, Knight R, STEPS group. Recruitment to randomised trials: strategies for trial enrollment and participation study. The STEPS study. Health Technol Assess 2007 Nov;11(48):iii, ix-iii,105 [FREE Full text] [doi: 10.3310/hta11480] [Medline: 17999843]
3. Ferreira IS, Pinto CB, Saleh Velez FG, Leffa DT, Vulcano de Toledo Piza P, Fregni F. Recruitment challenges in stroke neurorecovery clinical trials. Contemp Clin Trials Commun 2019 Sep;15:100404 [FREE Full text] [doi: 10.1016/j.conctc.2019.100404] [Medline: 31360793]
4. Boxall L, Hemsley A, White N. Exploring recruitment issues in stroke research: a qualitative study of nurse researchers' experiences. Nurse Res 2016 May;23(5):8-14. [doi: 10.7748/nr.23.5.8.s3] [Medline: 27188567]
5. Stroke facts. Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/stroke/facts.htm [accessed 2021-08-10]
6. Hesse S, Werner C. Poststroke motor dysfunction and spasticity: novel pharmacological and physical treatment strategies. CNS Drugs 2003;17(15):1093-1107. [doi: 10.2165/00023210-200317150-00004] [Medline: 14661987]
7. Sun JH, Tan L, Yu JT. Post-stroke cognitive impairment: epidemiology, mechanisms and management. Ann Transl Med 2014 Aug;2(8):80 [FREE Full text] [doi: 10.3978/j.issn.2305-5839.2014.08.05] [Medline: 25333055]
8. Stinear CM, Lang CE, Zeiler S, Byblow WD. Advances and challenges in stroke rehabilitation. Lancet Neurol 2020 Apr;19(4):348-360. [doi: 10.1016/S1474-4422(19)30415-6] [Medline: 32004440]
9. Bertani R, Melegari C, De Cola MC, Bramanti A, Bramanti P, Calabrò RS. Effects of robot-assisted upper limb rehabilitation in stroke patients: a systematic review with meta-analysis. Neurol Sci 2017 Sep;38(9):1561-1569. [doi: 10.1007/s10072-017-2995-5] [Medline: 28540536]
10. Cervera MA, Soekadar SR, Ushiba J, Millán JDR, Liu M, Birbaumer N, et al. Brain-computer interfaces for post-stroke motor rehabilitation: a meta-analysis. Ann Clin Transl Neurol 2018 May;5(5):651-663. [doi: 10.1002/acn3.544] [Medline: 29761128]

11. Morone G, Paolucci S, Cherubini A, De Angelis D, Venturiero V, Coiro P, et al. Robot-assisted gait training for stroke patients: current state of the art and perspectives of robotics. Neuropsychiatr Dis Treat 2017;13:1303-1311. [doi: 10.2147/NDT.S114102] [Medline: 28553117]
12. NIH's definition of a clinical trial. National Institutes of Health. URL: https://grants.nih.gov/policy/clinical-trials/definition.htm [accessed 2021-08-10]
13. American COSM, Chodzko-Zajko WJ, Proctor DN, Fiatarone SMA, Minson CT, Nigg CR, et al. American College of Sports Medicine position stand. Exercise and physical activity for older adults. Med Sci Sports Exerc 2009 Jul;41(7):1510-1530. [doi: 10.1249/MSS.0b013e3181a0c95c] [Medline: 19516148]
14. Shaghaghi A, Bhopal RS, Sheikh A. Approaches to recruiting "hard-to-reach" populations into research: a review of the literature. Health Promot Perspect 2011;1(2):86-94 [FREE Full text] [doi: 10.5681/hpp.2011.009] [Medline: 24688904]
15. Bonevski B, Randell M, Paul C, Chapman K, Twyman L, Bryant J, et al. Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. BMC Med Res Methodol 2014 Mar 25;14:42 [FREE Full text] [doi: 10.1186/1471-2288-14-42] [Medline: 24669751]
16. Liljas AEM, Walters K, Jovicic A, Iliffe S, Manthorpe J, Goodman C, et al. Strategies to improve engagement of 'hard to reach' older people in research on health promotion: a systematic review. BMC Public Health 2017 Apr 21;17(1):349 [FREE Full text] [doi: 10.1186/s12889-017-4241-8] [Medline: 28431552]
17. Heller C, Balls-Berry JE, Nery JD, Erwin PJ, Littleton D, Kim M, et al. Strategies addressing barriers to clinical trial enrollment of underrepresented populations: a systematic review. Contemp Clin Trials 2014 Nov;39(2):169-182. [doi: 10.1016/j.cct.2014.08.004] [Medline: 25131812]
18. Amon KL, Campbell AJ, Hawke C, Steinbeck K. Facebook as a recruitment tool for adolescent health research: a systematic review. Acad Pediatr 2014;14(5):439-447.e4. [doi: 10.1016/j.acap.2014.05.049] [Medline: 25169155]
19. Topolovec-Vranic J, Natarajan K. The use of social media in recruitment for medical research studies: a scoping review. J Med Internet Res 2016 Nov 7;18(11):e286 [FREE Full text] [doi: 10.2196/jmir.5698] [Medline: 27821383]
20. Kim AS, Poisson SN, Easton JD, Johnston SC. A cross-sectional study of individuals seeking information on transient ischemic attack and stroke symptoms online: a target for intervention? PLoS One 2012;7(10):e47997 [FREE Full text] [doi: 10.1371/journal.pone.0047997] [Medline: 23118919]
21. Watson NL, Mull KE, Heffner JL, McClure JB, Bricker JB. Participant recruitment and retention in remote ehealth intervention trials: methods and lessons learned from a large randomized controlled trial of two web-based smoking interventions. J Med Internet Res 2018 Aug 24;20(8):e10351 [FREE Full text] [doi: 10.2196/10351] [Medline: 30143479]
22. Davies B, Kotter M. Lessons from recruitment to an internet-based survey for degenerative cervical myelopathy: comparison of free and fee-based methods. JMIR Res Protoc 2018 Feb 05;7(2):e18 [FREE Full text] [doi: 10.2196/resprot.6567] [Medline: 29402760]
23. Bracken K, Hague W, Keech A, Conway A, Handelsman DJ, Grossmann M, et al. Recruitment of men to a multi-centre diabetes prevention trial: an evaluation of traditional and online promotional strategies. Trials 2019 Jun 19;20(1):366 [FREE Full text] [doi: 10.1186/s13063-019-3485-2] [Medline: 31217024]
24. Lane TS, Armin J, Gordon JS. Online recruitment methods for web-based and mobile health studies: a review of the literature. J Med Internet Res 2015;17(7):e183 [FREE Full text] [doi: 10.2196/jmir.4359] [Medline: 26202991]
25. Morgan AJ, Jorm AF, Mackinnon AJ. Internet-based recruitment to a depression prevention intervention: lessons from the Mood Memos study. J Med Internet Res 2013 Feb 12;15(2):e31 [FREE Full text] [doi: 10.2196/jmir.2262] [Medline: 23403043]
26. Langbaum JB, Karlawish J, Roberts JS, Wood EM, Bradbury A, High N, et al. GeneMatch: a novel recruitment registry using at-home APOE genotyping to enhance referrals to Alzheimer's prevention studies. Alzheimers Dement 2019 Apr;15(4):515-524 [FREE Full text] [doi: 10.1016/j.jalz.2018.12.007] [Medline: 30772251]
27. Nash EL, Gilroy D, Srikusalanukul W, Abhayaratna WP, Stanton T, Mitchell G, et al. Facebook advertising for participant recruitment into a blood pressure clinical trial. J Hypertens 2017 Dec;35(12):2527-2531. [doi: 10.1097/HJH.0000000000001477] [Medline: 28704263]
28. Lam NH, Woo BK. Digital media recruitment for fall prevention among older Chinese-American individuals: observational, cross-sectional study. JMIR Aging 2018 Nov 01;1(2):e11772 [FREE Full text] [doi: 10.2196/11772] [Medline: 31518249]
29. Berg K. Measuring balance in the elderly: preliminary development of an instrument. Physiotherapy Canada 1989 Nov;41(6):304-311. [doi: 10.3138/ptc.41.6.304]
30. Lin JH, Hsu MJ, Hsu HW, Wu HC, Hsieh CL. Psychometric comparisons of 3 functional ambulation measures for patients with stroke. Stroke 2010 Sep;41(9):2021-2025. [doi: 10.1161/STROKEAHA.110.589739] [Medline: 20671244]
31. Powell LE, Myers AM. The Activities-specific Balance Confidence (ABC) Scale. J Gerontol A Biol Sci Med Sci 1995 Jan;50A(1):M28-M34. [doi: 10.1093/gerona/50a.1.m28] [Medline: 7814786]
32. Benjamin EJ, Virani SS, Callaway CW, Chamberlain AM, Chang AR, Cheng S, et al. Heart disease and stroke statistics—2018 update: a report from the American Heart Association. Circulation 2018 Mar 20;137(12):e67-e492. [doi: 10.1161/CIR.0000000000000558]
33. Darmawan I, Bakker C, Brockman TA, Patten CA, Eder M. The role of social media in enhancing clinical trial recruitment: scoping review. J Med Internet Res 2020 Oct 26;22(10):e22810 [FREE Full text] [doi: 10.2196/22810] [Medline: 33104015]

34. Choi I, Milne DN, Glozier N, Peters D, Harvey SB, Calvo RA. Using different Facebook advertisements to recruit men for an online mental health study: engagement and selection bias. Internet Interv 2017 Jun;8:27-34 [FREE Full text] [doi: 10.1016/j.invent.2017.02.002] [Medline: 30135825]
35. Lee S, Torok M, Shand F, Chen N, McGillivray L, Burnett A, et al. Performance, cost-effectiveness, and representativeness of Facebook recruitment to suicide prevention research: online survey study. JMIR Ment Health 2020 Oct 22;7(10):e18762 [FREE Full text] [doi: 10.2196/18762] [Medline: 33090115]
36. Whitaker C, Stevelink S, Fear N. The use of Facebook in recruiting participants for health research purposes: a systematic review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]
37. Yu J, Paranagama D, Parasuraman S. Recruitment strategies and geographic representativeness for patient survey studies in rare diseases: experience from the living with myeloproliferative neoplasms patient survey. PLoS One 2020;15(12):e0243562. [doi: 10.1371/journal.pone.0243562] [Medline: 33382745]
38. Sanchez C, Grzenda A, Varias A, Widge AS, Carpenter LL, McDonald WM, et al. Social media recruitment for mental health research: a systematic review. Compr Psychiatry 2020 Nov;103:152197 [FREE Full text] [doi: 10.1016/j.comppsych.2020.152197] [Medline: 32992073]
39. Spahrkäs SS, Looijmans A, Sanderman R, Hagedoorn M. Recruiting participants for an international mHealth study via Facebook Ads: experiences from the Untire App RCT. Internet Interv 2021 Mar;23:100362. [doi: 10.1016/j.invent.2021.100362] [Medline: 33489782]
40. Cowie JM, Gurney ME. The use of Facebook advertising to recruit healthy elderly people for a clinical trial: baseline metrics. JMIR Res Protoc 2018 Jan 24;7(1):e20 [FREE Full text] [doi: 10.2196/resprot.7918] [Medline: 29367186]
41. Faro JM, Nagawa CS, Orvek EA, Smith BM, Blok AC, Houston TK, et al. Comparing recruitment strategies for a digital smoking cessation intervention: technology-assisted peer recruitment, social media, ResearchMatch, and smokefree.gov. Contemp Clin Trials 2021 Feb 08;103:106314. [doi: 10.1016/j.cct.2021.106314] [Medline: 33571687]
42. Muñoz RF, Leykin Y, Barrera AZ, Dunn LB, Gutierrez R, Curland RA, et al. Who comes to a self-help depression prevention website? Characteristics of Spanish- and English-speaking visitors. Internet Interv 2021 Mar;23:100359 [FREE Full text] [doi: 10.1016/j.invent.2020.100359] [Medline: 33520668]

Abbreviations

BBS: Berg Balance Scale
IRB: institutional review board
SM: social media
USC: University of Southern California

Edited by R Kukafka; submitted 01.04.21; peer-reviewed by K Amon, S Kardes, J McCullough, N Catherine; comments to author 28.05.21; revised version received 10.08.21; accepted 12.08.21; published 13.10.21.

Please cite as:
Reuter K, Liu C, Le N, Angyan P, Finley JM
General Practice and Digital Methods to Recruit Stroke Survivors to a Clinical Mobility Study: Comparative Analysis
J Med Internet Res 2021;23(10):e28923
URL: https://www.jmir.org/2021/10/e28923
doi:10.2196/28923
PMID:34643544

Original Paper

Sharing Government Health Data With the Private Sector: Community Attitudes Survey

Annette Braunack-Mayer[1], BMedSci(Hons), PhD; Belinda Fabrianesi[1], BSc, GradCert(HlthRes), GradDip(Edu); Jackie Street[1], BSc(Hons), GradDipPHC, PhD; Pauline O'Shaughnessy[2], BActSt, MActSt, PhD; Stacy M Carter[1], MPH(Hons), PhD; Lina Engelen[3], MSc, PhD; Lucy Carolan[1], BPubHlth(Hons); Rebecca Bosward[1], BPubHlth(Hons); David Roder[4], BDS, MPH, DDSc; Kylie Sproston[5], MEng(Hons), MSc, CEng, GAICD

[1]Australian Centre for Health Engagement, Evidence and Values, School of Health and Society, University of Wollongong, Wollongong, Australia

[2]School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, Australia

[3]School of Health and Society, University of Wollongong, Wollongong, Australia

[4]University of South Australia, Adelaide, Australia

[5]Bellberry Limited, South Australia, Australia

Corresponding Author:
Annette Braunack-Mayer, BMedSci(Hons), PhD
Australian Centre for Health Engagement, Evidence and Values
School of Health and Society
University of Wollongong
Northfields Avenue
Wollongong, 2522
Australia
Phone: 61 2 4298 1228
Email: abmayer@uow.edu.au

Abstract

Background: The use of government health data for secondary purposes, such as monitoring the quality of hospital services, researching the health needs of populations, and testing how well new treatments work, is increasing. This increase in the secondary uses of health data has led to increased interest in what the public thinks about data sharing, in particular, the possibilities of sharing with the private sector for research and development. Although international evidence demonstrates broad public support for the secondary use of health data, this support does not extend to sharing health data with the private sector. If governments intend to share health data with the private sector, knowing what the public thinks will be important. This paper reports a national survey to explore public attitudes in Australia toward sharing health data with private companies for research on and development of therapeutic drugs and medical devices.

Objective: This study aims to explore public attitudes in Australia toward sharing government health data with the private sector.

Methods: A web-based survey tool was developed to assess attitudes about sharing government health data with the private sector. A market research company was employed to administer the web-based survey in June 2019.

Results: The survey was completed by 2537 individuals residing in Australia. Between 51.8% and 57.98% of all participants were willing to share their data, with slightly fewer in favor of sharing to improve health services (51.99%) and a slightly higher proportion in favor of sharing for research and development (57.98%). There was a preference for opt-in consent (53.44%) and broad support for placing conditions on sharing health information with private companies (62% to 91.99%). Wide variability was also observed in participants' views about the extent to which the private sector could be trusted and how well they would behave if entrusted with people's health information. In their qualitative responses, the participants noted concerns about private sector corporate interests, corruption, and profit making and expressed doubt about the Australian government's capacity to manage data sharing safely. The percentages presented are adjusted against the Australian population.

Conclusions: This nationally representative survey provides preliminary evidence that Australians are uncertain about sharing their health data with the private sector. Although just over half of all the respondents supported sharing health data with the private sector, there was also strong support for strict conditions on sharing data and for opt-in consent and significant concerns about how well the private sector would manage government health data. Addressing public concern about sharing government

health data with the private sector will require more and better engagement to build community understanding about how agencies can collect, share, protect, and use their personal data.

(J Med Internet Res 2021;23(10):e24200) doi:10.2196/24200

KEYWORDS

big data; health information systems; health data; private sector; data linkage; public opinion; consent; trust; public interest; social license

Introduction

Background

Every day, people produce large amounts of health information about themselves through their interactions with health professionals, hospitals, and other government and nongovernment agencies. Beyond being a record of their health care, this information can be collated for a wide range of *secondary* uses, such as monitoring the quality of hospital services, researching the health needs of populations, and testing how well new treatments work.

As the secondary use of health data increases, so does the interest in what the public thinks about such data sharing [1-5]. This interest is related in part to growing public awareness of the risks associated with secondary use of health data, accentuated through recent data breaches and public controversies [6-10]. These events call attention to the fragility of public trust in the institutions that collect, hold, and use health data and highlight the need to understand what health data sharing the public will support, under what circumstances, for what purposes, and with whom.

Evidence from systematic and narrative reviews demonstrates broad public support for the secondary use of health data, particularly for health research [11-15]. However, research also shows that this support may not extend to sharing health data with the private sector, particularly if there is scope for commercial gain from such use [2,4,16-18]. The risks related to sharing health data, such as the potential for privacy violations, inaccuracy, misuse, discrimination, reputational damage, and embarrassment, are generally regarded as greater when sharing data with the private sector, even if it is for the purposes of research and development.

Public reticence about sharing health data with the private sector does not seem to be matched by similar concerns among governments. Rather, sharing health data with the private sector has become a component of many governments' health and economic strategies [19-21]. For example, the use of large public data sets to support all stages of therapeutic development is one area of focus in Australia's 2016 National Research Infrastructure Roadmap [22]. Internationally, many other countries have made similar moves through regulatory changes to increase access to and use of large public data sets [19,20]. If governments intend to share health data with the private sector, it is essential to know what the public considers important.

Aim

This paper reports a national survey that aimed to explore public attitudes in Australia toward sharing health data with private companies for research on and development of therapeutic drugs and medical devices.

Methods

Ethics Approval

This study was approved by the University of Wollongong Ethics Committee. All participants provided consent before participating in the study.

Survey Instrument

To develop the survey, we carried out an extensive review of the literature and identified demographic and sociocultural factors that might influence how the public view sharing personal health information with the private sector. We searched the peer-reviewed literature for tools to measure public attitudes toward data sharing. We developed a new instrument by combining questions from pre-existing tools with new questions and drawing on insights from the literature [16,23].

We used Survey Monkey (Momentive Inc) software to design a web-based version of the instrument [24]. To support instrument readability, the survey was piloted with a convenience sample of the general population (n=10) aged ≥14 years. We selected pilot participants to provide a diverse group with respect to age, gender, education, ethnicity, and the presence or absence of long-term illness. These participants provided feedback on the meaning of each question, the design and layout as a whole, and how long it took to complete the survey. We then refined the survey instrument, with the final survey taking approximately 9 minutes to complete. The survey was then programmed by McNair yellowSquares on the Web Survey Creator survey platform and checked for usability and technical functionality before launching.

The 11-page survey instrument included a half-page summary explaining the concepts of data linkage and sharing, including potential benefits and risks. We mentioned research and development of new drugs and medical devices and ended the introduction with the following statement: "We would like to know what you think about sharing this information with private companies such as drug companies and medical device manufacturers where the goal is to support the development of new treatments for diseases and disabilities."

This was followed by a 29-item instrument covering sociodemographic and health-related information about participants; support for sharing health information with private

companies; general views about private companies; and experience with health data collection, consent, and conditions on sharing (the survey instrument is provided in Multimedia Appendix 1). A single open-ended question at the end of the survey invited additional comments. To help participants understand that each question referred to *deidentified* government health data, the following banner appeared at the top of each page: "The questions below are about your government health information which has personal information removed, e.g. no name, no address, no date of birth, no Medicare number."

Recruitment and Procedures

An experienced market research company, McNair yellowSquares, recruited an opt-in sample of 2500 participants drawn from its online panel. McNair yellowSquares was asked to secure a sample that would be nationally representative by age, gender, and location. The company purposively selected participants from its panel to meet this requirement. Although potentially less ideal than probability sampling, this methodology had the practical advantage of ease of implementation and was considered appropriate for this exploratory study [25].

Australian participants of the online panel were emailed an invitation to participate in the closed survey via a unique one-time use link. Once the survey was completed, the link was disabled to prevent duplicates and the panel was regularly checked for duplication with various data points. The survey was not advertised in any manner. Up to 2 reminder emails were sent over the 3-week period during which the survey was open (May 17, 2019 to June 7, 2019). Upon completing the introductory section to establish the quotas, participants were directed to the participant information sheet, which described the researchers, purpose of the study, risks and benefits, time needed for completion, and data protection and storage. Participants were asked to indicate that they understood the participant information sheet; on assenting to this, they were directed to the first page of the survey. All questions were mandatory, and some items provided "I do not know" and "I prefer not to answer" as options. Participants were not able to view their responses by moving backward. There was no randomization of items, and all responses were captured on the McNair yellowSquares Web Survey Creator survey platform. Participation was voluntary, and participants received a small reward upon completion of all items in the survey.

McNair yellowSquares provided deidentified participant responses to the research team. All data and analyses were stored on a secure University of Wollongong server, only accessed by the research team.

Statistical Analysis

IBM SPSS [26] was used to analyze the data. Only completed survey data were analyzed. First, we provided a descriptive summary of the survey outcomes by showing a frequency table with relative frequencies for each question of interest. The association between variables is given in cross tabulations, and *P* values are provided to answer the specific questions of interest. To support population inference, we analyzed the survey data using poststratification gender-by-age-by-state weights. We used the 2016 Australian Bureau of Statistics census data to obtain the Australian population characteristics of gender, age, and state and calculate the survey weights based on the realized sample characteristics after combining categories with small sample counts. All results except for participant demographic in this paper are obtained using the SPSS Complex Samples procedure. Raw proportions are reported to show the distributions of demographic information about the participants of this study (Table 1).

The open-ended question was analyzed inductively in NVivo (QSR International) [27]. Coding was conducted by 1 author (RB), with a second author (LC) coding half of the responses to ensure coding integrity. The authors compared coding and resolved differences before presenting the analysis to the entire research team for further discussion. The quotes in the *Results* section present examples of the diversity of responses in relation to different parts of the survey and indicate how respondents interpreted this question. A thorough analysis of this question will be presented in a separate paper.

Table 1. Demographics of participants (N=2537).

Characteristics	Values, n (%)
Gender (N=2537)	
Male	1243 (48.99)
Female	1285 (50.65)
Other	9 (0.38)
Geographical area (N=2537)	
Metropolitan	1682 (66.29)
Nonmetropolitan	855 (33.7)
Employment (n=2526)	
Full-time and part-time employed	1481 (58.63)
Unemployed	120 (4.75)
Home duties	250 (9.9)
Student	112 (4.43)
Retired	456 (18.05)
Unable to work	107 (4.24)
Age (years; N=2537)	
<29	552 (21.76)
30-49	873 (34.41)
50-64	652 (25.7)
≥65	460 (18.13)
Highest level of education (n=2525)	
No formal qualifications	45 (1.80)
Year 10 or school certificate	265 (10.5)
Finished high school	422 (16.71)
Vocational education (trade or technical education)	840 (33.27)
University	953 (37.74)
Self-rated health (N =2537)	
Poor or fair	758 (29.88)[a]
Good	991 (39.06)[a]
Very good or excellent	788 (31.06)[a]

[a]The Australian population—adjusted proportion.

Results

Participant Demographics

This data set contains responses from participants recruited by a market research company who completed the full survey. A total of 2537 participants were recruited for this study. The market research company employed to recruit participants set the location, age, and gender quota matrix to +1%. This allows for additional participants in the case of individuals being removed after data checks (eg, not located in Australia). Fewer potential participants were removed than was expected by the authors, resulting in a data set with slightly more participants than initially planned. There were equal number of male and female participants, with approximately two-third residing in metropolitan areas and 59.99% being employed. More than 20% of the participants were aged <29 years, approximately 35% were aged 30-49 years, 25% were aged 50-64 years, and 18% were aged ≥65 years. A total of 71.04% of the participants had a university or vocational qualification, and a similar proportion of the participants rated their health as good, very good, or excellent. Demographic information about the participants is provided in Table 1.

Participants were also asked if they had a My Health Record [28]. The My Health Record is a web-based summary of one's health status, which was first launched as an opt-in record in Australia in 2012 and then controversially amended to an opt-out model in 2018. Approximately 1 in 10 Australians opted out of the system when it was amended. Just over 40.99% of

participants in our survey said they had a My Health Record, 35.98% said they did not, and 23.02% were unsure.

Sharing Health Information With Private Companies

Overall, participants were ambivalent about whether or not to share their health data with private companies. Between 50.02% and 59.99% of all participants were willing to share their data, with a slightly fewer proportion in favor of sharing to improve health services and a slightly higher proportion in favor of sharing for research and development. Figure 1 shows the degree of support for sharing health data for various purposes. The range of views was reflected in participants' comments at the end of the survey:

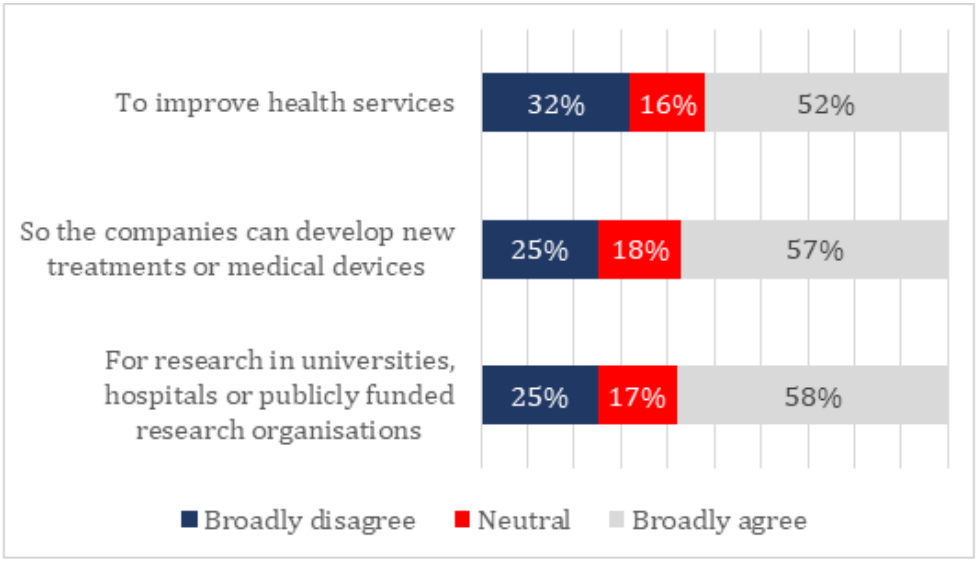
I am happy to share my information if it benefits me and others. [Male, 55-59 years, metropolitan]

Sharing health information with private companies is ok if the goal is to support the development of new treatments for diseases and disabilities. [Male, 18-24 years, metropolitan]

I don't think that information is any use to anybody for developing new drugs or procedures. [Male, ≥75 years, metropolitan]

I don't like my information being used by a private company. [Male, 60-64 years, metropolitan]

Figure 1. Support for sharing government health data with the private sector (N=2537): "To what extent do you agree with the government sharing your health information with private companies, such as drug companies or medical device manufactures?".

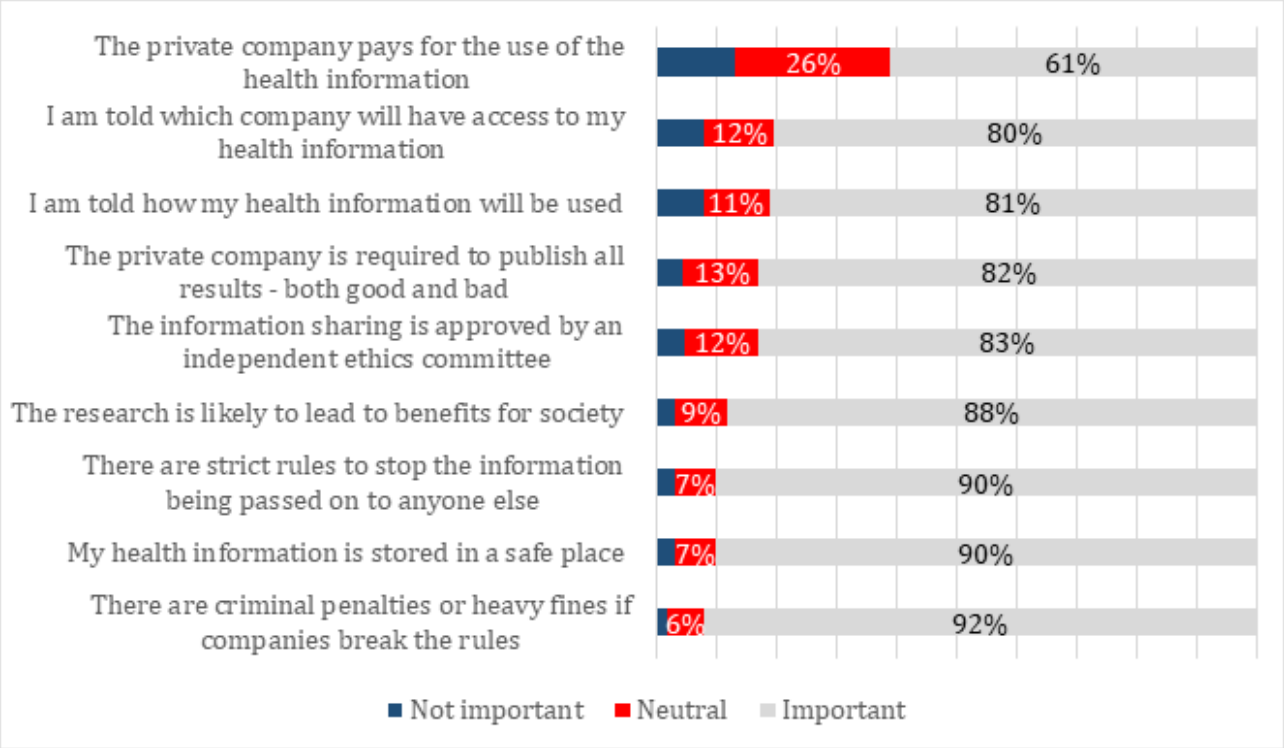


Conditions on Sharing

Participants were asked to assume that the government had decided to share their health information with a private company and to indicate the importance of various conditions on sharing. The participants responded on a scale of 1-7 with the anchors *Not important at all* and *Very important* and 4 in the neutral position. For all statements except one, 80.02% or more of the participants agreed that the condition was important. For one statement—private companies should pay for the use of the information—a small majority of participants (61.01%) considered the condition to be important. Figure 2 shows participants' responses to the conditions on sharing government health data with private companies.

For the aforementioned conditions, we compared the responses of participants who had previously indicated that they were willing to share government health data (for all three purposes) with those who were neutral or not willing to share data (Multimedia Appendix 2). In general, participants who were willing to share data were more concerned that the conditions be met; the differences were relatively small (7.64% to 28.54%), but *P* values suggest that the differences are statistically significant. The exceptions to this finding were being told which companies would have access to health information (for all 3 purposes), how information would be used (for development and research purposes), and whether the company would pay for the data (for research purposes). For these conditions, there were no differences between the 2 groups.

Figure 2. Conditions on sharing government health data with the private companies—adjusted percentages of (N=2537): "How important are various conditions if governments are to share data with private companies?".



Views About Private Companies

A series of statements were designed to assess participants' views about what private companies could or would do if they had access to government health information (Figure 3). Participants reported their level of agreement using a 7-point Likert scale ranging from *strongly disagree* to *strongly agree*. In reporting, these have been collapsed to *broadly disagree* (1-3), *neutral* (4), and *broadly agree* (5-7). Figure 3 shows the level of broad agreement for each of the statements.

There was wide variability in participants' responses to these statements (Figure 3). Over one-third of the participants considered that private companies could be trusted to act for the good of society or would store information safely, but these views were almost equally balanced by participants who thought the opposite. Approximately 59.99% of the participants thought that the government could not stop private companies from misusing information or control how they used it, but, again, approximately 1 out of 5 participants disagreed. Just under half of the participants said that their data could be reidentified, but at the other end of the scale, 23.02% of the participants did not think that reidentification was possible. Over half of the participants thought that private companies should not be allowed to make a profit from using the information, but one-fourth of the participants disagreed. For each statement, at least 1 in 5 participants was undecided.

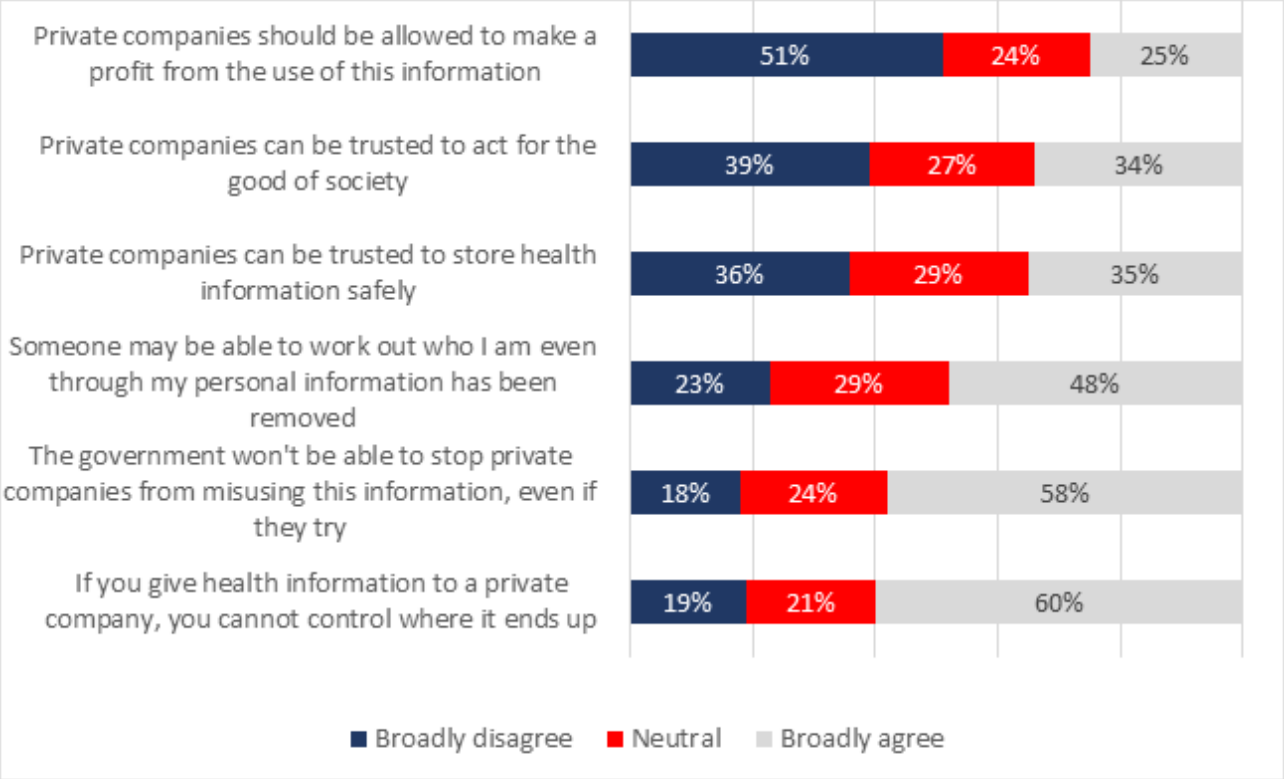
The comments at the end of the survey illustrated this range of views, and concerns about corporate interests, corruption, and profit making were recurrent themes:

> *I think private companies will inevitably use our information for profit rather than for the greater good.* [Male, 25-29 years, metropolitan]
>
> *The idea of greed preventing progress and a cure is 99% of my concerns.* [Female, 18-24 years, rural]
>
> *I just worry that my information will not be safe.* [Female, 30-34 years, metropolitan]

Although this was not a survey about how the public sector held and used health data, a number of participants were equally as concerned that the government could not keep their information secure:

> *The current government's record of online information processing has not been good. Look at what happened with the census.* [Male, 45-49 years, metropolitan]
>
> *Government is not very good at stopping anything in the past, e.g. bin full of census papers.* [Male, 60-64 years, rural]

Figure 3. Views on sharing government health data with private companies—adjusted percentages of (N=2537): "To what extent do you agree with the following statements about private companies using government health information to support development of new treatments?".



We examined the relationship between the participants' willingness to share government health data (for all 3 purposes) and their views about the private sector (Multimedia Appendix 3). The participants who had indicated that they were willing to share health data were more likely to say that private companies could be trusted (by between 51.3% and 60.82%), and that they should be able to make a profit from using government health data (51.31% to 45.62%). They were also slightly less concerned about the risk of identification (0.6% to 5.9%). Both groups of participants, who had indicated a willingness to share data and those who were neutral or did not wish to share data, were equally likely to think that any controls on data release would not work and that the government would not be able to control misuse by the private sector.

Consent Preferences

The consent preferences of the participants are shown in Multimedia Appendix 4. There was a preference for *opt-in* consent (54.98%): it was 3 times more popular than any other option. The participants' comments at the end of the survey reflected this view:

> *I would want total control over how, when and to whom my information is used and or shared with me giving the say so.* [Female, 60-64 years, rural]

> *It MUST be voluntary and OPT IN only.* [Male, 65-69 years, rural]

Each of the three other options for consent—refuse to share information at all, opt out, and *don't need to know*—attracted approximately 13% of the participants. For those who wanted opt-in consent, 62.51% requested that they be asked *every time* and 23.58% wanted to give general consent and then be recontacted from time to time, whereas the rest wanted to give consent just once. Multimedia Appendix 4 shows the adjusted percentages of consent preferences.

Participants in the opt-in group were slightly more likely to rate as important the conditions that could be placed on sharing their health information than those in the opt-out group (Table 2). The largest difference was related to how their health information would be used (89.98% stating that this was important compared with 81% in the opt-out group), and the smallest difference was related to payment for use of information (62.98% compared with 59.99%).

Opt-in and opt-out participants held similar views about private companies, with differences between the 2 groups very small (2.7% to 6.6%) and nonsignificant for all but one statement (Table 3).

Table 2. Relationship between participants' views on consent (opt in vs opt out) and level of agreement with the conditions on sharing data.

Conditions on sharing	Type of consent		*P* value
	Opt in (n=1356), n (%)	Opt out (n=352), n (%)	
I am told how my health information will be used	1215 (89.6)	284 (80.6)	<.001[a]
I am told which company will have access to my health information	1189 (87.7)	284 (80.6)	.001[a]
My health information is stored in a safe place	1285 (94.8)	312 (88.7)	<.001[a]
The private company pays for the use of the health information	848 (62.5)	210 (59.7)	.38
The information sharing is approved by an independent ethics committee	1187 (87.5)	286 (81.2)	.007[a]
The private company is required to publish all results—both good and bad	1180 (87.0)	288 (81.9)	.02[a]
The research is likely to lead to benefits for society	1246 (91.9)	304 (86.3)	.004[a]
There are strict rules to stop the information being passed on to anyone else	1275 (94.0)	310 (88.1)	<.001[a]

[a]Indicates level of significance at $P<.05$.

Table 3. Relationship between participants' views on consent (opt in vs opt out) and views about private companies.

Views about private companies	Type of consent		*P* value
	Opt in (n=1356), n (%)	Opt out (n=352), n (%)	
Private companies can be trusted to store health information safely	494 (36.4)	119 (33.7)	.38
Private companies should be allowed to make profit from the use of this information	346 (25.5)	98 (27.8)	.41
Private companies can be trusted to act for the good of society	473 (34.9)	105 (29.9)	.10
If you give health information to a private company, you cannot control where it ends up	868 (64.0)	202 (57.4)	.04[a]
Someone may be able to work out who I am even though my personal information has been removed	679 (50.1)	157 (44.6)	.09
The government won't be able to stop private companies from misusing this information, even if they try	818 (60.3)	199 (56.6)	.24

[a]Indicates level of significance at $P<.05$.

Sociodemographic Patterning of Responses

We investigated the impact of various sociodemographic factors on participants' views about whether health information should be shared and the conditions under which sharing might be acceptable (Multimedia Appendices 5-7). In general, demographic factors seemed to have only a small impact on participants' views, with differences being less than 5.99% for most demographic factors. There were a small number of exceptions. *Older people* (aged >65 years) were more willing than the youngest age group (60.2%-70.1% compared with 49.2%-56.4%) to share their health information with private companies. They were slightly less troubled than younger people about knowing which companies would have access to their data and more committed to publishing negative results (89% compared with 78.01%). The 3 oldest age groups were more supportive of criminal penalties, and the youngest age group was least likely to agree that ethics committee oversight was needed.

Across all measures, differences between people living in *metropolitan* and *nonmetropolitan areas* were small, with the largest difference (5.01%) between the groups showing nonmetropolitan dwellers slightly less likely to support data sharing for research.

The participants' *level of education* was related to their views, but only for some domains (Multimedia Appendices 5-7). The participants' level of education was not related to the degree of support for sharing government health data. However, participants with higher levels of education were generally more concerned about having conditions placed on the release of data, with differences between the least and most well-educated groups ranging from 16.98% to 40.01%. For example, 81.99% of the participants with university-level education wanted ethics committee oversight of data sharing, compared with 54% for participants with only year 10-level education. Compared with participants with year 10-level education, participants with university-level education were also more likely to want to know how their information would be used (81% compared with 46.98%), which company would access their data (81% compared with 50.02%), and that all results would be published (81.98% compared with 42.01%). A history of employment in the health sector or research did not appear to influence participants' responses.

Participants with *poorer self-reported healthstatus* were slightly less likely to support (5.01% to 5.99%) sharing their health data with the private sector, as were those who *took prescribed medications* (2.99% to 5.01%). However, participants (5.01% to 7.02%) who reported *having a chronic condition* were slightly more likely to support sharing data with the private sector.

Participants who said they had a *My Health Record* were between 17.2% and 20.4% more likely than those who said they did not have a record to support sharing data with private companies for health services improvement, development, or research.

Open-ended Question

The final question in the survey asked, "Is there anything else you would like to tell us about your views on sharing government health information with private companies where the goal is to support the development of new treatments for diseases and disabilities?" Approximately 18.01% of all respondents provided comments, primarily describing concerns about sharing government health information and the conditions under which they would support sharing or indicating support for data sharing.

Lack of trust in both private companies and the government was the most common concern. The participants cited corporate interests, corruption, and profit making as the main reasons for their distrust of private companies. They also referenced the poor track record of the government in handling data, and they questioned the ability of the government to keep their data secure and prevent misuse. Support for regulated access to health information was linked to respondents' concerns about security:

> *There have been recorded cases of information being misused, be it metadata to health information. The current government's record of online information processing has not been good. Look at what happened with the census.* [Male, 45-49 years, metropolitan]
>
> *Not in favour at all as I don't trust private companies with any sort of information & same goes for this bloody lying, corrupt government!!!* [Female, ≥75 years, metropolitan]
>
> *I think private companies will inevitably use our information for profit rather than for the greater good.* [Male, 25-29 years, metropolitan]

The respondents explained that if government health information is to be shared with private companies, certain conditions need to be met. The most common requirement was anonymization of health information and a guarantee that all personal information be removed. In addition, a large subset of participants believed that data sharing needs to deliver public benefits or support the common good. They provided examples of public benefits, including developing new treatments, finding cures, or improving the health of society. Giving consent was a prerequisite to sharing health information for many participants and the right to *opt in* rather than *opt out* was highlighted by a subset.

Discussion

Principal Findings

This nationally representative survey provides preliminary evidence that Australians are uncertain about sharing their health data with the private sector. Although just over half of all respondents supported sharing health data with the private sector, there was also strong support for strict conditions on sharing data and for opt-in consent. These views were reinforced by participants' ambivalence about the roles, motives, and actions of the private sector with respect to health data. Although, as a short survey, it represents relatively uninformed positions, it does indicate how people might react initially to reports of data sharing with the private sector in the news media or in public documents.

The findings of this survey demonstrate how difficult it may be to achieve policy change in this area in directions that are also acceptable to the community. Some of the conditions that participants wanted to impose on data sharing, such as using opt-in consent and providing information about each instance of use to each person who has provided data, are also conditions that some advocates of sharing would argue cannot be implemented [23,29-31]. Some intuitively attractive conditions, such as ensuring safe storage or compelling private companies to publish findings, may be difficult to enact through legislation and even more difficult to police. For example, despite decades of lament about publication bias in health research [32], relatively little headway has been made to change the practice [33-35].

A second set of challenges for policy makers may lie in identifying exactly which members of the community are concerned about what aspects of data sharing. In this survey, sociodemographic differences in views were generally small, and there were widely divergent views about what private companies could or would do if they had access to government health information. The participants who were willing to share health data were more cautious about the conditions under which they would be willing to share, but they were also more willing to trust the private sector and more willing to allow the private sector to take profits. The reasons for these findings are unclear, but they could suggest that participants had variable understandings of the private sector when answering the survey or that they had particular companies in mind. Whatever the reason, educating people about why it might be acceptable for the private sector to use public administrative data is unlikely to resonate equally across the community.

In its 2017 report on data availability and use in Australia, the Productivity Commission concluded that Australia lags behind other countries in its use of public sector data, particularly in the private sector [36]. In the Commission's view, Australia's foot-dragging has multiple causes, with limited community understanding and fragile trust at the top of the list, closely followed by legislative complexity, lack of leadership, data breaches, and poor data quality.

Our survey findings support the Commission's concern about the lack of community trust in data sharing. The participants in

our survey were uncertain about whether the private sector could be trusted, with at least one-third of the participants doubting the motivations and behaviors of the private sector when it came to their health information. However, many participants also agreed that sharing their health information with the private sector could yield public benefits, with just over half of all participants supporting the use of health information by private companies.

Recent scholarly studies of public views on using health data for secondary purposes also emphasize the importance of these 2 domains of trust and public benefit [15,37-39]. These recent studies cohere with our findings that understanding the benefits that can arise from using health data is necessary, but not sufficient, if the public is to entrust its health data to the private sector [10,40]. Many people are still uncomfortable with the idea of private companies accessing their government health data [41], and they have particular concerns about data privacy [42] and passing information on to marketers or insurers [43].

Building trust is not just a matter for the private sector. Trust in government is also important because it is the government that collects, holds, and releases health information in the first place [44]. This disquiet was reflected in the comments in our survey, although we did not actually ask participants to tell us their views on how well governments manage health data. Recent studies have also highlighted public misgiving about the public sector's ability to implement and manage data sharing and linkage safely, both in general and with private companies [37,41,42]. In her article examining Australian women's views and experiences of the My Health Record, Lupton [42] highlighted a number of well-publicized data breaches in Australia that may have contributed to participants' cynicism about their government's ability to keep health data protected.

At the time of this survey, we found no other Australian studies that provide a quantitative estimate of public support for sharing data with the private sector. The small number of international studies placed support for data sharing between 15% and 65% [45-47], a much larger range compared with our finding of 52% to 58%. These point estimates are helpful, but different research approaches are needed to reveal what lies beneath these numbers. Studies using focus groups, particularly in vulnerable populations; engagement and feedback through publicly focused websites; and deliberative methods such as citizens' juries will all help explain why participants are reticent to share their health data. For some topics, the use of different methods may yield different answers. For example, we found strong support for opt-in consent in this survey, whereas deliberative studies suggest that people may become less concerned about consent when they understand that shifting to opt-in consent for the secondary use of administrative health data would make the conduct of most big data research impractical and the findings untrustworthy [16,48]. The participants in our study were probably not all that different from other people in struggling to understand how data sharing, deidentification, and data linkage work or even how administrative health data could be used for research and development.

Limitations

This survey was conducted with an online panel of members of the public who had signed up to participate in research questionnaires, and it, therefore, has a number of limitations. In particular, as the participants were members of a panel who expressed interest and willingness to participate in research surveys, they may be more likely to be supportive of research, or at least more interested than the general public in research. The participants also probably had a reasonable level of confidence in using information technology and the internet, although what this meant for their attitudes to sharing their health data with the private sector was unclear. In addition, although participants were directed to focus on pharmaceutical companies and medical device manufacturers, it is possible that they also had other private health industries such as marketing and insurance companies in mind. This may have influenced the participants' responses.

Conclusions

Although there is broad public support for the secondary use of health data, our survey findings suggest that this support only extends to sharing health data with the private sector under tightly controlled circumstances. However, significant concerns are likely to remain. Addressing public concern about sharing government health data with the private sector will require more and better engagement to build community understanding about how agencies can collect, share, protect, and use their personal data.

Acknowledgments

This research was conducted on behalf of, and with the support of, the Population Health Research Network. The content is solely the responsibility of the authors and does not represent the views of the Population Health Research Network.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Community attitudes survey.
[PDF File (Adobe PDF File), 145 KB - jmir_v23i10e24200_app1.pdf]

Multimedia Appendix 2

The proportion of participants who were willing to share government health data by proportion of participants who agreed or disagreed that specific conditions should be met before sharing could occur.
[PNG File , 66 KB - jmir_v23i10e24200_app2.png]

Multimedia Appendix 3
The proportion of participants who were willing to share government health data by proportion of participants who agreed or disagreed on the views of private companies.
[PNG File , 63 KB - jmir_v23i10e24200_app3.png]

Multimedia Appendix 4
Consent preferences—adjusted percentages (N=2573): “What do you think about your health information being used by private companies for the development of new medicines or devices?”.
[PNG File , 21 KB - jmir_v23i10e24200_app4.png]

Multimedia Appendix 5
Adjusted percentages of willingness to share government health data with private companies by sociodemographic pattering (N=2537): “To what extent do you agree with the government sharing your health information with private companies, such as drug companies or medical device manufacturers?”.
[PDF File (Adobe PDF File), 122 KB - jmir_v23i10e24200_app5.pdf]

Multimedia Appendix 6
Adjusted percentages of conditions on sharing government health data with private companies by sociodemographic patterning (N=2537): “How important is it that each of the following conditions be met when information is shared with the private sector?”.
[PDF File (Adobe PDF File), 153 KB - jmir_v23i10e24200_app6.pdf]

Multimedia Appendix 7
Adjusted percentages of views on sharing government health data with private companies by sociodemographic patterning (N=2537): “To what extent do you agree with the following statements about private companies using government health information to support development of new treatments?”.
[PDF File (Adobe PDF File), 140 KB - jmir_v23i10e24200_app7.pdf]

References

1. Aitken M, Cunningham-Burley S, Pagliari C. Moving from trust to trustworthiness: Experiences of public engagement in the Scottish Health Informatics Programme. Sci Public Policy 2016 Oct;43(5):713-723 [FREE Full text] [doi: 10.1093/scipol/scv075] [Medline: 28066123]
2. Aitken M, Porteous C, Creamer E, Cunningham-Burley S. Who benefits and how? Public expectations of public benefits from data-intensive health research. Big Data & Society 2018 Dec 06;5(2):1-12 [FREE Full text] [doi: 10.1177/2053951718816724]
3. Lee K, Lim K, Jung SY, Ji H, Hong K, Hwang H, et al. Perspectives of patients, health care professionals, and developers toward blockchain-based health information exchange: qualitative study. J Med Internet Res 2020 Nov 13;22(11):e18582 [FREE Full text] [doi: 10.2196/18582] [Medline: 33185553]
4. Jones KH, Daniels H, Squires E, Ford DV. Public views on models for accessing genomic and health data for research: mixed methods study. J Med Internet Res 2019 Aug 21;21(8):e14384 [FREE Full text] [doi: 10.2196/14384] [Medline: 31436163]
5. Lin R, Cheng Y, Jiang Y. Exploring public awareness of overwork prevention with big data from google trends: retrospective analysis. J Med Internet Res 2020 Jun 26;22(6):e18181 [FREE Full text] [doi: 10.2196/18181] [Medline: 32589160]
6. Taylor M. Information governance as a force for good? Lessons to be learnt from Care.data. SCRIPTed 2014 Apr;11(1):1-8. [doi: 10.2966/scrip.110114.1]
7. Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why care.data ran into trouble. J Med Ethics 2015 May;41(5):404-409 [FREE Full text] [doi: 10.1136/medethics-2014-102374] [Medline: 25617016]
8. Publication of MBS/PBS data, Commissioner initiated investigation report. In: Australian Government. Canberra: Office of the Australian Information Commissioner; 2018:A.
9. Brown A. “Should I Stay or Should I Leave?”: Exploring (dis)continued Facebook use after the Cambridge analytica scandal. Soc Media Soc 2020 Mar 20;6(1):205630512091388 [FREE Full text] [doi: 10.1177/2056305120913884]
10. Chiauzzi E, Wicks P. Digital trespass: Ethical and terms-of-use violations by researchers accessing data from an online patient community. J Med Internet Res 2019 Feb 21;21(2):e11985 [FREE Full text] [doi: 10.2196/11985] [Medline: 30789346]

11. Aitken M, de St Jorre J, Pagliari C, Jepson R, Cunningham-Burley S. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. BMC Med Ethics 2016 Nov 10;17(1):73 [FREE Full text] [doi: 10.1186/s12910-016-0153-x] [Medline: 27832780]
12. Stockdale J, Cassell J, Ford E. "Giving something back": A systematic review and ethical enquiry into public views on the use of patient data for research in the United Kingdom and the Republic of Ireland. Wellcome Open Res 2018 Jan 17;3:6 [FREE Full text] [doi: 10.12688/wellcomeopenres.13531.2] [Medline: 30854470]
13. Garrison NA, Sathe NA, Antommaria AH, Holm IA, Sanderson SC, Smith ME, et al. A systematic literature review of individuals' perspectives on broad consent and data sharing in the United States. Genet Med 2015 Nov 19:663-671. [doi: 10.1038/gim.2015.138] [Medline: 26583683]
14. Kalkman S, van Delden J, Banerjee A, Tyl B, Mostert M, van Thiel G. Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. J Med Ethics 2019 Nov 12:105651 [FREE Full text] [doi: 10.1136/medethics-2019-105651] [Medline: 31719155]
15. Adjekum A, Blasimme A, Vayena E. Elements of trust in digital health systems: scoping review. J Med Internet Res 2018 Dec 13;20(12):e11254 [FREE Full text] [doi: 10.2196/11254] [Medline: 30545807]
16. The One-Way Mirror: Public attitudes to commercial access to health data. Data Visualization: Ipsos MORI / Wellcome Trust. URL: https://wellcome.figshare.com/articles/journal_contribution/The_One-Way_Mirror_Public_attitudes_to_commercial_access_to_health_data/5616448/1 [accessed 2021-07-27]
17. Kim T, Choi M. Older adults' willingness to share their personal and health information when adopting healthcare technology and services. Int J Med Inform 2019 Jun;126:86-94 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.03.010] [Medline: 31029268]
18. Paprica PA, de Melo MN, Schull MJ. Social licence and the general public's attitudes toward research based on linked administrative health data: a qualitative study. CMAJ Open 2019;7(1):40-46 [FREE Full text] [doi: 10.9778/cmajo.20180099] [Medline: 30718354]
19. Frequently asked questions: Public-Private Partnership (PPP) for Big Data. Press Release from European Commission. 2014. URL: https://ec.europa.eu/commission/presscorner/detail/en/MEMO_14_583 [accessed 2021-07-27]
20. Scottish Government. Joined Up Data for Better Decisions: A Strategy for Improving Data Access and Analysis. Edinburgh: Scottish Government; 2012. URL: https://www.gov.scot/publications/joined-up-data-better-decisions-strategy-improving-data-access-analysis/documents/
21. Pagliari C, Davidson S, Cunningham-Burley S, Laurie G, Aitken M, Sethi N. Public Acceptability of Data Sharing Between the Public, Private and Third Sectors for Research Purposes. Edinburgh: Scottish Government - Social Research Series; 2013.
22. 2016 National Research Infrastructure Roadmap released. Australian Government. 2017. URL: https://www.chiefscientist.gov.au/2017/05/2016-national-research-infrastructure-roadmap-released [accessed 2021-07-27]
23. Berry JG, Ryan P, Duszynski KM, Braunack-Mayer AJ, Carlson J, Xafis V, Vaccine Assessment using Linked Data (VALiD) Working Group. Parent perspectives on consent for the linkage of data to evaluate vaccine safety: a randomised trial of opt-in and opt-out consent. Clin Trials 2013;10(3):483-494. [doi: 10.1177/1740774513480568] [Medline: 23568940]
24. SurveyMonkey. SurveyMonkey Inc, San Mateo, California, US. URL: https://www.surveymonkey.com/ [accessed 2021-07-27]
25. Pennay DW, Neiger D, Lavrakas PJ, Borg K. The Online Panels Benchmarking Study: a Total Survey Error Comparison of Findings From Probability-based Surveys and Non-probability Online Panel Surveys in Australia. Canberra: Centre for Social Research and Methods, The Australian National University; 2018:A.
26. IBM SPSS software. IBM. URL: https://www.ibm.com/au-en/analytics/spss-statistics-software [accessed 2021-07-27]
27. Unlock insights in your data with powerful analysis. Nvivo. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2021-07-27]
28. My Health Record. Australian Government, Australian Digital Health Agency. URL: https://www.myhealthrecord.gov.au/news-and-media/australians-to-have-my-health-record [accessed 2021-07-27]
29. Cheung AS. Moving beyond consent for citizen science in big data health and medical research. Northwest J Technol Intellect Property 2018;16(1):A [FREE Full text]
30. Lipworth W, Mason PH, Kerridge I, Ioannidis JP. Ethics and epistemology in big data research. J Bioeth Inq 2017 Dec 20;14(4):489-500. [doi: 10.1007/s11673-017-9771-3] [Medline: 28321561]
31. Richards NM, King JH. Big Data Ethics. Wake Forest Law Review 2014 May 19:393-432 [FREE Full text]
32. Song F, Parekh S, Hooper L, Loke Y, Ryder J, Sutton A, et al. Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess 2010 Feb;14(8):iii, ix-iixi, 1 [FREE Full text] [doi: 10.3310/hta14080] [Medline: 20181324]
33. Vera-Badillo F, Napoleone M, Krzyzanowska M, Alibhai S, Chan A, Ocana A, et al. Bias in reporting of randomised clinical trials in oncology. Eur J Cancer 2016 Jul;61:29-35 [FREE Full text] [doi: 10.1016/j.ejca.2016.03.066] [Medline: 27151552]
34. DeVito NJ, Goldacre B. Catalogue of bias: publication bias. BMJ Evid Based Med 2019 Apr 06;24(2):53-54. [doi: 10.1136/bmjebm-2018-111107] [Medline: 30523135]

35. Howard B, Scott JT, Blubaugh M, Roepke B, Scheckel C, Vassar M. Systematic review: Outcome reporting bias is a problem in high impact factor neurology journals. PLoS One 2017;12(7):e0180986 [FREE Full text] [doi: 10.1371/journal.pone.0180986] [Medline: 28727834]
36. Data availability and use - Report No. 82. Australian Government - Productivity Commission. 2017. URL: https://www.pc.gov.au/inquiries/completed/data-access#report [accessed 2021-07-27]
37. Sheehan M, Friesen P, Balmer A, Cheeks C, Davidson S, Devereux J, et al. Trust, trustworthiness and sharing patient data for research. J Med Ethics 2020 May 18:106048. [doi: 10.1136/medethics-2019-106048] [Medline: 32424061]
38. Hawley K. Trustworthy groups and organisations. In: Faulkner P, Simpson T, editors. The Philosophy of Trust. Oxford, United Kingdom: Oxford University Press; 2017:230-249.
39. Kerasidou A. Trust me, I'm a researcher!: The role of trust in biomedical research. Med Health Care Philos 2017 Mar;20(1):43-50 [FREE Full text] [doi: 10.1007/s11019-016-9721-6] [Medline: 27638832]
40. Engaging consumers in their health data journey. Consumers Health Forum of Australia and NPS MedicineWise. 2018. URL: https://www.nps.org.au/assets/35dfcb232f9195d3-fc4272585b90-Engaging-Consumers-Health-Data-Report.pdf [accessed 2021-07-27]
41. Biddle N, Edwards B, Gray M, McEachern S. Public attitudes towards data governance in Australia. Centre for Social Research and Methods, Australian National University. 2019. URL: https://csrm.cass.anu.edu.au/research/publications/public-attitudes-towards-data-governance-australia-1 [accessed 2021-07-27]
42. Lupton D. 'I'd like to think you could trust the government, but I don't really think we can': Australian women's attitudes to and experiences of My Health Record. Digit Health 2019 Apr 29;5:2055207619847017 [FREE Full text] [doi: 10.1177/2055207619847017] [Medline: 31069106]
43. Robinson G, Dolk H, Dowds L, Given J, Kane F, Nelson E. Public attitudes to data sharing in Northern Ireland: Findings from the 2015 Northern Ireland life and times survey. Ulster University. 2018. URL: https://pure.ulster.ac.uk/en/publications/public-attitudes-to-data-sharing-in-northern-ireland-findings-fro [accessed 2021-07-27]
44. Ekeland AG, Linstad LH. Elaborating models of eHealth governance: qualitative systematic review. J Med Internet Res 2020 Oct 28;22(10):e17214. [doi: 10.2196/17214]
45. Darquy S, Moutel G, Lapointe A, D'Audiffret D, Champagnat J, Guerroui S, et al. Patient/family views on data sharing in rare diseases: study in the European LeukoTreat project. Eur J Hum Genet 2016 Mar 17;24(3):338-343. [doi: 10.1038/ejhg.2015.115] [Medline: 26081642]
46. Goodman D, Johnson CO, Bowen D, Smith M, Wenzel L, Edwards K. De-identified genomic data sharing: the research participant perspective. J Community Genet 2017 Jul 5;8(3):173-181 [FREE Full text] [doi: 10.1007/s12687-017-0300-1] [Medline: 28382417]
47. Zalin A, Papoutsi C, Shotliff K, Majeed A, Marston C, Reed J. The use of information for diabetes research and care: patient views in West London. Pract Diab 2016 Apr 20;33(3):81-86. [doi: 10.1002/pdi.2008]
48. Tully MP, Bozentko K, Clement S, Hunn A, Hassan L, Norris R, et al. Investigating the extent to which patients should control access to patient records for research: a deliberative process using citizens' juries. J Med Internet Res 2018 Mar 28;20(3):e112 [FREE Full text] [doi: 10.2196/jmir.7763] [Medline: 29592847]

Edited by R Kukafka, G Eysenbach; submitted 14.09.20; peer-reviewed by M Aitken, R Hendricks-Sturrup; comments to author 06.11.20; revised version received 21.12.20; accepted 19.05.21; published 01.10.21.

Please cite as:
Braunack-Mayer A, Fabrianesi B, Street J, O'Shaughnessy P, Carter SM, Engelen L, Carolan L, Bosward R, Roder D, Sproston K
Sharing Government Health Data With the Private Sector: Community Attitudes Survey
J Med Internet Res 2021;23(10):e24200
URL: https://www.jmir.org/2021/10/e24200
doi:10.2196/24200
PMID:34596573

Original Paper

A Web-Based Intervention (Germ Defence) to Increase Handwashing During a Pandemic: Process Evaluations of a Randomized Controlled Trial and Public Dissemination

Sascha Miller[1], MSc; Ben Ainsworth[2], PhD; Mark Weal[3], PhD; Peter Smith[4], PhD; Paul Little[5], PhD; Lucy Yardley[1,6], PhD; Leanne Morrison[1,5], PhD

[1]Center for Clinical and Community Applications of Health Psychology, Department of Psychology, University of Southampton, Southampton, United Kingdom

[2]Bath Centre for Mindfulness and Compassion, Department of Psychology, University of Bath, Bath, United Kingdom

[3]Web and Internet Science Group, Electronics and Computer Science, University of Southampton, Southampton, United Kingdom

[4]Department of Social Statistics and Demography, School of Economic, Social and Political Scientces, University of Southampton, Southampton, United Kingdom

[5]Primary Care and Population Sciences, School of Medicine, University of Southampton, Southampton, United Kingdom

[6]Centre for Academic Primary Care, School of Psychological Science, University of Bristol, Bristol, United Kingdom

Corresponding Author:
Sascha Miller, MSc
Center for Clinical and Community Applications of Health Psychology
Department of Psychology
University of Southampton
Building 44 Highfield Campus
Southampton, SO17 1BJ
United Kingdom
Phone: 44 02380595000
Email: sm2t07@soton.ac.uk

Abstract

Background: Washing hands helps prevent transmission of seasonal and pandemic respiratory viruses. In a randomized controlled trial (RCT) during the swine flu outbreak, participants with access to a fully automated, digital intervention promoting handwashing reported washing their hands more often and experienced fewer respiratory tract infections than those without access to the intervention. Based on these findings, the intervention was adapted, renamed as "Germ Defence," and a study was designed to assess the preliminary dissemination of the intervention to the general public to help prevent the spread of seasonal colds and flu.

Objective: This study compares the process evaluations of the RCT and Germ Defence dissemination to examine (1) how web-based research enrollment procedures affected those who used the intervention, (2) intervention usage in the 2 contexts, and (3) whether increased intentions to wash hands are replicated once disseminated.

Methods: The RCT ran between 2010 and 2012 recruiting participants offline from general practices, with restricted access to the intervention (N=9155). Germ Defence was disseminated as an open access website for use by the general public from 2016 to 2019 (N=624). The process evaluation plan was developed using Medical Research Council guidance and the framework for Analyzing and Measuring Usage and Engagement Data. Both interventions contained a goal-setting section where users self-reported current and intended handwashing behavior across 7 situations.

Results: During web-based enrolment, 54.3% (17,511/32,250) of the RCT participants dropped out of the study compared to 36.5% (358/982) of Germ Defence users. Having reached the start of the intervention, 93.8% (8586/9155) of RCT users completed the core section, whereas 65.1% (406/624) of Germ Defence users reached the same point. Users across both studies selected to increase their handwashing in 5 out of 7 situations, including before eating snacks (RCT mean difference 1.040, 95% CI 1.016-1.063; Germ Defence mean difference 0.949, 95% CI 0.766-1.132) and after blowing their nose, sneezing, or coughing (RCT mean difference 0.995, 95% CI 0.972-1.019; Germ Defence mean difference 0.842, 95% CI 0.675-1.008).

Conclusions: By comparing the preliminary dissemination of Germ Defence to the RCT, we were able to examine the potential effects of the research procedures on uptake and attrition such as the sizeable dropout during the RCT enrolment procedure that

may have led to a more motivated sample. The Germ Defence study highlighted the points of attrition within the intervention. Despite sample bias in the trial context, the intervention replicated increases in intentions to handwash when used "in the wild." This preliminary dissemination study informed the adaptation of the intervention for the COVID-19 health emergency, and it has now been disseminated globally.

Trial Registration: ISRCTN Registry ISRCTN75058295; https://www.isrctn.com/ISRCTN75058295

(J Med Internet Res 2021;23(10):e26104) doi:10.2196/26104

KEYWORDS

behavior; infection; prevention; respiratory tract infection; internet; evaluation studies; pandemic; COVID-19; transmission; virus; influenza; respiratory; intervention; digital intervention; dissemination

Introduction

Pandemic respiratory viruses present a global health threat, leading to more deaths across a wider spread of the population than seasonal flu, as seen through outbreaks such as SARS-CoV (severe acute respiratory syndrome), swine flu (H1N1 influenza), MERS-CoV (Middle East Respiratory Syndrome), as well as COVID-19 (SARS-CoV-2) [1]. However, for more vulnerable groups such as the older adults or seriously ill, seasonal flu can still present a serious health risk and increased likelihood of needing medical care [1]. Handwashing is an accessible and a simple infection control behavior [2,3] that has been promoted to the general public for many decades to slow the spread of cold and flu viruses both seasonally and during pandemics [4,5]. The COVID-19 pandemic has seen renewed calls for increased handwashing from governments and health organizations around the world as a means to control the spread of the virus [6,7]. Yet, despite high levels of health promotion and public awareness of handwashing, evidence from prior pandemics (ie, SARS-CoV and swine flu) suggests that increases in the reported levels of handwashing were low [8,9]. An effective, evidence-based behavioral intervention to support increased handwashing within the home in the event of pandemics such as COVID-19 and for seasonal respiratory tract infections is urgently required [10-13]. Digital interventions provide the advantages of quick dissemination and flexibility so that contents can be updated to reflect the changes during a pandemic.

PRIMIT (PRImary care trial of a website based infection control intervention to Modify Influenza-like illness and respiratory infection Transmission) was commissioned and funded by the UK Medical Research Council in 2008 for use in the event of a pandemic [14]. A theory-based, stand-alone, digital intervention to increase handwashing targeting the general population was developed using the person-based approach [15]. This included extensive mixed-methods evaluation of the usability, functionality, and acceptability of the intervention [3,16,17]. A randomized controlled trial (RCT) of over 20,000 participants was carried out in the United Kingdom while swine flu was circulating in the community. The findings of this RCT established that participants who had access to the intervention reported washing their hands more than those who did not: they and the people they lived with had lesser respiratory tract infections and users who contracted respiratory tract infections were ill for lesser time [14]. A behavioral analysis of the data collected during the RCT demonstrated that viewing each of the 4 available sessions led to additive increases in handwashing levels [18]. However, completing the first session was deemed to be effective engagement; after viewing this session, the majority of users increased their handwashing to sufficient levels to lower the transmission of viruses [18,19]. This session contained a range of motivational messages and included a goal-setting behavior change technique (BCT). The goal-setting section required users to consider their current handwashing frequency across a range of specific situations (eg, after going to the toilet, before eating a meal) and then make a plan to increase handwashing in the future. Based on the findings from the behavioral analysis [18], in 2016, the architecture of the intervention was adapted to enable access to all the intervention contents in 1 session (rather than 4 sequential sessions). The intervention was renamed as "Germ Defence" and disseminated to the general public to lower the transmission of seasonal cold and flu viruses [20]. The data generated by the PRIMIT and Germ Defence studies provide us with the opportunity to compare the uptake and usage of the intervention in the RCT and "in the wild" contexts.

RCTs are considered to be the gold standard for evaluating intervention efficacy. However, they have been criticized for producing results that are not replicated once an intervention is freely disseminated to the intended population [21-23]. RCTs require artificial conditions that differ from how an intervention may be used "in the wild." For example, RCTs of digital interventions typically involve effortful web-based research procedures such as lengthy enrollment processes, followed by in-depth baseline, interim, and follow-up data collection across several weeks, if not months [24]. These demanding procedures may affect both the type of people who take part in an RCT and how they use the intervention [21,25,26]. Participants with lower levels of education or health literacy are more likely to drop out of a study during these processes [27], meaning that these groups are unintentionally excluded. Conversely, this can lead to volunteer bias within the sample [26], whereby people who are more highly motivated to perform the behavior or have higher levels of health literacy/education are more likely to participate in the trial. In addition, research suggests that effortful procedures may increase participants' sense of support or accountability, leading to artificially high levels of engagement [21].

Calls have been made to evaluate digital interventions so that their effectiveness may be established beyond trial conditions [24]. However, a review of studies of publicly available digital interventions found that few reported in-depth usage data and

only 1 intervention was identified as having been empirically examined through an RCT as well [21]. Having established the effectiveness of the PRIMIT intervention in a trial context, the Germ Defence study was subsequently devised as a novel preliminary dissemination of the intervention to the general public. Comparing intervention usage across these 2 contexts provides the opportunity to examine the impact of the rigors of RCTs on intervention usage and effectiveness and to assess whether the intervention maintains efficacy when accessed "in the wild" [28,29].

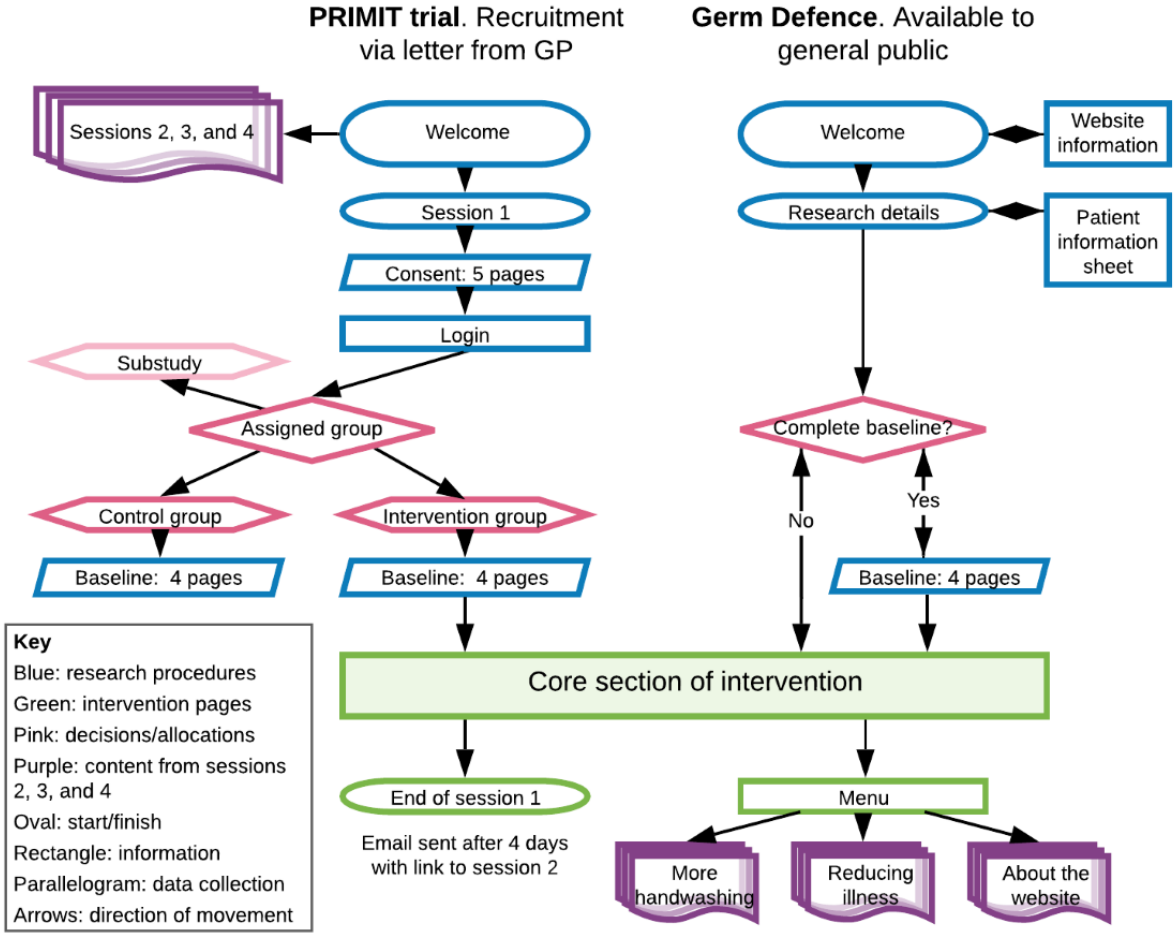
This paper reports process evaluations [30] of the PRIMIT and Germ Defence studies. The aim of this study was to provide insights that would help to optimize the intervention and maximize reach in the event of a pandemic and to provide a comparison between RCT and "in the wild" contexts. This has proved timely, as the Germ Defence intervention has now been adapted and widely disseminated for use during the COVID-19 pandemic [31,32]. By comparing trial and "in the wild" contexts, we were able to examine the effect of the RCT research procedures on uptake and usage. Examining attrition during trial procedures has the potential to inform RCT design and produce findings that will translate when implemented at scale within the community. Further, examining the usage of intervention content across the 2 contexts provides insight into how to design interventions that are likely to be more effective and engaging when disseminated at scale. As the intervention is effective where other public health campaigns have had less success [8,9], it is important to understand how increased handwashing was achieved. The Germ Defence study enabled us to examine whether the goal-setting section continued to support increased intentions for handwashing in specific situations when the intervention is not supported by the trial context. The specific objectives of this study were to compare (1) completion of the web-based research enrollment procedures, (2) usage of the intervention content, and (3) self-reported intentions to wash hands in specific situations.

Methods

Study Procedures

The PRIMIT trial ran across 3 winters from 2010 to 2012. Participants were randomized equally to either the intervention or control groups. Both groups were required to give consent, log in, and complete baseline measures to enroll in the study. Participants in the intervention group were given immediate access to the first session of the website (Figure 1). This trial is registered with the ISRCTN registry number ISRCTN75058295; for full details of the trial, see [14].

Figure 1. Research procedures for PRIMIT (PRImary care trial of a website based infection control intervention to Modify Influenza-like illness and respiratory infection Transmission) and Germ Defence studies.



Germ Defence was available from 2016 to 2019. On the webpages prior to accessing the intervention, users were provided with details of the study and asked if they would like to participate by providing baseline measures or go directly to

the intervention. For users volunteering to complete baseline measures, the consent process was adapted to provide minimal but sufficient information for informed consent, with full details available as an optional click-through.

Study Design

A process evaluation plan to compare usage of PRIMIT and Germ Defence studies and interventions was developed while adapting and preparing Germ Defence for dissemination in 2016. The plan was structured using the Analyzing and Measuring Usage and Engagement Data (AMUsED) framework [20,33-35] (see Multimedia Appendices 1-3) and is in line with Medical Research Council process evaluation guidance [30]. The process evaluation team combined expertise in psychology, primary care, statistical analyses, and computer science. Most team members were already familiar with the intervention having worked on the PRIMIT study development [3,16,17] or the RCT [14,18].

Participants

PRIMIT study participants were recruited to the RCT by letters sent from National Health Service general practices to patients aged 18 years or older living with at least one other person who was willing to report their illnesses for the study and had internet access. Patients with severe mental problems, who were terminally ill, or who had a skin complaint that restricts handwashing were excluded from this study. Access to the website was restricted so that only users who received specific login details were able to enroll and consent to the study online. Participants were not compensated for taking part in this study.

The Germ Defence study was accessible to members of the general public who indicated being aged 16 years or older. Website details were distributed to various health sector organizations for use with their clients (eg, health support charities for people at risk from respiratory tract infections and local council public health organizations). Germ Defence is endorsed by the National Institute for Health and Care Excellence antimicrobial resistance guidelines and a link to the intervention is provided on their website. The intervention was also promoted directly to the general public (eg, through a chain of high-street pharmacies, play groups, social media, student intranet).

Ethical Approval

The PRIMIT study was approved by a multicenter research ethics committee (08/H0502/14). The Germ Defence study was approved by the University of Southampton, School of Psychology ethics committee (19399).

Measures

The PRIMIT study and Germ Defence websites were developed using LifeGuide software (University of Southampton) [36]. The software collected self-report measures and automatically recorded user interaction with the intervention such as time, date, pages viewed, and the order they were viewed in. For the Germ Defence study, web-based, self-reported baseline measures of user characteristics and behaviors were streamlined in comparison to those used in the PRIMIT study (see Table 1). Measures were selected based on the findings from the PRIMIT study [14] and completion of the AMUsED framework checklists [20]. By making the baseline measures voluntary, it was possible to split the sample into 2 groups: those who chose to complete these measures and those who did not.

Table 1. Germ Defence and equivalent PRIMIT web-based study measures.

Measures, items	Description	Response options	PRIMIT[a] study	Germ Defence
Baseline web-based self-report questionnaires				
Gender	Please select one	Male, Female, Other	✓	✓
Age	Select age from dropdown box.	PRIMIT study: 18-99 years, Germ Defence: 16-99 years	✓	✓
Recruitment	How did you hear about the website?	health support group, school or children's group, search engine, social media, news article, word of mouth, other.		✓
Current daily handwashing	How many times a day do you wash your hands, including using antibacterial hand gel, on average?	0-2, 3-4, 5-6, 7-9, 10 times or more.	✓	✓
Intended daily handwashing	How many times a day do you intend to wash your hands in the future, including using antibacterial hand gel, on average?	0-2, 3-4, 5-6, 7-9, 10 times or more.	✓	✓
Perceived risk of catching a respiratory tract infection				
Perceived likelihood of users catching a respiratory tract infection	PRIMIT: 2 items (eg, My chances of catching a cold or flu are high if I don't take action to prevent it).	1-7 Likert scale ranging from "strongly disagree" to "strongly agree"	✓	✓
	Germ Defence: My chances of catching a cold, flu, or stomach bug are high if I don't take action to prevent it.	1-7 Likert scale ranging from "strongly disagree" to "strongly agree."		
Perceived severity of catching a respiratory tract infection	If I catch a cold, flu, or stomach bug, I am likely to become seriously ill.	1-7 Likert scale ranging from "strongly disagree" to "strongly agree."		✓
Goal-setting section within the intervention				
Current behavior: Over the last week, I washed my hands	C1: Before I ate a meal	1=almost never, 2=sometimes, 3=quite often, 4=very often, 5=almost always	✓	✓
	C2: Before I ate snacks	1=almost never, 2=sometimes, 3=quite often, 4=very often, 5=almost always	✓	✓
	C3: After I went to the toilet	1=almost never, 2=sometimes, 3=quite often, 4=very often, 5=almost always	✓	✓
	C4: When I came in to the house	1=almost never, 2=sometimes, 3=quite often, 4=very often, 5=almost always	✓	✓
	C5: After I had been close to someone with a cold, flu, or upset stomach	1=almost never, 2=sometimes, 3=quite often, 4=very often, 5=almost always Germ Defence only: 0=not applicable	✓	✓
	C6: After blowing my nose or sneezing/coughing on my hands	1=almost never, 2=sometimes, 3=quite often, 4=very often, 5=almost always Germ Defence only: 0=not applicable	✓	✓
	C7: After touching something with lots of germs on	Germ Defence only: 1=almost never, 2=sometimes, 3=quite often, 4=very often, 5=almost always 0=not applicable		✓
Intended behavior: I will try to wash my hands	Items as for current behavior	Responses as for current behavior		

[a]PRIMIT: PRImary care trial of a website based infection control intervention to Modify Influenza-like illness and respiratory infection Transmission.

Intervention Structure

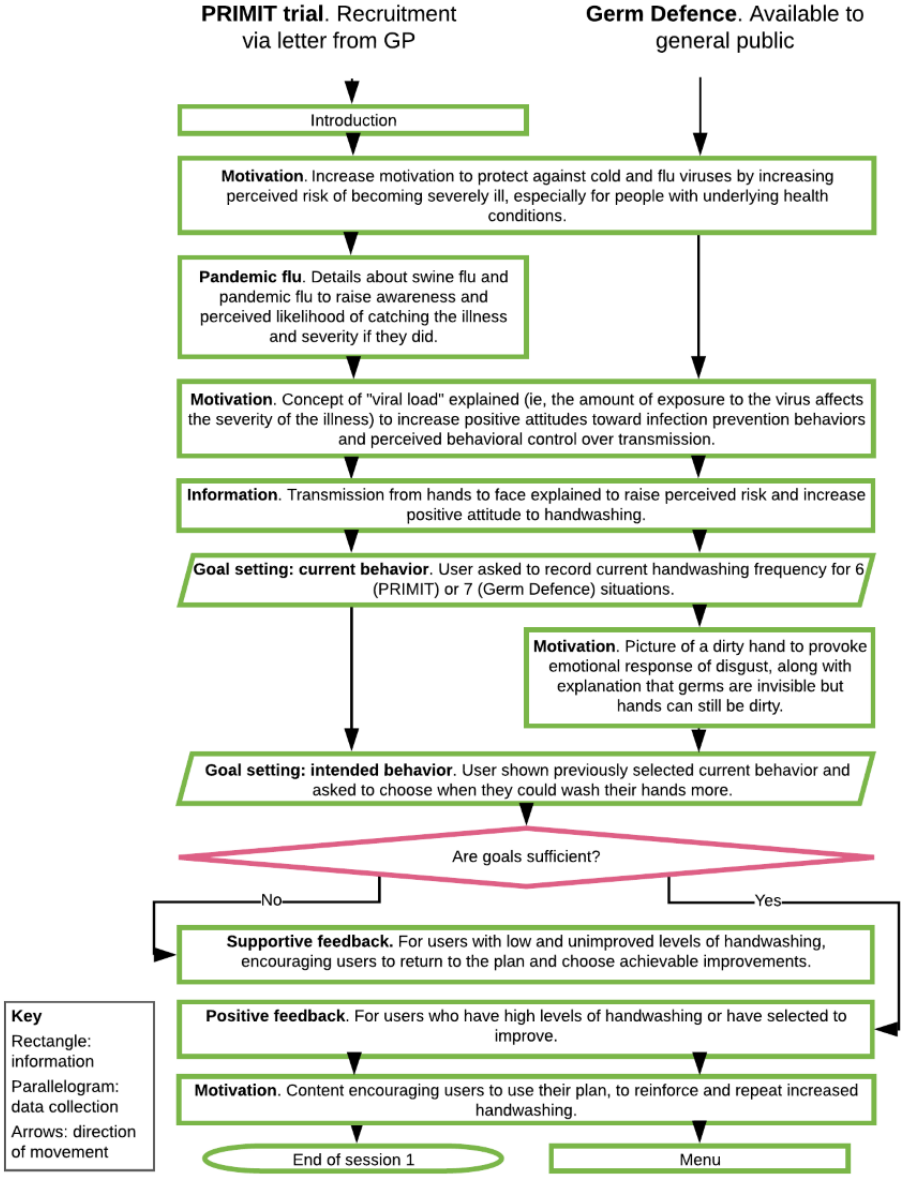
The intervention used for the PRIMIT and Germ Defence studies was developed by researchers at the University of Southampton. Full details of intervention structure and development are reported elsewhere [3,14,16,17] and archived copies are available [37] (see PRIMIT and Germ Defence v1). The intervention content for both the studies was "frozen" during the trial and dissemination periods.

PRIMIT study users had access to 4 sessions of theory-based and evidence-based content released sequentially over 3.5 weeks. Having established that viewing the first session of the PRIMIT study intervention represented effective engagement

[18], Germ Defence was also structured with the same content as the core section (Figure 2). After completing the core section, all information from the following 3 sessions was then available, allowing users to access the whole intervention in 1 visit, with no further sessions necessary. These core pages focused on informing, supporting, and motivating users to wash their hands more [3,17] and were "tunneled" so that users had to move through pages in a specific order. The appearance was updated for Germ Defence, but only 2 content changes were made to the core section: details about pandemic flu [17] and particularly the swine flu outbreak that started in 2009 were removed and a motivational message from the original "session 2" was added in between the goal-setting pages.

Within this core section was a goal-setting BCT encouraging handwashing in situations where there is increased risk of infection through spreading the virus inside the home (eg, when I come into the house) or transmission of viruses from hands to face (eg, before I eat snacks). These are the only pages in the intervention that require user interaction, and users were only able to progress further if they selected current and intended future handwashing for all situations. Users were provided with feedback after completing their plan: positive for high or improved levels of handwashing; supportive for low and unimproved levels of handwashing, suggesting users' revisit and change of their plan. Having finished these pages, PRIMIT study users had completed session 1. For Germ Defence users, after this core section was a menu allowing users to access 3 further components with information from sessions 2-4 of the PRIMIT study. These included further support for handwashing and information on other infection prevention behaviors (ie, social distancing, not touching your face, wearing a mask, cleaning surfaces).

Figure 2. PRIMIT (PRImary care trial of a website based infection control intervention to Modify Influenza-like illness and respiratory infection Transmission) study and Germ Defence intervention structure for core section.



Data Analysis

Anonymous usage data collected automatically through interaction with the intervention was examined for all Germ Defence users and PRIMIT participants in the intervention group. Usage of the core section of Germ Defence is reported for the whole sample and the subsample of users who completed baseline measures. Responses to self-report measures (ie, at baseline and goal-setting) were analyzed for the PRIMIT study intervention group and Germ Defence baseline respondents. As part of the PRIMIT trial requirement, a small proportion of users was placed into a substudy and they completed different baseline measures that excluded handwashing behaviors. As this group cannot be included in analyses with other users, they have been removed from this process evaluation.

SPSS for Windows version 25 (IBM Corp) was used for all statistical analyses. The frequency distribution of the scores for self-report measures was visually assessed for normality, and all measures of user characteristics at baseline were nonnormally distributed. All analyses were two-tailed. Owing to the difference in sample sizes, analyses of usage data are reported as percentages to enable more meaningful comparisons.

Completion of Web-Based Research Enrollment Procedures

To examine the completion of the web-based research enrollment procedures, usage data were analyzed for the proportions of users who had viewed each page, continued on to another page, and left the study. Owing to changes in the data collection across the 3 winters, usage data covering research enrollment for the first 2 winters are not available (intervention usage data were unaffected). Therefore reported usage percentages for enrollment pages are taken from users recruited in the third winter, representing over 75% of the total sample. The number of views for the welcome pages (PRIMIT study n=209,852, Germ Defence n=12,106) included multiple views by the same participants, people who arrived at the page unintentionally, and bots (automated software programs). In addition, PRIMIT study participants who were returning to the intervention to view later sessions are included in this number. To enable equivalent comparisons between the 2 research procedures and to ensure that usage was intentional, the first page has been separated and proportions of participants have been calculated based on the sample who reached the second page.

Intervention Usage

Intervention content usage was examined for the pages viewed and attrition in the core section of the intervention. Confidence intervals at 95% were used to compare the means of baseline behavioral and psychological measures to determine whether there were any differences between the intervention users in the two contexts. Logistic regression analyses were carried out to examine whether baseline user characteristics predicted completion of the core section of the intervention.

Handwashing Behaviors

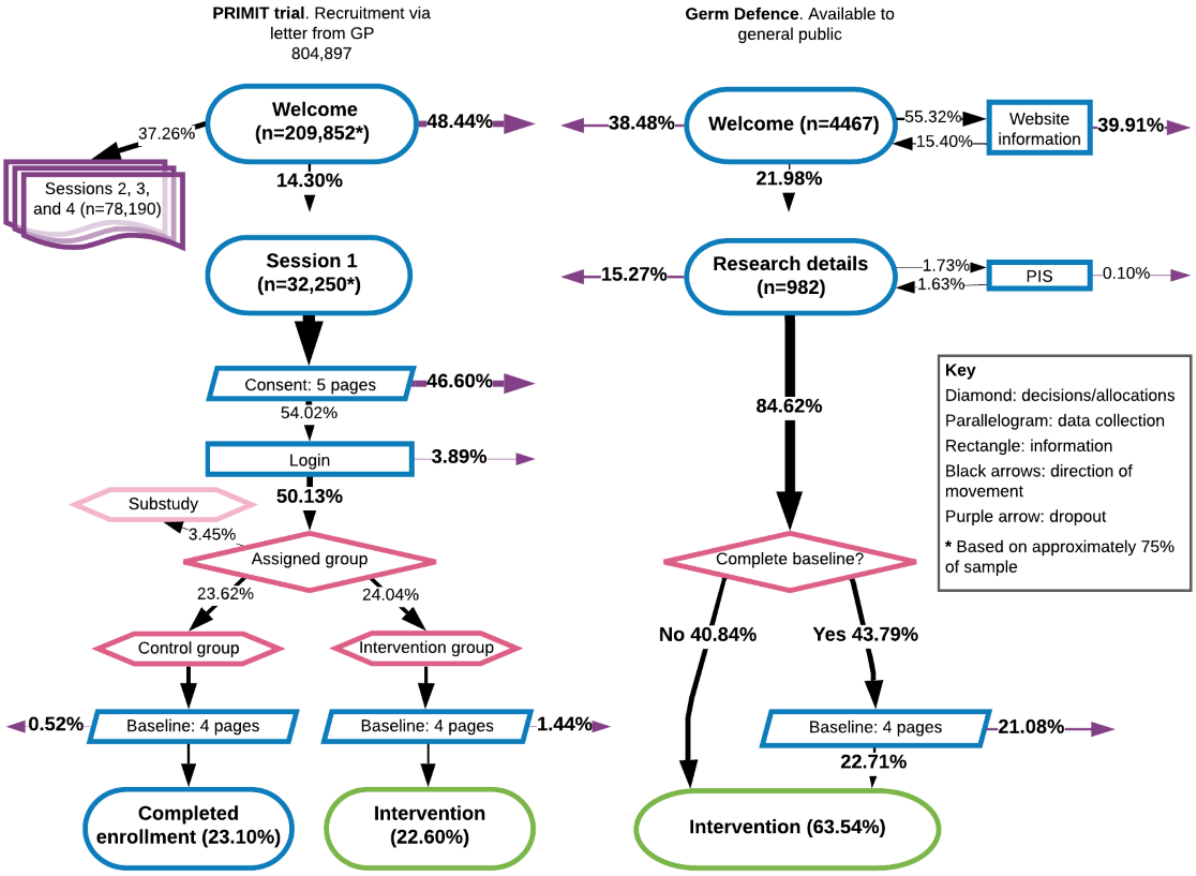
Confidence intervals at 95% were also used to compare the scores for intentions to handwash in the goal-setting situations. The first completion of intended handwashing responses for the goal-setting items was analyzed. Confidence intervals comparing individual handwashing situations are reported for practical significance, with mean differences >0.3 on a scale of 1-5. Paired sample cases items were excluded listwise to reduce any bias due to attrition, and equal variances were assumed for independent samples.

Results

Completion of Web-Based Research Enrollment Procedures

Letters inviting patients to take part in the PRIMIT study were sent to 804,897 individuals, of which 2.5% (20,042/804,897) provided web-based consent, completed baseline measures, and were assigned to a group (Figure 3). The Germ Defence website received 12,106 visits to the first page. The majority of these were identified as bots (7639/12,106, 63.1%). Of the remaining 4467 visits, 4.9% (223/4467) continued to provide baseline measures and access the intervention and 8.9% (401/4467) accessed the intervention directly.

Figure 3. Flowcharts of usage during PRIMIT (PRImary care trial of a website based infection control intervention to Modify Influenza-like illness and respiratory infection Transmission) study and Germ Defence research procedures. Owing to the difference in sample sizes, analyses of usage data are reported as percentages to enable more meaningful comparisons. GP: general practice; PIS: patient information sheet.



Intervention Usage

User Characteristics

The sample population in Germ Defence study was younger (mean 42 years) and predominantly female (184/243, 75.7%) compared to that in PRIMIT (mean 57 years, 5002/8943, 55.9%). The samples were similar in most of the self-reported behaviors and beliefs at baseline (Table 2). PRIMIT study users reported higher intended levels of daily handwashing at baseline (mean 3.97) compared to current daily levels (mean 3.85; mean difference 0.118, 95% CI 0.107-0.130), as did Germ Defence users (intention mean 4.11, current mean 3.87; mean difference 0.239, 95% CI 0.131-0.348). Comparisons of scores between PRIMIT study and Germ Defence users for current (mean difference –0.022, 95% CI –0.172 to 0.127) and intended (mean difference –0.150, 95% CI –0.296 to –0.004) daily handwashing showed no practical significance. Germ Defence users perceived themselves as more likely to contract a respiratory tract infection (mean 5.53) than users in the PRIMIT study intervention group (mean 5.11; mean difference 0.418, 95% CI 0.199-0.637). None of the user characteristics recorded at baseline (eg, age, daily handwashing behavior) predicted completion of the core section for either PRIMIT or Germ Defence (Table S1 in Multimedia Appendix 4).

When asked how they had heard about the Germ Defence website, 236 users responded the following sources: news (41/236, 17.4%), word of mouth (21/236, 8.9%), social media (16/236, 6.8%), search engine (15/236, 6.4%), school or children's group (14/236, 5.9%), and health support group (4/236, 1.7%); 125 (52.97%) users selected "other." When asked to provide further details, 67 (28.4%) users identified themselves as working in health care–related professions. Owing to health care professionals' (HCPs) familiarity with infection prevention behaviors, the characteristics of the identified HCPs were compared with those of non-HCPs within the Germ Defence study to examine any potential sample bias. When the sample was split by profession, HCPs indicated higher daily levels of current handwashing (mean 4.26) than non-HCPs (mean 3.71; mean difference 0.543, 95% CI 0.237-0.849). HCPs also selected higher daily intentions to wash their hands in the future (mean 4.51) compared to non-HCPs (mean 3.96; mean difference 0.550, 95% CI 0.243-0.856). Although HCPs perceived themselves as more likely to contract a respiratory tract infection (mean 6.00) than non-HCPs (mean 5.33; mean difference 0.667, 95% CI 0.213-1.121), the perceived likelihood of becoming very ill was similar across all Germ Defence baseline responders (HCPs mean 3.25; non-HCPs mean 3.62; mean difference 0.038, 95% CI –0.472 to 0.549).

Table 2. Characteristics of PRIMIT study intervention group and Germ Defence study baseline responders.

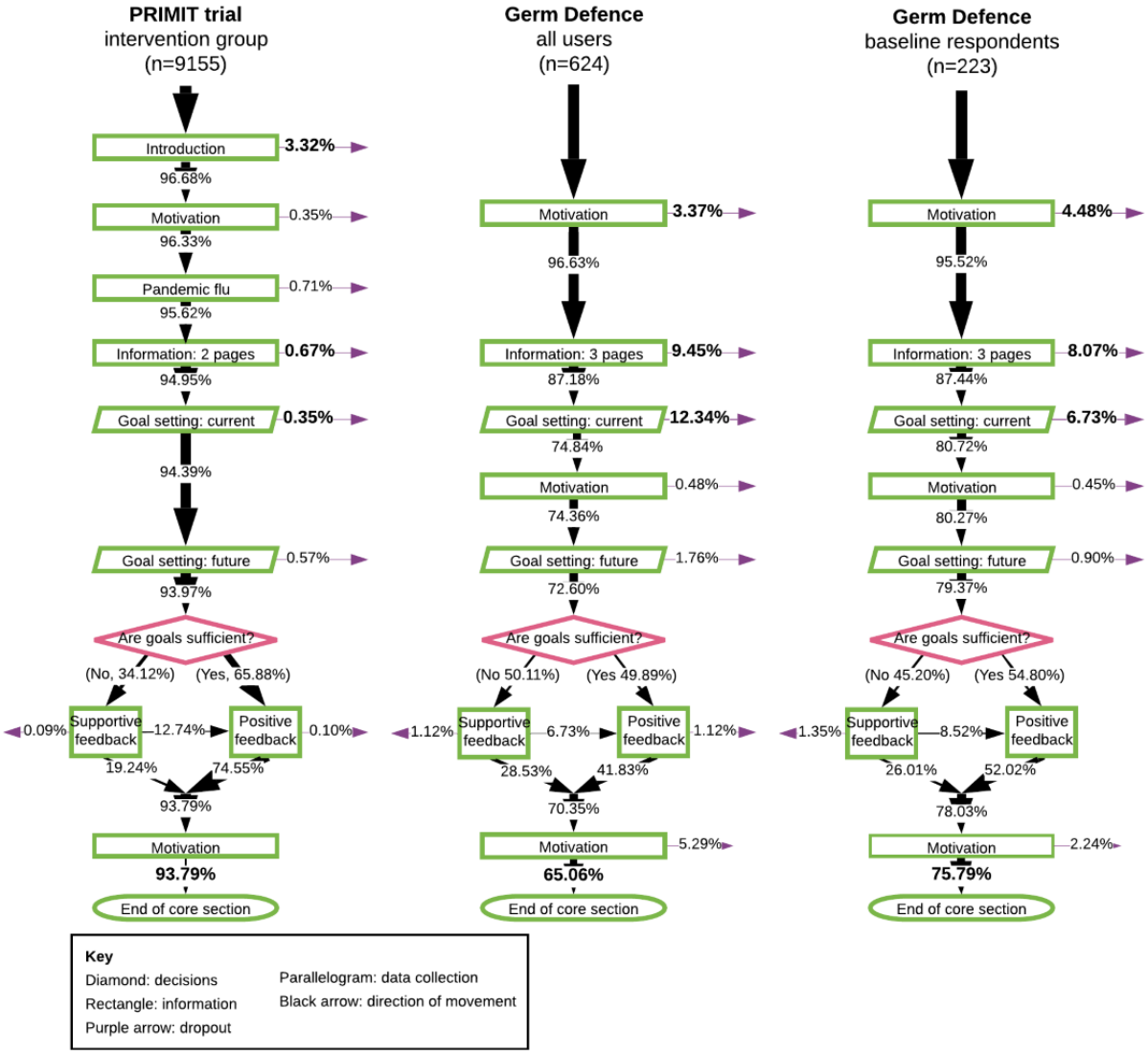
Measure	Participants (n)	Mean (SD)	Min-Max
Responders			
PRIMIT[a] study	8959	N/A[b]	N/A
Germ Defence	250	N/A	N/A
Gender split (female)			
PRIMIT study (n=8943)	5002	N/A	N/A
Germ Defence (n=243)	184	N/A	N/A
Age (years)			
PRIMIT study	8945	56.64 (13.631)	18-94
Germ Defence	250	42.11 (13.035)	16-74
Current daily handwashing			
PRIMIT study	8945	3.85 (1.150)	1-5
Germ Defence	234	3.87 (1.094)	1-5
Intended daily handwashing			
PRIMIT study	8944	3.97 (1.117)	1-5
Germ Defence	231	4.11 (1.090)	1-5
Perceived likelihood of user becoming ill			
PRIMIT study	8837	5.11 (1.652)	1-7
Germ Defence	223	5.52 (1.582)	1-7
Perceived severity for user			
PRIMIT study	N/A	N/A	N/A
Germ Defence	223	3.61 (1.747)	1-7

[a]PRIMIT: PRImary care trial of a website based infection control intervention to Modify Influenza-like illness and respiratory infection Transmission.
[b]N/A: not applicable.

Intervention Content Usage

Attrition was lower in PRIMIT, with fewer users failing to complete the core section (568/9155, 6.2%) compared to Germ Defence users (218/624, 34.9%) (see Figure 4). The pages that saw the most attrition in Germ Defence were those leading up to and including the first page of the goal-setting section, which saw a total of 25.2% (157/624) of users leave the intervention, accounting for 72.0% (157/218) of all attrition. Germ Defence users who had volunteered to complete baseline measures had lower levels of attrition across all of the core section (54/223, 24.2%) and up to the start of the goal-setting section (43/223, 19.3%). The introduction page to the PRIMIT study saw the highest level of attrition (304/9155, 3.3%), meaning that 53.5% (304/568) of the total attrition occurred on the first page. Completion of the core section of Germ Defence by users who had identified themselves as HCPs was 79.1% (53/67), which was similar to that of all baseline responders (169/223, 75.8%). After completing the core section, 38.8% (242/624) of all Germ Defence users accessed one or more of the 3 further components (see Figure 1) compared to 47.5% (106/223) of the Germ Defence baseline respondents.

Figure 4. Flowcharts of percentage of usage for the core sections of PRIMIT (PRImary care trial of a website based infection control intervention to Modify Influenza-like illness and respiratory infection Transmission) and Germ Defence. Owing to the difference in sample sizes, analyses of usage data are reported as percentages to enable more meaningful comparisons.

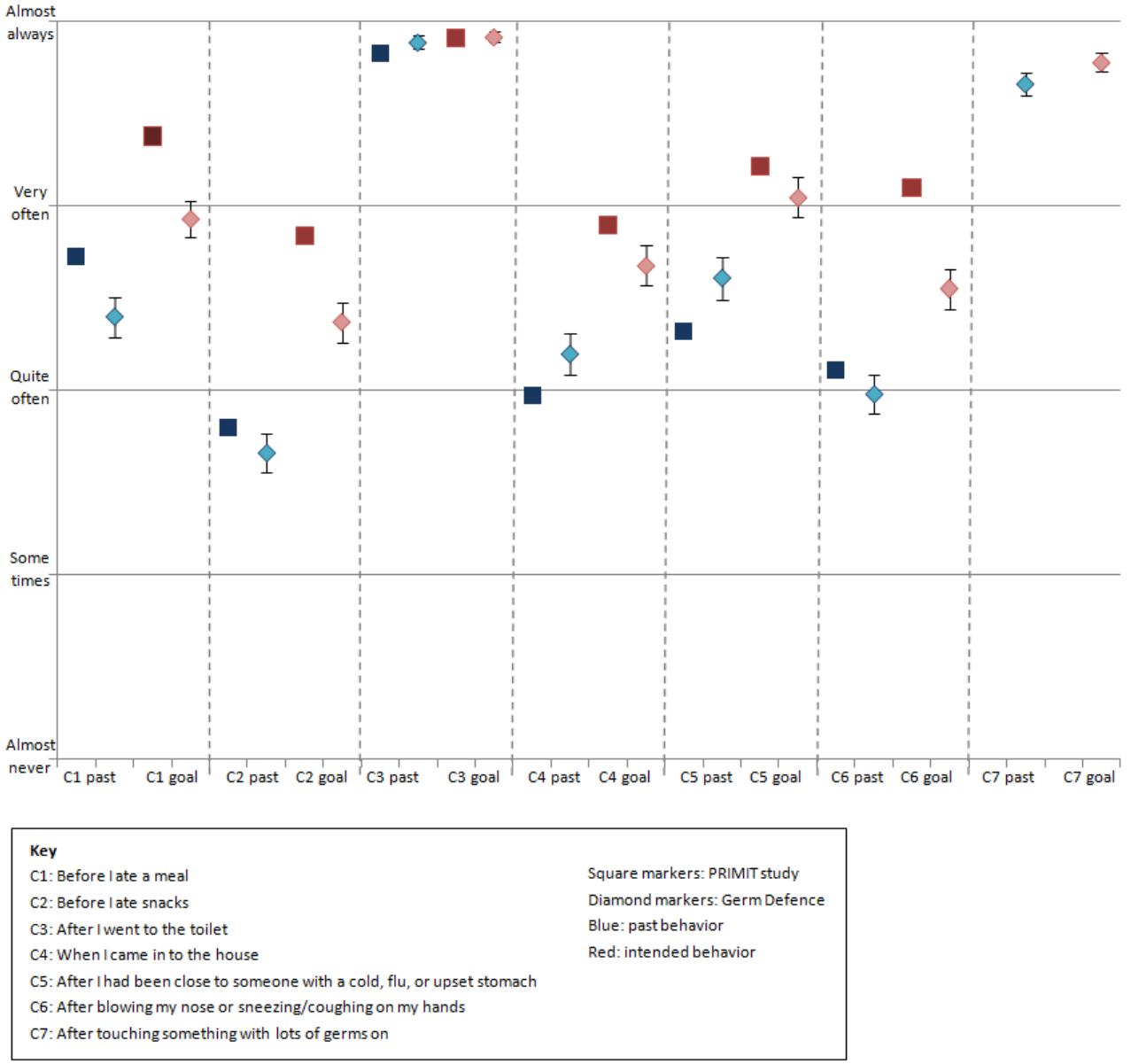


Handwashing Behaviors

Germ Defence baseline responders and PRIMIT study users both chose to increase their intended handwashing compared to their current levels for 5 of the 7 situations (see Figure 5 and Tables S2-S4 in Multimedia Appendix 5 for mean differences and confidence intervals). Intentions to wash hands after going to the toilet did not increase in either context (Table S2 in Multimedia Appendix 5). Germ Defence users also selected similar current and intended levels for "after touching something with germs on." Scores for HCPs using Germ Defence were not different from those of non-HCPs for both current and intended behavior (Table S4 in Multimedia Appendix 5). Owing to the large sample size, mean ranges for the PRIMIT study are smaller than those for the Germ Defence sample.

When comparing current behavior between the 2 studies, the score for "before I ate a meal" was the only practically significant situation (mean difference 0.320, 95% CI 0.121-0.518), with PRIMIT study users washing their hands more frequently (mean 3.73) than Germ Defence users (mean 3.41) (Table S3 in Multimedia Appendix 5). Scores for intended behavior showed that PRIMIT study users had selected higher frequencies (mean 4.10) for "after blowing my nose, sneezing/coughing on my hands" (mean difference 0.381, 95% CI 0.206-0.557) compared to Germ Defence users (mean 3.72).

Figure 5. Means and confidence intervals for goal-setting behavior change technique for PRIMIT (PRImary care trial of a website based infection control intervention to Modify Influenza-like illness and respiratory infection Transmission) study and Germ Defence.



Discussion

Overview of This Study

This paper presents process evaluations of an intervention applied in 2 different contexts to lower the transmission of viruses at home by increasing handwashing. The intervention was used in an RCT by National Health Service patients recruited via their general practices (PRIMIT study) and as an open access intervention available to the general public (Germ Defence). By comparing the 2 contexts, we hope to better understand any effect that the design of the RCT had on who accessed the intervention, how the intervention was used, and whether the intervention continued to increase handwashing intentions once disseminated to the public.

Implications for Web-Based Research Enrollment Procedures

Half the PRIMIT RCT participants dropped out during the web-based research enrollment process prior to reaching the intervention. This intervention had been developed in line with the person-based approach; accessibility and engagement with content and design were optimized through think-aloud interviews and survey responses [3]. However, although baseline measures were assessed quantitatively to ensure they were valid [17], additional patient and public involvement or qualitative assessment during development would have provided information on the impact and acceptability of these research processes. This would have allowed the opportunity to improve them and reduce the risk of losing a substantial number of potential users before they have even reached the intervention. The Germ Defence study lost a third of the participants during research enrollment. Considerable effort was invested to develop and optimize the consent and baseline measures for Germ

Defence. The process was made easier and quicker than that for PRIMIT with optional rather than compulsory baseline measures. These pages were constructed in line with the design values and language used for the original intervention [3,17] and were additionally tested for usability by 8 researchers experienced in person-based digital intervention development. As a preliminary dissemination study, these procedures were necessary to compare participant characteristics and handwashing behaviors of the Germ Defence study to those from the PRIMIT study. As with the PRIMIT RCT, the potential impact of this on uptake had not been assessed in advance. However, by making the process voluntary, we were able to demonstrate that over 60% of Germ Defence users preferred not to complete baseline measures at all. Almost 80% of the first visits to the Germ Defence website did not progress past the first page. The data available do not offer explanation for the high initial dropout. One possible explanation could be that the intervention was first developed in 2008 and intended to be used on personal computers and larger screen tablets; the website was unable to responsively adapt to mobile viewing (eg, users needed to enlarge or scroll across the page if viewing content on a mobile phone screen). This may have discouraged a considerable number of users. In addition, the number of visits at this point may include multiple visits by the same people; consequently, users who visited the first page and then decided to return at a later point to view the whole intervention cannot be identified within the attrition rate.

Implications for Intervention Usage

High levels of attrition during research enrollment procedures are a concern as they may lead to selection bias in the remaining sample. Comparison of intervention usage between the 2 contexts highlighted this effect: over 90% of PRIMIT study users completed the core section of the intervention compared to 65% of all Germ Defence users. This suggests that effortful enrollment research procedures and high levels of dropout may have resulted in a remaining sample that was more motivated to use the intervention. This is supported within Germ Defence where the completion rate increased to 75% for users who had participated in baseline measures, and they were more likely to view the additional components within the intervention. As none of the user characteristics recorded at baseline (eg, handwashing behavior, perceived likelihood of becoming ill) were associated with completing the core section of the intervention, explanations for this increased usage may be explained by alternative characteristics (eg, motivation to learn about infection prevention behaviors, altruism). As suggested above, extending qualitative methods used to develop intervention content to include research procedures may also help identify characteristics associated with engagement with the research as well as with target behavior.

The majority of attrition from the intervention occurred across the first few pages in both studies. Germ Defence users may have chosen to leave as they were dissatisfied with the content, or alternatively, given that these pages received positive feedback during development, users may have felt they had found sufficient information and did not need to continue. Further research is required to understand this attrition, particularly for the introduction page of the PRIMIT study. This page contained no “active ingredients,” yet saw the highest proportion of users leave despite having completed the effortful enrolment process. The page with the highest proportion of attrition in Germ Defence was the first page of the goal-setting section. This is the first time users were required to interact with the intervention, and they were unable to progress to the next page without entering current handwashing frequencies. Although users may have decided to leave rather than complete an activity they considered to be too effortful, the goal-setting section was seen to be effective at raising intentions to wash hands more in the future.

Implications for Promoting Increased Handwashing

When comparing PRIMIT study and Germ Defence handwashing selections for the goal-setting BCT, users in both contexts intended to improve their frequency of handwashing in the future in 5 situations (ie, before meals, before snacks, after coming in to the home, after being close to someone who is ill, after sneezing or coughing). Minimal improvement was seen for washing hands after going to the toilet and touching something with germs on. This is probably due to the high levels already reported for current behavior. Interestingly, users reported only washing their hands “quite often” on average after blowing their nose, sneezing, or coughing, despite repeated public health campaigns targeting this specific behavior in the United Kingdom [38]. Qualitative research during intervention development highlighted that people found this situation difficult to carry out due to lack of control [2,3]. However, users did chose to improve on this behavior in the future having viewed the intervention.

The goal-setting section helped Germ Defence users to plan improvements in handwashing in the same way as had been seen in the PRIMIT study RCT for 5 of the 6 shared situations. This suggests that the intervention mechanisms work as effectively when disseminated to the general public as they did in the RCT and that the reduction of illnesses seen in PRIMIT study sample is likely to be replicated in the Germ Defence sample. These similar levels of intended handwashing occurred despite the RCT experiencing greater levels of selection bias as discussed above, leading to a more motivated sample. An explanation for this may be found in the Germ Defence sample where almost a quarter of baseline responders were HCPs. At baseline, HCPs indicated higher amounts of both current and intended daily handwashing than non-HCPS. Yet, when completing the goal-setting section, HCPs and non-HCPs showed similar levels of current and increased intended future handwashing for the specific situations within the BCT. This suggests that although HCPs wash their hands more frequently, they might be overlooking handwashing opportunities in the home that are important for reducing infection transmission. If this effect was seen with HCPs using Germ Defence, then it is likely that using these specific situations was also effective for highly motivated users in the PRIMIT study.

Application of the Findings to Disseminate Germ Defence for Use in a Pandemic

In March 2020, Germ Defence was adapted and disseminated globally for use in the COVID-19 pandemic [31,32]. The findings from this preliminary dissemination study were applied

to the intervention to maximize uptake and reach. Having established that even minimal research enrolment procedures may act as a barrier to usage, all baseline measures were removed for the Germ Defence COVID-19 intervention so that users could access the intervention as quickly and easily as possible [31]. Instead, after completing the core section, users are invited to take part in a survey where some demographic information is collected along with questions about using the intervention. While the sample that reaches this stage of the intervention will be biased, the survey has still collected valuable and insightful information [39]. In addition, the goal-setting section embedded within the intervention provides a measure of past/intended behavior. To reduce attrition within the goal-setting section, an additional message explaining the value of completing the measures has been added, which is hoped will encourage participation.

Limitations

Comparing the large sample in the PRIMIT study to the much smaller number of users who accessed Germ Defence is problematic because they are likely to differ in many respects, which could not be evaluated. However, comparisons are often made between implementation of interventions in different contexts, and given the difference in contexts and probably also populations, it is notable that beliefs at baseline and self-reported behaviors were similar in the trial and the implementation. The difference in the number of participants is likely to reflect recruitment methods: the PRIMIT study sent letters directly to over 800,000 people, whereas Germ Defence relied heavily on health organizations to distribute the details. The low level of Germ Defence users suggests that prevention of seasonal colds and flu may not be a strong motivation for many people (Germ Defence for COVID-19 has seen considerably higher usage) [30]. This means that people who did use Germ Defence may have been more highly motivated to practice good infection prevention behaviors, and given the low proportion of uptake for PRIMIT, the same may be said of that sample. Yet, despite this, the intervention group for the PRIMIT study experienced lesser respiratory tract infections, and selections for the goal-setting section demonstrated the ability of the intervention to help users in both studies to increase their handwashing.

The large proportion of dropouts seen on the first page of Germ Defence is unexplained. This is a common phenomenon of intervention usage in the community [40], and as such, including qualitative data collection in future process evaluations would provide some insight for this problem and inform further refinement of the first page to maximize usage.

Conclusions

This study provides an example of how interventions assessed through RCTs can be examined and adapted to try to optimize their usage when disseminated to the general public. By comparing RCT data to a novel preliminary dissemination study, we were able to examine our 3 aims. First, this research demonstrates that the Germ Defence intervention continues to raise users' intentions to wash their hands more in specific situations where the risk of virus transmission is high, replicating the behavior changes reported during the PRIMIT RCT. The responses show that using a goal-setting BCT was helpful for increasing intended handwashing, including in situations identified as being particularly difficult, despite having been already targeted by long-term public health messages. Second, we established that the effortful web-based enrolment procedures required for the PRIMIT RCT led to participant dropout, acting as a barrier to accessing health information. Third, the effect of this was apparent through our aim of comparing intervention usage, as the remaining PRIMIT users showed higher levels of usage than those accessing the intervention in the Germ Defence study. By designing the Germ Defence study to observe and compare the preliminary dissemination of the intervention, we were able to establish points of increased attrition within the intervention. The findings provided the opportunity to adapt and improve Germ Defence for wider dissemination and during the onset of COVID-19 for a public health emergency.

Acknowledgments

We would like to thank the Economic and Social Research Council for funding this paper (award 1692136, Understanding digital intervention engagement: Making sense of large-scale usage data). The PRIMIT study was funded by the Medical Research Council. The intervention used for the PRIMIT and Germ Defence studies was developed using LifeGuide, which was partly funded by the National Institute for Health Research (NIHR) Biomedical Research Centre, Southampton. LY is an NIHR Senior Investigator and is affiliated to the NIHR Applied Research Collaboration West and the NIHR Health Protection Research Unit in Behavioural Science and Evaluation of Interventions at the University of Bristol in partnership with Public Health England.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Stage 1 checklist of AMUsED (Analyzing and Measuring Usage and Engagement Data) framework: Familiarization with the data. Completed for the PRIMIT (PRImary care trial of a website based infection control intervention to Modify Influenza-like illness and respiratory infection Transmission) study and Germ Defence.

[DOCX File , 24 KB - jmir_v23i10e26104_app1.docx]

Multimedia Appendix 2

Stage 2 checklist: selecting usage variables and generating research questions.
[DOCX File , 20 KB - jmir_v23i10e26104_app2.docx]

Multimedia Appendix 3
Stage 3 checklist: preparation for analysis.
[DOCX File , 15 KB - jmir_v23i10e26104_app3.docx]

Multimedia Appendix 4
Results of the logistic regression analysis of baseline user characteristics predicting the completion of core sections of PRIMIT (PRImary care trial of a website based infection control intervention to Modify Influenza-like illness and respiratory infection Transmission) and Germ Defence intervention.
[DOCX File , 14 KB - jmir_v23i10e26104_app4.docx]

Multimedia Appendix 5
Mean differences and 95% confidence intervals for goal-setting section on handwashing frequencies.
[DOCX File , 18 KB - jmir_v23i10e26104_app5.docx]

References

1. What is a pandemic? World Health Organization (2010). URL: https://www.who.int/csr/disease/swineflu/frequently_asked_questions/pandemic/en/ [accessed 2020-05-06]
2. Morrison LG, Yardley L. What infection control measures will people carry out to reduce transmission of pandemic influenza? A focus group study. BMC Public Health 2009 Jul 23;9:258 [FREE Full text] [doi: 10.1186/1471-2458-9-258] [Medline: 19627568]
3. Yardley L, Miller S, Teasdale E, Little P, Primit Team. Using mixed methods to design a web-based behavioural intervention to reduce transmission of colds and flu. J Health Psychol 2011 Mar;16(2):353-364. [doi: 10.1177/1359105310377538] [Medline: 20929941]
4. About handwashing. Global Handwashing Partnership. URL: https://globalhandwashing.org/about-handwashing/ [accessed 2020-09-29]
5. Hand hygiene: why, how, and when? World Health Organization (2009). 2009. URL: https://www.who.int/gpsc/5may/Hand_Hygiene_Why_How_and_When_Brochure.pdf [accessed 2020-09-29]
6. Public information campaign focuses on handwashing. Department of Health and Social Care, UK Government. URL: https://www.gov.uk/government/news/public-information-campaign-focuses-on-handwashing [accessed 2020-09-14]
7. Handwashing: clean hands save lives. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/handwashing/index.html [accessed 2020-09-14]
8. Lau JTF, Kim JH, Tsui HY, Griffiths S. Anticipated and current preventive behaviors in response to an anticipated human-to-human H5N1 epidemic in the Hong Kong Chinese general population. BMC Infect Dis 2007 Mar 15;7:18 [FREE Full text] [doi: 10.1186/1471-2334-7-18] [Medline: 17359545]
9. Rubin GJ, Amlôt R, Page L, Wessely S. Public perceptions, anxiety, and behaviour change in relation to the swine flu outbreak: cross sectional telephone survey. BMJ 2009 Jul 02;339:b2651 [FREE Full text] [doi: 10.1136/bmj.b2651] [Medline: 19574308]
10. Wu JT, Riley S, Fraser C, Leung GM. Reducing the impact of the next influenza pandemic using household-based public health interventions. PLoS Med 2006 Sep;3(9):e361 [FREE Full text] [doi: 10.1371/journal.pmed.0030361] [Medline: 16881729]
11. West R, Michie S, Rubin GJ, Amlôt R. Applying principles of behaviour change to reduce SARS-CoV-2 transmission. Nat Hum Behav 2020 May;4(5):451-459. [doi: 10.1038/s41562-020-0887-9] [Medline: 32377018]
12. Gold N, Hu X, Denford S, Xia R, Towler L, Groot J, et al. Effectiveness of digital interventions to improve household and community infection prevention and control behaviours and to reduce incidence of respiratory and/or gastro-intestinal infections: a rapid systematic review. BMC Public Health 2021 Jun 21;21(1):1180 [FREE Full text] [doi: 10.1186/s12889-021-11150-8] [Medline: 34154553]
13. Michie S, West R. Behavioural, environmental, social, and systems interventions against covid-19. BMJ 2020 Jul 28;370:m2982. [doi: 10.1136/bmj.m2982] [Medline: 32723732]
14. Little P, Stuart B, Hobbs FDR, Moore M, Barnett J, Popoola D, et al. An internet-delivered handwashing intervention to modify influenza-like illness and respiratory infection transmission (PRIMIT): a primary care randomised trial. Lancet 2015 Oct 24;386(10004):1631-1639. [doi: 10.1016/S0140-6736(15)60127-1] [Medline: 26256072]
15. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]

16. Yardley L, Miller S, Schlotz W, Little P. Evaluation of a Web-based intervention to promote hand hygiene: exploratory randomized controlled trial. J Med Internet Res 2011 Dec 09;13(4):e107 [FREE Full text] [doi: 10.2196/jmir.1963] [Medline: 22155673]
17. Miller S, Yardley L, Little P, PRIMIT team. Development of an intervention to reduce transmission of respiratory infections and pandemic flu: measuring and predicting hand-washing intentions. Psychol Health Med 2012;17(1):59-81. [doi: 10.1080/13548506.2011.564188] [Medline: 21644184]
18. Ainsworth B, Steele M, Stuart B, Joseph J, Miller S, Morrison L, et al. Using an Analysis of Behavior Change to Inform Effective Digital Intervention Design: How Did the PRIMIT Website Change Hand Hygiene Behavior Across 8993 Users? Ann Behav Med 2017 Jun;51(3):423-431 [FREE Full text] [doi: 10.1007/s12160-016-9866-9] [Medline: 27909944]
19. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
20. Miller S, Ainsworth B, Yardley L, Milton A, Weal M, Smith P, et al. A Framework for Analyzing and Measuring Usage and Engagement Data (AMUsED) in Digital Interventions: Viewpoint. J Med Internet Res 2019 Feb 15;21(2):e10966 [FREE Full text] [doi: 10.2196/10966] [Medline: 30767905]
21. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the Trial: Systematic Review of Real-World Uptake and Engagement With Digital Self-Help Interventions for Depression, Low Mood, or Anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
22. Clay RA. More than one way to measure. American Psychological Association. URL: https://www.apa.org/monitor/2010/09/trials [accessed 2020-05-18]
23. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. Soc Sci Med 2018 Aug;210:2-21 [FREE Full text] [doi: 10.1016/j.socscimed.2017.12.005] [Medline: 29331519]
24. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
25. Van der Mispel C, Poppe L, Crombez G, Verloigne M, De Bourdeaudhuij I. A Self-Regulation-Based eHealth Intervention to Promote a Healthy Lifestyle: Investigating User and Website Characteristics Related to Attrition. J Med Internet Res 2017 Jul 11;19(7):e241 [FREE Full text] [doi: 10.2196/jmir.7277] [Medline: 28698168]
26. Tripepi G, Jager KJ, Dekker FW, Zoccali C. Selection bias and information bias in clinical research. Nephron Clin Pract 2010;115(2):c94-c99 [FREE Full text] [doi: 10.1159/000312871] [Medline: 20407272]
27. Reinwand DA, Crutzen R, Elfeddali I, Schneider F, Schulz DN, Smit ES, et al. Impact of Educational Level on Study Attrition and Evaluation of Web-Based Computer-Tailored Interventions: Results From Seven Randomized Controlled Trials. J Med Internet Res 2015 Oct 07;17(10):e228 [FREE Full text] [doi: 10.2196/jmir.4941] [Medline: 26446779]
28. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
29. Murray E, Treweek S, Pope C, MacFarlane A, Ballini L, Dowrick C, et al. Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. BMC Med 2010 Oct 20;8:63 [FREE Full text] [doi: 10.1186/1741-7015-8-63] [Medline: 20961442]
30. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ 2015 Mar 19;350:h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
31. Ainsworth B, Miller S, Denison-Day J, Stuart B, Groot J, Rice C, et al. Infection Control Behavior at Home During the COVID-19 Pandemic: Observational Study of a Web-Based Behavioral Intervention (Germ Defence). J Med Internet Res 2021 Feb 25;23(2):e22197 [FREE Full text] [doi: 10.2196/22197] [Medline: 33566791]
32. Germ Defence. URL: https://www.germdefence.org/index.html [accessed 2021-09-27]
33. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
34. Rippetoe PA, Rogers RW. Effects of components of protection-motivation theory on adaptive and maladaptive coping with a health threat. J Pers Soc Psychol 1987 Mar;52(3):596-604. [doi: 10.1037//0022-3514.52.3.596] [Medline: 3572727]
35. Tariman JD, Berry DL, Halpenny B, Wolpin S, Schepp K. Validation and testing of the Acceptability E-scale for web-based patient-reported outcomes in cancer care. Appl Nurs Res 2011 Feb;24(1):53-58 [FREE Full text] [doi: 10.1016/j.apnr.2009.04.003] [Medline: 20974066]
36. Williams S, Yardley L, Weal M, Willis G. Introduction to lifeguide: Open-source software for creating online interventions for health care. Med-e-Tel 2010. URL: https://tinyurl.com/z4bauh8a [accessed 2021-09-27]
37. Germ Defence Archive. URL: http://archive.germdefence.org/ [accessed 2021-09-27]
38. Catch it, bin it, kill it: respiratory and hand hygiene campaign (2011). Department of Health and Social Care UK Government. URL: https://www.gov.uk/government/publications/catch-it-bin-it-kill-it-respiratory-and-hand-hygiene-campaign-2011 [accessed 2020-09-29]

39. Morton K, Ainsworth B, Miller S, Rice C, Bostock J, Denison-Day J, et al. Adapting Behavioral Interventions for a Changing Public Health Context: A Worked Example of Implementing a Digital Intervention During a Global Pandemic Using Rapid Optimisation Methods. Front Public Health 2021;9:668197 [FREE Full text] [doi: 10.3389/fpubh.2021.668197] [Medline: 33981669]
40. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the Uptake of and Engagement With Health and Well-Being Smartphone Apps: Systematic Review. J Med Internet Res 2020 May 29;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]

Abbreviations

AMUsED: Analyzing and Measuring Usage and Engagement Data
BCT: behavior change technique
HCP: health care professional
NIHR: National Institute for Health Research
PRIMIT: PRImary care trial of a website based infection control intervention to Modify Influenza-like illness and respiratory infection Transmission
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 27.11.20; peer-reviewed by B White, S Wang; comments to author 01.02.21; revised version received 17.05.21; accepted 30.05.21; published 05.10.21.

Please cite as:
Miller S, Ainsworth B, Weal M, Smith P, Little P, Yardley L, Morrison L
A Web-Based Intervention (Germ Defence) to Increase Handwashing During a Pandemic: Process Evaluations of a Randomized Controlled Trial and Public Dissemination
J Med Internet Res 2021;23(10):e26104
URL: https://www.jmir.org/2021/10/e26104
doi:10.2196/26104
PMID:34519661

Tutorial

Understanding Uptake of Digital Health Products: Methodology Tutorial for a Discrete Choice Experiment Using the Bayesian Efficient Design

Dorothy Szinay[1], MSc; Rory Cameron[2,3], PhD; Felix Naughton[1], PhD; Jennifer A Whitty[2,3], PhD; Jamie Brown[4,5], PhD; Andy Jones[2], PhD

[1]Behavioural and Implementation Science Group, School of Health Sciences, University of East Anglia, Norwich, United Kingdom
[2]Norwich Medical School, University of East Anglia, Norwich, United Kingdom
[3]National Institute for Health Research, Applied Research Collaboration East of England, Cambridge, United Kingdom
[4]Department of Behavioural Science and Health, University College London, London, United Kingdom
[5]SPECTRUM Consortium, London, United Kingdom

Corresponding Author:
Dorothy Szinay, MSc
Behavioural and Implementation Science Group
School of Health Sciences
University of East Anglia
Norwich Research Park Earlham Road
Norwich, NR4 7TJ
United Kingdom
Phone: 44 1603593064
Email: d.szinay@uea.ac.uk

Abstract

Understanding the preferences of potential users of digital health products is beneficial for digital health policy and planning. Stated preference methods could help elicit individuals' preferences in the absence of observational data. A discrete choice experiment (DCE) is a commonly used stated preference method—a quantitative methodology that argues that individuals make trade-offs when engaging in a decision by choosing an alternative of a product or a service that offers the greatest utility, or benefit. This methodology is widely used in health economics in situations in which revealed preferences are difficult to collect but is much less used in the field of digital health. This paper outlines the stages involved in developing a DCE. As a case study, it uses the application of a DCE to reveal preferences in targeting the uptake of smoking cessation apps. It describes the establishment of attributes, the construction of choice tasks of 2 or more alternatives, and the development of the experimental design. This tutorial offers a guide for researchers with no prior knowledge of this research technique.

(J Med Internet Res 2021;23(10):e32365) doi:10.2196/32365

KEYWORDS

discrete choice experiment; stated preference methods; mHealth; digital health; quantitative methodology; uptake; engagement; methodology; preference; Bayesian; design; tutorial; qualitative; user preference

Introduction

Understanding how the public values different aspects of digital health tools, such as smoking cessation or physical activity apps, can help providers of the tools to identify functionality that is important to users, which may improve uptake (ie, selection, download, and installation of apps) [1]. This is important because uptake of digital tools is generally low. More information regarding the preferences of users when selecting a digital health tool, for example via an app store, may allow providers to present their products in such a way that may increase their uptake. However, pragmatic challenges, such as examining how each potentially modifiable aspect of a digital health product (eg, presentation, design, and features that it offers) or intervention design will impact preference or the choice of uptake, often mean this is not feasible or practical [2]. Therefore, increasing attention is being paid toward stated preference methods to understand preferences when designing digital health products and services, with examples including

COVID-tracing apps [3,4], sun protection apps to prevent skin cancer [5], and the uptake of health apps in general [6].

Stated preference methods are survey-based methods aiming to elicit individuals' preferences toward a specific behavior, particularly those that are not well understood. The most widely used type of stated preference method is the discrete choice experiment (DCE) [7]. According to Spinks et al [8], Louviere and Hensher (1982) and Louviere and Woodworth (1983) originally developed DCEs to study the marketing and economics of transport, and the fields of psychology and economics have profoundly influenced the DCE methodology since it was developed. In recent years, DCEs have been increasingly used in health and health care settings [9,10], as well as in addiction research [11] and digital health [4-6]. The increasing number of DCEs in digital health highlights their potential, although they are currently underused.

Discrete choice differentiates from other stated preference methods in the way that responses are elicited [12]. The DCE uses a survey-based experimental design, where participants are presented with a series of hypothetical scenarios. In these scenarios, participants are shown situations, known as *choice tasks*. Attempting to mimic real-world decision making, in each choice task, participants then have to choose a product or a service from two or more options, known as *alternatives* [13]. Each alternative consists of a set of characteristics, known as *attributes*, with at least two types, known as *attribute levels* [13]. Participants are asked to choose a preferred alternative in each choice task, which allows researchers to quantify the relative strength of preferences for improvements in certain attributes [8,14].

The outputs from statistical models developed using DCE data can be beneficial for estimating uptake of new products or services, including digital health tools, where observational data are not available or are difficult to obtain otherwise [15,16]. Lack of observational data often implies a requirement to seek scientific views and comments from experts in order to generate predictions of a target behavior [17]. However, DCEs can provide an empirical alternative to expert opinions, while accounting for possible interactions between attributes (eg, design of a product and brand name), which are otherwise often ignored [18].

In our research, we wanted to understand how to present health apps on curated health app portals to increase their uptake. This paper describes the development of a DCE in digital health that aims to elicit potential user preferences on smoking cessation app uptake. It explains how the attributes and their levels are selected and describes the construction of choice tasks and the experimental design. The study protocol of the research this paper is based on is registered on the Open Science Framework [19].

Development of a DCE

The development of a DCE should follow published recommendations, including the checklist for good research practices [9], guides on the development of a DCE [13,20], recommendations on how to construct the experimental design [7,20-23], and which statistical methods can be used [24].

Establishing Attributes

An important step in designing a DCE is the identification of the relevant attributes for the subject matter. Attributes in a DCE can be quantitative, such as cost, or qualitative, such as the design of a product [25]. The identification of attributes is typically based on primary and secondary data collection to ensure that the DCE is tailored to the study setting [13]. It should ideally commence with a literature review that will inform qualitative research to identify relevant attributes [26]. Although there is no set limit on the number of attributes that can be included in a DCE, to ensure that the cognitive load of the participants is manageable, it should be less than 10 [13], with a general expectation to include 5-7 attributes [27].

Our DCE was based on a comprehensive systematic review investigating factors influencing the uptake and engagement with health and well-being smartphone apps [28] and a qualitative research component that consisted of a think-aloud and interview study to examine further the previously identified factors or attributes [29]. The importance of qualitative research lies in ensuring inclusion of attributes that are relevant to most participants [25]. Of the 14 factors initially identified as being relevant for the uptake of health and well-being apps, 5 were retained and included in the DCE: the monthly price of the app, who developed the app, the star ratings of the app, the description of the app, and images shown. These factors were chosen due to their perceived importance during our previous qualitative research and for pragmatic reasons, including how easily measurable and presentable they were within the DCE.

An important step in designing a DCE is in ensuring the content validity of the instrument: the identification of relevant attributes for the subject matter. Following administration of the survey, methods are available for the measurement and assessment of the content validity of the instrument, although their use is not widely reported [30].

Establishing Attribute Levels

The next step is to establish attribute levels. The level of an attribute must also be of a range that ensures a trade-off between attributes. A trade-off is defined as an exchange in which a participant gives up some amount of one attribute to gain more of another. It has been suggested that increasing the number of levels for an attribute increases the relative importance of that attribute [31] and that imbalance in the numbers of levels across attributes raises the importance of the attributes with higher levels [32]. Yang et al [32] suggested that a balance exists between simpler designs with lower numbers of levels, which reduce the respondent burden (and consequently measurement error) and are useful for identifying attribute rankings, and more complex designs with higher levels (and higher statistical precision) and is more sensitive to identifying trade-offs between attributes. Based on this, and the commonly adopted practices in the research field, we aimed to include at least three levels for each attribute.

If a range is not suitable, participants might consider the differences between levels unimportant [25]. For example, the

difference between the star ratings of 4.8 and 4.7 for a smoking cessation app is not as relevant as the difference between 4.8 and 4. In our research, to refine attribute levels, a survey was conducted with 34 participants. In the survey, the levels of two attributes we were unsure of (the monthly price of the app and the ratings) were carefully considered in order to specify at a sufficiently wide range so that the difference between the levels would likely make a difference in response. When a range is not wide enough, there is a risk that participants could ignore the attributes because they judge the difference between levels to be insignificant [20]. See Figure 1 for the final list of attributes and levels included in our DCE.

Figure 1. Attributes and attribute levels in our DCE. DCE: discrete choice experiment.

Attributes	Attribute levels
1. The monthly price of the app	• £0 • £2.99 • £5.99 • £8.99
2. Who developed the app	• Doesn't say • Mhealth Essentials Ltd. • NHS Digital
3. The ratings of the app	• Doesn't show • 3.2 ★★★☆☆ • 4.0 ★★★★☆ • 4.8 ★★★★★
4. App description	• **Generic**, to create a rough idea of what the app is about without getting into details of app features • **Short**, with some details about app features • **Long**, with a detailed description of the app and its features
5. Images	• Shows the logo of the app • Shows the screenshot(s) of the app • Shows the logo and screenshot(s) of the app

Choice Tasks

Once the attributes and their levels are identified, the decision to develop full- or partial-profile tasks with or without an opt-out option needs to be made. A full profile refers to the display of all five attributes in both alternatives in each choice set. A partial profile DCE will not present certain attributes for certain alternatives. For example, if a DCE is used to investigate the trade-off between a higher number of attributes (eg, a total of nine attributes), it could be beneficial to limit the number of attributes shown at one time (eg, five attributes) to limit participant cognitive load. Five attributes are generally considered low enough to complete a full-profile choice task, which consequently maximizes the information about trade-offs [33]. Hence, in our research, we applied a full-profile DCE.

A neutral option ("Neither of these 2"), known as an opt-out alternative, was included, in addition to selecting alternative apps. The opt-out option has the potential to make the choices more realistic [34] by simulating a real-world context where individuals can exercise their right not to take up an app, given the apps on offer [20]. In our DCE, a participant had the option to choose or reject the hypothetical uptake of a smoking cessation app. However, when a participant selects the opt-out option, no information is provided on how they trade-off attribute levels or alternatives [13]. In some situations, a *forced-choice* scenario can be included, where participants who chose the opt-out option are prompted to make a choice regardless. An example of a scenario with an opt-out option is shown in Figure 2.

Figure 2. An example of a scenario with an opt-out option used in our DCE. DCE: discrete choice experiment.

You wish to quit smoking, and you decide to select a smartphone app to do that. Please look at the options carefully and decide which app (App 1 or App 2) you think you would likely want to download and use to help you quit smoking. You could also choose "Neither of these two" if you do not like either option and would not choose to download either app.

Take your time to make a decision. Please select an option and click on the arrow to continue.

	App 1	App 2
The monthly price of the app	£8.99	£0
Who developed the app	Mhealth Essentials Ltd.	NHS Digital
The ratings of the app	4.8 ★★★★★	4.0 ★★★★☆
App description	Generic, to create a rough idea of what the app is about without getting into details of app features	Short, with some details about app features
Images shown	Logo and screenshot(s) of the app	Logo of the app

Experimental Design

An experimental design is a systematic method of generating choice sets that are presented to respondents. This enables the specification of the choice sets that respondents see, with the objective of obtaining a high-quality data set [7]. When creating the experimental design, there are several aspects that need to be taken into consideration, including (1) the analytical model specification, (2) whether the aim is to estimate main effects only or interaction effects as well, (3) whether the design is labeled or unlabeled, (4) the number of choice tasks and blocking options to be used, (5) which type of design of the choice matrix to use (eg, full factorial or fractional factorial, orthogonal or efficient), and (6) how the attribute-level balance will be achieved. These are now considered.

Analytical Model Specification

The first step in the generation of an experimental design is to specify the analytical model to estimate the parameters of the DCE. This step is an important component of choosing the type of choice matrix design, described later in this paper. The approach selected here needs to be accounted for when generating the structure of the experimental design.

A discrete choice model describes the probability that an individual will choose a specific alternative. This probability is expressed as a function of measured attribute levels specific to the alternative and of characteristics of the individual making the choice. This probability is represented by the dependent variable (the *choice variable*), which indicates the choice made by participants [8]. In this modeling framework, the attributes are the independent variables [8,13].

As part of the analytical model specification, knowing what type of statistical analysis will be used is key. Data analysis involves regression modeling in a random utility framework [8]. The random utility model conventionally used is also based on the Lancaster theory of consumer demand [35], which together assume that individuals make trade-offs when making a decision and would choose an option that offers the greatest utility [36], determined by how much importance they place on the attributes associated with the product [37].

The multinomial logit (MNL) model has been previously described as the "workhorse" of DCE estimation [38,39], and it typically serves as a starting point for basic model estimation (although alternative models, such as probit, may be used). It is important to note that MNL requires some important assumptions and limitations—for example, independence of irrelevant alternatives, homogeneity of preferences, and independence of observed choices [40,41]. Extensions of MNL (eg, nested logit, mixed logit, and latent class models) may be used to account for these limitations [39,40].

Based on the model specified in our DCE, the underlying utility function for alternative *j* [38] is shown in Textbox 1.

Textbox 1. The utility function used in our DCE research. DCE: discrete choice experiment.

$U_j = (\beta_{cost} \times X_{jcost}) + (\beta_{developer} \times X_{jdeveloper}) + (\beta_{ratings} \times X_{j\ ratings}) + (\beta_{description} \times X_{jdescription}) + (\beta_{images} \times X_{njimages}) + \varepsilon$

Note:

1) U is the overall utility derived from alternative j.

2) β is the coefficient attached to X_j estimated in the analysis and represents the part-worth utility attached to each attribute level.

3) ε is the random error of the model—in other words, the unmeasured factors influencing the variation of preferences.

Main Effects or Interaction Effects

The next step in model specification is deciding whether main effects or interaction effects will be investigated. The main effects, the most commonly used, investigate the effect of each attribute level on the choice variable. The effect on the choice variable gained by combining two or more attribute levels (eg, app developer and the app's monthly cost) refers to an interaction effect [13]. In our DCE, given the novel nature of the research on the uptake of health apps and the lack of empirical evidence to suggest the presence of potential interactions between attributes, we decided to only look at main effects.

Labeled or Unlabeled Experiment

In a labeled experiment, the alternatives are specific and different (eg, smartphone app-based smoking cessation intervention vs website-based smoking cessation intervention) and alternative specific attributes could be used (eg, some attributes relevant only for apps and others for websites). This is in contrast to an unlabeled experimental design, where the alternatives are unspecified (eg, smoking cessation app alternative 1 vs smoking cessation app alternative 2) and also must have the same attributes. Given that a DCE model estimates parameters for each of the alternatives being considered, these alternative specific parameters must be included in the structure of the experimental design (described in the next section) in a labeled experiment; in an unlabeled experiment, because alternative specific parameters are arbitrary, they are excluded [22,42,43]. In health economics, the unlabeled approach is the most common. In our DCE, the unlabeled approach was deemed logical here as we were comparing different presentations of the same app. Therefore, our DCE design applied an unlabeled approach.

Generation of the Structure of the Experimental Design

Once the model is specified, the structure of the experimental design can be generated. For this stage, hypothetical alternatives are generated and combined to form choice tasks, based on the chosen attributes and their levels. Several different software packages may be used to generate the experimental design of a DCE, such as Ngene, SAS, SPEED, SPSS, and Sawtooth. For our DCE, Ngene software was used [44].

Number of Choice Tasks and Blocking

The next step in the generation of an experimental design is to decide on the choice task and blocking. To minimize respondent and cognitive burden, and the risk of participants losing interest during the DCE task, consideration must be paid to the target population, the number of tasks, and their complexity [13]. The higher the number of attributes, alternatives, and choice tasks, the higher the task complexity [20]. The literature suggests that a feasible limit is 18 choice sets per participant [45,46]. In the review by Marshall et al [27], most studies included between 7 and 16 choice sets. In our DCE, we administered 12 choice tasks per participant, which were deemed a number low enough to avoid excessive cognitive load but high enough to establish sufficient statistical precision.

We developed 48 choice tasks and blocked them into 4 survey versions (12 choice tasks for each). Each block represented a separate survey, and participants were randomly assigned to one of the four survey versions. Blocking is a technique widely used in DCEs to reduce cognitive burden by partitioning large experimental designs into subsets of equal size, thereby reducing the number of choice tasks that any one respondent is required to complete [47]. Blocks were generated in Ngene software, which allows for the minimization of the average correlation between the versions and attributes' levels [48]. For the blocking to be successful, the number of choice tasks included in one block must be divisible by the number of attribute levels; in our DCE, attributes had either three or four levels.

It is noteworthy that to undertake the sample size calculation, it is crucial to know the number of alternatives per choice set, the largest number of levels of any attribute (for DCEs looking at main effects only) or the largest level of any two attributes (for a DCE looking at interaction effects), and the number of blocks [38]. Therefore, DCEs using blocking require a larger sample size [47].

Type of Choice Matrix Design

Depending on the number of attributes and their levels, a full- or fractional-factorial design can be applied. A full-factorial design would include all possible combinations of the attributes' levels and allow the estimation of all main effects and interaction effects independent of one another [20]. However, this type of design is often considered impractical due to the high number of choice tasks required [20]. To illustrate this, the formula of calculation of the possible unique choice alternatives for a full-factorial design is L^A, where L represents the number of levels and A the number of attributes [39]. If the attributes in the DCE have a different number of levels, these need to be calculated separately and multiplied together. To reduce response burden, in our DCE, we generated a fractional-factorial design in Ngene [44], representing a sample of possible alternatives from the full-factorial design. This way, we were able to reduce the total 432 alternatives in the full design (given by $L^A = 4^2 \times 3^3$) to a fractional sample of 96 alternatives, arranged in 48 choice pairs.

Systematic approaches for generation of fractional-factorial designs may be further categorized into orthogonal design and

efficient design. An orthogonal design is a column-based design based on orthogonal arrays that present properties of orthogonality (attributes are statistically independent of one another) and level balance (levels of attributes appear an equal number of times) and does not introduce correlation between the attributes [38]. An orthogonal array is an optimal design that is often used for DCEs examining main effects when the number of attributes and their levels is small.

For studies with five or more attributes with two or more levels, an orthogonal design may not be practical. There has therefore been a recent change in thinking toward a nonorthogonal and statistically more efficient design [38]. When perfect orthogonality and balance cannot be achieved or are not desirable, an efficient design can be applied [20]. In contrast to an orthogonal design, an efficient design aims to increase the precision of parameter estimates for a given sample size (ie, minimizing the standard errors of the estimated coefficients), while allowing some limited correlation between attributes. The most widely used efficiency measure is the D-error, which may be easily estimated using various software packages, such as Ngene, and refers to the efficiency of the experimental design in extracting information from respondents [21]. Experimental designs generated using this approach are known as D-efficient designs. A D-efficient experimental design is also recommended to maximize statistical efficiency and minimize the variability of parameter estimates [7].

An efficient design requires that known prior information about the parameters (known as priors) be made available to the algorithm and also requires the analyst to specify the analytical model specification, as described previously. Depending on what information is available, one of three types of D-efficient design can be generated [21]:

1. *D_z-efficient* design (*z* stands for zero priors): If no prior information about the magnitude or directions of the parameters is available. D_z-efficient design is an orthogonal design. This design assumes the parameters are zero.
2. *D_p-efficient* design (*p* stands for priors): This assumes a fixed, certain value and direction for the parameters.
3. *D_b-efficient* design (*b* stands for Bayesian): A Bayesian approach is whereby the parameter is not known with certainty but may be described by its probability distribution.

The best practice is to pilot the DCE. For the pilot phase, there is limited information available and using the D_z-efficient or D_p-efficient design is sensible. In our DCE, we chose to apply a D_p-efficient design, as the direction of priors of the app was known from the previously conducted survey, to narrow down the attribute levels and to provide prior estimates of the parameters for the attribute levels. For example, we knew that a trusted organization will likely positively influence uptake and cost estimated negatively so. The direction of priors was assumed to be a small near-zero negative or a positive value for the design.

The pilot phase provided the estimation that we used to generate a D_b-efficient design for the final DCE. It is noteworthy that when the parameter priors are different from zero, the efficient design generated produces smaller prediction errors than orthogonal designs [21,49,50]. Hence, a D-efficient design will outperform an orthogonal design, and (given reliable priors), a D_p-efficient design will outperform a D_z-efficient design [21]. Further, when reasonable assumptions about the distributions are made, a D_b-efficient design will outperform a D_p-efficient design. Therefore, it may be advisable to start piloting with a D_p-efficient design and to generate a D_b-efficient design for the final DCE. The DCE literature provides a detailed and more comprehensive description of orthogonal and efficient designs [21] and the approximation of the Bayesian efficient design [23].

Attribute-Level Balance in the Model

The attribute-level balance aims to ensure all attribute levels ideally appear an equal number of times in the experimental design. The allocation of the attribute levels within the experimental design can affect statistical power; if a certain level is underrepresented in the choice sets generated, then the coefficient for that level cannot be easily estimated. How attributes levels are distributed is therefore an important consideration when designing the choice sets. Dominant alternatives, where all attribute levels of one alternative are more desirable than all attribute levels in the others, do not provide information about how trade-offs are made, as individuals usually would select the dominant alternatives. Therefore, avoiding dominant alternatives in the experimental design is important and can be achieved by consulting the software manual to ensure the correct algorithm is used. The syntax used in Ngene to generate choice sets of the pilot phase and more information about the algorithm used can be accessed on the Open Science Framework [19].

Piloting the DCE and Generating the Bayesian Design

In addition to providing estimations for the choice matrix design described above, piloting offers an opportunity to ensure that the information is presented clearly and that the choices are realistic and meaningful. It also provides insight into how cognitively demanding it is for respondents to complete. This can be achieved by gathering feedback on the survey completion process. The findings of the pilot may suggest that the DCE needs to be amended, such as reducing the number of choice sets or the number of attributes, so that the responses are a better reflection of the participants' preferences and improve the precision in the parameter estimates [13].

There is no formal guidance on how large the pilot sample should be, and this is largely guided by the budget and complexity of the experimental design. Accuracy of the priors will improve with increasing sample size, but as few as 30 responses may be sufficient to generate useable data [44]. In our pilot study conducted with 49 individuals, feedback from the participants suggested that with the initial order of the attributes, there was a tendency to ignore the last two attributes, app description and images of the app, the most text-heavy attributes. This may have compromised the examination of the relative importance of those two attributes (app description and images of the app). Therefore, we decided to change the final order of the attributes from (1) *monthly price of the app*, (2) *the*

ratings of the app, (3) *who developed the app*, (4) *the description*, and (5) *images shown* to the one listed in Figures 1 and 2. The longest completion time for the survey was under 12 min. Thus, we concluded that the number of choice tasks did not need to be reduced.

In our research, the data from the pilot phase were analyzed using the freely available Apollo package in R software [51]. The coefficients and their standard errors from the output were used as priors to generate the final choice sets using the Bayesian efficient design following the steps described previously. The syntax used in R used to analyze the pilot data and that used to generate the Bayesian efficient design in Ngene can be accessed on the Open Science Framework [19].

Internal Validity

Assessing the internal validity of a DCE can help with understanding the consistency and trade-off assumptions made by participants [52]. There are several ways to examine the internal validity of a DCE. For example, in the *stability validity test*, a choice task would be repeated later in the sequence to investigate the consistency of the participants' decision, whether they would choose the same alternative [52]. Another way to test internal validity is the *within-set dominated pairs* type of internal validity, in which one alternative is a dominant alternative in which all attributes are the most desirable ones. The choice sets designed to measure internal validity are excluded from the analysis. There are several internal validity tests that are built into software packages such as MATLAB [52], although these can be produced manually as well. In our research, we used the stability validity test to check the internal validity by repeating a randomly generated choice task (in our case, it was the fourth). Therefore, participants were shown 12 choice tasks, plus an additional hold-out task. The data from the randomly generated hold-out task were excluded from the analysis.

Although internal validity checks provide some measure of data quality, it should be noted that answering a repeat choice inconsistently is not a violation of random utility theory [53]. Furthermore, there is no consensus on what to do with the data from responses that fail validity tests. Following the advice of Lancsar and Louviere [54], we did not exclude participants who failed the internal validity check, as that might have caused statistical bias or affected statistical efficiency. However, we reported data on internal validity to enable the reader to make a judgement on likely biases.

All additional study materials used in our example, including the full data set and the results of the DCE, can be accessed on Open Science Framework [19].

Discussion

Summary

This paper describes the development of a DCE, following the stages required to establish attributes and their levels, construct choice tasks, define the utility model, decide on labeled and unlabeled choices to apply, decide on the number of choice tasks that need to be generated, and make decisions on the structure of the experimental design, how to achieve attribute-level balance, how to assess the internal model validity, and how to pilot-test. In doing so, the intention is to advance methodological awareness of the application of stated preference methods in the field of digital health, as well as to provide researchers with an overview of their application using a case study of a DCE of smoking cessation app uptake.

Although DCEs are widely used to understand patient and provider choices in health care [8,10,15,55], they have only recently started to gain popularity in digital health [4-6] and as such represent an underused approach in digital health. With the growing evidence of the benefit of digital health initiatives, there are clear benefits to widening the application of DCEs so that they may more routinely inform digital health development, inform digital tool presentation, and, most importantly, predict uptake and engagement with digital products. Although several attempts have been made to measure engagement with digital tools using a wide range of methodologies [56-58], the insights we have from them that can be translated to uptake are limited. One plausible explanation is that uptake of digital tools is difficult to empirically measure.

Benefits and Limitations of DCEs

DCEs bring several benefits to help overcome the issue of measuring uptake in digital health or in other areas where the measurement of the predictors of uptake in a good or service is required. For example, as illustrated by the case study here, they enable the researcher to gain measurable insights into situations in which quantitative measures are hard to otherwise obtain, such as the factors impacting the uptake of health apps on curated health app portals. A DCE also helps to quantify preferences to support more complex decisions [59]. An example would be the consideration of how to plan the development of an app that would provide appealing looks or features that would promote uptake. The DCE methodology is also considered a convenient approach to investigate the uptake of new interventions, including digital health interventions [38], for example, digital behavior change interventions using a health and well-being smartphone app. Therefore, DCEs can be used in hypothetical circumstances, enabling the measurement of preferences for a potential policy change or digital health system change before it is implemented [13], such as the recent investigation of the uptake of a COVID-19 test-and-trace health app [3,4]. The experimental nature of the DCE also means that participants' preferences can be recorded based on controlled experimental conditions, where attributes are systematically varied by researchers to obtain insight into the marginal effect of attribute changes on individuals' choices [7].

Despite their benefits, the application of DCEs presents several challenges. As with all expressed preference methodologies, the hypothetical nature of the DCE choice set raises concerns about external validity and the degree to which real-world decisions might equate to those made by study participants under experimental conditions, a phenomenon known as the intention-behavior gap [60]. As such, participants may believe they would choose a scenario presented and described in a choice task, but in real life, there might be other factors that would influence their behaviors, such as the aesthetics of the app [28]. This limitation can at least partially be overcome by

developing convincing and visually appealing choice tasks. Nevertheless, to date, there has been limited progress in testing for external validity due to the difficulty in investigating preferences in the real world [38]. Indeed, a recent systematic review of the literature on DCEs in health care reported that only 2% of the included studies (k=7) report details of the investigation of external validity [47], while an earlier systematic review and meta-analysis (k=6) found that DCEs have only a moderate level of accuracy in predicting behaviors of health choices [61]. To our knowledge, no study has been published that investigates the external validity of a DCE developed in digital health. One potential opportunity to undertake some testing would be through a curated health app portal, where the same health app is presented in two or more different ways. With the help of website analytics, actual user behavior could be measured in this situation.

A final significant concern associated with the use of a DCE is that any single choice set is unlikely to be able to present the user with all relevant attributes, regardless of how well it has been developed [61]. Choosing the most relevant attributes to test in a DCE, therefore, requires comprehensive preparatory research, which can lengthen the time required to undertake the development phase of any piece of work.

Conclusion

In summary, DCEs have significant potential in digital health research and can serve as an important decision-making tool in a field where observational data are lacking. We hope that the content of this paper provides a useful introduction and guide to those interested in developing such experiments in digital health.

Acknowledgments

We are grateful to two experts in discrete choice experiments, Prof. Michiel Bliemer from the University of Sydney, Australia, a co-developer of Ngene software, and Prof. Stephane Hass from the University of Leeds, United Kingdom, a co-developer of the Apollo package in R software, for their advice on the syntax used to generate the choice tasks in Ngene and on the code used in the Apollo package. JAW and RC received funding from the National Institute for Health Research (NIHR) Applied Research Collaboration East of England. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

Authors' Contributions

DS prepared the manuscript. All authors have reviewed the draft for important intellectual content and approved the final version.

Conflicts of Interest

JB has received unrestricted funding to study smoking cessation from Pfizer and J&J, who manufacture smoking cessation medications.

References

1. Danner M, Hummel JM, Volz F, van Manen JG, Wiegard B, Dintsios C, et al. Integrating patients' views into health technology assessment: analytic hierarchy process (AHP) as a method to elicit patient preferences. Int J Technol Assess Health Care 2011 Oct;27(4):369-375. [doi: 10.1017/S0266462311000523] [Medline: 22004779]
2. Hall J, Viney R, Haas M, Louviere J. Using stated preference discrete choice modelling to evaluate health care programs. Journal of Business Research 2002 Jun;57(9):1026-1032. [doi: 10.1016/S0148-2963(02)00352-1]
3. Wiertz C, Banerjee A, Acar OA, Ghosh A. Predicted adoption rates of contact tracing app configurations: insights from a choice-based conjoint study with a representative sample of the UK population. SSRN Journal 2020 Apr 30:1-19. [doi: 10.2139/ssrn.3589199]
4. Jonker M, de Bekker-Grob E, Veldwijk J, Goossens L, Bour S, Rutten-Van Mölken M. COVID-19 contact tracing apps: predicted uptake in the Netherlands based on a discrete choice experiment. JMIR Mhealth Uhealth 2020 Oct 09;8(10):e20741-e20715 [FREE Full text] [doi: 10.2196/20741] [Medline: 32795998]
5. Nittas V, Mütsch M, Braun J, Puhan MA. Self-monitoring app preferences for sun protection: discrete choice experiment survey analysis. J Med Internet Res 2020 Nov 27;22(11):e18889 [FREE Full text] [doi: 10.2196/18889] [Medline: 33245282]
6. Leigh S, Ashall-Payne L, Andrews T. Barriers and facilitators to the adoption of mobile health among health care professionals from the United Kingdom: discrete choice experiment. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e17704 [FREE Full text] [doi: 10.2196/17704] [Medline: 32628118]
7. Reed Johnson F, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. Value Health 2013;16(1):3-13 [FREE Full text] [doi: 10.1016/j.jval.2012.08.2223] [Medline: 23337210]
8. Spinks J, Chaboyer W, Bucknall T, Tobiano G, Whitty JA. Patient and nurse preferences for nurse handover-using preferences to inform policy: a discrete choice experiment protocol. BMJ Open 2015 Nov 11;5(11):e008941 [FREE Full text] [doi: 10.1136/bmjopen-2015-008941] [Medline: 26560060]

9. Bridges JFP, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint analysis applications in health—a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value Health 2011 Jun;14(4):403-413 [FREE Full text] [doi: 10.1016/j.jval.2010.11.013] [Medline: 21669364]
10. Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete choice experiments in health economics: a review of the literature. Pharmacoeconomics 2014 Sep;32(9):883-902. [doi: 10.1007/s40273-014-0170-x] [Medline: 25005924]
11. Kotnowski K, Fong GT, Gallopel-Morvan K, Islam T, Hammond D. The impact of cigarette packaging design among young females in Canada: findings from a discrete choice experiment. Nicotine Tob Res 2016 May;18(5):1348-1356. [doi: 10.1093/ntr/ntv114] [Medline: 26014454]
12. Lambooij MS, Harmsen IA, Veldwijk J, de Melker H, Mollema L, van Weert YWM, et al. Consistency between stated and revealed preferences: a discrete choice experiment and a behavioural experiment on vaccination behaviour compared. BMC Med Res Methodol 2015 Mar 12;15:19 [FREE Full text] [doi: 10.1186/s12874-015-0010-5] [Medline: 25887890]
13. Mangham LJ, Hanson K, McPake B. How to do (or not to do) ... designing a discrete choice experiment for application in a low-income country. Health Policy Plan 2009 Mar;24(2):151-158. [doi: 10.1093/heapol/czn047] [Medline: 19112071]
14. Trapero-Bertran M, Rodríguez-Martín B, López-Bastida J. What attributes should be included in a discrete choice experiment related to health technologies? A systematic literature review. PLoS One 2019;14(7):e0219905 [FREE Full text] [doi: 10.1371/journal.pone.0219905] [Medline: 31318926]
15. Hall J, Kenny P, King M, Louviere J, Viney R, Yeoh A. Using stated preference discrete choice modelling to evaluate the introduction of varicella vaccination. Health Econ 2002 Jul;11(5):457-465. [doi: 10.1002/hec.694] [Medline: 12112494]
16. Fiebig DG, Knox S, Viney R, Haas M, Street DJ. Preferences for new and existing contraceptive products. Health Econ 2011 Sep;20 Suppl 1:35-52. [doi: 10.1002/hec.1686] [Medline: 21809412]
17. Terris-Prestholt F, Hanson K, MacPhail C, Vickerman P, Rees H, Watts C. How much demand for new HIV prevention technologies can we really expect? Results from a discrete choice experiment in South Africa. PLoS One 2013;8(12):e83193 [FREE Full text] [doi: 10.1371/journal.pone.0083193] [Medline: 24386160]
18. Terris-Prestholt F, Quaife M, Vickerman P. Parameterising user uptake in economic evaluations: the role of discrete choice experiments. Health Econ 2016 Feb;25 Suppl 1:116-123 [FREE Full text] [doi: 10.1002/hec.3297] [Medline: 26773825]
19. Szinay D, Rory C, Jones A, Whitty J, Chadborn T, Brown J, et al. Adult Smokers' Preferences for the Uptake of Smoking Cessation Apps: A Discrete Choice Experiment. 2021 Mar 12. URL: https://osf.io/5439x/ [accessed 2020-11-01]
20. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. Pharmacoeconomics 2008;26(8):661-677. [doi: 10.2165/00019053-200826080-00004] [Medline: 18620460]
21. Rose JM, Bliemer MCJ. Constructing efficient stated choice experimental designs. Transport Rev 2009 Sep;29(5):587-617. [doi: 10.1080/01441640902827623]
22. de Bekker-Grob EW, Hol L, Donkers B, van Dam L, Habbema JDF, van Leerdam ME, et al. Labeled versus unlabeled discrete choice experiments in health economics: an application to colorectal cancer screening. Value Health 2010;13(2):315-323 [FREE Full text] [doi: 10.1111/j.1524-4733.2009.00670.x] [Medline: 19912597]
23. Bliemer MC, Rose JM, Hess S. Approximation of bayesian efficiency in experimental choice designs. J Choice Model 2008;1(1):98-126. [doi: 10.1016/s1755-5345(13)70024-1]
24. Hauber AB, González JM, Groothuis-Oudshoorn CGM, Prior T, Marshall DA, Cunningham C, et al. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR Conjoint Analysis Good Research Practices Task Force. Value Health 2016 Jun;19(4):300-315 [FREE Full text] [doi: 10.1016/j.jval.2016.04.004] [Medline: 27325321]
25. Kløjgaard ME, Bech M, Søgaard R. Designing a stated choice experiment: the value of a qualitative process. J Choice Model 2012;5(2):1-18. [doi: 10.1016/s1755-5345(13)70050-2]
26. Buchanan J, Blair E, Thomson KL, Ormondroyd E, Watkins H, Taylor JC, et al. Do health professionals value genomic testing? A discrete choice experiment in inherited cardiovascular disease. Eur J Hum Genet 2019 Nov 11;27(11):1639-1648 [FREE Full text] [doi: 10.1038/s41431-019-0452-z] [Medline: 31186546]
27. Marshall D, Bridges JFP, Hauber B, Cameron R, Donnalley L, Fyie K, et al. Conjoint analysis applications in health: how are studies being designed and reported? An update on current practice in the published literature between 2005 and 2008. Patient 2010 Dec 01;3(4):249-256. [doi: 10.2165/11539650-000000000-00000] [Medline: 22273432]
28. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 May 29;22(5):e17572-e17523 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
29. Szinay D, Perski O, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of health and well-being apps and curated app portals: think-aloud and interview study. JMIR Mhealth Uhealth 2021 Apr 27;9(4):e27173 [FREE Full text] [doi: 10.2196/27173] [Medline: 33904827]
30. Rakotonarivo OS, Schaafsma M, Hockley N. A systematic review of the reliability and validity of discrete choice experiments in valuing non-market environmental goods. J Environ Manage 2016 Dec 01;183:98-109 [FREE Full text] [doi: 10.1016/j.jenvman.2016.08.032] [Medline: 27576151]
31. Ratcliffe J, Longworth L. Investigating the structural reliability of a discrete choice experiment within health technology assessment. Int J Technol Assess Health Care 2002;18(1):139-144.

32. Yang J, Reed SD, Hass S, Skeen MB, Johnson FR. Is easier better than harder? An experiment on choice experiments for benefit-risk tradeoff preferences. Med Decis Making 2021 Feb;41(2):222-232. [doi: 10.1177/0272989X20979833] [Medline: 33463397]
33. Mühlbacher A, Johnson FR. Choice experiments to quantify preferences for health and healthcare: state of the practice. Appl Health Econ Health Policy 2016 Jun;14(3):253-266. [doi: 10.1007/s40258-016-0232-7] [Medline: 26992386]
34. Watson V, Becker F, de Bekker-Grob E. Discrete choice experiment response rates: a meta-analysis. Health Econ 2017 Jun;26(6):810-817. [doi: 10.1002/hec.3354] [Medline: 27122445]
35. Ryan M, Gerard K, Amaya-Amaya M. Discrete choice experiments in a nutshell. In: Ryan M, Gerard K, Amaya-Amaya M, editors. Using Discrete Choice Experiments to Value Health and Health Care. Dordrecht, the Netherlands: Springer; 2008:13-46.
36. McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P, editor. Frontiers in Econometrics. Cambridge, MA: Academic Press; 1973:105-142.
37. Potoglou D, Burge P, Flynn T, Netten A, Malley J, Forder J, et al. Best-worst scaling vs. discrete choice experiments: an empirical comparison using social care data. Soc Sci Med 2011 May;72(10):1717-1727. [doi: 10.1016/j.socscimed.2011.03.027] [Medline: 21530040]
38. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ 2012 Feb;21(2):145-172. [doi: 10.1002/hec.1697] [Medline: 22223558]
39. Hensher DA, Rose JM, Green W. Applied Choice Analysis: A Primer. Cambridge, UK: Cambridge University Press; 2005:1.
40. Hensher DA, Greene WH. The mixed logit model: the state of practice. Transportation 2003;30(2):133-176. [doi: 10.1023/A:1022558715350]
41. Train K. Logit. Discrete Choice Methods with Simulation. 2 ed. Cambridge: Cambridge University Press; 2009:34-75.
42. Kruijshaar ME, Essink-Bot M, Donkers B, Looman CW, Siersema PD, Steyerberg EW. A labelled discrete choice experiment adds realism to the choices presented: preferences for surveillance tests for Barrett esophagus. BMC Med Res Methodol 2009 May 19;9(1):31 [FREE Full text] [doi: 10.1186/1471-2288-9-31] [Medline: 19454022]
43. Jin W, Jiang H, Liu Y, Klampfl E. Do labeled versus unlabeled treatments of alternatives' names influence stated choice outputs? Results from a mode choice study. PLoS One 2017;12(8):e0178826 [FREE Full text] [doi: 10.1371/journal.pone.0178826] [Medline: 28806764]
44. ChoiceMetrics. NGene.1.1.1 User Manual & Reference Guide A. Online: ChoiceMetrics; 2012:1.
45. Christofides NJ, Muirhead D, Jewkes RK, Penn-Kekana L, Conco DN. Women's experiences of and preferences for services after rape in South Africa: interview study. BMJ 2006 Jan 28;332(7535):209-213 [FREE Full text] [doi: 10.1136/bmj.38664.482060.55] [Medline: 16330476]
46. Hanson K, McPake B, Nakamba P, Archard L. Preferences for hospital quality in Zambia: results from a discrete choice experiment. Health Econ 2005 Jul;14(7):687-701. [doi: 10.1002/hec.959] [Medline: 15619273]
47. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete choice experiments in health economics: past, present and future. Pharmacoeconomics 2019 Feb;37(2):201-226 [FREE Full text] [doi: 10.1007/s40273-018-0734-2] [Medline: 30392040]
48. Janssen EM, Hauber AB, Bridges JFP. Conducting a discrete-choice experiment study following recommendations for good research practices: an application for eliciting patient preferences for diabetes treatments. Value Health 2018 Jan;21(1):59-68 [FREE Full text] [doi: 10.1016/j.jval.2017.07.001] [Medline: 29304942]
49. Ferrini S, Scarpa R. Designs with a priori information for nonmarket valuation with choice experiments: a Monte Carlo study. J Environ Econ Manage 2007 May;53(3):342-363. [doi: 10.1016/j.jeem.2006.10.007]
50. Kessels R, Jones B, Goos P, Vandebroek ML. An efficient algorithm for constructing Bayesian optimal choice designs. SSRN J 2006:1 [FREE Full text] [doi: 10.2139/ssrn.968620]
51. Hess S, Palma D. Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application. J Choice Model 2019 Sep;32:100170. [doi: 10.1016/j.jocm.2019.100170]
52. Johnson FR, Yang J, Reed SD. The internal validity of discrete choice experiment data: a testing tool for quantitative assessments. Value Health 2019 Feb;22(2):157-160 [FREE Full text] [doi: 10.1016/j.jval.2018.07.876] [Medline: 30711059]
53. Hess S, Daly A, Batley R. Revisiting consistency with random utility maximisation: theory and implications for practical work. Theory Decis 2018 Jan 2;84(2):181-204 [FREE Full text] [doi: 10.1007/s11238-017-9651-7] [Medline: 31983783]
54. Lancsar E, Louviere J. Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? Health Econ 2006 Aug;15(8):797-811. [doi: 10.1002/hec.1104] [Medline: 16615039]
55. Quaife M, Terris-Prestholt F, Eakle R, Cabrera Escobar MA, Kilbourne-Brook M, Mvundura M, et al. The cost-effectiveness of multi-purpose HIV and pregnancy prevention technologies in South Africa. J Int AIDS Soc 2018 Mar;21(3):1 [FREE Full text] [doi: 10.1002/jia2.25064] [Medline: 29537654]
56. Perski O, Blandford A, Garnett C, Crane D, West R, Michie S. A self-report measure of engagement with digital behavior change interventions (DBCIs): development and psychometric evaluation of the "DBCI Engagement Scale". Transl Behav Med 2020 Feb 03;10(1):267-277 [FREE Full text] [doi: 10.1093/tbm/ibz039] [Medline: 30927357]

57. Craig Lefebvre R, Tada Y, Hilfiker SW, Baur C. The assessment of user engagement with eHealth content: the eHealth Engagement Scale. J Comput Mediat Commun 2010 Jul 01;15(4):666-681. [doi: 10.1111/j.1083-6101.2009.01514.x]
58. O'Brien HL, Toms EG. The development and evaluation of a survey to measure user engagement. J. Am. Soc. Inf. Sci 2009 Oct 19;61(1):50-69. [doi: 10.1002/asi.21229]
59. Brett Hauber A, Fairchild AO, Reed Johnson F. Quantifying benefit-risk preferences for medical interventions: an overview of a growing empirical literature. Appl Health Econ Health Policy 2013 Aug;11(4):319-329. [doi: 10.1007/s40258-013-0028-y] [Medline: 23637054]
60. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
61. Quaife M, Terris-Prestholt F, Di Tanna GL, Vickerman P. How well do discrete choice experiments predict health choices? A systematic review and meta-analysis of external validity. Eur J Health Econ 2018 Nov;19(8):1053-1066. [doi: 10.1007/s10198-018-0954-6] [Medline: 29380229]

Abbreviations

DCE: discrete choice experiment
MNL: multinomial logit

Edited by G Eysenbach; submitted 24.07.21; peer-reviewed by HL Tam, R Marshall; comments to author 16.08.21; revised version received 24.08.21; accepted 18.09.21; published 11.10.21.

Please cite as:
Szinay D, Cameron R, Naughton F, Whitty JA, Brown J, Jones A
Understanding Uptake of Digital Health Products: Methodology Tutorial for a Discrete Choice Experiment Using the Bayesian Efficient Design
J Med Internet Res 2021;23(10):e32365
URL: https://www.jmir.org/2021/10/e32365
doi:10.2196/32365
PMID:34633290

Original Paper

Brain Mechanisms of Virtual Reality Breathing Versus Traditional Mindful Breathing in Pain Modulation: Observational Functional Near-infrared Spectroscopy Study

Xiao-Su Hu[1], MSc, PhD; Katherine Beard[1], MSc, DDS; Mary Catherine Sherbel[1,2], MSc, DDS; Thiago D Nascimento[1], MSc, DDS; Sean Petty[3], BS; Eddie Pantzlaff[1], DDS; David Schwitzer[1], DDS; Niko Kaciroti[1,4], PhD; Eric Maslowski[5], BS; Lawrence M Ashman[1,6], DDS; Stephen E Feinberg[1,6], DDS, PhD; Alexandre F DaSilva[1], DDS, DMedSc

[1]Headache & Orofacial Pain Effort Lab, Biologic and Materials Sciences & Prosthodontics Department, University of Michigan School of Dentistry, Ann Arbor, MI, United States

[2]Department of Orthodontics and Pediatric Dentistry, University of Michigan School of Dentistry, Ann Arbor, MI, United States

[3]3D Lab, Digital Media Commons, University of Michigan, Ann Arbor, MI, United States

[4]Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, United States

[5]Moxytech Inc, Ann Arbor, MI, United States

[6]Department of Oral & Maxillofacial Surgery, University of Michigan School of Dentistry, Ann Arbor, MI, United States

Corresponding Author:
Alexandre F DaSilva, DDS, DMedSc
Headache & Orofacial Pain Effort Lab
Biologic and Materials Sciences & Prosthodontics Department
University of Michigan School of Dentistry
205 Zina Pitcher Pl, room 1021
Ann Arbor, MI, 48109-5720
United States
Phone: 1 734 615 3807
Fax: 1 734 763 3453
Email: adasilva@umich.edu

Abstract

Background: Pain is a complex experience that involves sensory-discriminative and cognitive-emotional neuronal processes. It has long been known across cultures that pain can be relieved by mindful breathing (MB). There is a common assumption that MB exerts its analgesic effect through interoception. Interoception refers to consciously refocusing the mind's attention to the physical sensation of internal organ function.

Objective: In this study, we dissect the cortical analgesic processes by imaging the brains of healthy subjects exposed to traditional MB (TMB) and compare them with another group for which we augmented MB to an outside sensory experience via virtual reality breathing (VRB).

Methods: The VRB protocol involved in-house–developed virtual reality 3D lungs that synchronized with the participants' breathing cycles in real time, providing them with an immersive visual-auditory exteroception of their breathing.

Results: We found that both breathing interventions led to a significant increase in pain thresholds after week-long practices, as measured by a thermal quantitative sensory test. However, the underlying analgesic brain mechanisms were opposite, as revealed by functional near-infrared spectroscopy data. In the TMB practice, the anterior prefrontal cortex uniquely modulated the premotor cortex. This increased its functional connection with the primary somatosensory cortex (S1), thereby facilitating the S1-based sensory-interoceptive processing of breathing but inhibiting its other role in sensory-discriminative pain processing. In contrast, virtual reality induced an immersive 3D exteroception with augmented visual-auditory cortical activations, which diminished the functional connection with the S1 and consequently weakened the pain processing function of the S1.

Conclusions: In summary, our study suggested two analgesic neuromechanisms of VRB and TMB practices—exteroception and interoception—that distinctively modulated the S1 processing of the ascending noxious inputs. This is in line with the concept of dualism (Yin and Yang).

(J Med Internet Res 2021;23(10):e27298) doi:10.2196/27298

KEYWORDS

virtual reality breathing; traditional mindful breathing; pain; functional near-infrared spectroscopy; mobile phone

Introduction

Background

With the development of functional neuroimaging, our understanding of pain has matured to a concept of multidimensional experience in which the brain integrates inputs from sensory-discriminative and cognitive-emotional systems as a central hub [1]. Pain neuroimaging has also proved that complementary medicine approaches, beyond pharmacological analgesic means, can modulate these central systems [2].

Mindful breathing (MB) is widely accepted as an authentic treatment for pain relief by patients and society in general [3]. The adoption of MB is a welcomed change in our clinical mindset. It decreases our tendency to rely exclusively on pain medications, which can sometimes escalate to dire side effects [4]. In addition, MB techniques are self-facilitated and easy to implement compared with other methods. MB requires learners to regulate their attention to the dynamic interoceptive nature of breathing. When other thoughts disrupt the focus, the learners need to recognize the disruption and refocus on their breathing. This practice gradually gains a learner's mental control and stabilization abilities. This ability has been proven to alleviate anxiety, stress, depression, and pain among patients [5].

However, the brain mechanisms for the MB practice in pain modulation are poorly understood [5]. Pain is believed to be represented in the brain via affective and sensory networks [6]. Briefly, the ascending noxious signal reaches the spinal trigeminal nucleus, thalamus, and sensory cortex [7,8]. The signal is also processed in the insular cortex and subjectively evaluated in the anterior cingulate cortex, prefrontal cortex (PFC), and other cognitive-emotional regions [9,10]. However, whether the two dimensions can be separately modulated is not well supported by the existing literature [2]. One study investigated a group of long-term Zen meditation practitioners with significantly higher pain thresholds and found increased activation in sensory-related regions (thalamus and insula) but reduced activation in pain-evaluation areas (medial PFC and anterior PFC [aPFC]) [11]. Moreover, direct associations were found between the level of PFC deactivation and meditation-induced pain reduction [11]. In addition, another study compared the effect of a real-meditation training program and a placebo relaxation program, which found increased connectivity between the posterior cingulate cortex and the dorsolateral PFC (DLPFC) [12,13]. Collectively, these findings suggest that mindfulness meditation may modulate pain through a unique mechanism (eg, high-level cortical function in the PFC).

An existing problem with current MB training is the difficulty of long-time attention focusing, especially for beginners, as this is a subjective interoception process. Therefore, there is an urgent need for a tangible method that can provide an immersive sensory guide to facilitate mental control and match the expectations of the current tech-savvy generation. Recently, the development of virtual reality (VR) has enabled the implementation of such methods. VR is a computer-simulated 3D and interactive experience [14]. Delivered by a visual-audio headset, the *virtual* experience modulates human sensory and emotional systems. The VR technology has effectively managed pain from burns, cancer, and dental procedures [15,16]. Researchers assumed that this process translocates the patients into immersive 3D auditory (eg, esoteric music and rain sound) and visual (eg, breaking waves and moving geometric patterns) contextual experiences [17], although the exact brain mechanisms remain unclear. For a long time, the VR-based pain modulation effect was understood as a distraction mechanism, in which limited attention is partially occupied by exteroceptive VR stimulation (auditory and visual) instead of pain [15,16]. A couple of previous studies found decreased activation of pain-related regions during VR sessions. However, recent studies have shown an after-VR effect rather than only distraction mechanisms [18].

Objectives

Although proven to be effective in pain modulation, the underlying brain mechanisms of the two processes—abstract sensory-interoception and VR-based sensory-exteroception—remain unclear. This study compared the effectiveness of the two methods in modulating the patients' pain thresholds in the same study design and clinical environment. We used a week-long protocol in which the participants practiced the traditional MB (TMB) and VR breathing (VRB) in the lab on the first and seventh day, respectively, intercalated by five daily MB practices at home. During the in-lab sessions, we measured the participants' pain threshold using a facial thermal quantitative sensory test (tQST) after their breathing practices. We used functional near-infrared spectroscopy (fNIRS) as a neuroimaging technique to measure participants' cortical connectivity and activation. fNIRS is a novel optical brain imaging technique that uses near-infrared light to monitor oxygen levels at multiple cortical locations [19]. Although with less spatial resolution and limited light penetration ability, studies have found that fNIRS signals are highly correlated with blood-oxygen-level-dependent signals [20]. It can be a promising substitute for fMRI in many particular scenarios, given that it is quiet, nonferromagnetic, and relatively motion tolerant. In addition, the higher resolution of fNIRS can provide more physiological information, such as heart rate variability (HRV) [21,22]. In the field of pain, an increasing number of fNIRS-based investigations have been developed [23,24]. In this study, owing to the noise level requirements of mindfulness meditation, and the use of electronic VR devices, we selected fNIRS as the neuroimaging technique to study the VRB and TMB practices by focusing on the key cortical regions for sensory processing and inhibition.

Methods

Participants

We recruited 40 healthy adult participants in this study (women: 21/40, 52%; age: mean 28 years, SD 4 years). Our exclusion criteria included significant hearing or visual impairment, a history of chronic pain or recent acute pain, significant medical conditions, or current evidence of respiratory distress or asthma. The recruited participants were randomly divided into 2 groups. The first group had interoceptive breathing focusing sessions using an in-house–developed, visual-auditory, 3D VR technology aid (VRB: n=20; women: 11/20, 55%; age: mean 26 years, SD 4 years), whereas the second group had abstract MB (TMB: n=20; women: 10/20, 50%; age: mean 29 years, SD 4 years). This study was approved by the institutional review board of the University of Michigan.

Experiment Protocol

We designed a week-long protocol for both groups. Within the seven days of the protocol, we scheduled each participant for two in-person appointments in the lab on the first and seventh days. From the second to the sixth day, we asked the participants to practice home self-guided exercises following the instructions (Multimedia Appendix 1).

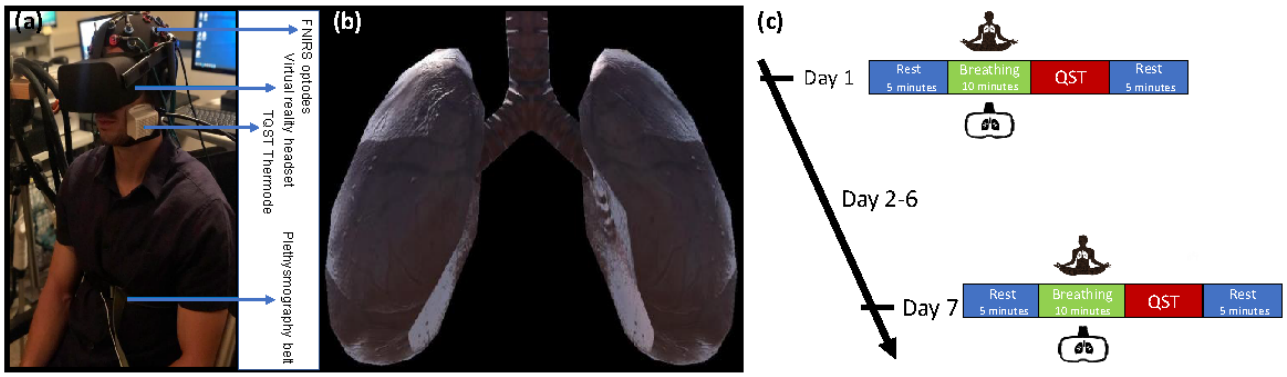
For both in-lab sessions, we asked each participant to complete a set of McGill Pain Questionnaire and the Positive and Negative Affect Schedule (PANAS) questionnaire. Participants were then seated in a dental chair, and we helped them with the comfortable placement of headphones, Oculus Rift virtual imaging equipment (Oculus VR), and a plethysmography belt. Finally, we set up the fNIRS imaging sensors and thermal quantitative sensory thermode, as indicated in Figure 1 (A and B).

As indicated in Figure 1C, we first asked the participants to relax and rest in the dental chair for 5 minutes while we collected the resting-state fNIRS data. The participants then underwent a 10-minute interoceptive breathing awareness practice (TMB or VRB protocol). Participants in the VRB group watched an in-house developed VR display of a 3D lungs image that was synchronized to their inhaling and exhaling cycles in real time via an Oculus Rift device (Oculus VR), as shown in Figure 1B. Meanwhile, they listened to their breathing sounds using headphones. In an early study, Abushakra and Faezipour [25] conceptualized a mobile app with synchronization based on the breath sounds picked by the microphone; however, in practice, their concept could not accurately discern between expiration (breathing out) and inspiration (breathing in) sounds. In our independent method, breathing synchronization was performed using a Braebon plethysmography belt (Great Lakes Neuro Technology). Alternatively, we asked the TMB group participants to *abstractly* imagine their breathing inflating and deflating. We collected fNIRS data during breathing practice for both groups.

Next, we administered 20 trials of the tQST (Medoc Pathway System). We used 20 times repeated measurements and used the averaged temperature thresholds for further analysis to ensure the test-retest. We placed a single unilateral thermode on the left mandibular nerve branch of the trigeminal cranial nerve (V3 division) for each participant. Within each trial of the 20 trials, the thermode temperature controlled by the controlling device increased from a baseline of 30 °C (86 °F) to a maximum temperature of 50 °C (122 °F), with an increase rate of 1 °C per second. We instructed participants to click the button on the mouse at the first detection of pain, as it stopped the temperature from increasing. The thermode temperature then returned to its baseline and gave the subject a 10-second rest period before the next thermal trial. We also collected fNIRS brain data during the tQST session.

Upon completion of all 20 tQST trials, we asked participants to relax for another 5 minutes for a final collection of resting-state fNIRS data. In addition, we asked participants to complete another set of PANAS and McGill pain questionnaires. Following the completion of the first session, we gave participants a sheet of at-home breathing practice prompt and instructed them to read and complete this exercise for 5 minutes three times a day (after waking, midday, and before bed). Finally, we asked participants to repeat the same protocol during their second in-lab visit on day 7.

Figure 1. (A) The experimental setup with all the technologies integrated. (B) The virtual 3D lungs in the Oculus Rift headset from participants' view, which moved in synchronization with their breathing cycles in real time (inhaling and exhaling). (C) The experiment protocol. FNIRS: functional near-infrared spectroscopy; QST: quantitative sensory test; TQST: thermal quantitative sensory test.

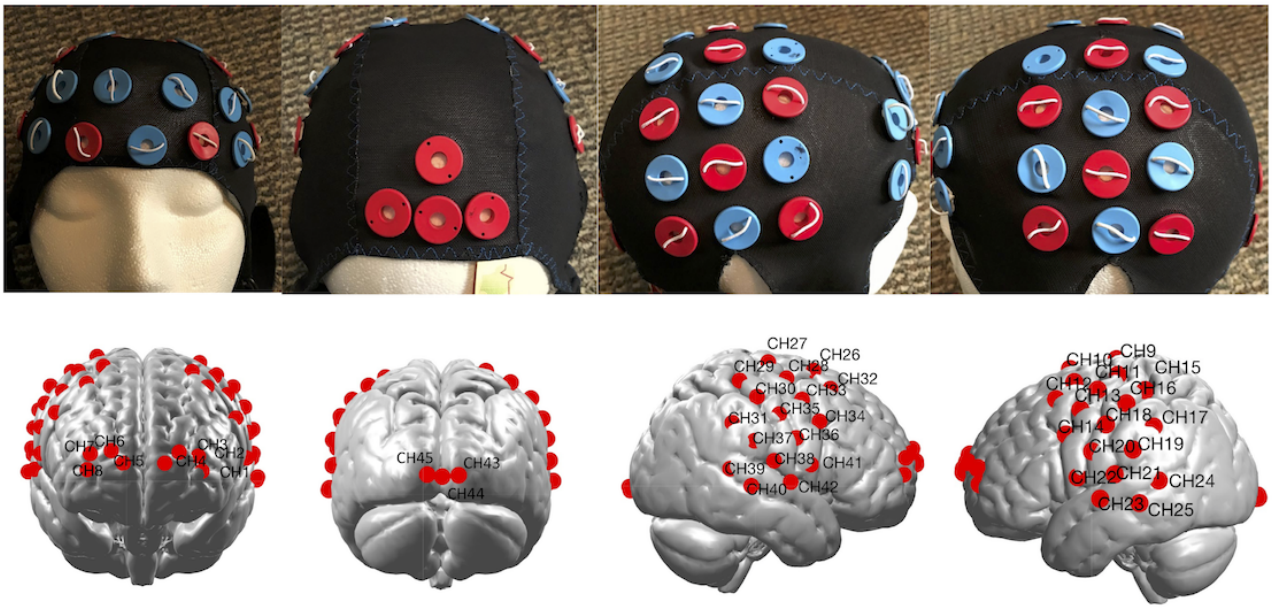


fNIRS Neuroimaging and Probe Localization

We used a TechEN-CW6 fNIRS system (TechEn, Inc) with wavelengths of 690 and 830 nm. The fNIRS cap setup included eight emitters of near-infrared light and 28 detectors spaced 3 cm apart, yielding 45 data channels (CHs) deployed at the bilateral aPFC, premotor cortex (PMC), supplementary motor area (SMA), motor cortex, primary somatosensory cortex (S1), and visual cortex (V1), as indicated in Figure 2. Neuroimaging data were collected at a sampling rate of 25 Hz throughout the entire experiment.

The probe holding cap was established and applied consistently for each participant using the international 10-10 transcranial system positioning [26]. We designed the cap in three sizes—56, 58, and 60, respectively (an example cap is shown in Figure 2), to account for head size variation. In addition, we applied a photogrammetry method to register all optodes and data CHs onto the cortical surface. The detailed method was described in our previous paper [27]. Briefly, we used the Structure Sensor (Occipital Inc) with an iPad (Apple Inc) to capture the 3D photos of the designed caps in three sizes. We then loaded the 3D photo in the MATLAB software (Mathworks) and pinpointed the locations of fNIRS optodes with five fiducial markers (Nasion, Inion, Cz, AR, and AL in the 10-10 system). The derived optodes coordinates were affinely transferred into the Montreal Neurological Institute space using the MATLAB-based AtlasViewerGUI toolbox (Citation). The midpoints between the source and detector (optodes) pairs were used as the coordinates for each CH. Finally, we matched the regions detected by each CH using the estimated center points in the neurosynth.org database. We also estimated the covering range for each CH with a voxel size of 10 mm using the WFU_pick atlas in the XJview toolbox [28].

Figure 2. The designed functional near-infrared spectroscopy cap for functional near-infrared spectroscopy light emitters and detectors with channel (emitter-detector pairs) localization estimation.



HRV Data Estimation

We estimated the HRV data using time-domain methods by calculating the SD of the normal-to-normal parameter using MATLAB software (MathWorks). We calculated the parameters based on the optical density data of 830 wavelengths, bandpass filtered with a cutoff frequency of 0.01-2 Hz. The parameters were then calculated using the formulas also available in the paper by Wang and Huang [29]:

where N is the total number of R peaks and  is the mean of the R-R intervals.

fNIRS Data Analysis

We analyzed fNIRS data using the near-infrared spectroscopy–toolbox [30] in MATLAB software (MathWorks) and a set of customized scripts. In this study, we focused only on oxyhemoglobin (HbO) and not deoxyhemoglobin (HbR). Quantitative analysis indicated that HbR signal changes contributed 16%-22%, whereas HbO signal changes contributed 73%-79% to the total changes measured by fNIRS in cortical hemoglobin concentrations [31].

Brain Activation Analysis During the tQST Session

We applied a generalized linear model with prewhitening and robust least squares [32] to analyze the data collected during the tQST session. Specifically, the raw fNIRS data were first down-sampled to 2 Hz and then converted into HbO and HbR using the modified Beer-Lambert law [33]. We then applied a CH-based generalized linear model regression to each participant's data, assuming a canonical hemodynamic response function model peaking at 6 seconds. The process can be expressed as follows:

$$Wy_{i,j} = Wx\beta_{i,j} + W\varepsilon_{i,j} \quad \textbf{(3)}$$

where W is the whitening matrix, y is the observed HbO data, x is the design matrix (model), β is the regression coefficient, ε is the residual, and i and j separately represent the participant and CH index, respectively.

Group-level analysis was conducted using a linear mixed-effects model based on the regression coefficients derived from the individual-level analysis. The model can be expressed as follows:

$$Y_g = X_g B + Z_g \theta + \varepsilon \textbf{(4)}$$

where Y_g is the regression coefficient obtained from the first level, X_g is the fixed effects term including the modeled brain response at the group level, Z_g is the random effects term counting for between-participant difference, B and θ are the fixed and random effects (coefficients) at the group level, and ε is the residual. Finally, we used a 2-tailed t test to examine the effect of a specific CH:

where Cov_{group} is the covariance of the group-level model.

Brain Connectivity Analysis

To study the brain mechanism during breathing practice, we calculated the associated functional connectivity patterns using the pipelines in the near-infrared spectroscopy–toolbox [30]. The calculation process was described in a previous study [34]. The raw fNIRS data were first down-sampled to 4 Hz. We converted the raw data into HbO and HbR using the modified Beer-Lambert law [33]. We then used bandpass filters to filter the HbO data into two frequency bands: high (0.5-1 Hz) and low (0.01-0.08 Hz) frequency bands. These two frequency bands were selected to avoid the physiological signal bandwidth, including the Mayer wave (0.1 Hz), respiratory (0.3-0.5 Hz), and cardiac (1-1.5 Hz) relevant fluctuations [35]. Next, we calculated the between-CH correlation at the individual level using the robust correlation method in the toolbox [34]. Then, the individual-level correlation coefficient was converted to a Z score using Fisher Z-transform [36]. Finally, a linear mixed-effect model was applied to obtain the group-level connectivity effect. This calculation was implemented on the data collected from both groups during the two lab visits.

Brain Connectivity-Temperature Correlation Analysis

We inspected the relationship between brain connectivity during breathing practice and the temperature threshold measured during the tQST session. We ran an elastic net regression to select the best region-to-region connections for the temperature thresholds. This selection process was performed using the Lasso toolbox in MATLAB (MathWorks). Specifically, we iteratively varied the weight controlling the lasso versus ridge optimization from 0 to 1 in increments of 0.1 to achieve a minimum squared error. With each weight, we applied 10-fold cross-validation with Monte Carlo repetitions (100 times) to guarantee a converged output. Next, we calculated the Pearson correlation coefficient (r) and Spearman rank correlation coefficient (ρ) between the selected connections and temperature thresholds. In addition to the Pearson correlation coefficient, we calculated the Spearman rank correlation coefficients to prevent outliers from affecting the correlation analysis [37].

Results

Clinical Measurement

Our first observation was that the average temperature threshold measured by the tQST increased from 45.4 to 46.0 °C (P=.001) for the TMB group, whereas the threshold increased from 46.5 to 47.1 for the VRB group (P=.02) from in-lab sessions 1 to 2 after one week (Figure 3A). Although the thresholds increased at approximately the same level, we did not find a between-group difference. In addition, Figure 3 (B and C) indicate that participants in the TMB group gained higher serenity scores after the practice (visit 1: P=.01; visit 2: P=.001). However, they felt more tired (visit 1: P=.03; visit 2: P=.01), according to the PANAS questionnaires.

Figure 4 shows the HRV estimation results, where panel A shows the heartbeat signal extracted from the fNIRS signal, and panel B shows the estimated SD of normal-to-normal parameters for different sessions. We did not find significant differences between the TMB and VRB groups, suggesting that there might be no different breathing patterns associated with the two breathing practices.

Figure 3. (A) Temperature thresholds measured in the thermal quantitative sensory test sessions for traditional mindful breathing and virtual reality breathing groups. (B) Serenity score (Positive and Negative Affect Schedule) in pre- and postbreathing practices in the traditional mindful breathing group. (C) Fatigue score (Positive and Negative Affect Schedule) in pre- and postbreathing practices in the traditional mindful breathing group. (D) Serenity score (Positive and Negative Affect Schedule) in pre- and postbreathing practices in the virtual reality breathing group. (E) Fatigue score (Positive and Negative Affect Schedule) in pre- and postbreathing practices in the virtual reality breathing group. TMB: traditional mindful breathing; VRB: virtual reality breathing. The asterisks indicate statistical difference between the scores collected from two groups.

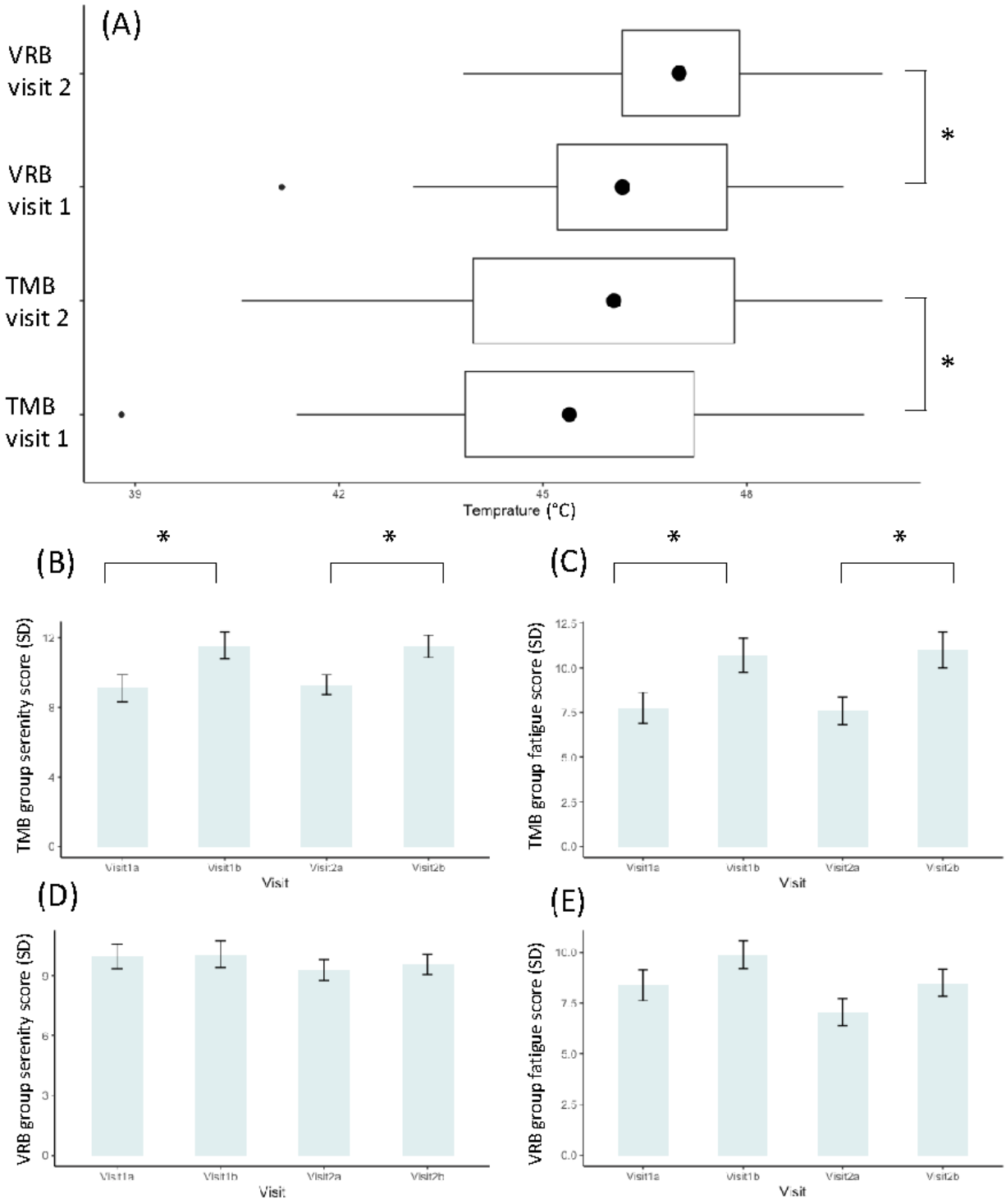
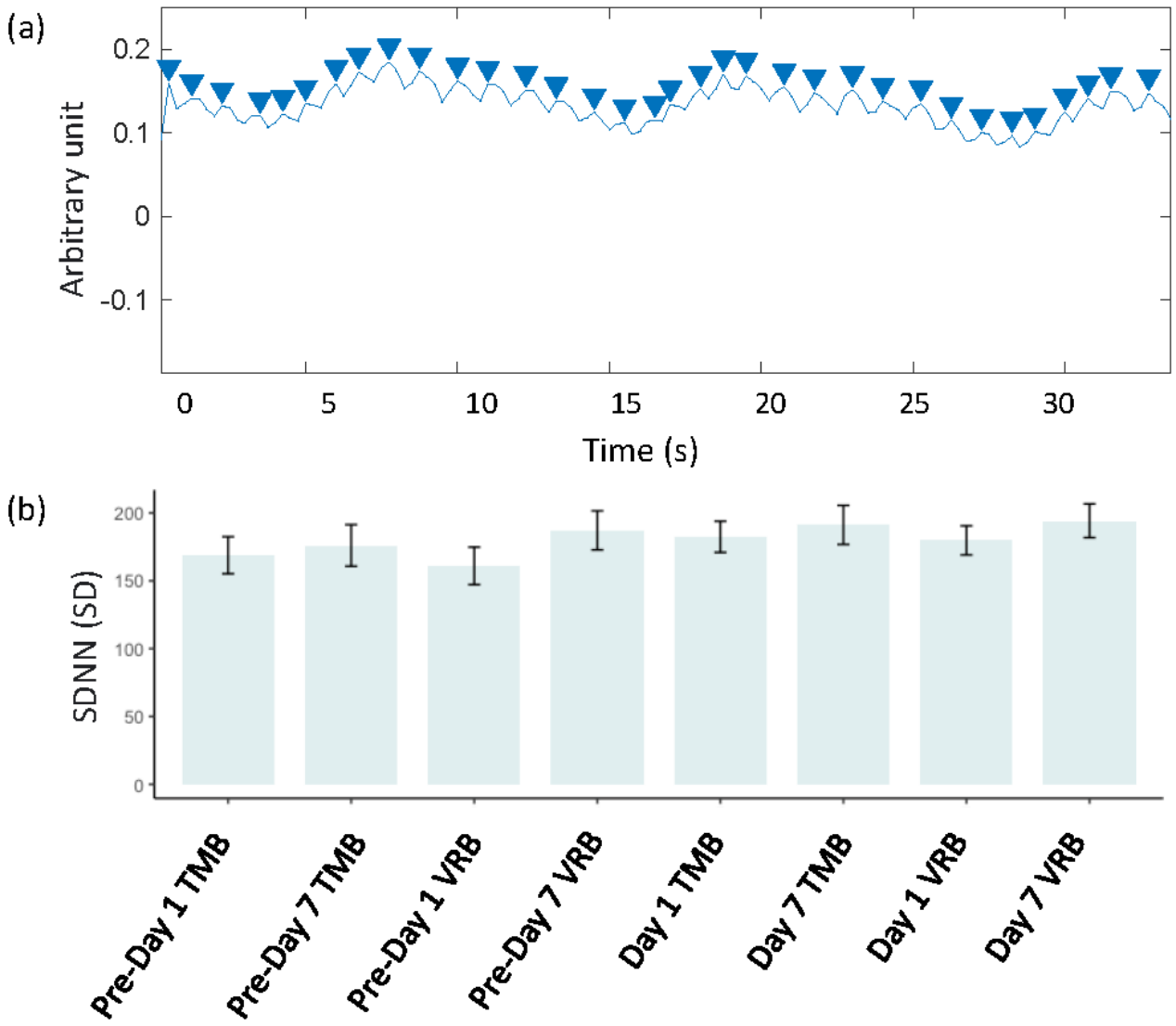


Figure 4. (A) The extracted heartbeat signal from the functional near-infrared spectroscopy signal. (B) The estimated SDNN parameter for the prebreathing resting-state sessions, traditional mindful breathing, and virtual reality breathing practice sessions. SDNN: SD of normal to normal; TMB: traditional mindful breathing; VRB: virtual reality breathing.

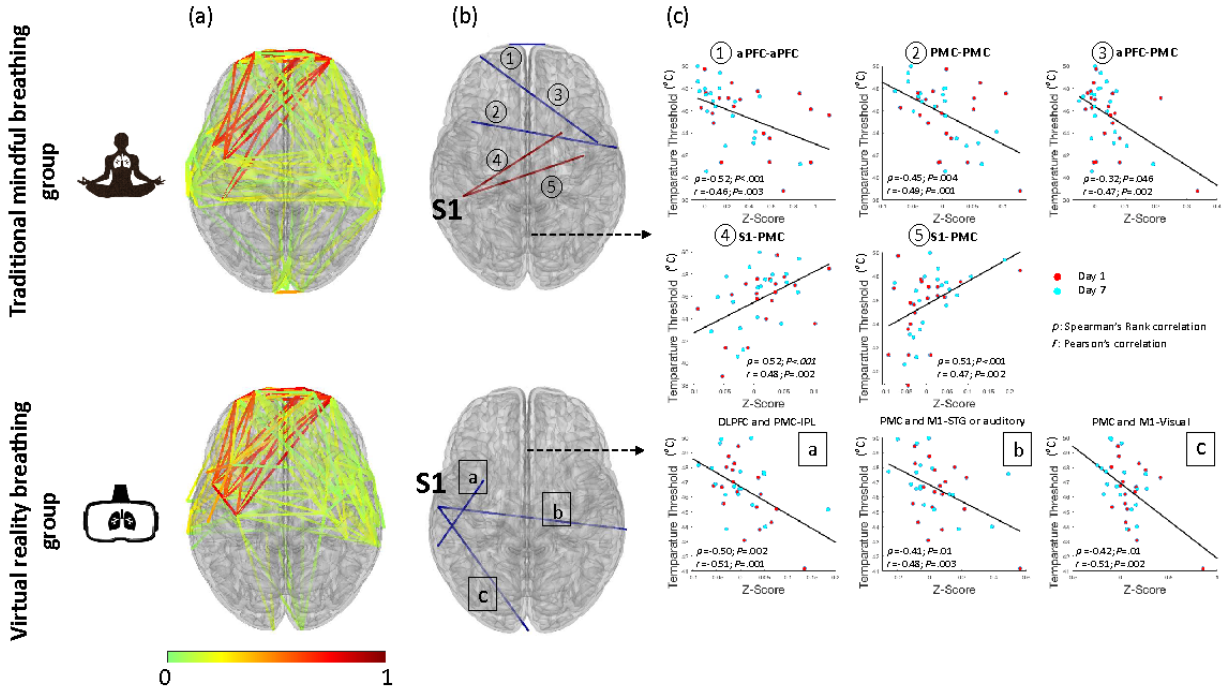


Brain Functional Connectivity

To study and compare the mechanisms of the two breathing practices, we then investigated the brain connectivity during both breathing practices (pre-tQST) and brain activation during the tQST sessions (thermal pain challenge). As shown in Figure 5A, during breathing practices, the TMB group demonstrated a denser functional connectivity pattern (P<.001) among areas such as the aPFC, PMC, SMA, the S1, and the auditory and visual regions than the VRB group.

We further investigated the correlation between brain connectivity and temperature thresholds across participants. As shown in Figure 5 (B and C), the connections within the aPFC (r=−0.46, P=.003; ρ=−0.52, P<.001), between the bilateral PMC (r=−0.49, P=.001; ρ=−0.45, P=.004), between the aPFC and PMC (r=−0.47, P=.002; ρ=0.32, P=.046), and between the S1 and PMC/SMA (r=+0.47,0.48, P=.002; ρ=0.52, 0.51, P<.001) were found to be associated with the temperature thresholds in the TMB group. Whereas in the VRB group, the connections between the PMC/DLPFC and the inferior parietal lobe (r=−0.51, P<.001; ρ=−0.50, P=.002), between the S1 and V1 (r=−0.51, P=.002; ρ=−0.42, P=.01), and between the S1 and superior temporal gyrus (STG; r=−0.48, P=.003; ρ=−0.41, P=.01) were found to be associated with the temperature thresholds in the VRB group.

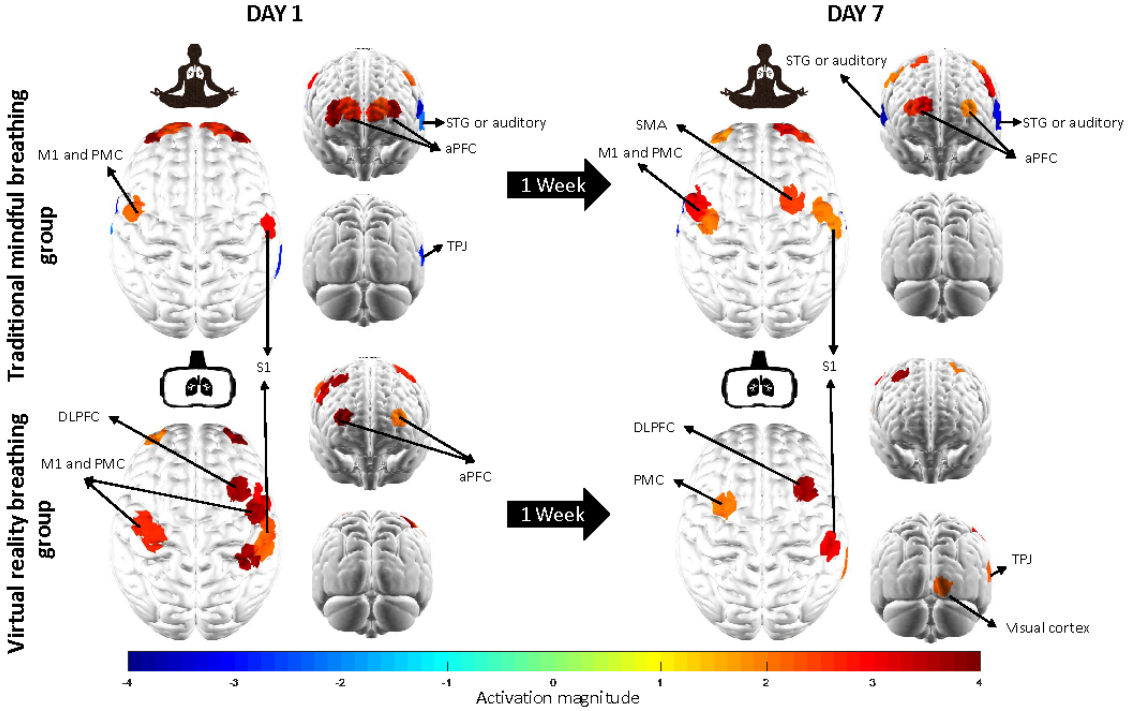
Figure 5. Functional connectivity patterns during the breathing practices. (A) Significant connectivity patterns (P<.001). (B) Connections directly correlated with pain thresholds. (C) Scatter plots of the connectivity-pain threshold relationship. aPFC: anterior prefrontal cortex; DLPFC: dorsolateral prefrontal cortex; IPL: inferior parietal lobe; M1: motor cortex; PMC: premotor cortex; S1: primary somatosensory cortex; STG: superior temporal gyrus.



Brain Activation

Next, we examined cortical activation during the tQST session to study how the brain processes pain after the two types of practices. As indicated in Figure 6, the analysis first confirmed contralateral S1 region activation in both groups at both visits (TMB group visit 1: CH 35, t_{72}=3.0, P=.004; TMB group visit 2: CH 35, t_{72}=2.1, P=.04; VRB group visit 1: CH 35, t_{72}=2.4, P=.02; CH 30, t_{72}=2.4 P=.02; CH 29, t_{72}=4.9, P<.001; VRB group visit 2: CH 30, t_{72}=3.4 P<.001).

Figure 6. Brain activation map for the thermal quantitative sensory test sessions on visit days 1 and 7. aPFC: anterior prefrontal cortex; DLPFC: dorsolateral prefrontal cortex; M1: motor cortex; PMC: premotor cortex; S1: primary somatosensory cortex; SMA: supplementary motor area; STG: superior temporal gyrus; TPJ: temporal-parietal junction.

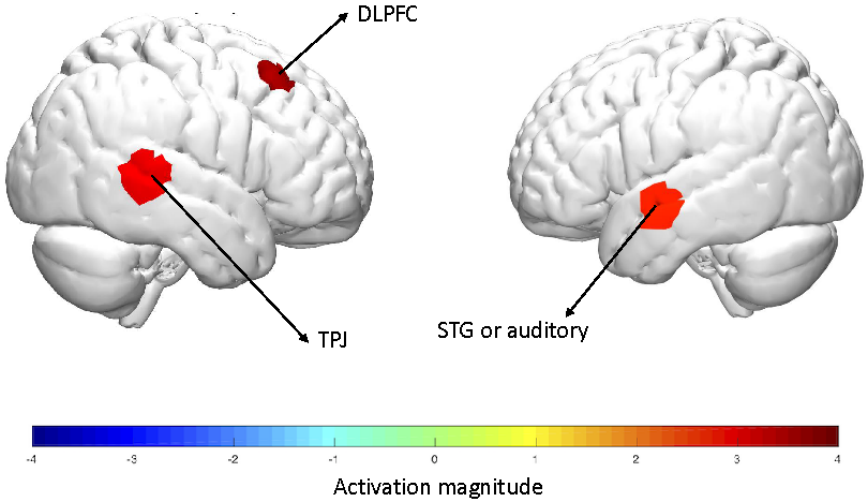


Our second observation was the activation of the aPFC region in both groups, as indicated in Figure 6 (TMB group, visit 1: CH 2-8, t_{72}=4.9 [*P*<.001], 2.4 [*P*=.02], 2.7 [*P*=.008], 2.7 [*P*=.009], 2.6 [*P*=.01], 4.7 [*P*<.001], 3.7 [*P*<.001], visit 2: CH 2, t_{72}=2.0 [*P*=.047], CH 6-7, t_{72}=3.2 [*P*=.002], 2.6 [*P*=.01]; VRB group, visit 1: CH 2, t_{72}=2.1 [*P*=.04], CH 7, t_{72}=4.6 [*P*<.001]).

Next, for the TMB group, we found activations in the PMC: visit 1, CH 13, t_{72}=2.4 (*P*=.02), visit 2: CH 11, t_{72}=2.2 (*P*=.03), CH 13, t_{72}=3.0 (*P*=.004), CH 14, t_{72}=3.4 (*P*=.001), CH 26, t_{72}=2.6 (*P*=.01), CH 33, t_{72}=2.3 (*P*=.03). We also observed deactivations in the bilateral auditory cortices (visit 1: CH 22, t_{72}=−3.7 [*P*<.001], CH 23, t_{72}=−2.3 [*P*=.02]; visit 2: CH 22, t_{72}=−3.2 [*P*=.002], CH 23, t_{72}=−3.6 [*P*<.001], CH 42, t_{72}=−4.0 [*P*<.001]) and right temporal-parietal junction (TPJ; visit 1: CH 39, t_{72}=−2.9 [*P*=.004]). For the VRB group, we observed activations in the PMC/SMA (visit 1: CH 9, t_{72}=2.8 [*P*=.007], CH 11, t_{72}=2.8 [*P*=.007], CH 33, t_{72}=4.3 [*P*<.001], CH 34, t_{72}=3.6 [*P*<.001], visit 2: CH 10, t_{72}=2.3 [*P*=.03], right DLPFC, visit 1: CH 32, t_{72}=7.0 [*P*<.001], visit 2, CH 32, t_{72}=4.6 [*P*<.001]) and V1 (visit 2 CH 43, t_{72}=2.2 [*P*=.03]).

Finally, we performed a contrast analysis between the two groups by combining the two visits, as shown in Figure 7. The results suggested that the VRB group showed greater activation in the right DLPFC (CH 32, t_{72}=5.0 [*P*<.001, false discovery rate (FDR) corrected]), right TPJ (CH 39; t_{72}=3.4 [*P*=.001, FDR corrected]), and left STG regions (CH 23, t_{72}=2.7 [*P*=.048, FDR corrected]). A detailed list of brain activation during the tQST sessions is provided in Multimedia Appendix 2.

Figure 7. Between-group contrast analysis results (virtual reality breathing group [+Red]; traditional mindful breathing group [−Blue]). DLPFC: dorsolateral prefrontal cortex; STG: superior temporal gyrus; TPJ: temporal-parietal junction.



Discussion

Principal Findings

It is commonly assumed that MB exerts its analgesic effect through abstract interoception [5]. Interoception refers to consciously refocusing the mind's attention to the physical sensation of organ function. On the other hand, VR provides participants with an immersive visual-auditory sensory-exteroceptive experience that modulates pain. In this study, we dissected these central analgesic processes by imaging the brains of 2 groups of healthy subjects using fNIRS, exposed to either TMB or a VRB protocol.

Our first finding was that both groups attained a raised pain threshold after one week of breathing practice without significantly different HRV measurements. The results suggested that both TMB and VRB techniques effectively increased pain thresholds in the participants. It is worth noting that the pain-evaluation (tQST) sessions were conducted after the breathing practices, which reproduced their application in the clinical environment, as before a medical/dental procedure. Thus, the analgesic effects were postbreathing effects. The 0.6 °C pain threshold increase in both groups is considered significant, as the human nociceptive sensation thresholds for warmth can be as low as 1.5 °C above the baseline at around 30 °C [38,39]. Nevertheless, the collected PANAS scores suggested that the analgesic mechanisms of the two breathing practices were different. We found that the serenity score increased in the TMB group, although it was accompanied by an increased fatigue score. In contrast, the VRB group showed no significant change in the serenity scores. Serenity is a mental state of being calm, peaceful, and untroubled [40]. This mental state has been shown to increase after meditation [41] and reduce pain [42].

Following the observed temperature threshold changes, we first analyzed the brain activation patterns during the tQST sessions. Our results first confirmed the activation of the contralateral S1 region to noxious thermal stimulation. The S1 has been studied intensively for its critical function in pain and intensity processing [43,44]. As expected, we observed consistent activations in the contralateral (right) S1 region during the tQST sessions in both groups, evoked by the noxious heat stimulation applied to the participants. Interestingly, we noticed a qualitative association between aPFC activation and average pain thresholds across visits and groups, as indicated in Figure 6. Specifically, the TMB group had the lowest pain threshold of 45.4 °C on day 1, accompanied by significant activation in seven out of eight data CHs, whereas the VRB group had the highest pain threshold of 47.1 °C on day 7, with aPFC activation. As one of the cortical executive regions, the aPFC plays an essential role in pain appraisal [45]. Structurally, previous studies reported that chronic pain led to gray matter loss [46-49], whereas meditation increased gray matter volume in the aPFC region [50]. Functionally, studies have found greater aPFC activation with meditation practice [51]. In the context of pain, aPFC activation was correlated with the unpleasantness aspect of pain [52]. However, during the pain process, fMRI imaging revealed reduced functional brain activation in the aPFC region [11,53]. In this study, the trend of less activation was associated with an increased pain threshold, suggesting that less appraisal of pain was induced by the week-long breathing practice. We also found right TPJ, right DLPFC, and V1 activation in the VRB group, whereas TPJ and STG (auditory) deactivation were observed in the TMB group. The TPJ region serves as a hub for integrating multisensory body-related information, including touch, visual, and auditory inputs [54]. In this study, different from the proposed distraction mechanisms [16,18], we observed a postbreathing effect, in which the participant's TPJ and visual activation were present under the pain condition, even without the VR experience. The right DLPFC possibly provided a modulatory effect on pain to achieve a higher pain threshold. The DLPFC is a critical region that directs attention away from pain [55,56] and inhibits both affective and sensory aspects of pain in the brain [57], as revealed by previous fMRI studies.

To further study the brain mechanisms during the two types of breathing practices, we then analyzed their associated functional connectivity. The TMB group demonstrated a closer working relationship among areas, including the aPFC, PMC, and S1 regions. In contrast, we observed fewer connections to the visual and auditory regions in the VRB group. The immersive 3D, sensory-exteroceptive, virtual experience reinforced the participants' cortical audio-visual activations, thus depriving the S1 processing of the ascending pain inputs. To study whether functional connectivity is related to pain sensitivity, we examined the correlations between the observed connectivity strength and the pain thresholds measured during the tQST session. Interestingly, we found that the connections among the aPFC, PMC, and S1 regions in the TMB group were associated with pain thresholds across visits. In contrast, in the VRB group, we found that the functional connections among the auditory/visual regions, PMC, inferior parietal lobe, and S1 were associated with temperature thresholds. A previous resting-state fMRI study found lower pain sensitivity in meditators with decoupled executive and pain-relevant brain regions [11]. Similarly, in our study, we found lower pain sensitivity in the TMB group participants with less connectivity between the aPFC and PMC-S1 regions. For the VRB group, lower pain sensitivity was observed in participants with decoupled visual-auditory-DLPFC and PMC regions.

On the basis of these findings, we propose two possible mechanisms for the TMB and VRB practice—in the TMB group, the aPFC modulated attention [58-60] and contextual evaluation of internal sensory events [61-63]. The PMC, as part of the mirror neuron system [64], increased its functional connection with the S1 to facilitate the sensory-interoceptive processing of breathing. This process inhibited the S1 in sensory-discriminative pain processing during later tQST sessions. In contrast, VR induced an immersive 3D exteroception with augmented visual-auditory cortical activations to diminish functional connection with the S1, consequently weakening the pain processing function of the S1.

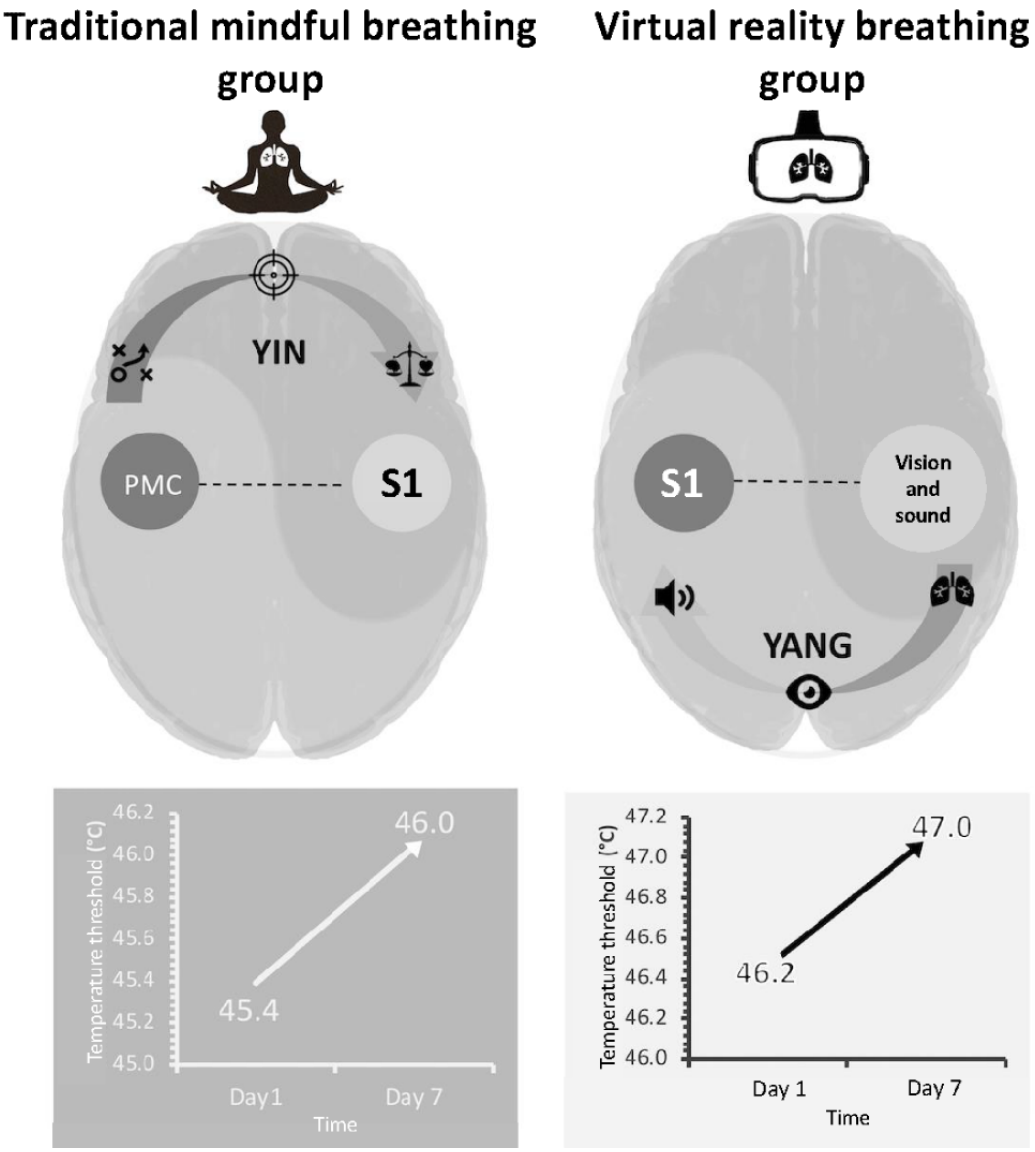
Limitations

There were some caveats in this study. First, we asked the participants to mimic the VR experience for at-home practices between day 1 and day 7 lab visits. However, the at-home practice was not as immersive as the in-lab VR practice (owing to the noncompletely portable apparatus of the technologies), which might have dampened its effect. We will use a mobile VR device that works with smartphones in our future studies to address this issue. Next, instead of a nonintervention control group, we used the TMB group as the active control group. Although we found different brain activation and connectivity patterns, in contrast to the VRB group, we still need to compare both groups with a nonintervention control group and evaluate the effects of both the TMB and VRB breathing practices in the future.

Conclusions

In conclusion, as shown in Figure 8, our study suggested two distinct analgesic mechanisms of VRB and TMB practices, following the concept of dualism, Yin and Yang, in ancient Asian philosophy [65]. On the Yin-side, the aPFC, activated by the TMB practice, modulated the PMC to maintain an uninterrupted sensory-interoception via the S1, which prevailed over its sensory-discriminative processing of the ascending pain. On the Yang-side, the VRB practice brought an immersive 3D sensory-exteroceptive VR experience via augmentation of cortical visual-auditory activations that overrode the pain processing function to raise the pain threshold.

Figure 8. Schematic plots of the mechanisms under traditional mindful breathing and virtual reality breathing accompanied by the temperature thresholds measured by the thermal quantitative sensory test in this study. S1: primary somatosensory cortex; PMC: premotor cortex.



Acknowledgments

The authors would like to thank Sumer Panesar and Nathan Wigington for their help with data collection and experiment preparation. The authors would also like to thank Jacqueline Dobson for her help with proofreading the manuscript.

Conflicts of Interest

The content described in this study was developed at the University of Michigan and disclosed to the University of Michigan Office of Technology Transfer. All intellectual property rights, including but not limited to patents or patent applications, trademarks, and copyright of software, algorithms, reports, displays, and visualizations, are owned by the Regents of the University of Michigan.

Multimedia Appendix 1
Breathing practice instructions.
[DOCX File , 199 KB - jmir_v23i10e27298_app1.docx]

Multimedia Appendix 2
Brain activation during the thermal quantitative sensory test session.
[DOCX File , 31 KB - jmir_v23i10e27298_app2.docx]

References

1. Tracey I. Can neuroimaging studies identify pain endophenotypes in humans? Nat Rev Neurol 2011 Mar;7(3):173-181. [doi: 10.1038/nrneurol.2011.4] [Medline: 21304481]
2. Talbot K, Madden V, Jones S, Moseley G. The sensory and affective components of pain: are they differentially modifiable dimensions or inseparable aspects of a unitary experience? A systematic review. Br J Anaesth 2019 Aug;123(2):e263-e272 [FREE Full text] [doi: 10.1016/j.bja.2019.03.033] [Medline: 31053232]
3. Dhanani NM, Caruso TJ, Carinci AJ. Complementary and alternative medicine for pain: an evidence-based review. Curr Pain Headache Rep 2011 Feb;15(1):39-46. [doi: 10.1007/s11916-010-0158-y] [Medline: 21063917]
4. Reuben DB, Alvanzo AA, Ashikaga T, Bogat GA, Callahan CM, Ruffing V, et al. National institutes of health pathways to prevention workshop: the role of opioids in the treatment of chronic pain. Ann Intern Med 2015 Feb 17;162(4):295-300. [doi: 10.7326/M14-2775] [Medline: 25581341]
5. Zeidan F, Vago DR. Mindfulness meditation-based pain relief: a mechanistic account. Ann N Y Acad Sci 2016 Jun;1373(1):114-127 [FREE Full text] [doi: 10.1111/nyas.13153] [Medline: 27398643]
6. Rainville P, Feine JS, Bushnell MC, Duncan GH. A psychophysical comparison of sensory and affective responses to four modalities of experimental pain. Somatosens Mot Res 1992;9(4):265-277. [doi: 10.3109/08990229209144776] [Medline: 1492527]
7. Coghill RC, McHaffie JG, Yen Y. Neural correlates of interindividual differences in the subjective experience of pain. Proc Natl Acad Sci U S A 2003 Jul 08;100(14):8538-8542 [FREE Full text] [doi: 10.1073/pnas.1430684100] [Medline: 12824463]
8. Tracey I, Mantyh PW. The cerebral signature for pain perception and its modulation. Neuron 2007 Aug 02;55(3):377-391 [FREE Full text] [doi: 10.1016/j.neuron.2007.07.012] [Medline: 17678852]
9. Oshiro Y, Quevedo AS, McHaffie JG, Kraft RA, Coghill RC. Brain mechanisms supporting discrimination of sensory features of pain: a new model. J Neurosci 2009 Nov 25;29(47):14924-14931 [FREE Full text] [doi: 10.1523/JNEUROSCI.5538-08.2009] [Medline: 19940188]
10. Lobanov O, Zeidan F, McHaffie J, Kraft R, Coghill R. From cue to meaning: brain mechanisms supporting the construction of expectations of pain. Pain 2014 Jan;155(1):129-136 [FREE Full text] [doi: 10.1016/j.pain.2013.09.014] [Medline: 24055334]
11. Grant JA, Courtemanche J, Rainville P. A non-elaborative mental stance and decoupling of executive and pain-related cortices predicts low pain sensitivity in Zen meditators. Pain 2011 Jan;152(1):150-156. [doi: 10.1016/j.pain.2010.10.006] [Medline: 21055874]
12. Zeidan F, Emerson NM, Farris SR, Ray JN, Jung Y, McHaffie JG, et al. Mindfulness meditation-based pain relief employs different neural mechanisms than placebo and sham mindfulness meditation-induced analgesia. J Neurosci 2015 Nov 18;35(46):15307-15325 [FREE Full text] [doi: 10.1523/JNEUROSCI.2542-15.2015] [Medline: 26586819]
13. Creswell JD, Taren AA, Lindsay EK, Greco CM, Gianaros PJ, Fairgrieve A, et al. Alterations in resting-state functional connectivity link mindfulness meditation with reduced interleukin-6: a randomized controlled trial. Biol Psychiatry 2016 Jul 01;80(1):53-61. [doi: 10.1016/j.biopsych.2016.01.008] [Medline: 27021514]
14. Cipresso P, Giglioli I, Raya M, Riva G. The past, present, and future of virtual and augmented reality research: a network and cluster analysis of the literature. Front Psychol 2018 Nov 6;9:2086 [FREE Full text] [doi: 10.3389/fpsyg.2018.02086] [Medline: 30459681]
15. Li A, Montaño Z, Chen VJ, Gold JI. Virtual reality and pain management: current trends and future directions. Pain Manag 2011 Mar;1(2):147-157 [FREE Full text] [doi: 10.2217/pmt.10.15] [Medline: 21779307]
16. Hoffman HG, Chambers GT, Meyer WJ, Arceneaux LL, Russell WJ, Seibel EJ, et al. Virtual reality as an adjunctive non-pharmacologic analgesic for acute burn pain during medical procedures. Ann Behav Med 2011 Apr;41(2):183-191 [FREE Full text] [doi: 10.1007/s12160-010-9248-7] [Medline: 21264690]
17. Ansado J, Chasen C, Bouchard S, Northoff G. How brain imaging provides predictive biomarkers for therapeutic success in the context of virtual reality cognitive training. Neurosci Biobehav Rev 2021 Jan;120:583-594. [doi: 10.1016/j.neubiorev.2020.05.018] [Medline: 32533997]
18. Gupta A, Scott K, Dukewich M. Innovative technology using virtual reality in the treatment of pain: does it reduce pain via distraction, or is there more to it? Pain Med 2018 Jan 01;19(1):151-159. [doi: 10.1093/pm/pnx109] [Medline: 29025113]
19. Ferrari M, Quaresima V. A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. Neuroimage 2012 Nov 01;63(2):921-935. [doi: 10.1016/j.neuroimage.2012.03.049] [Medline: 22510258]
20. Cui X, Bray S, Bryant DM, Glover GH, Reiss AL. A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. Neuroimage 2011 Feb 14;54(4):2808-2821 [FREE Full text] [doi: 10.1016/j.neuroimage.2010.10.069] [Medline: 21047559]
21. Perdue KL, Westerlund A, McCormick SA, Nelson CA. Extraction of heart rate from functional near-infrared spectroscopy in infants. J Biomed Opt 2014 Jun;19(6):067010 [FREE Full text] [doi: 10.1117/1.JBO.19.6.067010] [Medline: 24972361]
22. Hakimi N, Setarehdan SK. Stress assessment by means of heart rate derived from functional near-infrared spectroscopy. J Biomed Opt 2018 Nov;23(11):1-12 [FREE Full text] [doi: 10.1117/1.JBO.23.11.115001] [Medline: 30392197]

23. Hu X, Nascimento T, DaSilva A. Shedding light on pain for the clinic: a comprehensive review of using functional near-infrared spectroscopy to monitor its process in the brain. Pain 2021 Apr 05 (forthcoming). [doi: 10.1097/j.pain.0000000000002293] [Medline: 33990114]
24. Karunakaran KD, Peng K, Berry D, Green S, Labadie R, Kussman B, et al. NIRS measures in pain and analgesia: fundamentals, features, and function. Neurosci Biobehav Rev 2021 Jan;120:335-353. [doi: 10.1016/j.neubiorev.2020.10.023] [Medline: 33159918]
25. Abushakra A, Faezipour M. Augmenting breath regulation using a mobile driven virtual reality therapy framework. IEEE J Biomed Health Inform 2014 May;18(3):746-752. [doi: 10.1109/JBHI.2013.2281195] [Medline: 24043409]
26. Jurcak V, Tsuzuki D, Dan I. 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. Neuroimage 2007 Feb 15;34(4):1600-1611. [doi: 10.1016/j.neuroimage.2006.09.024] [Medline: 17207640]
27. Hu X, Wagley N, Rioboo AT, DaSilva AF, Kovelman I. Photogrammetry-based stereoscopic optode registration method for functional near-infrared spectroscopy. J Biomed Opt 2020 Sep;25(9):095001 [FREE Full text] [doi: 10.1117/1.JBO.25.9.095001] [Medline: 32880124]
28. xjView: a viewing program for SPM. xjView. URL: https://www.alivelearn.net/xjview/ [accessed 2021-09-03]
29. Wang H, Huang S. SDNN/RMSSD as a surrogate for LF/HF: a revised investigation. Model Simul Eng 2012;2012:1-8. [doi: 10.1155/2012/931943]
30. Santosa H, Zhai X, Fishburn F, Huppert T. The NIRS Brain AnalyzIR Toolbox. Algorithms 2018 May 16;11(5):73. [doi: 10.3390/a11050073]
31. Gagnon L, Yücel MA, Dehaes M, Cooper RJ, Perdue KL, Selb J, et al. Quantification of the cortical contribution to the NIRS signal over the motor cortex using concurrent NIRS-fMRI measurements. Neuroimage 2012 Feb 15;59(4):3933-3940 [FREE Full text] [doi: 10.1016/j.neuroimage.2011.10.054] [Medline: 22036999]
32. Barker JW, Aarabi A, Huppert TJ. Autoregressive model based algorithm for correcting motion and serially correlated errors in fNIRS. Biomed Opt Express 2013 Jul 17;4(8):1366-1379 [FREE Full text] [doi: 10.1364/BOE.4.001366] [Medline: 24009999]
33. Jöbsis FF. Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. Science 1977 Dec 23;198(4323):1264-1267. [doi: 10.1126/science.929199] [Medline: 929199]
34. Santosa H, Aarabi A, Perlman SB, Huppert TJ. Characterization and correction of the false-discovery rates in resting state connectivity using functional near-infrared spectroscopy. J Biomed Opt 2017 May 01;22(5):55002 [FREE Full text] [doi: 10.1117/1.JBO.22.5.055002] [Medline: 28492852]
35. Pinti P, Scholkmann F, Hamilton A, Burgess P, Tachtsidis I. Current status and issues regarding pre-processing of fNIRS neuroimaging data: an investigation of diverse signal filtering methods within a general linear model framework. Front Hum Neurosci 2019 Jan 11;12:505 [FREE Full text] [doi: 10.3389/fnhum.2018.00505] [Medline: 30687038]
36. Fisher RA. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika 1915 May;10(4):507-521. [doi: 10.2307/2331838]
37. Mukaka M. Statistics corner: a guide to appropriate use of correlation coefficient in medical research. Malawi Med J 2012 Sep;24(3):69-71 [FREE Full text] [Medline: 23638278]
38. Lautenbacher S, Strian F. Sex differences in pain and thermal sensitivity: the role of body size. Percept Psychophys 1991 Aug;50(2):179-183. [doi: 10.3758/bf03212218] [Medline: 1945739]
39. Green BG, Akirav C. Threshold and rate sensitivity of low-threshold thermal nociception. Eur J Neurosci 2010 May;31(9):1637-1645 [FREE Full text] [doi: 10.1111/j.1460-9568.2010.07201.x] [Medline: 20525076]
40. Watson D, Clark L. The PANAS-X: manual for the positive and negative affect schedule - expanded form. Ames: The University of Iowa 1994:1-99. [doi: 10.17077/48vt-m4t2]
41. van Hooff M, Baas M. Recovering by means of meditation: the role of recovery experiences and intrinsic motivation. Appl Psychol 2012 Jan 20;62(2):185-210 [FREE Full text] [doi: 10.1111/j.1464-0597.2011.00481.x]
42. Finan PH, Garland EL. The role of positive affect in pain and its treatment. Clin J Pain 2015 Feb;31(2):177-187 [FREE Full text] [doi: 10.1097/AJP.0000000000000092] [Medline: 24751543]
43. Apkarian AV, Bushnell MC, Treede R, Zubieta J. Human brain mechanisms of pain perception and regulation in health and disease. Eur J Pain 2005 Aug;9(4):463-484. [doi: 10.1016/j.ejpain.2004.11.001] [Medline: 15979027]
44. Bushnell MC, Duncan GH, Hofbauer RK, Ha B, Chen JI, Carrier B. Pain perception: is there a role for primary somatosensory cortex? Proc Natl Acad Sci U S A 1999 Jul 06;96(14):7705-7709 [FREE Full text] [doi: 10.1073/pnas.96.14.7705] [Medline: 10393884]
45. Peng K, Steele SC, Becerra L, Borsook D. Brodmann area 10: collating, integrating and high level processing of nociception and pain. Prog Neurobiol 2018 Feb;161:1-22 [FREE Full text] [doi: 10.1016/j.pneurobio.2017.11.004] [Medline: 29199137]
46. Elsenbruch S, Schmid J, Kullmann JS, Kattoor J, Theysohn N, Forsting M, et al. Visceral sensitivity correlates with decreased regional gray matter volume in healthy volunteers: a voxel-based morphometry study. Pain 2014 Feb;155(2):244-249. [doi: 10.1016/j.pain.2013.09.027] [Medline: 24099953]
47. Yuan C, Shi H, Pan P, Dai Z, Zhong J, Ma H, et al. Gray matter abnormalities associated with chronic back pain: a meta-analysis of voxel-based morphometric studies. Clin J Pain 2017 Nov;33(11):983-990. [doi: 10.1097/AJP.0000000000000489] [Medline: 28234752]

48. Krause T, Asseyer S, Taskin B, Flöel A, Witte AV, Mueller K, et al. The cortical signature of central poststroke pain: gray matter decreases in somatosensory, insular, and prefrontal cortices. Cereb Cortex 2016 Jan;26(1):80-88. [doi: 10.1093/cercor/bhu177] [Medline: 25129889]
49. Obermann M, Rodriguez-Raecke R, Naegel S, Holle D, Mueller D, Yoon M, et al. Gray matter volume reduction reflects chronic pain in trigeminal neuralgia. Neuroimage 2013 Jul 01;74:352-358. [doi: 10.1016/j.neuroimage.2013.02.029] [Medline: 23485849]
50. Babu MG, Kadavigere R, Koteshwara P, Sathian B, Rai KS. Rajyoga meditation induces grey matter volume changes in regions that process reward and happiness. Sci Rep 2020 Sep 30;10(1):16177 [FREE Full text] [doi: 10.1038/s41598-020-73221-x] [Medline: 32999361]
51. Miyashiro S, Yamada Y, Muta T, Ishikawa H, Abe T, Hori M, et al. Activation of the orbitofrontal cortex by both meditation and exercise: a near-infrared spectroscopy study. PLoS One 2021 Feb 23;16(2):e0247685 [FREE Full text] [doi: 10.1371/journal.pone.0247685] [Medline: 33621250]
52. Zeidan F, Martucci KT, Kraft RA, Gordon NS, McHaffie JG, Coghill RC. Brain mechanisms supporting the modulation of pain by mindfulness meditation. J Neurosci 2011 Apr 6;31(14):5540-5548 [FREE Full text] [doi: 10.1523/JNEUROSCI.5791-10.2011] [Medline: 21471390]
53. Grant JA, Rainville P. Pain sensitivity and analgesic effects of mindful states in Zen meditators: a cross-sectional study. Psychosom Med 2009 Jan;71(1):106-114. [doi: 10.1097/PSY.0b013e31818f52ee] [Medline: 19073756]
54. Calvert GA, Campbell R, Brammer MJ. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr Biol 2000 Jun 01;10(11):649-657 [FREE Full text] [doi: 10.1016/s0960-9822(00)00513-3] [Medline: 10837246]
55. Pariente J, White P, Frackowiak RS, Lewith G. Expectancy and belief modulate the neuronal substrates of pain treated by acupuncture. Neuroimage 2005 May 01;25(4):1161-1167. [doi: 10.1016/j.neuroimage.2005.01.016] [Medline: 15850733]
56. Peyron R, Laurent B, García-Larrea L. Functional imaging of brain responses to pain. A review and meta-analysis (2000). Neurophysiol Clin 2000 Oct;30(5):263-288. [doi: 10.1016/s0987-7053(00)00227-6] [Medline: 11126640]
57. Lorenz J, Minoshima S, Casey KL. Keeping pain out of mind: the role of the dorsolateral prefrontal cortex in pain modulation. Brain 2003 May;126(Pt 5):1079-1091. [doi: 10.1093/brain/awg102] [Medline: 12690048]
58. Kuusinen V, Cesnaite E, Peräkylä J, Ogawa KH, Hartikainen KM. Orbitofrontal lesion alters brain dynamics of emotion-attention and emotion-cognitive control interaction in humans. Front Hum Neurosci 2018 Nov 1;12:437 [FREE Full text] [doi: 10.3389/fnhum.2018.00437] [Medline: 30443211]
59. Hunt LT, Malalasekera WM, de Berker AO, Miranda B, Farmer SF, Behrens TE, et al. Triple dissociation of attention and decision computations across prefrontal cortex. Nat Neurosci 2018 Oct;21(10):1471-1481 [FREE Full text] [doi: 10.1038/s41593-018-0239-5] [Medline: 30258238]
60. Toplak ME, Jain U, Tannock R. Executive and motivational processes in adolescents with Attention-Deficit-Hyperactivity Disorder (ADHD). Behav Brain Funct 2005 Jun 27;1(1):8 [FREE Full text] [doi: 10.1186/1744-9081-1-8] [Medline: 15982413]
61. Eippert F, Veit R, Weiskopf N, Erb M, Birbaumer N, Anders S. Regulation of emotional responses elicited by threat-related stimuli. Hum Brain Mapp 2007 May;28(5):409-423 [FREE Full text] [doi: 10.1002/hbm.20291] [Medline: 17133391]
62. O'Doherty J, Kringelbach ML, Rolls ET, Hornak J, Andrews C. Abstract reward and punishment representations in the human orbitofrontal cortex. Nat Neurosci 2001 Jan;4(1):95-102. [doi: 10.1038/82959] [Medline: 11135651]
63. Peters J, Büchel C. Neural representations of subjective reward value. Behav Brain Res 2010 Dec 01;213(2):135-141. [doi: 10.1016/j.bbr.2010.04.031] [Medline: 20420859]
64. Rizzolatti G, Sinigaglia C. The mirror mechanism: a basic principle of brain function. Nat Rev Neurosci 2016 Dec;17(12):757-765. [doi: 10.1038/nrn.2016.135] [Medline: 27761004]
65. Religions in the Modern World Traditions and Transformations. Oxfordshire England: Routledge; 2016.

Abbreviations

aPFC: anterior prefrontal cortex
CH: channel
DLPFC: dorsolateral prefrontal cortex
FDR: false discovery rate
fNIRS: functional near-infrared spectroscopy
HbO: oxyhemoglobin
HbR: deoxyhemoglobin
HRV: heart rate variability
MB: mindful breathing
PANAS: Positive and Negative Affect Schedule
PFC: prefrontal cortex
PMC: premotor cortex

S1: primary somatosensory cortex
SMA: supplementary motor area
STG: superior temporal gyrus
TMB: traditional mindful breathing
TPJ: temporal-parietal junction
tQST: thermal quantitative sensory test
V1: visual cortex
VR: virtual reality
VRB: virtual reality breathing

Edited by R Kukafka; submitted 20.01.21; peer-reviewed by J Koudys, N Khalili-Mahani; comments to author 31.05.21; revised version received 24.07.21; accepted 27.07.21; published 12.10.21.

Please cite as:
Hu XS, Beard K, Sherbel MC, Nascimento TD, Petty S, Pantzlaff E, Schwitzer D, Kaciroti N, Maslowski E, Ashman LM, Feinberg SE, DaSilva AF
Brain Mechanisms of Virtual Reality Breathing Versus Traditional Mindful Breathing in Pain Modulation: Observational Functional Near-infrared Spectroscopy Study
J Med Internet Res 2021;23(10):e27298
URL: https://www.jmir.org/2021/10/e27298
doi:10.2196/27298
PMID:34636731

Original Paper

Effects of Virtual Reality Training on Upper Limb Function and Balance in Stroke Patients: Systematic Review and Meta-Meta-Analysis

Jinlong Wu[1*], MA; Aihua Zeng[2*], MA; Ziyan Chen[1], MA; Ye Wei[3], MA; Kunlun Huang[1], MA; Jiafeng Chen[1], MA; Zhanbing Ren[1], PhD

[1]Department of Physical Education, Shenzhen University, Shenzhen, China

[2]Shanwei Polytech, Shanwei, China

[3]Nanshan District Culture, Radio, Television, Tourism and Sports Bureau, Shenzhen, China

[*]these authors contributed equally

Corresponding Author:
Zhanbing Ren, PhD
Department of Physical Education
Shenzhen University
No 3588 Nanhai Road
Shenzhen, 518061
China
Phone: 86 13318881919
Email: rzb@szu.edu.cn

Abstract

Background: Virtual reality (VR) training is a promising intervention strategy that has been utilized in health care fields like stroke rehabilitation and psychotherapy. Current studies suggest that VR training is effective in improving the locomotor ability of stroke patients.

Objective: This is the first meta-meta-analysis of the effects of VR on motor function in stroke patients. This study aimed to systematically summarize and quantify the present meta-analyses results of VR training and produce high-quality meta-meta-analysis results to obtain a more accurate prediction.

Methods: We searched 4 online databases (Web of Science, Scopus, PubMed, and Chinese National Knowledge Infrastructure) for meta-analysis studies. After accounting for overlap, 10 studies (accounting for almost 550 stroke patients) were obtained. Based on the meta-meta-analysis of these patients, this study quantified the impact of VR training on stroke patients' motor performance, mainly including upper limb function, balance, and walking ability. We combined the effects under the random effect model and pooled the estimates as standardized mean differences (SMD).

Results: The results of the meta-meta-analysis showed that VR training effectively improves upper limb function (SMD 4.606, 95% CI 2.733-6.479, P<.05) and balance (SMD 2.101, 95% CI 0.202-4.000, P<.05) of stroke patients. However, the results showed considerable heterogeneity and thus, may need to be treated with caution. Due to the limited research, a meta-meta-analysis of walking ability was not performed.

Conclusions: These findings represent a comprehensive body of high-quality evidence that VR training is more effective at improving upper limb function and balance of stroke patients.

(J Med Internet Res 2021;23(10):e31051) doi:10.2196/31051

KEYWORDS

virtual reality; stroke; motor function; review; meta-meta-analysis; mental health; motor skills; rehabilitation; digital intervention; health care; stroke patients

Introduction

Stroke is the most common cause of chronic physical disabilities, such as dyskinesia. Most stroke patients have dyskinesia, which causes different degrees of impediment to upper limb function, walking ability, and balance. Because of limited exercise ability, these patients cannot participate in daily activities; thus, their quality of life is reduced [1,2]. Stroke rehabilitation mainly aims to help patients return to society and work [3]. Therefore, it is important to comprehensively understand the severity of stroke; improve treatment methods; reduce the incidence, disability, and mortality of stroke; and find safe and effective treatment for stroke patients.

In the past decade, virtual reality (VR), as a means of neurological rehabilitation for stroke, has gradually become popular in the field of rehabilitation because of the continuous improvement in virtual systems and the substantial reduction in cost of virtual equipment [4]. VR technology is a system that can simulate the environment, scene, and activity in real time and allow users to interact through multiple sensory modalities. The system can be combined with a treadmill, bionic gloves, or robots to provide better feedback for users [5]. Moreover, in the virtual rehabilitation scene created by VR technology, the content, duration, and intensity of the exercise can be manipulated, and even timely feedback can be obtained so as to provide users sufficient and personalized exercise [6]. The following VR-related technologies are widely used in the treatment of stroke patients: an innovative exoskeleton, VR telerehabilitation system, IREX immersion VR systems, Xbox Kinect, keyboard with VR, VR combined with gloves, Nintendo Wii, and virtual surfaces [6]. Compared with traditional rehabilitation, the main advantage of VR training is that stroke patients can think of it as an exciting game rather than as treatment; VR training can help users focus their attention completely on the task, thus improving motivation and treatment compliance, which can be of great benefit in recovering from poststroke trauma [7].

In the last 20 years, a large number of studies have confirmed that VR training has certain advantages in improving the condition of people with dyskinesia [8,9]. Among these studies, nearly 300 experimental and more than 60 meta-analysis studies of stroke patients have been published in international journals. However, between the different meta-analysis studies, there is an inconsistency in the effect size of VR training to improve the different exercise abilities of stroke patients. The purpose of this study was to aggregate high-quality evidence from randomized trials and quantify the effects of VR training on the exercise performance of these patients. Meta-meta-analysis is the meta-analysis of meta-analyses and follows the primary outcome research. Moreover, the overlap of primary studies in a meta-analysis is also fully considered in order to clearly illustrate the effectiveness of VR training in improving motor performance. These findings are expected to be the highest quality of evidence to date. Meanwhile, the methodological quality of the included meta-analyses was evaluated to provide valuable information for future research and practice and to help clinical rehabilitation practitioners better understand the potential benefits of VR training.

Methods

Search Strategy

Two researchers independently searched for meta-analysis articles published in the following databases: Web of Science, Scopus, PubMed, and China National Knowledge Infrastructure. They used 3 sets of keywords: (1) meta, meta-analysis; (2) stroke, poststroke; (3) virtual reality, virtual game, virtual video, Nintendo Wii, Kinect, Xbox, exergame. The Chinese versions of the aforementioned keywords were retrieved from the Chinese database. The search time frame was from the establishment of the database to November 25, 2020. A complete record of search strings is provided in Multimedia Appendix 1, and the PubMed search strategy is provided as an example. In addition, the reference lists of eligible studies were manually searched to ensure that all meta-analysis studies were included as much as possible.

Study Selection

In this study, 2 researchers independently screened articles by title, abstract, and full text. Before reaching a consensus, differences between the 2 researchers were discussed. If no agreement was reached, a third researcher made the final decision after a group discussion. First, the titles and abstracts were screened, and the following studies were excluded: (1) not a meta-analysis; (2) did not involve related topics, including upper limb function, balance ability, and walking ability (hand function was excluded from upper limb function for research homogeneity); (3) focused on nonstroke patients; (4) used languages other than English or Chinese; or (5) was not published in peer-reviewed journals (ie, excluding conference abstracts, book chapters, and reviews). Next, the full text was screened to exclude meta-analyses with unclear participants or topics. The included meta-analyses had to be randomized controlled trials; other related meta-analyses, such as observational research and longitudinally designed research, were excluded. VR technology can only be used as a general term at present. However, in order to further homogenize the included meta-analyses, it was necessary to define VR technology. The intervention measures of VR technology had to involve building a VR platform. Patients can interact with the electronic screen through motion sensors, accelerometers, gyroscopes, machine gloves, and other devices, and there is no need for them to grasp or manipulate any real objects. In addition to excluding non-VR, computer program–aided meta-analyses, those with unclear statistics were also excluded.

Data Extraction

A table was designed to extract relevant information. When the 2 researchers had differences during the data extraction process, they discussed and resolved these differences between themselves. If the data were not transparent or lacked relevant information, the author was contacted as much as possible to obtain specific information and improve the quality of this research. If the sample size was not directly stated, the researcher calculated the sample size as accurately as possible from the experimental study. If only the total sample size was reported, these samples were divided into 2 parts to determine

the experimental group's and control group's approximate values.

If there were multiple measurements for the final indicators, we used the Fugl-Meyer Assessment Upper Extremity Scale (FMA-UE) as the primary measurement for upper limb function, the Berg Balance Scale (BBS) for balance, and Timed Up and Go for walking ability. In addition, in order to obtain more reliable data, the first choice for this study was to extract data comparing the effects of VR training with those of conventional rehabilitation training.

Quality Assessment

The meta-meta-analysis quality was evaluated by 2 researchers who independently coded using an 11-item multisystem evaluation tool (Assessment of Multiple Systematic Reviews) [10], which proved to be of the right consistency, reliability, and content validity. The scoring criteria were 1 for "yes" and 0 for "not met," "not applicable," or "not reported." The total score of each meta-analysis was the sum of the 11 items, and the quality evaluation was set as low quality (0-4), medium quality (5-8), or high quality (9-11). The final score for each meta-analysis was the average score between the 2 researchers, and the meta-analyses assessed as low quality were excluded. After the included studies were coded, the researchers discussed the coding divergence and used SPSS Version 22 (IBM Corp, Armonk, NY) to assess the intraclass correlation coefficients (ICCs) in order to determine the reliability between the evaluators: 1.00 means complete agreement, and 0.00 means absolute difference [11].

Correction for Overlap of Basic Research

According to the research results of Munder et al [12] and considering that the overlap of research in a meta-meta-analysis may lead to distortion of the results, this study took the overlap of the main results into account and corrected for the overlap of initial studies. When an experimental study was included in multiple meta-analysis studies, it only contributed to this meta-meta-analysis. This was mainly completed by determining the number of meta-analyses of each significant study. For each meta-meta-analysis, the uniqueness of each included meta-analysis was added to set the adjusted research number: kadj. Finally, a meta-analysis with $k_{adj} \leq 3$ was excluded.

Data Analysis

The information included in this study and the conclusions on the benefits of VR training in improving stroke patients' motor performance were from the meta-analyses we rated as medium or high quality. A meta-analysis of the effect size of VR training was conducted under the random effect model of multiple motor abilities. Only 1 effect size was extracted for each meta-analysis. Comprehensive Meta-Analysis (CMA) software 2.0 was used to combine the effect size. CMA is a weighted estimation of effect size based on the method by DerSimonian and Laird. The effect size of each meta-analysis was converted into a standardized mean difference (SMD) to obtain the overall effect size. The effect size was quantified as large (SMD >0.8), medium (SMD 0.5-0.8), or small (SMD 0.2-0.5) [13,14].

The weighted estimation of heterogeneity in all of the meta-analyses was quantified as I^2 (because heterogeneity is not accidental, but instead is the percentage of total variation among studies), with 25%, 50%, and 75% representing low, medium, and high, respectively. For those using the Q value, the heterogeneity score needed to be re-estimated for cross-review comparison. Larger values indicated higher heterogeneity, and $P<.05$ indicated significant differences in heterogeneity [15]. Because of the small number of included studies, further subgroup analysis and sensitivity analysis were not conducted in this study.

Publication Bias

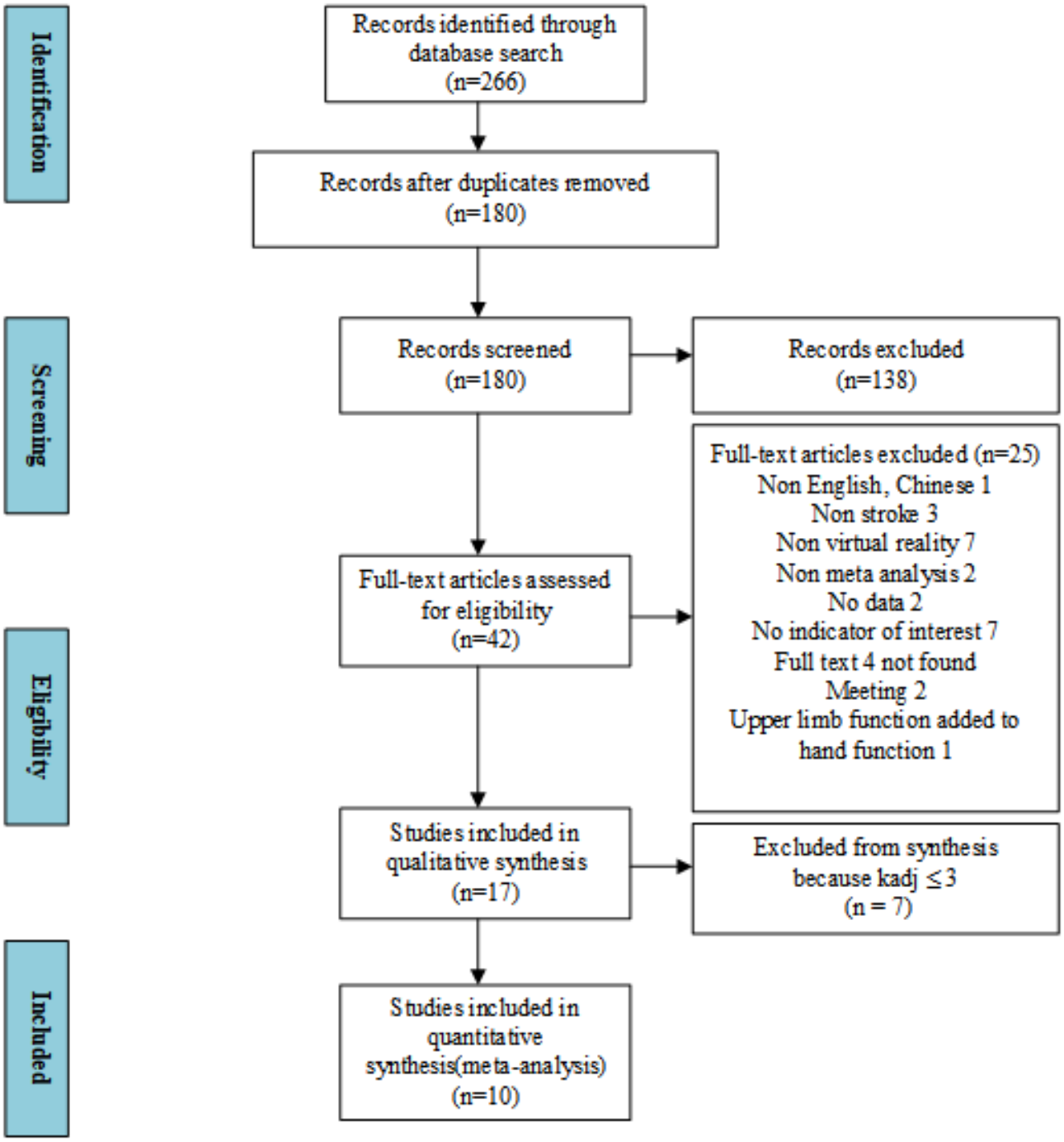
Publication bias is defined as "the bias caused by the nature and direction of research that influences the decision whether to publish or otherwise distribute it" [16], which leads to an unknown number of unpublished negative results. Since there were less than 10 meta-analysis articles for each result in this study, no funnel plot was generated, and the Egger test was used to evaluate the publication bias [17].

Results

Literature Search

A total of 226 records was retrieved from the database, and 180 records remained after removing the duplicate records. A total of 138 records was deleted because of their unqualified titles and abstracts. A more detailed full-text screening was performed with the remaining 42 studies, of which 25 were excluded because they did not meet the established criteria. Therefore, the qualitative research included 17 studies. An additional 7 studies were excluded because k_{adj} was ≤3, and therefore, the final quantitative analysis included a total of 10 meta-analysis studies (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for studies included in and excluded from the meta-meta-analysis.



Research Characteristics

The characteristics of all the included studies are shown in Tables 1-3. A total of 10 studies was obtained: 6 studies evaluated the rehabilitation of upper limb function (Table 1), 4 studies evaluated the rehabilitation of balance (Table 2), and 2 studies evaluated the rehabilitation of walking function (Table 3).

Table 1. Characteristics of the included studies for upper limb function.

Author(s), year	Aim	k[a]	Sample size	Measuring instrument	Effect size			I^2	k_{adj}
					Effect size	95% CI (LL[b] to UL[c])	*P* value		
Yang et al, 2013 [18]	Effect of VR[d] technology on the rehabilitation of upper and lower limb motor function in stroke patients	8	EG[e]: 164; CG[f]: 143	FMA-UE[g]	4.27	2.47 to 6.06	≤.001	0	3.033
Laver et al, 2017 [19]	To determine the efficacy of VR compared with an alternative intervention or no intervention on upper limb function and activity	10	EG: 191; CG: 172	FMA-UE	3.30	1.29 to 5.32	.001	0	4.867
Maier et al, 2019 [20]	To evaluate the efficacy of specific VR and nonspecific VR systems for rehabilitating upper limb function and activity after stroke	17	EG: 704; CG: 551	FMA-UE	0.21	0.08 to 0.33	.002	12	7.950
Zhong et al, 2019 [21]	To evaluate the VR technology on the clinical curative effect of upper limb function in patients undergoing cerebral apoplexy hemiplegia rehabilitation	17	EG: 283; CG: 274	FMA-UE	5.33	3.54 to 7.13	≤.001	22	8.117
Domínguez-Téllez et al, 2020 [2]	To evaluate game-based VR interventions to improve upper limb motor function and quality of life after stroke	9	EG: 154; CG: 148	FMA-UE	1.53	0.51 to 2.54	.003	92	3.417
Mekbib et al, 2020 [22]	To evaluate the overall effectiveness of VR therapies compared with that of conventional therapies in the recovery of upper limb functions across the 3 ICF[h] domains	19	EG: 363; CG: 345	FMA-UE	3.84	0.93 to 6.75	.01	64	9.950

[a]k: number of primary studies.

[b]LL: lower limit.

[c]UL: upper limit.

[d]VR: virtual reality.

[e]EG: experimental group.

[f]CG: control group.

[g]FMA-UE: Fugl-Meyer Assessment Upper Extremity Scale.

[h]ICF: International Classification of Functioning, Disability and Health.

Table 2. Characteristics of the included studies for balance.

Author(s), year	Aim	k[a]	Sample size	Measuring instrument	Effect size			I^2	k_{adj}
					Effect size	95% CI (LL[b] to UL[c])	*P* value		
Gibbons et al, 2016 [23]	Effects of VR[d] on lower limb outcomes in stroke patients	9	EG[e]: 104; CG[f]: 95	BBS[g], FRT[h]	0.42	0.11 to 0.73	.42	10	4.077
Iruthayarajah et al, 2017 [24]	To evaluate the effectiveness of VR interventions in improving balance in a chronic stroke population	12	EG: 132; CG: 142	BBS	0.506	0.259 to 0.753	≤.001	2.622	5.012
Mohammadi et al, 2019 [25]	To evaluate the effect of VR on balance as compared with that of conventional therapy alone poststroke	13	EG: 161; CG: 153	BBS, FRT, PAS[i]	0.64	0.36 to 0.92	.083	36.7	5.530
Liang et al, 2020 [26]	To evaluate the effectiveness of VR technology in promoting balance and walking function rehabilitation in stroke	17	EG: 215; CG: 217	FRT, PAS	4.09	2.20 to 5.97	≤.001	84.5	8.513

[a]k: number of primary studies.

[b]LL: lower limit.

[c]UL: upper limit.

[d]VR: virtual reality.

[e]EG: experimental group.

[f]CG: control group.

[g]BBS: Berg Balance Scale.

[h]FRT: functional reach test.

[i]PAS: Postural Assessment Scale.

Table 3. Characteristics of the included studies for walking ability.

Author(s), year	Aim	k[a]	Sample size	Measuring instrument	Effect size			I^2	k_{adj}
					Effect size	95% CI (LL[b] to UL[c])	*P* value		
Iruthayarajah et al, 2017 [24]	To evaluate the effectiveness of VR[d] interventions in improving balance in a chronic stroke population	13	EG[e]: 145; CG[f]: 166	TUG[g]	0.367	0.134 to 0.601	.002	0	5.467
Liang et al, 2020 [26]	To evaluate the effectiveness of VR technology in promoting balance and walking function rehabilitation in stroke	18	EG: 271; CG: 271	TUG	–2.79	–4.88 to –0.69	.009	86.2	12.267

[a]k: number of primary studies.

[b]LL: lower limit.

[c]UL: upper limit.

[d]VR: virtual reality.

[e]EG: experimental group.

[f]CG: control group.

[g]TUG: Timed Up and Go.

Methodological Quality of Included Studies

In this study, the methodological quality of 10 meta-analysis articles was evaluated. Their scores were all >4, which was a satisfactory result. The specific scores are shown in Table 4. Moreover, the reliability of the quality evaluation was high (ICC=0.96, 95% CI 0.93-0.98, *P*<.001).

Table 4. Methodological quality of the included studies.

	1[a]	2[b]	3[c]	4[d]	5[e]	6[f]	7[g]	8[h]	9[i]	10[j]	11[k]	Total score
Yang et al, 2013 [18]	1	0	1	?[l]	1	1	1	?	1	1	0	8
Gibbons et al, 2016 [23]	1	1	1	1	1	1	1	?	1	0	0	8
Laver et al, 2017 [19]	1	1	1	1	0	0	1	?	?	0	0	5
Iruthayarajah et al, 2017 [24]	1	0	1	1	1	1	1	?	?	0	0	6
Mohammadi et al, 2019 [25]	1	1	1	1	1	1	1	?	1	0	1	9
Maier et al, 2019 [20]	0	1	0	0	1	1	1	?	1	1	1	6
Zhong et al, 2019 [21]	1	1	1	?	1	1	1	?	1	1	1	7
Domínguez-Téllez et al, 2020 [2]	1	1	1	1	1	1	1	?	1	0	1	9
Liang et al, 2020 [26]	1	1	1	?	1	1	1	?	1	1	0	8
Mekbib et al, 2020 [22]	1	0	1	0	0	0	1	?	1	0	1	5

[a]Whether the preliminary design scheme was provided.

[b]Whether the included studies and data extraction were repetitive.

[c]Whether an extensive and comprehensive literature search was carried out.

[d]Whether the publication was considered.

[e]Whether the list of included and excluded studies was provided.

[f]Whether the characteristics of the included studies were described

[g]Whether the scientificity of the included studies was evaluated and reported.

[h]Whether the scientificity of the included studies was appropriately used in the derivation of conclusions.

[i]Whether it was appropriate to summarize the research results.

[j]Whether the possibility of publication bias was evaluated.

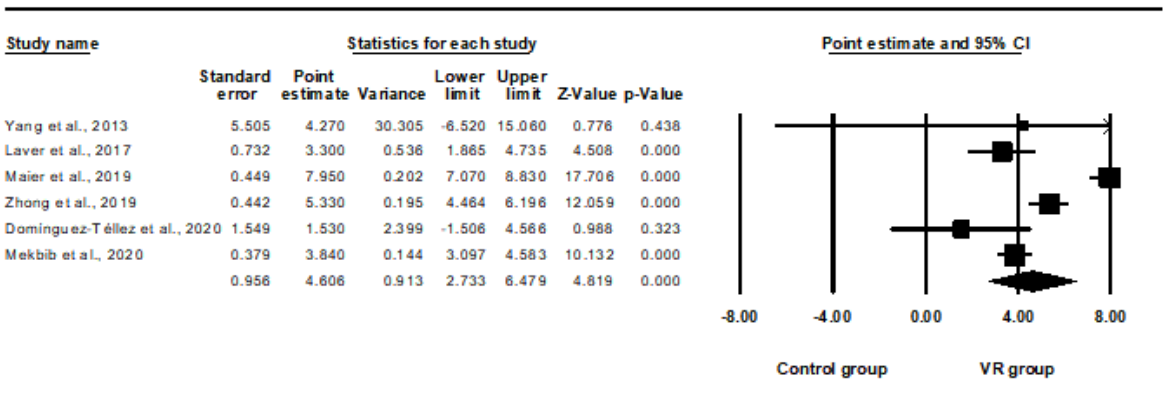
[k]Whether the relevant conflicts of interest were explained.

[l]Unknown.

Upper Limb Function

Evidence for VR training helping to improve stroke patients' upper limb function was provided by 7 meta-analysis studies, and significant differences were found in these studies. After accounting for overlap, 6 studies were finally confirmed. As shown in Figure 2, the meta-meta-analysis showed that the effect of VR-based intervention on stroke had a statistically large effect size (SMD 4.606, 95% CI 2.733-6.479, *P*<.05; Figure 2), and considerable heterogeneity was found (Q=62.851, I^2=92.045). The intercept of the Egger Test for publication bias was –1.44, *P*=.671.

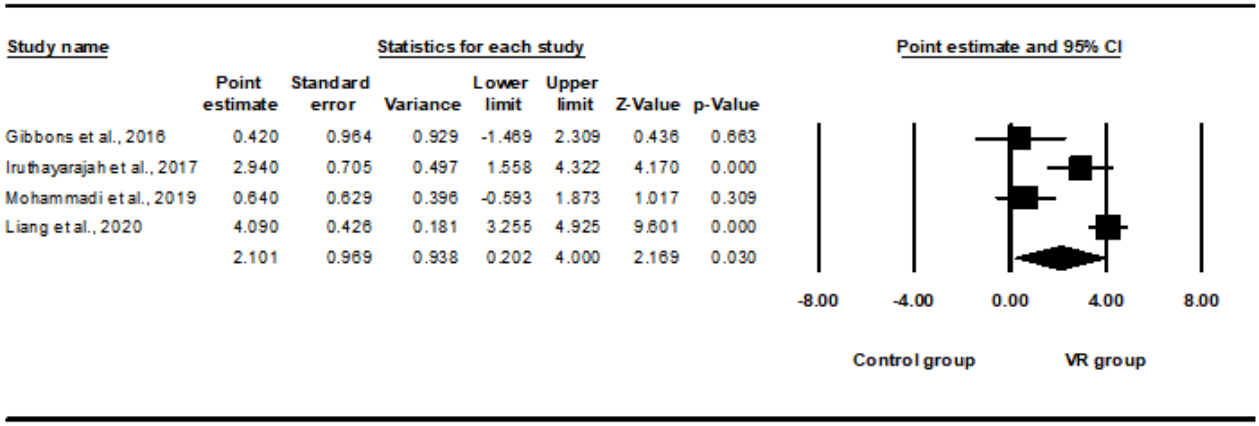
Figure 2. Meta-meta-analysis of the effects of virtual reality (VR) interventions on upper limb function. The bottom row describes a combined overall effect of treatment which random-effects models were used to estimate.



Balance Function

Evidence for VR training helping to improve stroke patients' balance ability was provided by 11 meta-analysis studies, and significant differences were found in 10 studies. After accounting for overlap, 4 studies were finally confirmed. As shown in Figure 3, the meta-meta-analysis showed that the efficacy of VR-based intervention on balance after stroke had a statistically large effect size (SMD 2.101, 95% CI 0.202-4.000, P<.05), and considerable heterogeneity was found (Q=27.061, I^2=88.914). The intercept of the Egger Test for publication bias was –7.09, P=.222.

Figure 3. Meta-meta-analysis of the effects of virtual reality (VR) training on balance. The bottom row describes a combined overall effect of treatment which random-effects models were used to estimate.



Lower Limb Function

Evidence for VR training helping to improve lower extremity function in stroke patients was provided by 8 meta-analyses, and 8 studies showed significant differences. After accounting for overlap, only 2 studies met the condition of k_{adj} >3, and therefore, the effect size was not summarized.

Discussion

Principal Findings

The purpose of this meta-meta-analysis was to quantify the impact of VR training on motor performance in patients poststroke by summarizing the current mid-quality and high-quality meta-analysis results. It found that VR training showed a significant improvement in upper limb function (SMD 4.606) and balance ability (SMD 2.101), which was consistent with the results of most previous meta-analysis articles. According to the included meta-analysis studies, there were 7 studies on the effect of VR training on the upper extremities and 10 studies on balance. Our conclusions provided strong evidence for VR-induced improvement and confirmed the potential therapeutic benefits of VR training.

In the included meta-analysis studies, the FMA-UE was used to evaluate the upper extremities. It includes 33 items related to proximal and distal upper extremity movement, mainly to evaluate the reflex activity, motor control, and muscle strength of the upper extremity on the hemiplegic side. FMA-UE is widely applied as the outcome index of rehabilitation tests and to record recovery after stroke [27]. Unfortunately, due to the limited number of studies, it is impossible to test the dose effect for stroke patients with different degrees of dyskinesia to determine the most effective intervention dose of VR. The BBS was used by 3 of 4 included studies, showing a greater degree of improvement. Balance ability is a crucial step to improve the motor ability of stroke patients, although rehabilitation often involves many functions, such as visual function, vestibular function, and somatosensory receptors. However, based on the existing meta-analysis studies, VR training can effectively ameliorate these functions, leading to a greater degree of improvement in balance. It is worth noting that a customized balance rehabilitation system seems to have the same curative effect as a commercial virtual balance system, suggesting that stroke patients can carry out rehabilitation training through commercial virtual games [24]. In this meta-meta-analysis, there was a specific deviation in the effect size of these meta-analysis articles that can be considered a result of the different studies included. In addition, statistical differences were not excluded. In terms of gait, 2 meta-analysis results in this study showed moderate effects on improving walking ability. Due to the lack of sufficient qualified meta-analysis studies, no meta-meta-analysis of walking ability was performed, but its influence cannot be ignored. Due to decreased walking ability and fear of falling, stroke patients' motor ability is weakened, which is unfavorable for recovery [28].

Optimizing and strengthening brain compensation mechanisms are essential for dyskinesia [29]. The virtual environment created by VR technology can promote the illusion of body movement, strengthen the activation of the motor brain area with a sense of immersion, mobilize changes in brain neuroplasticity, reconstruct the cell synapses of the nervous system, and realize the direct training of the central nervous system, which plays a significant role in the reorganization and recovery of neural structures after stroke [30,31]. Compared with other interventions, VR training has many potential advantages. For example, the cost of current VR equipment is relatively low.

Only a screen and VR system are needed for the completion of the intervention. Second, repetitive task training has been proven to be effective in improving the motor ability of stroke patients. Current VR equipment is more convenient to carry; thus, patients can carry out regular training in their own homes. Third, current VR systems are compatible with other systems, which makes it more convenient for training. All of these cannot be realized by conventional therapy. Without increasing the number of rehabilitation therapists, these are conducive to improving the efficacy of rehabilitation and reducing patients' medical burden [32].

VR training includes 4 components that could work together to ensure success: intensive therapy, motivating therapy through exercise games, stimulation of motor learning, and positive feedback between the stimulus and the response. Therefore, a single mechanism cannot explain the impact of VR training on stroke patients, which may be the result of both psychological mechanisms (such as self-efficacy, reward mechanism, and emotion) and physiological mechanisms (vestibular and somatosensory receptors) [33-35]. With the development of medical imaging, researchers have begun to explore the effects of VR training on the brain functions of stroke patients. Although the current research is still in its infancy, with the deepening of research, the effectiveness of VR training will soon benefit from more direct evidence [18,19,36-38].

This study conducted a meta-meta-analysis on the current meta-analysis data for VR training in improving stroke, which strengthened the evidence on this topic. After a series of controlled experiments and strict methodological quality assessment, confidence in applying VR technology in clinical rehabilitation was increased. Nevertheless, there were still some limitations in this study. First, the retrieval language was confined to English or Chinese; thus, publications in other languages may have been missed. Second, since this meta-meta-analysis was based on previously published meta-analysis articles, primary research without meta-analysis may be omitted. Finally, our primary problem was that the results showed a high degree of statistical heterogeneity. Therefore, the results should be interpreted with caution.

Conclusion

In general, the current evidence supports that VR, which has a medium to large effect size, is beneficial for stroke patients' motor ability, especially in the upper extremities and for balance. However, no specific rehabilitation treatment has been formulated to date, likely due to different motor ability levels. Consequently, future research on this topic requires randomized controlled trials with larger sample sizes and longer duration to verify that VR training is the best treatment for improving stroke patients' motor performance and to verify the optimal type, frequency, duration, and cycle of VR training for patients with different motor abilities.

Acknowledgments

This research was funded by Humanities and Social Science, ShenZhen University, Young Teacher Award (grant number 17QNFC59).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 18 KB - jmir_v23i10e31051_app1.docx]

References

1. McEwen D, Taillon-Hobson A, Bilodeau M, Sveistrup H, Finestone H. Virtual reality exercise improves mobility after stroke: an inpatient randomized controlled trial. Stroke 2014 Jun;45(6):1853-1855. [doi: 10.1161/STROKEAHA.114.005362] [Medline: 24763929]
2. Domínguez-Téllez P, Moral-Muñoz JA, Salazar A, Casado-Fernández E, Lucena-Antón D. Game-Based Virtual Reality Interventions to Improve Upper Limb Motor Function and Quality of Life After Stroke: Systematic Review and Meta-analysis. Games Health J 2020 Feb 01;9(1):1-10. [doi: 10.1089/g4h.2019.0043] [Medline: 32027185]
3. Corbetta D, Imeri F, Gatti R. Rehabilitation that incorporates virtual reality is more effective than standard rehabilitation for improving walking speed, balance and mobility after stroke: a systematic review. J Physiother 2015 Jul;61(3):117-124 [FREE Full text] [doi: 10.1016/j.jphys.2015.05.017] [Medline: 26093805]
4. Zhang L, Pi Y, Shen C, Zhu H, Li X, Ni Z, et al. Expertise-Level-Dependent Functionally Plastic Changes During Motor Imagery in Basketball Players. Neuroscience 2018 Jun 01;380:78-89. [doi: 10.1016/j.neuroscience.2018.03.050] [Medline: 29634999]
5. Wu J, Ma Y, Ren Z. Rehabilitative Effects of Virtual Reality Technology for Mild Cognitive Impairment: A Systematic Review With Meta-Analysis. Front Psychol 2020 Sep 25;11:1811 [FREE Full text] [doi: 10.3389/fpsyg.2020.01811] [Medline: 33101098]

6. Merians AS, Jack D, Boian R, Tremaine M, Burdea G, Adamovich S, et al. Virtual reality-augmented rehabilitation for patients following stroke. Phys Ther 2002 Sep;82(9):898-915. [Medline: 12201804]
7. Proffitt R, Lange B. Considerations in the efficacy and effectiveness of virtual reality interventions for stroke rehabilitation: moving the field forward. Phys Ther 2015 Mar 24;95(3):441-448 [FREE Full text] [doi: 10.2522/ptj.20130571] [Medline: 25343960]
8. Ren Z, Wu J. The Effect of Virtual Reality Games on the Gross Motor Skills of Children with Cerebral Palsy: A Meta-Analysis of Randomized Controlled Trials. Int J Environ Res Public Health 2019 Oct 14;16(20):3885 [FREE Full text] [doi: 10.3390/ijerph16203885] [Medline: 31614990]
9. Wu J, Loprinzi PD, Ren Z. The Rehabilitative Effects of Virtual Reality Games on Balance Performance among Children with Cerebral Palsy: A Meta-Analysis of Randomized Controlled Trials. Int J Environ Res Public Health 2019 Oct 28;16(21):4161 [FREE Full text] [doi: 10.3390/ijerph16214161] [Medline: 31661938]
10. Burda BU, Holmer HK, Norris SL. Limitations of A Measurement Tool to Assess Systematic Reviews (AMSTAR) and suggestions for improvement. Syst Rev 2016 Apr 12;5(1):58 [FREE Full text] [doi: 10.1186/s13643-016-0237-1] [Medline: 27072548]
11. Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. Tutor Quant Methods Psychol 2012 Feb 01;8(1):23-34 [FREE Full text] [doi: 10.20982/tqmp.08.1.p023] [Medline: 22833776]
12. Munder T, Brütsch O, Leonhart R, Gerger H, Barth J. Researcher allegiance in psychotherapy outcome research: an overview of reviews. Clin Psychol Rev 2013 Jun;33(4):501-511. [doi: 10.1016/j.cpr.2013.02.002] [Medline: 23500154]
13. Rebar AL, Stanton R, Geard D, Short C, Duncan MJ, Vandelanotte C. A meta-meta-analysis of the effect of physical activity on depression and anxiety in non-clinical adult populations. Health Psychol Rev 2015 Jul 03;9(3):366-378. [doi: 10.1080/17437199.2015.1022901] [Medline: 25739893]
14. Tchero H, Tabue Teguo M, Lannuzel A, Rusch E. Telerehabilitation for Stroke Survivors: Systematic Review and Meta-Analysis. J Med Internet Res 2018 Oct 26;20(10):e10867 [FREE Full text] [doi: 10.2196/10867] [Medline: 30368437]
15. Weber L, Kamp-Becker I, Christiansen H, Mingebach T. Treatment of child externalizing behavior problems: a comprehensive review and meta-meta-analysis on effects of parent-based interventions on parental characteristics. Eur Child Adolesc Psychiatry 2019 Aug 8;28(8):1025-1036. [doi: 10.1007/s00787-018-1175-3] [Medline: 29948228]
16. Zhai Y, Zhu W, Cai Y, Sun D, Zhao J. Clinical- and cost-effectiveness of telemedicine in type 2 diabetes mellitus: a systematic review and meta-analysis. Medicine (Baltimore) 2014 Dec;93(28):e312 [FREE Full text] [doi: 10.1097/MD.0000000000000312] [Medline: 25526482]
17. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane handbook for systematic reviews of interventions, 2nd edition. Chichester, United Kingdom: John Wiley & Sons; 2019.
18. Yang YJ, Yue YS, Guo JB. Effect of Virtual Reality on Motor Function in Patients with Stroke: A Systematic Review of Randomized Controlled Trials. China Journal of Rehabilitation Theory and Practice 2013;18(6):1. [doi: 10.3969/j.issn.1006-9771.2013.08.002]
19. Laver KE, Lange B, George S, Deutsch JE, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. Cochrane Database Syst Rev 2017 Nov 20;11:CD008349 [FREE Full text] [doi: 10.1002/14651858.CD008349.pub4] [Medline: 29156493]
20. Maier M, Rubio Ballester B, Duff A, Duarte Oller E, Verschure PFMJ. Effect of Specific Over Nonspecific VR-Based Rehabilitation on Poststroke Motor Recovery: A Systematic Meta-analysis. Neurorehabil Neural Repair 2019 Feb;33(2):112-129 [FREE Full text] [doi: 10.1177/1545968318820169] [Medline: 30700224]
21. Zhong C, Hongchen H, Shuangshuang H, Rui M, Jingjing W, Chengqi H. The effectiveness of virtual reality therapy in relieving upper limb hemiparesis after stroke:A meta-analysis of randomized and controlled trials. Chinese Journal of Physical Medicine and Rehabilitation 2019;41(6):463-468. [doi: 10.3760/cma.j.issn.0254-1424.2020.07.011]
22. Mekbib DB, Han J, Zhang L, Fang S, Jiang H, Zhu J, et al. Virtual reality therapy for upper limb rehabilitation in patients with stroke: a meta-analysis of randomized clinical trials. Brain Inj 2020 Mar 20;34(4):456-465. [doi: 10.1080/02699052.2020.1725126] [Medline: 32064964]
23. Gibbons EM, Thomson AN, de Noronha M, Joseph S. Are virtual reality technologies effective in improving lower limb outcomes for patients following stroke - a systematic review with meta-analysis. Top Stroke Rehabil 2016 Dec;23(6):440-457. [doi: 10.1080/10749357.2016.1183349] [Medline: 27237336]
24. Iruthayarajah J, McIntyre A, Cotoi A, Macaluso S, Teasell R. The use of virtual reality for balance among individuals with chronic stroke: a systematic review and meta-analysis. Top Stroke Rehabil 2017 Jan 16;24(1):68-79. [doi: 10.1080/10749357.2016.1192361] [Medline: 27309680]
25. Mohammadi R, Semnani AV, Mirmohammadkhani M, Grampurohit N. Effects of Virtual Reality Compared to Conventional Therapy on Balance Poststroke: A Systematic Review and Meta-Analysis. J Stroke Cerebrovasc Dis 2019 Jul;28(7):1787-1798. [doi: 10.1016/j.jstrokecerebrovasdis.2019.03.054] [Medline: 31031145]
26. Liang M, Zhen W, Rong X, Jing H, Jing M, Jing T, et al. A meta-analysis of the effectiveness of virtual reality training for improving balance and walking after a stroke. Chinese Journal of Physical Medicine and Rehabilitation 2021;42(7):632-639. [doi: 10.3760/cma.j.issn.0254-1424.2020.07.011]

27. Woytowicz EJ, Rietschel JC, Goodman RN, Conroy SS, Sorkin JD, Whitall J, et al. Determining Levels of Upper Extremity Movement Impairment by Applying a Cluster Analysis to the Fugl-Meyer Assessment of the Upper Extremity in Chronic Stroke. Arch Phys Med Rehabil 2017 Mar;98(3):456-462 [FREE Full text] [doi: 10.1016/j.apmr.2016.06.023] [Medline: 27519928]
28. Wing JJ, Burke JF, Clarke PJ, Feng C, Skolarus LE. The role of the environment in falls among stroke survivors. Arch Gerontol Geriatr 2017 Sep;72:1-5 [FREE Full text] [doi: 10.1016/j.archger.2017.04.007] [Medline: 28482268]
29. Lloréns R, Noé E, Colomer C, Alcañiz M. Effectiveness, usability, and cost-benefit of a virtual reality-based telerehabilitation program for balance recovery after stroke: a randomized controlled trial. Arch Phys Med Rehabil 2015 Mar;96(3):418-425.e2. [doi: 10.1016/j.apmr.2014.10.019] [Medline: 25448245]
30. van Nimwegen M, Speelman AD, Hofman-van Rossum EJM, Overeem S, Deeg DJH, Borm GF, et al. Physical inactivity in Parkinson's disease. J Neurol 2011 Dec;258(12):2214-2221 [FREE Full text] [doi: 10.1007/s00415-011-6097-7] [Medline: 21614433]
31. Chen Y, Gao Q, He C, Bian R. Effect of Virtual Reality on Balance in Individuals With Parkinson Disease: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. Phys Ther 2020 Jun 23;100(6):933-945. [doi: 10.1093/ptj/pzaa042] [Medline: 32157307]
32. Ruiz-González L, Lucena-Antón D, Salazar A, Martín-Valero R, Moral-Munoz JA. Physical therapy in Down syndrome: systematic review and meta-analysis. J Intellect Disabil Res 2019 Aug 20;63(8):1041-1067. [doi: 10.1111/jir.12606] [Medline: 30788876]
33. Wegner M, Helmich I, Machado S, Nardi AE, Arias-Carrion O, Budde H. Effects of exercise on anxiety and depression disorders: review of meta- analyses and neurobiological mechanisms. CNS Neurol Disord Drug Targets 2014;13(6):1002-1014. [doi: 10.2174/18715273136661406121028 41] [Medline: 24923346]
34. Matta Mello Portugal E, Cevada T, Sobral Monteiro-Junior R, Teixeira Guimarães T, da Cruz Rubini E, Lattari E, et al. Neuroscience of exercise: from neurobiology mechanisms to mental health. Neuropsychobiology 2013;68(1):1-14 [FREE Full text] [doi: 10.1159/000350946] [Medline: 23774826]
35. Chen F, Leng Y, Ge J, Wang D, Li C, Chen B, et al. Effectiveness of Virtual Reality in Nursing Education: Meta-Analysis. J Med Internet Res 2020 Sep 15;22(9):e18290 [FREE Full text] [doi: 10.2196/18290] [Medline: 32930664]
36. You SH, Jang SH, Kim Y, Hallett M, Ahn SH, Kwon Y, et al. Virtual reality-induced cortical reorganization and associated locomotor recovery in chronic stroke: an experimenter-blind randomized study. Stroke 2005 Jun;36(6):1166-1171. [doi: 10.1161/01.STR.0000162715.43417.91] [Medline: 15890990]
37. August K, Lewis J, Chandar G, Merians A, Biswal B, Adamovich S. FMRI analysis of neural mechanisms underlying rehabilitation in virtual reality: activating secondary motor areas. Conf Proc IEEE Eng Med Biol Soc 2006;2006:3692-3695. [doi: 10.1109/IEMBS.2006.260144] [Medline: 17946197]
38. Wang Z, Wang P, Xing L, Mei L, Zhao J, Zhang T. Leap Motion-based virtual reality training for improving motor functional recovery of upper limbs and neural reorganization in subacute stroke patients. Neural Regen Res 2017 Nov;12(11):1823-1831 [FREE Full text] [doi: 10.4103/1673-5374.219043] [Medline: 29239328]

Abbreviations

BBS: Berg Balance Scale
CMA: Comprehensive Meta-Analysis
FMA-UE: Fugl-Meyer Assessment Upper Extremity Scale
ICC: intraclass correlation coefficient
SMD: standardized mean difference
VR: virtual reality

Edited by R Kukafka; submitted 08.06.21; peer-reviewed by L Zhang, Y Zhang; comments to author 28.06.21; revised version received 07.08.21; accepted 12.08.21; published 12.10.21.

Please cite as:
Wu J, Zeng A, Chen Z, Wei Y, Huang K, Chen J, Ren Z
Effects of Virtual Reality Training on Upper Limb Function and Balance in Stroke Patients: Systematic Review and Meta-Meta-Analysis
J Med Internet Res 2021;23(10):e31051
URL: https://www.jmir.org/2021/10/e31051
doi:10.2196/31051
PMID:34636735

Original Paper

Validity Evidence Based on Relations to Other Variables of the eHealth Literacy Questionnaire (eHLQ): Bayesian Approach to Test for Known-Groups Validity

Christina Cheng[1,2*], PhD; Gerald Elsworth[1,2*], PhD; Richard H Osborne[1,2*], PhD

[1]Centre for Global Health and Equity, School of Health Sciences, Swinburne University of Technology, Hawthorn, Australia

[2]School of Health and Social Development, Faculty of Health, Deakin University, Burwood, Australia

[*]all authors contributed equally

Corresponding Author:
Christina Cheng, PhD
Centre for Global Health and Equity
School of Health Sciences
Swinburne University of Technology
Room 907, Level 9, AMDC Building
453/469-477 Burwood Road
Hawthorn, Victoria 3122
Australia
Phone: 61 392145470
Email: cccheng@swin.edu.au

Abstract

Background: As health resources and services are increasingly delivered through digital platforms, eHealth literacy is becoming a set of essential capabilities to improve consumer health in the digital era. To understand eHealth literacy needs, a meaningful measure is required. Strong initial evidence for the reliability and construct validity of inferences drawn from the eHealth Literacy Questionnaire (eHLQ) was obtained during its development in Denmark, but validity testing for varying purposes is an ongoing and cumulative process.

Objective: This study aims to examine validity evidence based on relations to other variables—using data collected with the known-groups approach—to further explore if the eHLQ is a robust tool to understand eHealth literacy needs in different contexts. A priori hypotheses are set for the expected score differences among age, sex, education, and information and communication technology (ICT) use for each of the 7 eHealth literacy constructs represented by the 7 eHLQ scales.

Methods: A Bayesian mediated multiple indicators multiple causes model approach was used to simultaneously identify group differences and test measurement invariance through differential item functioning across the groups, with ICT use as a mediator. A sample size of 500 participants was estimated. Data were collected at 3 diverse health sites in Australia.

Results: Responses from 525 participants were included for analysis. Being older was significantly related to lower scores in 4 eHLQ scales, with *3. Ability to actively engage with digital services* having the strongest effect (total effect –0.37; P<.001), followed by *1. Using technology to process health information* (total effect –0.32; P<.001), *5. Motivated to engage with digital services* (total effect –0.21; P=.01), and *7. Digital services that suit individual needs* (total effect –0.21; P=.02). However, the effects were only partially mediated by ICT use. Higher education was associated with higher scores in *1. Using technology to process health information* (total effect 0.22; P=.01) and *3. Ability to actively engage with digital services* (total effect 0.25; P<.001), with the effects mostly mediated by ICT use. Higher ICT use was related to higher scores in all scales except *2. Understanding health concepts and language* and *4. Feel safe and in control*. Either no or ignorable cases of differential item functioning were found across the 4 groups.

Conclusions: By using a Bayesian mediated multiple indicators multiple causes model, this study provides supportive validity evidence for the eHLQ based on relations to other variables as well as established evidence regarding internal structure related to measurement invariance across the groups for the 7 scales in the Australian community health context. This study also demonstrates that the eHLQ can be used to gain valuable insights into people's eHealth literacy needs to help optimize access and use of digital health and promote health equity.

(J Med Internet Res 2021;23(10):e30243) doi:10.2196/30243

KEYWORDS

eHealth; digital health; health literacy; health equity; questionnaire design; health literacy questionnaire; validity evidence; mediation effect; mobile phone

Introduction

Background

eHealth literacy, also known as digital health literacy, has been described as a set of essential capabilities to improve consumer health in the digital era [1,2]. As health resources and services continue to move to digital platforms, people need "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem" [1]. Digital health will have limited value if people do not have adequate eHealth literacy to effectively engage with these resources [1,3,4]. Hence, an understanding of eHealth literacy needs is paramount to ensure that digital health resources are aligned with such needs and avoid the potential widening of health inequities. However, current research and insights into the eHealth literacy needs of populations are limited, and the results can be inconsistent [5], possibly because of the lack of a rigorous theoretical framework to measure eHealth literacy [6].

To understand eHealth literacy needs, a useful and valid measurement of eHealth literacy is needed. With the introduction of this concept in 2006, the study by Norman and Skinner [7] developed the eHealth Literacy Scale (eHEALS) to assess people's ability to engage with eHealth, with the purpose of informing clinical decisions and health promotion planning. Initial validity testing of the tool in Canada demonstrated good internal consistency (coefficient α=.88), with principal component analysis suggesting a single-factor solution [7]. However, recent validation studies have cast doubts on the tool's dimensionality. The tool was found to be a better fit for a two-factor model in 5 studies, but there was no consensus on the items for the 2 subscales among these studies [8-12], whereas 3 studies reported that the eHEALS consisted of 3 dimensions [13-15]. Nevertheless, the tool has been widely used in eHealth literacy studies across the world in various settings to understand the eHealth literacy of various population groups, examine the association of eHealth literacy and sociodemographic factors, measure the effects of eHealth literacy on health outcomes, and use as an outcome measure of eHealth literacy interventions [6,16]. Yet, the linkages of the findings of these studies to specific eHealth recommendations were usually vague [6]. In contrast, the study by Norman [17] acknowledged that the digital landscape had evolved since 2006, especially around the interactivity and expanded capabilities of information and communication technologies (ICTs), and called for revision of the concept as well as its measurement tool.

Using a grounded validity-driven approach [18], the study by Norgaard et al [19] developed the eHealth Literacy Framework by integrating the perspectives and experiences of a wide range of eHealth stakeholders, including patients, health care providers, health informatics professionals, public health researchers, and computer scientists. Through concept-mapping workshops and international web-based surveys, 7 domains of eHealth literacy were identified [19]. On the basis of the eHealth Literacy Framework and on the back of the widely used and tested Health Literacy Questionnaire [20], the eHealth Literacy Questionnaire (eHLQ) was subsequently developed comprising 7 scales and representing the following 7 eHealth literacy constructs:

1. Using technology to process health information
2. Understanding of health concepts and language
3. Ability to actively engage with digital services
4. Feel safe and in control
5. Motivated to engage with digital services
6. Access to digital services that work
7. Digital services that suit individual needs [21]

Each eHLQ scale has 4 to 6 items relating to a 4-point ordinal scale, ranging from strongly disagree to strongly agree. The results are 7 scale scores with a range of 1-4, calculated by averaging the item scores within each scale with equal weighting. Initial validity testing of the eHLQ involved extensive discussion of the test content in the Australian and Danish contexts by an international multidisciplinary team experienced in questionnaire development and cognitive interviewing with community members from different cultural and educational backgrounds to ensure that the items were understood as intended. The items were then administered to 475 Danish participants randomly approached by trained interviewers in the broader-community locations, including libraries, workplaces, hospitals, nursing homes, health centers, and an outpatient clinic. Bayesian confirmatory factor analysis supported the seven-factor model, with all items loading strongly on their relevant factors and no statistically significant cross-loadings. Composite scale reliability (ranging from 0.75 to 0.87) demonstrated good internal consistency. Item response theory analysis confirmed that there were no disordered thresholds, and differential item functioning (DIF) testing established evidence of measurement invariance for age and sex [21]. The eHLQ has since been used to investigate the eHealth literacy of nursing students and pregnant migrant women in Denmark [22,23], as well as to examine the association of eHealth literacy and digital health service use in both Denmark and Australia [24,25]. The tool has also been used in Australia to understand the eHealth literacy needs of community members, leading to the generation of numerous concrete solutions to address the identified needs [26].

Study Aim

According to the *Standards for Educational and Psychological Testing* (the *Standards*) [27], the authoritative reference used to develop, use, and interpret educational and psychological measurements, validity testing is a continuous process and involves the examination of 5 sources of evidence to support the interpretation and use of the scores, including test content,

response process, internal structure, relations to other variables, and consequences of testing [27]. This study aims to examine evidence based on relations to other variables to further evaluate the eHLQ as a tool used to understand eHealth literacy needs.

Evidence of relations to other variables refers to an analysis of the relationship between the eHLQ scores and other variables with which the scores are predicted by theory or past research to be associated. The evidence may include association of the scores with certain demographic groups, relationships with predicted outcomes, or relationships between the scores and other external instruments that measure the same construct [27,28]. As there is no consensus on the dimensionality of the eHEALS, which is the most commonly used eHealth literacy tool, comparing the scores of the 2 instruments would be problematic. Therefore, this study focused on the testing of the association of the eHLQ scores with certain demographic groups, which is usually described as known-groups validity. A seminal paper on validity by Cronbach and Meehl [29] discussed that "If our understanding of a construct leads us to expect 2 groups to differ on the test, this expectation may be tested directly" [29]. Hence, group differences can be used to examine if an instrument is sensitive enough to discriminate "between these groups" [30]. However, the paper by Cronbach and Meehl [29] further noted that only moderate association should be expected because members of the groups were expected to overlap on the test, whereas failure to find a difference would also have serious implications for the test [29].

Hypotheses Setting

Literature Review

To evaluate known-groups validity, hypotheses based on theoretical and empirical evidence need to be set up and tested. As this is an emergent field of research, studies on the predictors of eHealth literacy are limited, and inconsistent results are common [5]. Nevertheless, it has been argued that inequalities due to sociodemographic factors will affect the use of technology, acquisition of skills, and digital literacy. Conventional evidence, both theoretical and empirical, generally suggests that age, sex, education, and ICT use are associated with the ability to use technology for health, which will in turn potentially link to a person's eHealth literacy [31-33]. Hence, a literature review was undertaken to generate hypotheses about the expected score differences across age, sex, education, and ICT use in relation to the 7 constructs or latent variables (ie, traits that cannot be directly observed or measured) representing the 7 scales of the eHLQ.

Age

People aged above 65 years are less likely than the younger generation to have had the chance to familiarize themselves with ICT either at school or at work [34,35]. Combined with the cognitive, motor, and sensory decline associated with aging, older adults face more barriers to and challenges in using technology for health than their younger counterparts [35-42]. With inadequate skill and ability, older people are more likely to experience computer anxiety [34,38], leading to less interest in using technology for health [34,38,43]. Slower processing of information and reduction of working memory caused by cognitive decline [39] can also lead to difficulty in understanding health concepts. In a systematic review of the use of digital health records among older adults, the 2 main barriers identified were privacy and security and access to, and ability to use, technology and the internet [42]. Hence, it was hypothesized as follows:

- H1: Age is negatively related to the scores on all latent variables representing the 7 scales.

Sex

Technology is traditionally perceived as a male-dominated domain, with men usually reporting higher levels of digital skills than women [32,44]. However, the study by Hargittai and Shafer [45] found no significant difference in the skill of web-based information searching by men and women in actual performance tests. Empirical findings indicated that women were more likely and more inclined to search for health information using the internet [38,46-48], with studies continuing to report that men tend to lag behind women in health knowledge [49-51]. The study by Brouwer et al [52] also found that women recorded a higher participation rate for a web-based health intervention than men, and women were more likely to engage in preventive activities related to health than men [49]. In terms of privacy concerns, no discussion of sex differences could be identified from the studies. These considerations led to the following two hypotheses:

- H2a: Being female is related to higher scores on the 3 latent variables representing the scales *1. Using technology to process health information, 2. Understanding of health concepts and language, and 5. Motivated to engage with digital services*.
- H2b: Sex is not related to score differences on the 4 latent variables representing the other 4 scales.

Education

Many studies have found education to be a predictor of ICT use and skills [32,34,35,38,53]. People with limited literacy, because of their limited ability to read and write, are likely to have less extensive health knowledge [49-51,54,55]. The generally higher-than-average reading level of web-based health information [56,57] may also disadvantage people with limited literacy. In addition, access to digital services to connect with health professionals generally requires some ability to read and write [58]. Besides, studies continue to find that searching for web-based health information, interpreting such information, and making decisions based on it is challenging for people with low literacy [36,59-61]. A further deterrence to using technology for health among people with limited literacy is that they tend to have greater privacy concerns because of mistrust of the internet and limited understanding of its capabilities [62]. Therefore, it was hypothesized as follows:

- H3: Education is positively related to the scores on all latent variables representing the 7 scales.

ICT Use

Higher ICT use is frequently found to be related to better digital skills and higher likelihood of searching for health information and using web-based health information and health apps

[35,43,53,63]. With skills and access, people are more likely to be motivated to adopt and use web-based health resources with ease [34]. Furthermore, frequent use of ICT will also improve skills to deal with digital privacy concerns [62]. However, the relationship between ICT use and health knowledge has been hardly explored in the literature. These findings led to the following hypothesis:

- H4: ICT use is positively related to the scores on the 6 latent variables representing scales 1, 3, 4, 5, 6, and 7 but not to the score on the latent variable representing *2. Understanding of health concepts and language*.

Language

Considering that Australia is a multicultural country, whether the nation's main language—English—was spoken at home was also included for analysis. However, because of the limited studies on the eHealth literacy of ethnic minorities [64] and because no studies could be identified targeting migrants who could speak the main language of their adopted country, no hypothesis was formulated for this group.

Methods

Data Collection

A cross-sectional survey was conducted across 3 health sites in Victoria, Australia, in 2018. The 3 sites included a private primary care medical clinic and a not-for-profit community health organization located in metropolitan areas as well as a private primary care medical clinic in a regional area. These sites were selected because they represented a mix of advantaged, disadvantaged, culturally diverse, metropolitan, and regional areas to ensure that the sample would capture people with different eHealth literacy levels. People attending the health sites were invited to participate if they were aged 18 years or older, with or without any health conditions, and were able to complete the eHLQ in paper-based format, web-based format, or face-to-face interview. The option of offering interviews allowed people with lower literacy who were not interested in reading to feel comfortable to participate in the survey, another strategy to ensure that people with potentially lower eHealth literacy were included. The exclusion criteria included people currently experiencing significant cognitive or mental health issues or too clinically unwell as deemed by their treating health care professionals and those with insufficient fluency in English to complete the survey because no family member or carer was present to assist them. The study was approved by the Deakin University Human Research Ethics Committee (approval number: HEAG-H 146_2017). Potential participants were provided information about the study, including that participation was voluntary. Returning the completed questionnaire was regarded as implied consent.

Demographic data collected for analysis included age, sex (male or female), education (less than secondary school, completed secondary school, certificate or diploma, or completed university or higher), language (spoke English at home or not), and ICT use. The classification of education into 4 categories was somewhat arbitrary. Other Australian studies of eHealth literacy and internet use only included the 3 education categories of secondary school or less, certificate or diploma, and university or higher [5,65]. Given that 17.9% (94/525) of the participants did not complete secondary school in this study (Table 1), it was decided that 4 categories of education would be appropriate. On the basis of existing studies, ICT use generally refers to access to, and use of, digital devices and the internet [43,47,60]. Therefore, ICT use was assessed by 3 survey questions, including number of digital devices used (range 0-4), number of ICT platforms used (range 0-10), and whether the participant had looked for web-based information in the last 3 months (yes or no). The number of digital devices used was determined by the question "Do you use any of the following devices?" with the answers including computer or laptop, mobile phone or smartphone, tablet, and other. The number of ICT platforms used was calculated by the participants' answer to the question "Do you use any of the following to connect with others?" with the answers including email, text message, Facebook, Twitter, Instagram, Snapchat, WhatsApp or WeChat, blogging, forum or chat room, and other.

Table 1. Participant characteristics (N=525).

Characteristics	Value
Age (years), mean (SD; range)	56.8 (18.6; 18-94)
Sex, n (%)	
Female	320 (61)
Male	203 (38.7)
Education, n (%)	
Less than secondary school	94 (17.9)
Completed secondary school	106 (20.2)
Certificate or diploma	141 (26.9)
Completed university or higher	175 (33.3)
Spoke English at home, n (%)	363 (69.1)
Ownership of digital device (a person may have more than one device), n (%)	
Computer or laptop	372 (71.2)
Mobile phone or smartphone	459 (87.4)
Tablet	241 (45.9)
Other	6 (1.1)
Average number of digital devices owned, mean (SD; range)	2.1 (0.9; 0-4)
Use of digital communication platform (a person may use more than one platform), n (%)	
Email	394 (75)
Text message	398 (75.8)
Facebook	266 (50.7)
Twitter	30 (5.7)
Instagram	104 (19.8)
Snapchat	51 (9.7)
WhatsApp or WeChat	112 (21.3)
Blogging	15 (2.9)
Forum or chat room	26 (5)
Other	9 (1.7)
Average number of digital platforms used, mean (SD; range)	2.7 (1.8; 0-10)
Looked for web-based information in the last 3 months, n (%)	392 (74.4)

Statistical Analysis

To evaluate known-groups validity, an important prerequisite for hypothesis testing is evidence of measurement equivalence or invariance across the groups [66], which refers to the stability of measurement across the different groups [67,68]. Measurement nonequivalence can occur when the characteristics of certain groups or grouping variables that are irrelevant to the construct being measured affect how people respond to the measurement [68]. Hence, group differences cannot be satisfactorily established if measurement invariance across group is not examined [69,70].

To evaluate measurement invariance, DIF is a common statistical observation that signals whether an item is functioning differentially across the grouping variables [71,72]. The presence of DIF indicates that there is a direct effect from a grouping variable on an item net of the association between the grouping variable and the latent construct. As such, the item is not measuring what it is intended to measure, and the estimated group differences are biased [68,73]. There are 2 types of DIF: uniform DIF and nonuniform DIF. Uniform DIF occurs when a group scores consistently and systematically higher or lower on a specific item than the other groups across all levels of ability, whereas nonuniform DIF is detected when the probability of endorsing an item among the groups varies across different ability levels [68,72,74]. According to the *Standards*, the main concern is uniform DIF because it can lead to "systematically different responses to a particular item" [27].

To ensure that DIF was considered, analysis using the multiple indicators multiple causes (MIMIC) model approach was chosen for this study. The MIMIC model is a type of structural equation modeling (SEM), which contains a *measurement model* that

describes the relationship of the latent variables and their observed variables and a *structural model* that describes the links among the latent variables [75]. Therefore, taking the SEM approach to known-groups analysis with latent variables will account for measurement error in the outcome variables [76]. This approach also "allows simultaneous factor analysis and regression of factor scores on covariates for the comparison of item functioning across groups, while accounting for differences in several personal characteristics" [77]. In a MIMIC model, covariates (represented by group membership) can be categorical or continuous such that continuous variables (eg, age) do not have to be divided into arbitrary groups as in other statistical methods [77,78]. The MIMIC model approach has also been shown to have superior performance even with small or uneven group sample sizes compared with using other approaches [70,74,79]. As the DIF and known-groups analysis for this study involved both categorical and continuous variables, and small and uneven sample sizes for certain groups existed, the MIMIC model was considered suitable. Although the MIMIC model approach only tests for uniform DIF, rather than nonuniform DIF, the method was still considered appropriate because uniform DIF is more likely to occur than nonuniform DIF [80], and the main purpose of collecting DIF evidence, according to the *Standards*, is to identify systematically different responses [27].

Furthermore, a Bayesian approach was used for the MIMIC model in this study. It has been argued that the Bayesian approach is a better reflection of substantive theories because it is less restrictive; in addition, it does not rely on data with normal distribution and performs well with a small sample size [80,81]. A few studies have pointed to several advantages of using Bayesian SEM modeling over the traditional *frequentist* approaches such as maximum likelihood [80,81]. Of particular relevance to this study, these studies have highlighted the benefit that Bayesian SEM offers for investigating covariances among item residuals and potential cross-loadings that can be hypothesized to be *approximately* zero rather than exactly zero as in traditional SEM [82]. The same strategy was used here by including uniform DIF in the model.

For a Bayesian MIMIC model, informative small variance priors are applied to the DIF paths. A prior variance of 0.01 or 0.02 means that 95% of the variation lies within the ranges of ±0.20 or ±0.28 [81,82]. In addition, informative priors are also given to the residual covariances using the inverse-Wishart distribution, "a standard prior distribution for covariance matrices in Bayesian analysis" [81]. The application involves testing a model with a large enough *df* of the inverse-Wishart distribution and gradually lowering the *df* parameter to find a more flexible model [82]. As such, several models with different informative priors are usually tested and compared to identify the model of interest, which is the model that is not rejected by the data and can be considered closest to the frequentist model that fits well enough [82]. Model fit in the Bayesian approach is evaluated by the posterior predictive P value (*PPP*) and 95% CI for the difference between the observed and replicated chi-square values. *PPP*<.05 and positive 95% CI indicate a misfit, and *PPP* of approximately .50 and a value of zero falling close to the middle of the 95% CI indicates an excellent fit [81]. Models can further be compared by examining model convergence and discrepancy information criterion with quicker convergence (potential scale reduction being consistently less than 1.05) and lower discrepancy information criterion value to be chosen as the model of interest [82].

For the MIMIC model, age, sex, education, ICT use, and language were included as covariates. However, given that research indicated that age, sex, and education were also associated with ICT use and skills [31,32,43], ICT use was also tested as a mediator. As such, a mediated MIMIC model—a model that allows for identifying DIF and group differences as well as providing insights into the underlying mechanism [83]—was set up for this study.

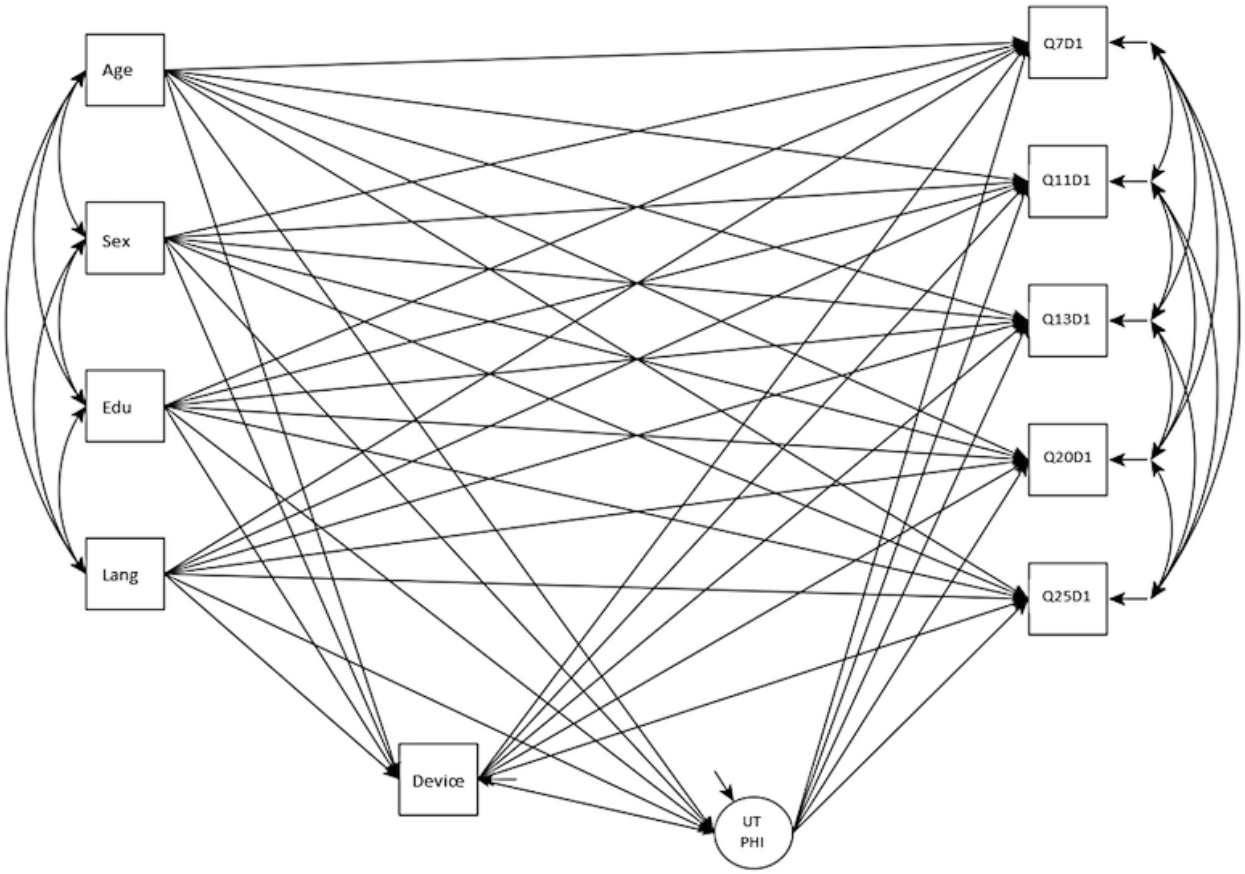
Given that there were 3 indicators (digital devices, ICT platforms, and search for web-based information) for ICT use, a final setup for the MIMIC model was to identify the single best indicator for ICT use. It has been argued that the single best indicator is sufficient for developing theoretically sophisticated models [84]. This was determined by running a Bayesian one-factor confirmatory factor analysis model of ICT use, with no prior, using the 3 indicators to find the one with the highest factor loading to represent ICT use. The result indicated an excellent model fit (*PPP*=.49, 95% CI for the difference between observed and replicated chi-square values –11.81 to 12.25), and number of devices was identified with the highest loading (0.81) on ICT use (Multimedia Appendix 1). Hence, *device* was used to represent ICT use in the final MIMIC model. The mediated DIF model using scale 1 of the eHLQ as an example is shown in Figure 1. As there is no consensus on sample size for Bayesian MIMIC model testing, which may range from 300 to more than 500 [79,85], a minimum sample size of 500 was estimated.

Descriptive statistics were conducted using SPSS, version 25.0 (IBM Corporation) [86], and the Bayesian MIMIC model testing was run using Mplus software version 8.3 [87]. For model selection, a sequence of one-factor models for each of the 7 factors was fitted to the data by varying the informative priors for *df* of the inverse-Wishart distribution=200, 150, 100, 80, and 60. The results would inform the use of the prior for residual covariance for the MIMIC model testing. Next, 3 models using the chosen prior combined with informative priors for DIF paths (variance=0.01, 0.015, and 0.02) were fitted to the data. Model estimation was performed with 50,000 iterations.

To determine the group differences, direct effect, indirect effect, and total effect produced in the Mplus outputs were examined. A significant total effect indicated significant group differences. A significant direct effect indicated that group differences existed independent of any mediating effect, whereas a significant indirect effect indicated that group differences were mediated through ICT use. Mplus produced the results of one-tailed P values and indicated that P<.025 was significant. However, because the hypotheses for age, sex, education, and ICT use were directional, P<.05 was considered significant, whereas P<.025 remained significant for either a positive or negative effect for language because no hypothesis was set a priori. A further calculation of mediation proportion was undertaken to gain deeper understanding of the extent of

mediation. Mediation proportion refers to the portion of effect on an outcome explained by an intermediate variable. It is calculated by dividing the indirect effect by the total effect [88]. It should be noted that computation of mediation proportion is considered not appropriate if the total effect is too small, that is, less than ±0.2, or in case of inconsistent mediation, that is, opposite signs for the estimates. If mediation proportion is at least 0.8, complete mediation can be claimed. As it was cautioned that statistical significance was sensitive to sample size and the effects should not be evaluated simply based on statistical significance, the significant group differences were further investigated by examining the size of the effect estimates to determine if the effects were practically significant or could be ignored because the size of the effect would have no appreciable bearing on the interpretation of group differences [89].

Figure 1. Bayesian multiple indicators multiple causes model for known-groups validity and differential item functioning testing with scale 1 of the eHealth Literacy Questionnaire as an example. Output from Mplus [87]: Age: range 18-94 years; Device: information and communication technology use represented by the number of devices used (range 0-4); Edu: Education: 1=less than secondary school, 2=completed secondary school, 3=certificate or diploma, and 4=completed university or higher; Lang: Language spoken at home: 0=English and 1=other languages; Q7D1, Q11D1, Q13D1, Q20D1, and Q25D1: eHealth Literacy Questionnaire items; Sex: 0=male and 1=female; UTPHI: eHealth Literacy Questionnaire scale 1: Using technology to process health information.



For the evaluation of DIF, a significant direct effect from the covariate on the observed variable, that is, the questionnaire item, indicated the presence of DIF [90-92], and one-tailed P<.025 was considered significant because no directional hypotheses were set up for DIF. If DIF was identified, the prior-posterior predictive P value (*PPPP*) from the model of interest needed to be examined. The *PPPP* is a value used "for the evaluation of hypotheses specifying small variance priors for the parameters of interest" [93]. It is about whether the informative priors for DIF can be considered approximate zero. If *PPPP*>.05, that is, it can be considered nonsignificant, the estimates of the DIF are considered approximate zero and are thus ignorable. Hence, if the model of interest has a variance prior of 0.01 and the *PPPP* is nonsignificant, then estimates within the range of ±0.20 could be considered ignorable [94].

Results

Participant Characteristics

A total of 525 responses were included for analysis. The mean age of the participants was 56.8 (SD 18.6) years. Of the 525 participants, 320 (61%) were women, 175 (33.3%) had a university education, and 162 (30.9%) spoke a language other than English at home. Ownership of digital devices was generally high, but of the 525 participants, 66 (12.6%) did not have a mobile phone and 153 (28.8%) did not have a computer or laptop, whereas 133 (25.6%) did not search for any web-based information (see Table 1 for participant characteristics). The scale scores are shown in Table 2. The results showed that the participants seemed to have relatively good knowledge about their health conditions (scale *2. Understanding of health concepts and language*: mean 2.95, SD 0.41), but they might

not always use technology for health (*1. Using technology to process health information*: mean 2.59, SD 0.61 and *5. Motivated to engage with digital services*: mean 2.63, SD 0.55). Although the participants were generally comfortable with the privacy and security of digital health systems (*4. Feel safe and in control*: mean 2.83, SD 0.49), they were less likely to consider that the systems met their individual needs (*7. Digital services that suit individual needs*: mean 2.43, SD 0.57).

Table 2. eHealth Literacy Questionnaire scale scores (N=525; score range 1-4).

Scale	Value, mean (SD)	Missing data
1. Using technology to process health information	2.59 (0.61)	0
2. Understanding of health concepts and language	2.95 (0.41)	0
3. Ability to actively engage with digital services	2.65 (0.68)	1
4. Feel safe and in control	2.83 (0.49)	5
5. Motivated to engage with digital services	2.63 (0.55)	0
6. Access to digital services that work	2.64 (0.45)	1
7. Digital services that suit individual needs	2.43 (0.57)	11

DIF Influence

Modeling testing identified *df*=60 as the prior for residual covariance, with subsequent testing of the 3 models all achieving good fit with similar results (Multimedia Appendix 2). Hence, the most restrictive model with prior variance for DIF path of 0.01 was chosen as the model of interest. The *PPP*s of the 7 scales ranged from .32 to .38, and all *PPPP*s were nonsignificant.

With the selected model of interest, significant direct effects were found for 2 items indicating possible DIF. However, both estimates were within the range of ±0.2; therefore, they were considered ignorable [94] (Multimedia Appendix 3). Thus, the results indicated no or ignorable DIF influence of ICT use, age, sex, education, and language on the scores of the 35 eHLQ items.

Known-Groups Validity

Mixed Evidence

The evidence on the relations of eHLQ scores to other variables based on known-groups validity is mixed, with some of the hypotheses supported (see Table 3 for estimated effects).

Table 3. Estimated effects of age, sex, education, language, and information and communication technology (ICT) use (device) on the 7 eHealth literacy latent variables.

eHealth literacy latent variable and eHealth Literacy Questionnaire scale	Total effect[a]	Direct effect[a]	Indirect effect[a]	Mediation proportion
Age[b]				
1. Using technology to process health information	*–0.32 (0.08)*[c,d]	*–0.22 (0.08)*	*–0.10 (0.02)*	0.31
2. Understanding of health concepts and language	–0.05 (0.12)	–0.01 (0.12)	–0.05 (0.03)	N/A[e]
3. Ability to actively engage with digital services	*–0.37 (0.07)*	*–0.26 (0.07)*	*–0.12 (0.02)*	0.32
4. Feel safe and in control	–0.01 (0.10)	0.01 (0.10)	–0.02 (0.03)	N/A
5. Motivated to engage with digital services	*–0.21 (0.09)*	–0.13 (0.09)	*–0.08 (0.03)*	0.38
6. Access to digital services that work	–0.08 (0.10)	–0.03 (0.10)	*–0.05 (0.03)*	N/A
7. Digital services that suit individual needs	*–0.21 (0.09)*	–0.13 (0.10)	*–0.08 (0.03)*	0.38
Sex[f]				
1. Using technology to process health information	–0.04 (0.05)	–0.04 (0.05)	–0.01 (0.02)	N/A
2. Understanding of health concepts and language	0.01 (0.07)	0.01 (0.07)	–0.00 (0.01)	N/A
3. Ability to actively engage with digital services	–0.06 (0.05)	–0.05 (0.05)	–0.01 (0.02)	N/A
4. Feel safe and in control	0.04 (0.06)	0.04 (0.06)	0.00 (0.01)	N/A
5. Motivated to engage with digital services	*–0.12 (0.06)*	*–0.11 (0.06)*	–0.00 (0.01)	N/A
6. Access to digital services that work	–0.01 (0.06)	–0.08 (0.06)	–0.00 (0.01)	N/A
7. Digital services that suit individual needs	–0.09 (0.06)	–0.09 (0.06)	–0.00 (0.01)	N/A
Education[g]				
1. Using technology to process health information	*0.22 (0.09)*	0.09 (0.09)	*0.13 (0.03)*	0.59
2. Understanding of health concepts and language	0.18 (0.13)	0.12 (0.13)	0.06 (0.04)	N/A
3. Ability to actively engage with digital services	*0.25 (0.08)*	0.11 (0.08)	*0.14 (0.03)*	0.56
4. Feel safe and in control	–0.03 (0.11)	–0.06 (0.11)	0.03 (0.03)	N/A
5. Motivated to engage with digital services	0.12 (0.10)	0.02 (0.10)	*0.11 (0.03)*	N/A
6. Access to digital services that work	–0.04 (0.11)	–0.11 (0.11)	*0.07 (0.03)*	N/A
7. Digital services that suit individual needs	0.11 (0.10)	0.01 (0.10)	*0.10 (0.03)*	N/A
Language[h]				
1. Using technology to process health information	–0.02 (0.05)	0.03 (0.05)	*–0.06 (0.02)*	N/A
2. Understanding of health concepts and language	*–0.15 (0.07)*	–0.12 (0.07)	–0.03 (0.02)	N/A
3. Ability to actively engage with digital services	–0.09 (0.05)	–0.03 (0.05)	*–0.07 (0.02)*	N/A
4. Feel safe and in control	–0.09 (0.06)	–0.08 (0.06)	–0.01 (0.02)	N/A
5. Motivated to engage with digital services	–0.01 (0.06)	0.04 (0.06)	*–0.05 (0.02)*	N/A
6. Access to digital services that work	–0.02 (0.06)	0.02 (0.06)	*–0.03 (0.02)*	N/A
7. Digital services that suit individual needs	–0.03 (0.06)	0.02 (0.06)	*–0.04 (0.02)*	N/A
ICT use (device)[i]				
1. Using technology to process health information	N/A	*0.38 (0.07)*	N/A	N/A
2. Understanding of health concepts and language	N/A	0.18 (0.11)	N/A	N/A
3. Ability to actively engage with digital services	N/A	*0.42(0.06)*	N/A	N/A
4. Feel safe and in control	N/A	0.09 (0.10)	N/A	N/A
5. Motivated to engage with digital services	N/A	*0.31 (0.08)*	N/A	N/A
6. Access to digital services that work	N/A	*0.20 (0.09)*	N/A	N/A

eHealth literacy latent variable and eHealth Literacy Questionnaire scale	Total effect[a]	Direct effect[a]	Indirect effect[a]	Mediation proportion
7. Digital services that suit individual needs	N/A	*0.30 (0.08)*	N/A	N/A

[a]Standardized estimates reported.

[b]Age: range 18-94 years.

[c]Posterior SD for estimates shown in parentheses.

[d]Italicized values indicated statistically significant differences with P<.05 for information and communication technology use (device), age, sex, and education and P<.025 for language.

[e]N/A: not applicable. For age, sex, education and language, *not applicable* is due to inconsistent mediation, total effect less than ±0.2, or lack of indirect effect [89]; for information and communication technology use (device), *not applicable* is due to the fact that it is treated as the mediator.

[f]Sex code: 0=male and 1=female.

[g]Education code: 1=less than secondary school, 2=completed secondary school, 3=certificate or diploma, and 4=completed university or higher.

[h]Language code: 0=spoke English at home and 1=spoke other language at home.

[i]Information and communication technology use (device): number of devices used, range 0-4.

Age

H1: Age is negatively related to the scores on all latent variables representing the 7 scales.

This hypothesis was supported for only 4 of the expected latent variables. Being older was most strongly related to lower scores in *3. Ability to actively engage with digital services*, with a total effect of –0.37 (posterior SD 0.07; P<.001). Age also had quite a strong negative effect on *1. Using technology to process health information*, with a total effect of –0.32 (posterior SD 0.08; P<.001). A total effect of –0.21 (posterior SD 0.09) was found for both *5. Motivated to engage with digital services* and *7. Digital services that suit individual needs*, with P=.01 and P=.02, respectively. For all 4 latent variables with significant total effect, approximately two-thirds was a direct effect.

Sex

H2a: Being female is related to higher scores on the 3 latent variables representing the scales *1. Using technology to process health information*, *2. Understanding of health concepts and language*, and *5. Motivated to engage with digital services*.

H2b: Sex is not related to score differences on the 4 latent variables representing the other 4 scales.

H2a was not supported, whereas H2b was supported. Sex was not related to the score differences in any of the latent variables. Although a significant total effect was found for the scale *5. Motivated to engage with digital services*, the estimate was –0.12, which was less than ±0.2 and was considered a *too small* effect [89] and therefore ignorable. It was also noted that the mediating effect of ICT use was 0 or close to 0 for all 7 latent variables.

Education

H3: Education is positively related to the scores on all latent variables representing the 7 scales.

This hypothesis was supported only for 2 of the 7 expected latent variables. Higher education was associated with higher scores in the latent variables representing *1. Using technology to process health information* (total effect 0.22, posterior SD 0.09; P=.01), with 59% of the effect mediated by ICT use, and *3. Ability to actively engage with digital services* (total effect 0.25, posterior SD 0.08; P<.001), with 56% of the effect mediated by ICT use.

ICT Use

H4: ICT use is positively related to the scores on the 6 latent variables representing scales 1, 3, 4, 5, 6, and 7 but not to the score on the latent variable representing *2. Understanding of health concepts and language*.

This hypothesis was supported except for the latent variable representing *4. Feel safe and in control*. Higher ICT use had the strongest relation to higher scores in *3. Ability to actively engage with digital services*, with a direct effect of 0.42 (posterior SD 0.06; P<.001), followed by *1. Using technology to process health information*, with a direct effect of 0.38 (posterior SD 0.07; P<.001). The other latent variables with significant positive effect included *5. Motivated to engage with digital services* (direct effect 0.31, posterior SD 0.08; P<.001), *6. Access to digital services that work* (direct effect 0.20, posterior SD 0.09; P=.02), and *7. Digital services that suit individual needs* (direct effect 0.30, posterior SD 0.08; P<.001).

Language

No group differences were found between the participants who spoke English and those who spoke a language other than English at home. Although a significant total effect was found for the latent variable representing *2. Understanding of health concepts and language*, the effect size was ignorable because the estimate of –0.15 was considered too small [89].

Discussion

Principal Findings

This study used a Bayesian mediated MIMIC model approach to collect evidence based on relations to other variables to evaluate the eHLQ as a tool to understand eHealth literacy needs, using data collected from known-groups validity in the Australian community health context. Hypotheses for the expected score differences for age, sex, education, ICT use, and speaking English at home or not were supported for some of the eHealth literacy latent variables represented by the relevant eHLQ scales but not all. The results also confirmed measurement invariance across 5 demographic groups. This is important because the presence of measurement invariance

indicates that when the eHLQ is applied to compare population groups with different demographic compositions in the Australian health setting, unbiased estimates of mean group differences are obtained, which provide reliable data to researchers, clinicians, and policy makers.

Although the *Standards*, the authoritative validity-testing reference, suggests that the expected score differences among groups can be an important source of evidence for validity [27], it was noted during hypotheses setting in this study that empirical findings on predictors of eHealth literacy are still limited and can be inconsistent. This is likely due to the lack of consensus on what is being measured and the use of a limited range of tools to measure the concept. Hence, the hypotheses for known-groups validity in this study need to be interpreted with caution. In contrast, this study is the first to explore group differences on eHealth literacy evaluated as a concept of 7 domains instead of simply focusing, as in other studies, on seeking or evaluating health information or using eHealth services.

This study is based on a contemporary statistical method, the Bayesian mediated MIMIC model, rather than methods frequently used in previous and current eHealth literacy research specifically and psychometric research in general. This could also be one of the reasons why the findings from this study are somewhat different from those of current studies. Common practice in testing for known-groups validity uses statistical methods such as the independent sample one-tailed or two-tailed *t* test or analysis of variance or relevant nonparametric tests to establish group differences. However, these methods do not take into account DIF as a potential confounding factor. Without establishing measurement invariance across demographic groups, any apparent group differences detected cannot be ascertained. Apart from being able to detect group differences as well as DIF in using a MIMIC model, this SEM approach accounts for measurement errors, and the use of the Bayesian approach also allows for an evaluation that is a better reflection of the real world, whereas the inclusion of a mediator adds valuable information to the underlying mechanism of the group differences detected. Unlike *t* tests or analysis of variance, age was not divided into arbitrary groups but was treated as a continuous variable. Therefore, the results of this study, in fact, provide new and unbiased insights into the predictors of eHealth literacy.

The findings in this study indicated that being older had lower scores in 4 scales, with *3. Ability to actively engage with digital services* having the strongest effect, followed by *1. Using technology to process health information*, *5. Motivated to engage with digital services*, and *7. Digital services that suit individual needs*. This result is generally consistent with existing studies. However, an interesting result from the mediated MIMIC model is that most of the effect of age was not mediated by ICT use, indicating that ICT use may have a limited role in these 4 domains of eHealth literacy. A possible factor is the cognitive, motor, and sensory decline associated with aging, as frequently suggested in the literature [35-42]. Such findings may also imply that simply providing ICT training may not be adequate with regard to enhancing eHealth literacy among older people, and other interventions are necessary. The focus on computer skills and ICT training as the main mechanism of current eHealth literacy interventions [95] not only may not be effective for older people, but other domains of eHealth literacy are also likely to be overlooked.

Another result of interest is the relationship between education and the eHealth literacy domains because education is only positively associated with the scales *1. Using technology to process health information* and *3. Ability to actively engage with digital services*. Unlike in the case of age, most of the effects were mediated by ICT use, confirming the role of ICT use in enhancing certain aspects of eHealth literacy. As such, providing ICT education and training is likely to benefit people with lower education in increasing their ability to engage with digital services and enhancing use of technology for health. Such a finding suggests that other efforts are required to address the eHealth literacy needs concerning privacy, motivation, or access to suitable digital services. In contrast, why education was not related to *2. Understanding of health concepts and language* in the digital context warrants further investigation.

Sex having no relationship with any of the eHLQ scale scores suggests that sex may not be a good predictor of eHealth literacy. This may be due to the narrowing gap in education between the sexes in recent years and the fact that technology use has become an indispensable part of modern-day life for most people. Although the hypothesis of ICT use is mostly supported, the results show that ICT use is not associated with the eHealth literacy domain of feeling safe and in control. This again reiterates that technical skills in ICT training are inadequate to address all eHealth literacy needs. These findings also call for assessment of eHealth literacy using unbiased multidimensional questionnaires such as the eHLQ so that eHealth literacy needs can be clearly identified and addressed.

Although the aim of this study is to collect evidence based on relations to other variables, the statistical method used also established the robustness of the internal structure of the eHLQ in terms of measurement invariance across 5 demographic groups. The presence of DIF indicates that a questionnaire item is not measuring what is intended, and thus the resulting scores may be biased [68]. Given the issue of health disparities among different sociodemographic groups, the presence of DIF in patient-reported outcome measures may lead to inaccurate interpretation of scores and inappropriate health care decisions [96]. Hence, providing clear evidence of measurement invariance across the 5 demographic groups is an important finding. The Danish validity testing also found no evidence of influence of age and sex on the item scores in the Danish setting [21]. With the accumulating evidence on DIF, users of the eHLQ in similar Australian and Danish contexts can be assured that the mean scores obtained from the eHLQ can be interpreted properly to address the different needs of different groups. As such, the validity evidence collected in this study supports the eHLQ as a tool to understand eHealth literacy needs and helps to inform the development of fit-for-purpose health interventions [26].

Limitations

A limitation of this study is that the hypotheses were based on limited empirical findings of eHealth literacy predictors such

that the results may need to be interpreted differently. Although the MIMIC model approach has the advantage of evaluating both group differences and DIF simultaneously, only uniform DIF was tested, and nonuniform DIF was not investigated. Nevertheless, uniform DIF is a more important threat to validity because it can lead to systematic score differences on account of group characteristics [27]. In examining evidence pertaining to relations to other variables, this study only focused on data collected from known-groups validity. Although the relationship with other instruments could not be tested because of the lack of an instrument to measure the same constructs, whether the test scores can be generalized to other settings or contexts is another potential source of evidence on relations to other variables [27] that warrants investigation in future studies. Because of limited resources and because the health sites were always busy during the data collection period, it was difficult to gauge the number of people coming through the door, and no response rate was recorded. However, the participants' characteristics demonstrated a generally well-represented sample. Unlike the Danish validity testing, which included the general population in various community settings, this study only focused on the community health setting, but data were nevertheless collected from different geographic locations, including both metropolitan and regional areas. Further testing of the eHLQ in other settings and cultures is required for the accumulation of validity evidence for the eHLQ.

Conclusions

With health resources and services increasingly delivered through digital technologies, eHealth literacy has become an essential capability in the digital age. This study provides robust validity evidence of the eHLQ in the Australian community health setting. The evidence demonstrates that the tool can be used by health care providers and policy makers to gain unbiased and valuable insights into people's diverse eHealth literacy needs so that tailored health interventions can be effectively developed in similar settings. The eHLQ can also be used to align the demand of any eHealth system with the eHealth literacy needs of users to optimize access and use of digital health among users and promote health equity.

Acknowledgments

The authors thank Dr Mukesh Haikerwal AC, Ms Jenny Ktenidis, and Ms Rori Plaza of Altona North Medical Group and Cirqit Health; Ms Janine Scott and Ms Olive Aumann of Carrington Health; and Dr Ewa Piejko, Dr Adel Asaid, Dr Remon Eskander, and Dr Poate Radrekusa of St Anthony Family Medical Practice for their generous support. The authors also thank Professor Lisa Hanna of Deakin University for her input and Dr Polina Putrik, Visiting Fellow at Deakin University, for her assistance in data collection. RHO was funded in part through the National Health and Medical Research Council of Australia Principal Research Fellowship #APP1155125.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Bayesian structural equation modelling for information and communication technology use.
[DOC File , 65 KB - jmir_v23i10e30243_app1.doc]

Multimedia Appendix 2
Bayesian model fit information of the eHealth Literacy Questionnaire for the testing of known-groups validity and differential item functioning of age, sex, education, language, and information and communication technology use (device), with model 1 as the model of interest.
[DOC File , 58 KB - jmir_v23i10e30243_app2.doc]

Multimedia Appendix 3
Estimates for the direct effect of eHealth Literacy Questionnaire items on information and communication technology use (device), age, sex, education, and language.
[DOC File , 77 KB - jmir_v23i10e30243_app3.doc]

References

1. Norman CD, Skinner HA. eHealth Literacy: Essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
2. Digital health literacy. World Health Organization. 2017. URL: https://www.who.int/global-coordination-mechanism/working-groups/digital_hl.pdf [accessed 2020-12-03]
3. Kayser L, Kushniruk A, Osborne RH, Norgaard O, Turner P. Enhancing the effectiveness of consumer-focused health information technology systems through eHealth literacy: A framework for understanding users' needs. JMIR Hum Factors 2015 May 20;2(1):e9 [FREE Full text] [doi: 10.2196/humanfactors.3696] [Medline: 27025228]

4. Monkman H, Kushniruk A. eHealth literacy issues, constructs, models, and methods for health information technology design and evaluation. Knowled Manag E-Learn 2015;7(4):541-549 [FREE Full text] [doi: 10.34105/j.kmel.2015.07.036]
5. Richtering SS, Hyun K, Neubeck L, Coorey G, Chalmers J, Usherwood T, et al. eHealth Literacy: Predictors in a population with moderate-to-high cardiovascular risk. JMIR Hum Factors 2017 Jan 27;4(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.6217] [Medline: 28130203]
6. Griebel L, Enwald H, Gilstad H, Pohl A, Moreland J, Sedlmayr M. eHealth literacy research-Quo vadis? Inform Health Soc Care 2018 Dec;43(4):427-442. [doi: 10.1080/17538157.2017.1364247] [Medline: 29045164]
7. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
8. Dale JG, Lüthi A, Skaraas B, Rundereim T, Dale B. Testing measurement properties of the Norwegian version of electronic health literacy scale (eHEALS) in a group of day surgery patients. J Multidiscip Healthc 2020;13:241-247 [FREE Full text] [doi: 10.2147/JMDH.S242985] [Medline: 32210568]
9. Gazibara T, Cakic J, Cakic M, Pekmezovic T, Grgurevic A. eHealth and adolescents in Serbia: psychometric properties of eHeals questionnaire and contributing factors to better online health literacy. Health Promot Int 2019 Aug 01;34(4):770-778. [doi: 10.1093/heapro/day028] [Medline: 29800141]
10. Neter E, Brainin E, Baron-Epel O. The dimensionality of health literacy and eHealth literacy. The European Health Psychologist. 2015. URL: https://www.ehps.net/ehp/index.php/contents/article/view/824 [accessed 2020-12-02]
11. Soellner R, Huber S, Reder M. The concept of ehealth literacy and its measurement. J Media Psychol 2014 Jan;26(1):29-38. [doi: 10.1027/1864-1105/a000104]
12. Richtering SS, Morris R, Soh S, Barker A, Bampi F, Neubeck L, et al. Examination of an eHealth literacy scale and a health literacy scale in a population with moderate to high cardiovascular risk: Rasch analyses. PLoS One 2017;12(4):e0175372 [FREE Full text] [doi: 10.1371/journal.pone.0175372] [Medline: 28448497]
13. Hyde LL, Boyes AW, Evans T, Mackenzie LJ, Sanson-Fisher R. Three-factor structure of the eHealth literacy scale among magnetic resonance imaging and computed tomography outpatients: A confirmatory factor analysis. JMIR Hum Factors 2018 Feb 19;5(1):e6 [FREE Full text] [doi: 10.2196/humanfactors.9039] [Medline: 29459356]
14. Paige SR, Miller MD, Krieger JL, Stellefson M, Cheong J. Electronic health literacy across the lifespan: Measurement invariance study. J Med Internet Res 2018 Jul 09;20(7):e10434 [FREE Full text] [doi: 10.2196/10434] [Medline: 29986848]
15. Sudbury-Riley L, FitzPatrick M, Schulz PJ. Exploring the Measurement Properties of the eHealth Literacy Scale (eHEALS) Among Baby Boomers: A Multinational Test of Measurement Invariance. J Med Internet Res 2017 Feb 27;19(2):e53 [FREE Full text] [doi: 10.2196/jmir.5998] [Medline: 28242590]
16. Karnoe A, Kayser L. How is eHealth literacy measured and what do the measurements tell us? A systematic review. Knowled Manag E-Learn 2015;7(4):576-600 [FREE Full text] [doi: 10.34105/j.kmel.2015.07.038]
17. Norman C. eHealth literacy 2.0: Problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec 23;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
18. Buchbinder R, Batterham R, Elsworth G, Dionne CE, Irvin E, Osborne RH. A validity-driven approach to the understanding of the personal and societal burden of low back pain: Development of a conceptual and measurement model. Arthritis Res Ther 2011;13(5):R152 [FREE Full text] [doi: 10.1186/ar3468] [Medline: 21933393]
19. Norgaard O, Furstand D, Klokker L, Karnoe A, Batterham R, Kayser L, et al. The e-health literacy framework: A conceptual framework for characterizing e-health users and their interaction with e-health systems. Knowled Manag E-Learn 2015;7(4):522-540 [FREE Full text] [doi: 10.34105/j.kmel.2015.07.035]
20. Osborne RH, Batterham RW, Elsworth GR, Hawkins M, Buchbinder R. The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). BMC Public Health 2013 Jul 16;13:658 [FREE Full text] [doi: 10.1186/1471-2458-13-658] [Medline: 23855504]
21. Kayser L, Karnoe A, Furstrand D, Batterham R, Christensen KB, Elsworth G, et al. A multidimensional tool based on the ehealth literacy framework: Development and initial validity testing of the eHealth Literacy Questionnaire (eHLQ). J Med Internet Res 2018 Feb 12;20(2):e36 [FREE Full text] [doi: 10.2196/jmir.8371] [Medline: 29434011]
22. Holt KA, Overgaard D, Engel LV, Kayser L. Health literacy, digital literacy and eHealth literacy in Danish nursing students at entry and graduate level: A cross sectional study. BMC Nurs 2020 Apr 10;19(1):22 [FREE Full text] [doi: 10.1186/s12912-020-00418-w] [Medline: 32308559]
23. Villadsen SF, Hadi H, Ismail I, Osborne RH, Ekstrøm CT, Kayser L. ehealth literacy and health literacy among immigrants and their descendants compared with women of Danish origin: a cross-sectional study using a multidimensional approach among pregnant women. BMJ Open 2020 May 07;10(5):e037076 [FREE Full text] [doi: 10.1136/bmjopen-2020-037076] [Medline: 32385065]
24. Holt KA, Karnoe A, Overgaard D, Nielsen SE, Kayser L, Røder ME, et al. Differences in the level of electronic health literacy between users and nonusers of digital health services: An exploratory survey of a group of medical outpatients. Interact J Med Res 2019 Apr 05;8(2):e8423 [FREE Full text] [doi: 10.2196/ijmr.8423] [Medline: 30950809]
25. Foley K, Freeman T, Ward P, Lawler A, Osborne R, Fisher M. Exploring access to, use of and benefits from population-oriented digital health services in Australia. Health Promot Int 2020 Dec 26:1105-1115. [doi: 10.1093/heapro/daaa145] [Medline: 33367568]

26. Cheng C, Elsworth GR, Osborne RH. Co-designing eHealth and Equity Solutions: Application of the Ophelia (Optimizing Health Literacy and Access) Process. Front Public Health 2020;8:604401 [FREE Full text] [doi: 10.3389/fpubh.2020.604401] [Medline: 33330344]
27. American Educational Research Association, American Pscyhological Association, National Council on Measurement in Education. Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association; 2014.
28. McCoach D, Gable R, Madura J. Evidence Based on Relations to Other Variables: Bolstering the Empirical Validity Arguments for Constructs. Instrument Development in the Affective Domain. 3rd Ed. New York: Springer; 2013:209-248.
29. Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull 1955;52(4):281-302 [FREE Full text] [doi: 10.1037/h0040957]
30. Fayers P, Machin D. Quality of Life: The Assessment, Analysis and Interpretation of Patient-Reported Outcomes. Second Edition. Chichester: John Wiley & Sons; 2007.
31. Bodie GD, Dutta MJ. Understanding health literacy for strategic health marketing: eHealth literacy, health disparities, and the digital divide. Health Mark Q 2008;25(1-2):175-203. [doi: 10.1080/07359680802126301] [Medline: 18935884]
32. Dodel M, Mesch G. Inequality in digital skills and the adoption of online safety behaviors. Info Commun Soc 2018 Feb 06;21(5):712-728. [doi: 10.1080/1369118x.2018.1428652]
33. Han JY, Wise M, Kim E, Pingree R, Hawkins RP, Pingree S, et al. Factors associated with use of interactive cancer communication system: An application of the comprehensive model of information seeking. J Comput Mediat Commun 2010 Apr;15(3):367-388 [FREE Full text] [doi: 10.1111/j.1083-6101.2010.01508.x] [Medline: 21760702]
34. Charness N, Boot WR. Aging and information technology use. Curr Dir Psychol Sci 2009 Oct 01;18(5):253-258. [doi: 10.1111/j.1467-8721.2009.01647.x]
35. van Deursen AJ, van Dijk JA, Peters O. Rethinking Internet skills: The contribution of gender, age, education, Internet experience, and hours online to medium- and content-related Internet skills. Poetics 2011 Apr;39(2):125-144. [doi: 10.1016/j.poetic.2011.02.001]
36. Agree EM, King AC, Castro CM, Wiley A, Borzekowski DL. "It's got to be on this page": Age and cognitive style in a study of online health information seeking. J Med Internet Res 2015 Mar 24;17(3):e79 [FREE Full text] [doi: 10.2196/jmir.3352] [Medline: 25831483]
37. Arning K, Ziefle M. Different perspectives on technology acceptance: The role of technology type and age. In: Holzinger A, Miesenberger K, editors. HCI and Usability for e-Inclusion. Berlin, Heidelberg: Springer; 2009:20-41.
38. Bidmon S, Terlutter R, Röttl J. What explains usage of mobile physician-rating apps? Results from a web-based questionnaire. J Med Internet Res 2014 Jun 11;16(6):e148 [FREE Full text] [doi: 10.2196/jmir.3122] [Medline: 24918859]
39. Echt KV, Burridge AB. Predictors of reported internet use in older adults with high and low health literacy: The role of socio-demographics and visual and cognitive function. Phys Occupat Ther Geriat 2011 Mar 02;29(1):23-43. [doi: 10.3109/02703181.2010.547657]
40. Austin R. EHealth literacy for older adults - Part II. Ania-Caring. 2012. URL: https://www.thefreelibrary.com/Ehealth+literacy+for+older+adults--part+II.-a0339637785 [accessed 2019-10-03]
41. Laberge JC, Scialfa CT. Predictors of web navigation performance in a life span sample of adults. Hum Factors 2005;47(2):289-302. [doi: 10.1518/0018720054679470] [Medline: 16170939]
42. Sakaguchi-Tang DK, Bosold AL, Choi YK, Turner AM. Patient portal use and experience among older adults: Systematic review. JMIR Med Inform 2017 Oct 16;5(4):e38 [FREE Full text] [doi: 10.2196/medinform.8092] [Medline: 29038093]
43. Arcury TA, Sandberg JC, Melius KP, Quandt SA, Leng X, Latulipe C, et al. Older adult internet use and ehealth literacy. J Appl Gerontol 2020 Feb;39(2):141-150 [FREE Full text] [doi: 10.1177/0733464818807468] [Medline: 30353776]
44. Park YJ. Do men and women differ in privacy? Gendered privacy and (in)equality in the Internet. Comput Hum Behav 2015 Sep;50:252-258. [doi: 10.1016/j.chb.2015.04.011]
45. Hargittai E, Shafer S. Differences in actual and perceived online skills: The role of gender*. Soc Sci Q 2006 Jun;87(2):432-448. [doi: 10.1111/j.1540-6237.2006.00389.x]
46. Escoffery C. Gender similarities and differences for e-Health behaviors among U.S. adults. Telemed J E Health 2018 May;24(5):335-343. [doi: 10.1089/tmj.2017.0136] [Medline: 28813630]
47. James DC, Harville C. eHealth literacy, online help-seeking behavior, and willingness to participate in mhealth chronic disease research among African Americans, Florida, 2014-2015. Prev Chronic Dis 2016 Nov 17;13:E156 [FREE Full text] [doi: 10.5888/pcd13.160210] [Medline: 27854421]
48. ÖZen H, Yaşi˙n B. Gender differences in the use of internet for health information search. Ege Acad Rev 2011 Apr 01;11(2):229. [doi: 10.21121/eab.2011219567]
49. Beier ME, Ackerman PL. Determinants of health knowledge: an investigation of age, gender, abilities, personality, and interests. J Pers Soc Psychol 2003 Feb;84(2):439-448. [Medline: 12585815]
50. Parmenter K, Waller J, Wardle J. Demographic variation in nutrition knowledge in England. Health Educ Res 2000 Apr;15(2):163-174 [FREE Full text] [doi: 10.1093/her/15.2.163] [Medline: 10751375]

51. Wilkinson AV, Vasudevan V, Honn SE, Spitz MR, Chamberlain RM. Sociodemographic characteristics, health beliefs, and the accuracy of cancer knowledge. J Cancer Educ 2009;24(1):58-64 [FREE Full text] [doi: 10.1080/08858190802664834] [Medline: 19259867]
52. Brouwer W, Oenema A, Raat H, Crutzen R, de Nooijer J, de Vries NK, et al. Characteristics of visitors and revisitors to an internet-delivered computer-tailored lifestyle intervention implemented for use by the general public. Health Educ Res 2010 Aug;25(4):585-595. [doi: 10.1093/her/cyp063] [Medline: 19897515]
53. Børøsund E, Cvancarova M, Ekstedt M, Moore SM, Ruland CM. How user characteristics affect use patterns in web-based illness management support for patients with breast and prostate cancer. J Med Internet Res 2013 Mar 01;15(3):e34 [FREE Full text] [doi: 10.2196/jmir.2285] [Medline: 23454601]
54. Hocking A, Laurence C, Lorimer M. Patients' knowledge of their chronic disease - the influence of socio-demographic characteristics. Aust Fam Physician 2013 Jun;42(6):411-416 [FREE Full text] [Medline: 23781550]
55. Macabasco-O'Connell A, DeWalt DA, Broucksou KA, Hawk V, Baker DW, Schillinger D, et al. Relationship between literacy, knowledge, self-care behaviors, and heart failure-related quality of life among patients with heart failure. J Gen Intern Med 2011 Sep;26(9):979-986 [FREE Full text] [doi: 10.1007/s11606-011-1668-y] [Medline: 21369770]
56. Cheng C, Dunn M. Health literacy and the internet: A study on the readability of Australian online health information. Aust N Z J Public Health 2015 Aug;39(4):309-314. [doi: 10.1111/1753-6405.12341] [Medline: 25716142]
57. McInnes N, Haglund BJ. Readability of online health information: Implications for health literacy. Inform Health Soc Care 2011 Dec;36(4):173-189. [doi: 10.3109/17538157.2010.542529] [Medline: 21332302]
58. DeMonte CM, DeMonte WD, Thorn BE. Future implications of eHealth interventions for chronic pain management in underserved populations. Pain Manag 2015;5(3):207-214. [doi: 10.2217/pmt.15.9] [Medline: 25971644]
59. Birru MS, Monaco VM, Charles L, Drew H, Njie V, Bierria T, et al. Internet usage by low-literacy adults seeking health information: An observational analysis. J Med Internet Res 2004 Sep 03;6(3):e25 [FREE Full text] [doi: 10.2196/jmir.6.3.e25] [Medline: 15471751]
60. Milne RA, Puts MT, Papadakos J, Le LW, Milne VC, Hope AJ, et al. Predictors of high eHealth literacy in primary lung cancer survivors. J Cancer Educ 2015 Dec;30(4):685-692. [doi: 10.1007/s13187-014-0744-5] [Medline: 25355524]
61. Werts N, Hutton-Rogers L. Barriers to achieving e-health literacy. Am J Health Sci 2013 Aug 14;4(3):115-120. [doi: 10.19030/ajhs.v4i3.8007]
62. Li X. Understanding eHealth literacy from a privacy perspective: eHealth literacy and digital privacy skills in American disadvantaged communities. Am Behav Scient 2018 Jul 09;62(10):1431-1449. [doi: 10.1177/0002764218787019]
63. Jensen JD, King AJ, Davis LA, Guntzviller LM. Utilization of internet technology by low-income adults: The role of health literacy, health numeracy, and computer assistance. J Aging Health 2010 Sep;22(6):804-826. [doi: 10.1177/0898264310366161] [Medline: 20495159]
64. Chesser A, Burke A, Reyes J, Rohrberg T. Navigating the digital divide: A systematic review of eHealth literacy in underserved populations in the United States. Inform Health Soc Care 2016;41(1):1-19. [doi: 10.3109/17538157.2014.948171] [Medline: 25710808]
65. Hanna K, Sambrook P, Armfield J, Brennan D. Internet use, online information seeking and knowledge among third molar patients attending public dental services. Aust Dent J 2017 Sep 31;62(3):323-330 [FREE Full text] [doi: 10.1111/adj.12509] [Medline: 28241385]
66. Gerdes KE, Geiger JM, Lietz CA, Wagaman MA, Segal EA. Examination of known-groups validity for the Empathy Assessment Index (EAI): Differences in EAI scores between social service providers and service recipients. J Society Soc Work Res 2012 Jan;3(2):94-112. [doi: 10.5243/jsswr.2012.7]
67. Meade AW, Lautenschlager GJ. A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. Organizat Res Methods 2016 Jun 29;7(4):361-388. [doi: 10.1177/1094428104268027]
68. Osterlind S, Everson H. In: Liao T, editor. Differential Item Functioning. London: SAGE Publication; 2009.
69. Ward JT, Ray JV, Fox KA. Exploring differences in self-control across sex, race, age, education, and language: Considering a bifactor MIMIC model. J Crim Just 2018 May;56:29-42. [doi: 10.1016/j.jcrimjus.2017.09.006]
70. Woods CM, Oltmanns TF, Turkheimer E. Illustration of MIMIC-Model DIF Testing with the schedule for nonadaptive and adaptive personality. J Psychopathol Behav Assess 2009;31(4):320-330 [FREE Full text] [doi: 10.1007/s10862-008-9118-9] [Medline: 20442793]
71. Teresi JA, Ramirez M, Lai J, Silver S. Occurrences and sources of Differential Item Functioning (DIF) in patient-reported outcome measures: Description of DIF methods, and review of measures of depression, quality of life and general health. Psychol Sci Q 2008;50(4):538 [FREE Full text] [Medline: 20165561]
72. Sireci SG, Rios JA. Decisions that make a difference in detecting differential item functioning. Edu Res Eval 2013 Apr;19(2-3):170-187. [doi: 10.1080/13803611.2013.767621]
73. Walker CM. What's the DIF? Why differential item functioning analyses are an important part of instrument development and validation. J Psychoedu Assess 2011 May 19;29(4):364-376 [FREE Full text] [doi: 10.1177/0734282911406666] [Medline: 25607262]

74. Teresi JA, Jones RN. Methodological issues in examining measurement equivalence in patient reported outcomes measures: Methods overview to the two-part series, "Measurement Equivalence of the Patient Reported Outcomes Measurement Information System (PROMIS) Short Forms". Psychol Test Assess Model 2016;58(1):37-78 [FREE Full text] [Medline: 28983448]
75. Byrne B. Structural Equation Modeling with Mplus. First edition. Oxfordshire: Routledge; 2013.
76. Chang C, Gardiner J, Houang R, Yu Y. Comparing multiple statistical software for multiple-indicator, multiple-cause modeling: an application of gender disparity in adult cognitive functioning using MIDUS II dataset. BMC Med Res Methodol 2020 Nov 12;20(1):275 [FREE Full text] [doi: 10.1186/s12874-020-01150-4] [Medline: 33183226]
77. Gallo JJ, Anthony JC, Muthén BO. Age differences in the symptoms of depression: A latent trait analysis. J Gerontol 1994 Nov;49(6):251-264. [doi: 10.1093/geronj/49.6.p251] [Medline: 7963280]
78. Grayson DA, Mackinnon A, Jorm AF, Creasey H, Broe GA. Item bias in the Center for Epidemiologic Studies Depression Scale: Effects of physical disorders and disability in an elderly community sample. J Gerontol B Psychol Sci Soc Sci 2000 Sep;55(5):273-282. [doi: 10.1093/geronb/55.5.p273] [Medline: 10985292]
79. Jamali J, Ayatollahi SM, Jafari P. The effect of small sample size on measurement equivalence of psychometric questionnaires in MIMIC model: A simulation study. Biomed Res Int 2017;2017:7596101 [FREE Full text] [doi: 10.1155/2017/7596101] [Medline: 28713828]
80. Finch WH, French BF. Bayesian modelling of differential item functioning: Type I error and power rates in the presence of non-normal ability distributions, impact, and anchor set contamination. Int J Quantit Res Edu 2013;1(4):341. [doi: 10.1504/ijqre.2013.058305]
81. Muthén B, Asparouhov T. Bayesian structural equation modeling: A more flexible representation of substantive theory. Psychol Methods 2012 Sep;17(3):313-335. [doi: 10.1037/a0026802] [Medline: 22962886]
82. Asparouhov T, Muthén B, Morin AJ. Bayesian structural equation modeling with cross-loadings and residual covariances. J Manag 2015 Jun 30;41(6):1561-1577. [doi: 10.1177/0149206315591075]
83. Cheng Y, Shao C, Lathrop QN. The Mediated MIMIC Model for understanding the underlying mechanism of DIF. Educ Psychol Meas 2016 Feb;76(1):43-63 [FREE Full text] [doi: 10.1177/0013164415576187] [Medline: 29795856]
84. Hayduk LA, Littvay L. Should researchers use single indicators, best indicators, or multiple indicators in structural equation models? BMC Med Res Methodol 2012 Oct 22;12:159 [FREE Full text] [doi: 10.1186/1471-2288-12-159] [Medline: 23088287]
85. Hancock GR. Effect size, power, and sample size determination for structured means modeling and mimic approaches to between-groups hypothesis testing of means on a single latent construct. Psychometrika 2001 Sep;66(3):373-388. [doi: 10.1007/bf02294440]
86. IBM SPSS Statistics for Mac, Version 25. Armonk, NY: IBM Corp; 2017.
87. Muthén L, Muthén B. Mplus Statistical Analysis with Latent Variables: User's Guide (Mac Version 8). Los Angeles, CA: Muthén & Muthén; 2017:1-944.
88. Ditlevsen S, Christensen U, Lynch J, Damsgaard MT, Keiding N. The mediation proportion: a structural equation approach for estimating the proportion of exposure effect on outcome explained by an intermediate variable. Epidemiology 2005 Jan;16(1):114-120. [doi: 10.1097/01.ede.0000147107.76079.07] [Medline: 15613954]
89. Kenny D. Mediation. 2019 May 27. URL: http://davidakenny.net/cm/mediate.htm [accessed 2020-12-02]
90. Bruyneel L, Li B, Squires A, Spotbeen S, Meuleman B, Lesaffre E, et al. Bayesian Multilevel MIMIC Modeling for studying measurement invariance in cross-group comparisons. Med Care 2017 Apr;55(4):25-35. [doi: 10.1097/MLR.0000000000000164] [Medline: 25170773]
91. Jensen R, King-Kallimanis B, Sexton E, Reeve B, Moinpour C, Potosky A, et al. Measurement properties of PROMIS Sleep Disturbance short forms in a large, ethnically diverse cancer cohort. Psychol Test Assess Model 2016:353-370 [FREE Full text]
92. Muthén B, Asparouhov T. Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. Mplus Web Notes: No. 4 Version 5. 2002. URL: https://www.statmodel.com/download/webnotes/CatMGLong.pdf [accessed 2019-03-26]
93. Hoijtink H, van de Schoot R. Testing small variance priors using prior-posterior predictive P values. Psychol Methods 2018 Sep;23(3):561-569. [doi: 10.1037/met0000131] [Medline: 28368177]
94. Asparouhov T, Muthén B. Prior-posterior predictive P-values. Mplus Web Notes: No. 22. 2017. URL: https://www.statmodel.com/download/PPPP.pdf [accessed 2019-10-02]
95. Watkins I, Xie B. eHealth literacy interventions for older adults: A systematic review of the literature. J Med Internet Res 2014 Nov 10;16(11):e225 [FREE Full text] [doi: 10.2196/jmir.3318] [Medline: 25386719]
96. Teresi JA, Fleishman JA. Differential item functioning and health assessment. Qual Life Res 2007;16 Suppl 1:33-42. [doi: 10.1007/s11136-007-9184-6] [Medline: 17443420]

Abbreviations

DIF: differential item functioning

eHEALS: eHealth Literacy Scale
eHLQ: eHealth Literacy Questionnaire
ICT: information and communication technology
MIMIC: multiple indicators multiple causes
PPP: posterior predictive P value
PPPP: prior-posterior predictive P value
SEM: structural equation modeling

Edited by R Kukafka; submitted 06.05.21; peer-reviewed by P Nguyen, S Weiwei, P Le; comments to author 28.06.21; revised version received 07.07.21; accepted 27.07.21; published 14.10.21.

Please cite as:
Cheng C, Elsworth G, Osborne RH
Validity Evidence Based on Relations to Other Variables of the eHealth Literacy Questionnaire (eHLQ): Bayesian Approach to Test for Known-Groups Validity
J Med Internet Res 2021;23(10):e30243
URL: https://www.jmir.org/2021/10/e30243
doi:10.2196/30243
PMID:34647897

Original Paper

Writing Practices Associated With Electronic Progress Notes and the Preferences of Those Who Read Them: Descriptive Study

Thomas H Payne[1], MD; Carolyn Keller[1], MD; Pallavi Arora[1], MD; Allison Brusati[1], MD; Jesse Levin[1], MD; Monica Salgaonkar[1], MHA; Xi Li[2], MS; Jennifer Zech[3], MSc; A Fischer Lees[1], MD

[1]Department of Medicine, University of Washington School of Medicine, Seattle, WA, United States

[2]University of Southern California, Los Angeles, CA, United States

[3]Columbia University, New York, NY, United States

Corresponding Author:
Thomas H Payne, MD
Department of Medicine
University of Washington School of Medicine
Box 359780
325 Ninth Ave
Seattle, WA, 98104-2499
United States
Phone: 1 206 744 1824
Email: tpayne@uw.edu

Abstract

Background: Hospital progress notes can serve as an important communication tool. However, they are criticized for their length, preserved content, and for the time physicians spend writing them.

Objective: We aimed to describe hospital progress note content, writing and reading practices, and the preferences of those who create and read them prior to the implementation of a new electronic health record system.

Methods: Using a sample of hospital progress notes from 1000 randomly selected admissions, we measured note length, similarity of content in successive daily notes for the same patient, the time notes were signed and read, and who read them. We conducted focus group sessions with note writers, readers, and clinical leaders to understand their preferences.

Results: We analyzed 4938 inpatient progress notes from 418 authors. The average length was 886 words, and most were in the Assessment & Plan note section. A total of 29% of notes (n=1432) were signed after 4 PM. Notes signed later in the day were read less often. Notes were highly similar from one day to the next, and 26% (23/88) had clinical risk associated with the preserved content. Note content of the highest value varied according to the reader's professional role.

Conclusions: Progress note length varied widely. Notes were often signed late in the day when they were read less often and were highly similar to the note from the previous day. Measuring note length, signing time, when and by whom notes are read, and the amount and safety of preserved content will be useful metrics for measuring how the new electronic health record system is used, and can aid improvements.

(J Med Internet Res 2021;23(10):e30165) doi:10.2196/30165

KEYWORDS

electronic documentation; electronic health records; hospital progress notes; copy-paste; EHR; patient records; workflow; human factors; clinical communication; physician communication; hospital

Introduction

Inpatient progress notes can serve as an important communication tool across the physician team, nurses, therapists, consultants, and the patient. Because safety problems that occur in hospitals can be traced to communication lapses [1], progress notes are vital to achieve safer care. However, some feel that few read progress notes, and that they therefore no longer serve the purpose of communication and are now primarily billing documents [2]. Many physicians feel they spend too much time writing progress notes [3,4] and employ methods to shorten writing time. These include copy-paste [5-8] and extensive templating with "note bloat," which can introduce

error [9-11], harm patients [12], and make it difficult for note readers to separate current from outdated content [13].

We have used our current electronic health record (EHR) system at the University of Washington (UW) Medical Center and the Harborview Medical Center for 17 years with little change in the format of the inpatient progress note used in the Medicine service. In preparation for the transition to a new inpatient EHR system, we conducted this analysis of progress notes on inpatient medical services to take stock of current practices. The purpose of this study was to describe current hospital progress note writing and reading practices, as well as the preferences of those who create progress notes and those who read them.

Methods

Overview

This work retrieved its data from the inpatient general medicine services at the UW Medical Center and the Harborview Medical Center, which are major teaching hospitals of the UW with approximately 35,000 combined admissions annually. The EHR system was installed in 2003, and the transition from paper to electronic notes occurred in 2006 using Cerner Millennium (Cerner Corp). Nearly all progress notes pertaining to these inpatient services are typed using the Clinical Notes Editor, based on templates that automatically import patient-specific data such as medication lists, vital signs, and laboratory results [14]. Daily progress notes are required by hospital bylaws and are mostly written by residents (usually interns) and attending hospitalists.

Selection of Notes

We randomly selected 1000 patient admissions to the general Medicine service of UW Medical Center and the Harborview Medical Center between July 1, 2016, and June 30, 2017. The Medicine services at both facilities share the same progress note template. We excluded patients admitted to subspecialty services with unique progress note templates (oncology, cardiology, and geriatrics). For each admission, we extracted the data for all daily progress notes as shown in Textbox 1 from the analytical data repository (Enterprise Data Warehouse, Caradigm), which contains a subset of EHR data extracted for research. Progress notes were identified by the title “Medicine - Inpt Record,” the note type used within our system for Medicine service daily progress notes. Given that this study was focused on practices around the use of progress notes by those who create and those who read them, we excluded all notes other than progress notes (admission notes, procedure notes, consult notes, interim summaries, discharge summaries, and other notes).

Textbox 1. Metadata obtained for each progress note.

Full text of the note and other data listed were gathered for each note. Authenticators are supervising physicians whose note cosignature finalizes the note.

Metadata obtained:

- Note date and time
- Note title
- Note authors and authenticators
- Note identifier
- Note text
- Note action log (provided by the electronic health record system), which includes the following:
- Action (eg, perform, transcribe, modify, sign and CC/review, verify)
- Performed by (name)
- Performed date and time
- Action status
- Comment
- Proxy personnel
- Requested by (name)

Note Analysis

We deidentified each note using published methods [15] and stored them securely. Using Python scripts written for this project, we determined the total number of words in each note and in each note section (Identification/Chief Concern, Interim History, etc) (Textbox 2). We also determined when notes were signed; before progress notes are signed, they cannot be viewed except by the author.

Textbox 2. Progress note sections.

Sections are automatically created using the template used to create Medicine progress notes.

HOSPITAL DAY

IDENTIFICATION/CHIEF CONCERN

INTERVAL HISTORY

INPATIENT PROBLEM LIST

ALLERGIES

SCHEDULED MEDICATIONS

PRN MEDICATIONS

PHYSICAL EXAM

LABS

- Micro

IMAGING

ASSESSMENT & PLAN

- Fluids/electrolytes/nutrition
- Prophylaxis
- Tubes/lines
- Disposition
- Code status
- Contacts

ATTENDING STATEMENT

- Additional diagnoses

Copy-Paste Analysis

We identified the sequence of daily notes written for each patient during their hospital stay and then determined the percentage of text within each and in the note as a whole that overlapped with the note written on that same patient the previous day using natural language processing methods [16,17].

To determine the clinical implications of copy-paste, we used methods described by Hammond et al [18] to highlight shared content across progress notes from one day compared to the previous day for the same patient. We then used the same 6-point scale as in that paper to rank the clinical importance of copy-paste. Since this required time-consuming manual review by clinicians, we performed this for a subset of notes.

Measuring Note Readership

Note reading practices were analyzed by extracting a note-viewing record using auditing software (P2Sentinel, Cerner Corporation). Each time a note was viewed, the username, user role (resident physician, attending physician, registered nurse, etc), and timestamp were recorded in the auditing database. To determine views within the same hospital day—which have unique potential to communicate the patient's current clinical state and today's plan—we assessed note views within 12 hours after the note was signed. Usernames were used to identify members of the patient's primary team. Views by physicians on the patient's primary team were assumed to be related to the note writing process, and were excluded from the note-reading analysis. Statistical testing was performed using Stata/IC 13 (StataCorp LLC).

Focus Groups

To understand different perspectives on the current use of progress notes, we conducted 3 sets of focus groups: *note authors* (Medicine interns and hospitalists), *note consumers* (nurses, therapists, and consultants who view Medicine progress notes that others had written), and *leaders* (hospital service leaders who form documentation policies and standards). Each focus group was led by a coinvestigator who followed a script and showed PowerPoint slides of the results of the note analysis and the copy-paste analyses. The 5 focus group sessions were recorded, and transcripts were made for all but one of the focus groups.

The UW Institutional Review Board approved this work and designated it as "minimal risk."

Results

From 1000 randomly selected hospital admissions to the Medicine service, we obtained 4938 inpatient progress notes written by 418 authors, an average of 4.9 daily progress notes per patient admission.

The results of the note analysis are summarized in Figures 1-4. The average note length was 886 words (median 827), and most of the length (~500 words) was in the Assessment & Plan section, which comprised on average two-thirds of the note. There was marked variation in note length—some notes contained over 2000 words in the Assessment & Plan section alone. The Interval History and Physical Exam sections were among the shortest sections. These findings reflect the common practice of copying one day's Assessment & Plan into the next day's note and appending each day's assessment to those of the previous days. (While the History and Physical sections are also frequently copied, the lack of appending new information to old information prevents these sections from lengthening over the hospital course.)

Progress note authors signed their notes at various times of the day as shown in Figure 1. A total of 29% (n=1432) of notes were signed after 4 PM, and some were signed as late as 10 PM.

Figure 1. The time at which progress notes were signed by the author.

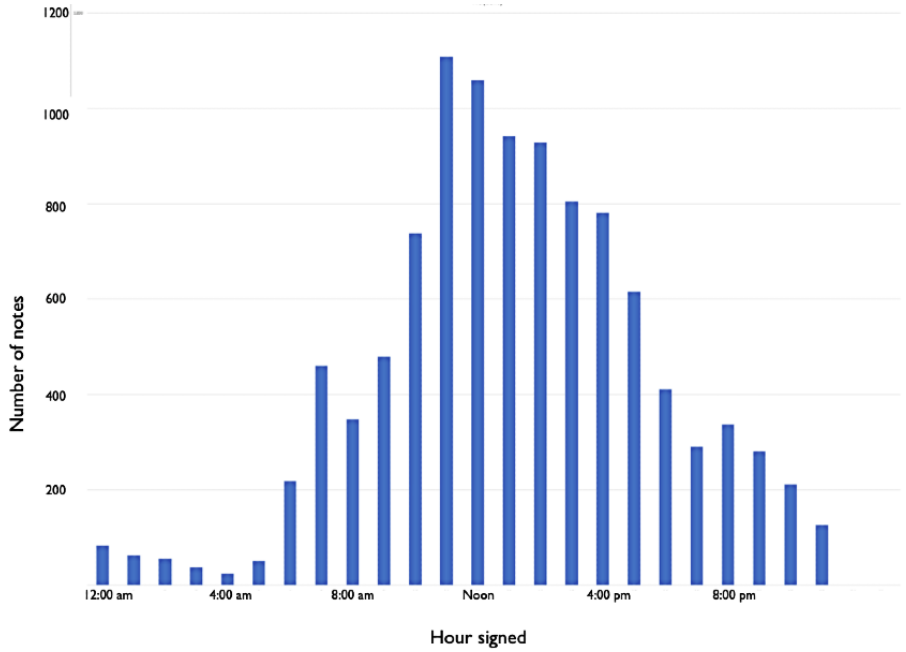


Figure 2. Distribution of note length in words (the programming code used to calculate note length is available from the authors).

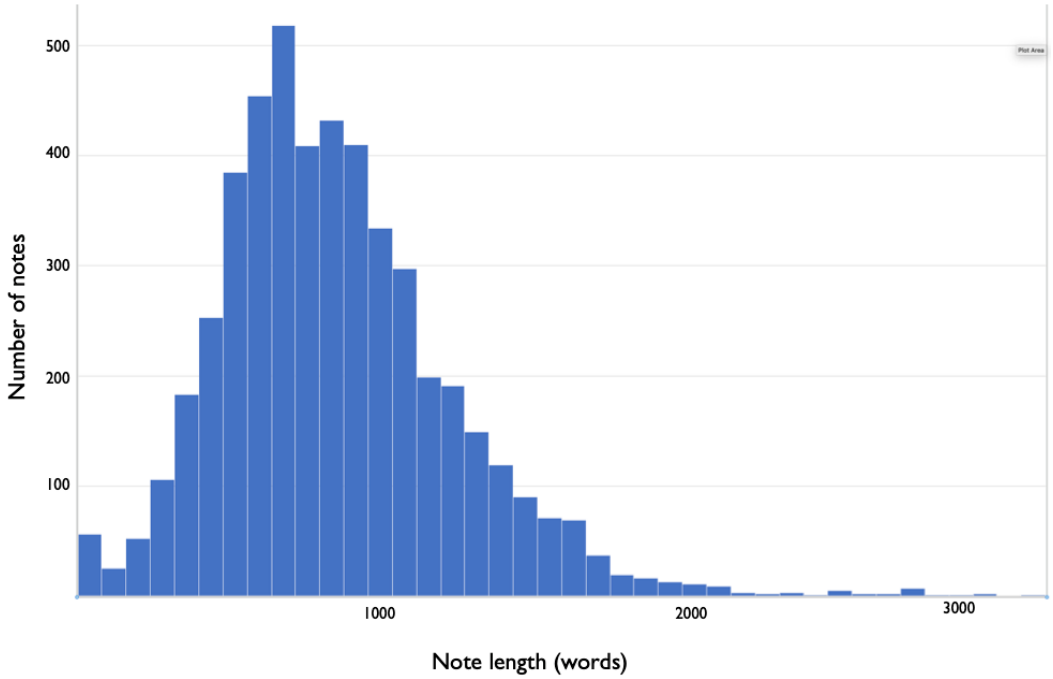


Figure 3. Length of note sections. Box and whisker plots of the number of words in each note section is shown. Box shows IQR. ID: identification.

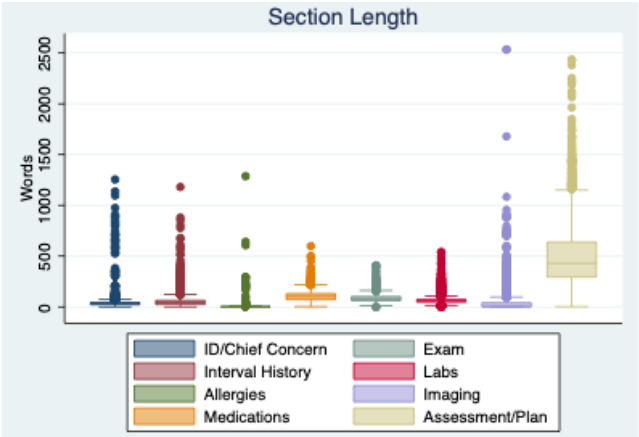
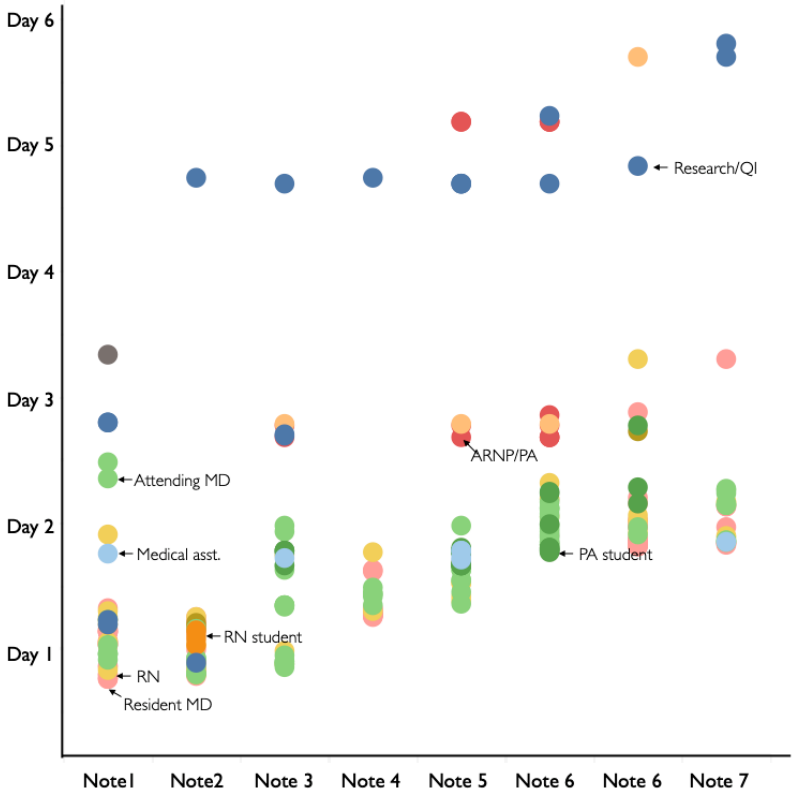


Figure 4. Display of note readership data for 7 notes. Each colored dot indicates who (role) read the note and when over several days. Different colors indicate different roles. Each column of dots shows the readership of one progress note. QI: quality improvement, ARNP: advanced registered nurse practitioner, PA: physician assistant, RN: registered nurse, MD: medical doctor.



Notes had high levels of similarity to the prior day's note on the same patient. The median note similarity was 66% using the methods described above. While note similarity was high for all author types (Figures 5 and 6), it was higher for trainees than attendings (P<.001), and higher when both notes were written by the same author rather than by different authors (P<.001) (Figure 6).

We conducted manual reviews of note pairs to assess the clinical importance of note similarity. Preserved content from one day to the next was visually highlighted, using the CopyFind program [18]. Physician reviewers then assessed the preserved content for risk using the Hammond scale [18] (Textbox 3). The results showed that 26% (23/88) of the pairs were assessed to have minimal or some risk because of human copying; a second set of reviews found 17% (5/29) of the pairs had minimal or some risk in the preserved content. In this sample then, about 1 in 5 notes had clinical risk associated with preserved content, which was very likely the result of copy-paste practices. An example of copying assessed to be of minimal risk (code 4) was including the phrase "Gen Surg to take to the OR today" when this happened the day before. An example of copying regarded to be of some risk (code 5) was when the History section was completely copied from the day before, incurring a legal risk of fraud. In the Plan section of the same note was the phrase "will obtain MRI" copied forward, but the MRI (magnetic resonance imaging) was obtained the day before as evidenced by the results of the MRI appearing elsewhere in the note.

Figure 5. Similarity by author training level. Box and whisker plots of the similarity of notes calculated using natural language processing methods [16,17] are shown. Box shows IQR.

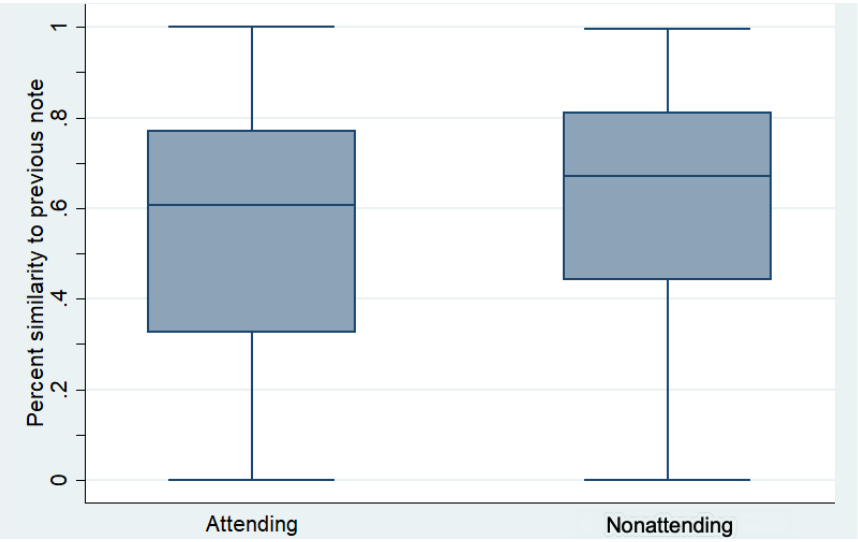
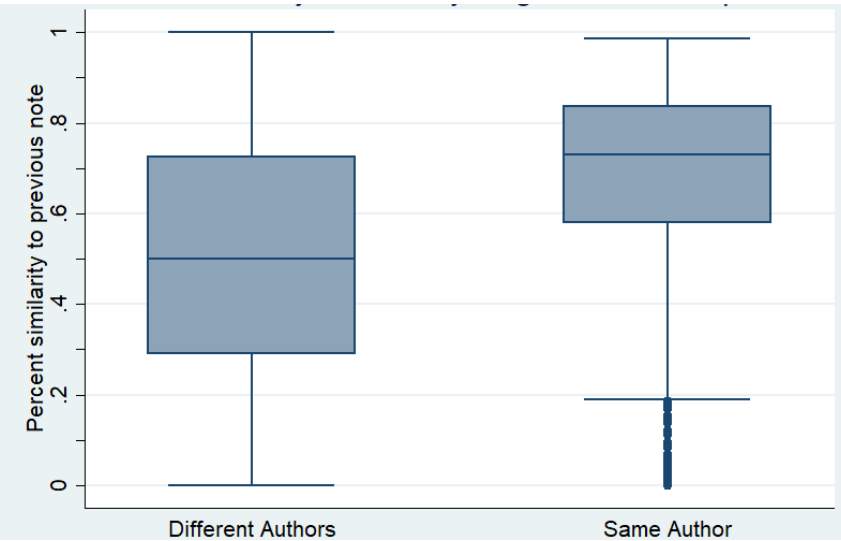


Figure 6. Similarity by original authorship. Box and whisker plots of the similarity of notes calculated using natural language processing methods [16,17] are shown. Box shows IQR.



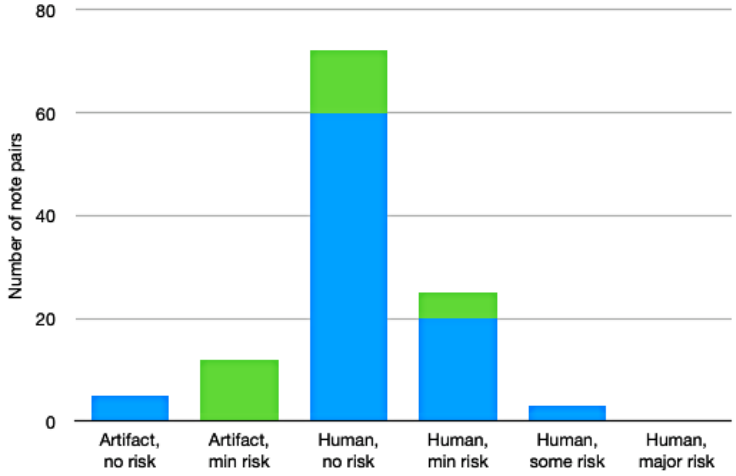
Textbox 3. Risk scale for duplicated material appearing in notes.

The scale below is used to assess the risk associated with duplicated note text, derived from Hammond et al [18] (Figure 7).

Code risk description:

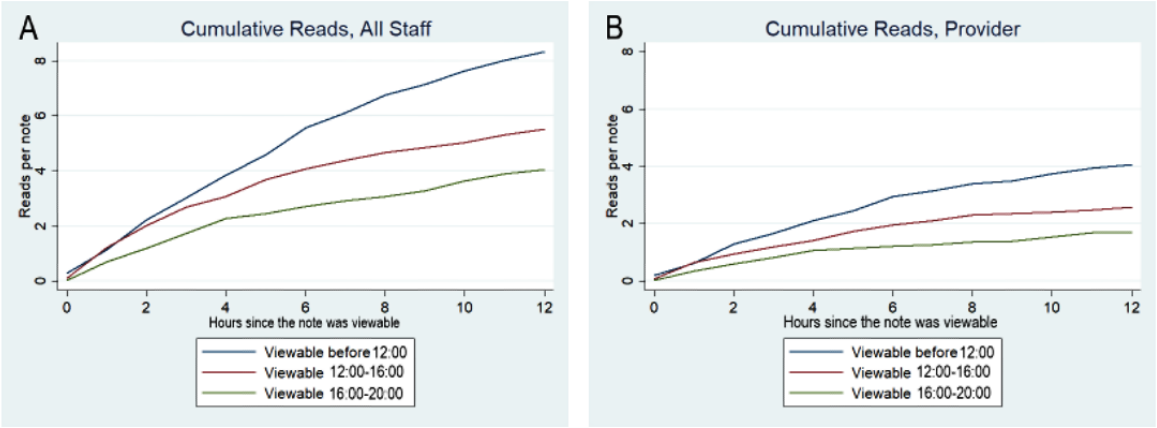
- 1 = Artifact, not misleading, no risk
- 2 = Artifact, minimally misleading, minimal risk
- 3 = Human, not misleading, no risk
- 4 = Human, minimally misleading, minimal risk
- 5 = Human, misleading, some risk
- 6 = Human, clinically misleading, major risk

Figure 7. Copy-paste risk, assessed using the Hammond risk scale [18]. Colors represent judgment of risk assigned by the 2 physicians who analyzed a sample of note pairs. One physician assigned scores to 100 note pairs (blue) and one assigned scores to 25 note pairs (green).



Note-reading practices were analyzed on a random subset of 250 notes (limited by auditing time constraints). The 250 notes in this sample were similar to the overall progress note library: the authors who wrote them were representative of the pool of authors in the larger library, and because they were randomly sampled, other note characteristics (eg, time written, length) were similar to the overall note library. This sample had 4036 note views, an average of 16.1 views per note, which includes many types of professionals beyond bedside nurses and physicians. Figures 8A and 8B describe note readership by note sign time. Notes signed before noon were read more than notes signed between noon and 4 PM (P=.002), and those signed between noon and 4PM were read more than those between 4 PM and 8 PM (P=.05), with successively fewer reads for notes signed between 4 PM and 8 PM. Analysis of note reading by role revealed that more notes were read by providers (physicians, nurse practitioners, physician assistants, medical students) when published before noon (ie, early in the day) (P=.003). Nursing readership was not correlated with the time the notes were signed (P=.95).

Figure 8. Note reads by published time. Plots show the cumulative number of all staff (A) and providers (B) who read notes as a function of the time since the note was viewable (signed).



Focus Groups

Note Authors

We learned that note authors vary widely in how they use progress notes (ie, section of note reviewed, such as Labs, Scheduled Medications, etc), and in the purposes for which they use them, such as regarding them to be a billing document, as a "note to self" to remember important items from one day to the next, and as a communication tool to colleagues. Note authors mentioned this as one reason there is so much heterogeneity in the length and content of inpatient progress notes.

Note Consumers

Nursing and other ancillary service staff said they value medical progress notes and prefer detailed notes. They regarded them to be an important communication tool that helps them understand what is going on with patients and the plan for the day and for the hospital stay. They frequently used the progress note to assist their communication with families. They reported wanting to see more details regarding the disposition plan and preferred to have the Labs and Radiology sections deleted.

Leaders

Leaders believed the collective memory in notes to be important to prevent "reinventing the wheel." They believed that the primary focus of a note should be communication and patient

safety, but that we should also make trainees aware of future billing pressures when they finish training and are able to bill. They felt, as a group, that the medication list is important to include in progress notes.

Readership

Figure 3 shows that notes are read by authorized users from a wide range of disciplines, including registered nurses, physicians, and administrators.

Discussion

Our analysis of a large set of randomly selected hospital progress notes found that notes were often long and highly similar to the prior day's note. Additionally, notes that were signed later were read less frequently; about one-third were signed after 4 PM when their value as a communication tool for other providers is diminished. Given the vast quantity of physician hours spent on their creation, inpatient progress notes have been relatively neglected as an area of study. Limitations in duty hours, increased patient complexity, and decline in physician morale suggest a need to re-evaluate our note-writing practices.

We found broad variation in note-writing and reading practices. Some notes in our sample were over 2000 words (approximately half the length of this paper); this may reflect the common practice of copying one day's assessment into the next day's note and adding each day's findings to those of the previous days. The progress notes often become a running summary of the hospitalization.

Our analysis showed a high degree of similarity between notes on the same patient on successive days. Though methods for measuring note content similarity vary across studies [19], it is interesting that others also noted this [20,21]. Note similarity could occur from either note templating or from copy-paste; we believe that the high similarity in our sample was mainly due to copy-paste, as the most heavily templated sections of our progress note (Allergies, Medications, and Labs) make up just 22% of the average note length. Our focus group with note authors provided further evidence that this was the case.

Note similarity was high across training levels, and while trainee notes had more similarity than attending notes, attending notes were still over 50% similar to the prior day's note. Perhaps equally concerning was a high degree of similarity between notes with distinct authors. While our Medicine community sees the progress note as more of a living, communal document (where the practice of copy-paste is not viewed as flagrant plagiarism), this practice nevertheless is a clear setup for miscommunication and medical error [22,23]. Indeed, manual review revealed that 1 in 5 notes had clinical risk associated with preserved content, which was very likely the result of copy-paste practices. Reassuringly, this was primarily in the minimal-risk category with no examples of major human risk captured in the small sample we reviewed. This may suggest that careful use of copy-paste with a diligent review (defined as a review sufficient to assure that the note is accurate on the day it is written) of the information that is carried forward is not associated with a major risk of error.

Progress notes in our sample were read frequently by nonprimary team members. Readership of timely clinical data (within 12 hours of note signing) was highly sensitive to the time of day the note was signed, with significantly more notes read if signed before noon. This was especially true of physicians, who tend to have day schedules on the acute care services and are thus unlikely to view a note that is published late in the evening (as many notes are). Nurse views were not sensitive to note-signing time, likely reflecting the presence of night-shift nurses, who also want to use the progress note as a means to communicate the plan of care. (However, if a night-shift nurse reads a note, the content from the morning rounds is over 12 hours old.)

In the focus groups, we heard many and different reasons nurses and physicians value progress notes. Some nurses read them soon after they were created even if it was late in the day and appreciated the detail of Medicine progress notes compared with other services. For nurses, these notes served as an important communication tool to help them understand the Medicine team's plan. Most expressed interest in more information tailored to their needs. Note authors—largely house staff and hospitalists—also valued notes, though what they valued differed by group. House staff are charged with tracking details, critical and minor, and used their notes for their "future selves"—as a way to remind them of what is to be done and for the collective memory of what has occurred. For hospitalists, there was wide variation in the attitude toward progress note writing, with some hospitalists favoring a more succinct note that concisely summarizes why the patient remains hospitalized and focuses on changes in patient status and plan for that day, while others (especially those who also worked nights) also focused on the importance of a note as a communication tool. Ultimately, the use of the note as a repository or collective memory of hospital course was a hotly contested topic among the hospitalists depending on their attitude toward and group practice with regards to interim summary writing. This suggests a need for direction from leadership regarding the use and role of medical inpatient progress notes versus interim summaries as collective memory of a patient's hospital course.

Recent literature offers additional perspectives and hopes for progress notes. We know that attendings and house staff physicians differ in their perception of note quality but agree that communication is an important purpose for them [24]. Some senior authors advocated restoring the story to clinical notes [25,26]. If this is an objective, using voice to create notes could fit this into the workflow because it may be faster than typing a more narrative note. Creating a wiki-style note, broadly adopting the APSO (Assessment, Plan, Subjective, Objective) format, utilizing vendor tools to create an ongoing hospital course, or drastically shortening notes were not explored in this study but may be embraced by a segment of physicians.

These results have implications for those who develop electronic documentation systems used in EHRs. Documentation methods that facilitate the completion of notes sooner in the day may result in greater readership, fulfilling an important communication objective for electronic notes. However, the opportunity cost of earlier documentation may be diverting physician hours away from other time-sensitive tasks, such as

contacting specialists or meeting with families. Permitting the simple creation of a summary of the hospital course, outside of the progress note itself, could result in more succinct notes that convey what progress occurred that day, and would be simpler to read. Developing note-writing tools that permit more rapid note creation, tailored to the aptitudes of the note author, such as use of voice [27,28] or scribes, could permit the notes to be signed earlier in the day, resulting in increased readership, which is an important objective of the daily progress note.

Templates can also speed note writing but is done so at the risk of retaining content of the history or physical exam, which is included in some templates. Many clinicians who read notes are most interested in what the note author has to say rather than what is within the template text. Every change to note-writing approaches has potential for side-effects. For example, methods that reduce the time to create notes and that permit them to be completed earlier in the day may also decrease note quality.

There is an important role for hospital leaders in setting the direction for this critical part of clinical care and training [29]. As a result of this work, our community has expressed agreement on increasing the communication value of progress notes, in part by making it possible for them to be completed earlier, and to reduce the time devoted to writing them.

Acknowledgments

We greatly appreciate the assistance of Xi Li, who performed the note content analysis, and of Jacki DeNonno and Johnathan Bubalo, Cerner Corporation, for obtaining the data on note readership.

This work was funded by a University of Washington Department of Medicine Accelerator Grant.

Conflicts of Interest

None declared.

References

1. Schriefer J, Leonard MS. Patient safety and quality improvement: an overview of QI. Pediatr Rev 2012 Aug 01;33(8):353-9; quiz 359. [doi: 10.1542/pir.33-8-353] [Medline: 22855927]
2. Siegler EL, Adelman R. Copy and paste: a remediable hazard of electronic health records. Am J Med 2009 Jun;122(6):495-496. [doi: 10.1016/j.amjmed.2009.02.010] [Medline: 19486708]
3. Sinsky C, Colligan L, Li L, Prgomet M, Reynolds S, Goeders L, et al. Allocation of Physician Time in Ambulatory Practice: A Time and Motion Study in 4 Specialties. Ann Intern Med 2016 Sep 06;165(11):753. [doi: 10.7326/m16-0961]
4. Sulmasy LS, López AM, Horwitch CA, , American College of Physicians Ethics, ProfessionalismHuman Rights Committee. Ethical Implications of the Electronic Health Record: In the Service of the Patient. J Gen Intern Med 2017 Aug;32(8):935-939 [FREE Full text] [doi: 10.1007/s11606-017-4030-1] [Medline: 28321550]
5. Hirschtick RE. A piece of my mind. Copy-and-paste. JAMA 2006 May 24;295(20):2335-2336. [doi: 10.1001/jama.295.20.2335] [Medline: 16720812]
6. Tsou A, Lehmann C, Michel J, Solomon R, Possanza L, Gandhi T. Safe Practices for Copy and Paste in the EHR. Appl Clin Inform 2017 Dec 20;26(01):12-34. [doi: 10.4338/aci-2016-09-r-0150]
7. Thornton JD, Schold JD, Venkateshaiah L, Lander B. Prevalence of Copied Information by Attendings and Residents in Critical Care Progress Notes*. Critical Care Medicine 2013;41(2):382-388. [doi: 10.1097/ccm.0b013e3182711a1c]
8. Weis JM, Levy PC. Copy, paste, and cloned notes in electronic health records: prevalence, benefits, risks, and best practice recommendations. Chest 2014 Mar 01;145(3):632-638. [doi: 10.1378/chest.13-0886] [Medline: 24590024]
9. Turchin A, Goldberg SI, Breydo E, Shubina M, Einbinder JS. Copy/paste documentation of lifestyle counseling and glycemic control in patients with diabetes: true to form? Arch Intern Med 2011 Aug 08;171(15):1393-1394 [FREE Full text] [doi: 10.1001/archinternmed.2011.219] [Medline: 21606091]
10. Thielke S, Hammond K, Helbig S. Copying and pasting of examinations within the electronic medical record. Int J Med Inform 2007 Jun;76 Suppl 1:S122-S128. [doi: 10.1016/j.ijmedinf.2006.06.004] [Medline: 16899403]
11. Yadav S, Kazanji N, Paudel S, Falatko J, Shoichet S, Maddens M, et al. Comparison of accuracy of physical examination findings in initial progress notes between paper charts and a newly implemented electronic health record. J Am Med Inform Assoc 2017 Jan;24(1):140-144 [FREE Full text] [doi: 10.1093/jamia/ocw067] [Medline: 27357831]
12. Roddy JT, Arora VM, Chaudhry SI, Rein LM, Banerjee A, Swenson SL, et al. The prevalence and implications of copy and paste: internal medicine program director perspectives. J Gen Intern Med 2018 Dec 31;33(12):2032-2033 [FREE Full text] [doi: 10.1007/s11606-018-4575-7] [Medline: 30066119]
13. Khorana AA. Physician As Typist. JCO 2010 Aug 20;28(24):3899-3900. [doi: 10.1200/jco.2010.29.4504]
14. Payne TH, Perkins M, Kalus R, Reilly D. The transition to electronic documentation on a teaching hospital medical service. AMIA Annu Symp Proc 2006:629-633 [FREE Full text] [Medline: 17238417]

15. Neamatullah I, Douglass MM, Lehman LH, Reisner A, Villarroel M, Long WJ, et al. Automated de-identification of free-text medical records. BMC Med Inform Decis Mak 2008 Jul 24;8(1):32 [FREE Full text] [doi: 10.1186/1472-6947-8-32] [Medline: 18652655]
16. Rehurek R, Sojka P. Software framework for topic modelling with large corpora. 2010 Presented at: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks; March; Valletta, Malta p. 46-50 URL: https://radimrehurek.com/lrec2010_final.pdf
17. RaRe-Technologies/gensim | gensim – Topic Modelling in Python. GitHub. URL: https://github.com/RaRe-Technologies/gensim [accessed 2021-01-13]
18. Hammond KW, Helbig ST, Benson CC, Brathwaite-Sketoe BM. Are electronic medical records trustworthy? Observations on copying, pasting and duplication. AMIA Annu Symp Proc 2003:269-273 [FREE Full text] [Medline: 14728176]
19. Wrenn JO, Stein DM, Bakken S, Stetson PD. Quantifying clinical narrative redundancy in an electronic health record. Journal of the American Medical Informatics Association 2010 Jan 01;17(1):49-53. [doi: 10.1197/jamia.m3390]
20. Zhang R, Pakhomov S, McInnes BT, Melton GB. Evaluating measures of redundancy in clinical texts. AMIA Annu Symp Proc 2011;2011:1612-1620 [FREE Full text] [Medline: 22195227]
21. Wang MD, Khanna R, Najafi N. Characterizing the Source of Text in Electronic Health Record Progress Notes. JAMA Intern Med 2017 Aug 01;177(8):1212-1213 [FREE Full text] [doi: 10.1001/jamainternmed.2017.1548] [Medline: 28558106]
22. Weir C, Hurdle J, Felgar M, Hoffman J, Roth B, Nebeker J. Direct Text Entry in Electronic Progress Notes. Methods Inf Med 2018 Feb 07;42(01):61-67. [doi: 10.1055/s-0038-1634210]
23. Singh H, Giardina TD, Meyer AND, Forjuoh SN, Reis MD, Thomas EJ. Types and origins of diagnostic errors in primary care settings. JAMA Intern Med 2013 Mar 25;173(6):418-425 [FREE Full text] [doi: 10.1001/jamainternmed.2013.2777] [Medline: 23440149]
24. Stewart E, Kahn D, Lee E, Simon W, Duncan M, Mosher H, et al. Internal medicine progress note writing attitudes and practices in an electronic health record. J Hosp Med 2015 Aug;10(8):525-529. [doi: 10.1002/jhm.2379] [Medline: 26138806]
25. Gantzer HE, Block BL, Hobgood LC, Tufte J. Restoring the Story and Creating a Valuable Clinical Note. Annals of Internal Medicine 2020 Sep 01;173(5):380-382. [doi: 10.7326/m20-0934]
26. Morgan M. Matt Morgan: When bullet points miss the heart. BMJ 2019 Oct 17;366:l5500. [doi: 10.1136/bmj.l5500] [Medline: 31530546]
27. Payne TH, Alonso WD, Markiel JA, Lybarger K, White AA. Using voice to create hospital progress notes: Description of a mobile application and supporting system integrated with a commercial electronic health record. J Biomed Inform 2018 Jan;77:91-96 [FREE Full text] [doi: 10.1016/j.jbi.2017.12.004] [Medline: 29233669]
28. Payne T, Alonso W, Markiel J, Lybarger K, Lordon R, Yetisgen M, et al. Using voice to create inpatient progress notes: effects on note timeliness, quality, and physician satisfaction. JAMIA Open 2018 Oct;1(2):218-226 [FREE Full text] [doi: 10.1093/jamiaopen/ooy036] [Medline: 31984334]
29. Siegler EL. Electronic medical records and hospital progress notes. JAMA 2012 Dec 12;308(22):2336; author reply 2337. [doi: 10.1001/jama.2012.14495] [Medline: 23232889]

Abbreviations

APSO: Assessment, Plan, Subjective, Objective
EHR: electronic health record
MRI: magnetic resonance imaging
UW: University of Washington

Edited by R Kukafka; submitted 03.05.21; peer-reviewed by G Schiff, R Marshall; comments to author 23.05.21; revised version received 13.07.21; accepted 27.07.21; published 06.10.21.

Please cite as:
Payne TH, Keller C, Arora P, Brusati A, Levin J, Salgaonkar M, Li X, Zech J, Lees AF
Writing Practices Associated With Electronic Progress Notes and the Preferences of Those Who Read Them: Descriptive Study
J Med Internet Res 2021;23(10):e30165
URL: https://www.jmir.org/2021/10/e30165
doi:10.2196/30165
PMID:34612825

Original Paper

Transforming Anesthesia Data Into the Observational Medical Outcomes Partnership Common Data Model: Development and Usability Study

Antoine Lamer[1,2,3], PhD; Osama Abou-Arab[4], MD, PhD; Alexandre Bourgeois[5], MSc; Adrien Parrot[2], MD; Benjamin Popoff[6], MD; Jean-Baptiste Beuscart[1], MD, PhD, Prof Dr; Benoît Tavernier[1,7], MD, Prof Dr; Mouhamed Djahoum Moussa[7], MD

[1]Univ. Lille, CHU Lille, ULR 2694 - METRICS: Évaluation des technologies de santé et des pratiques médicales, Lille, France

[2]InterHop, Paris, France

[3]Univ. Lille, Faculté Ingénierie et Management de la Santé, Lille, France

[4]Department of Anaesthesiology and Critical Care Medicine, Amiens Picardie University Hospital, Amiens, France

[5]Department of Anesthesiology and Critical Care Medicine, Regional University Hospital of Nancy, Nancy, France

[6]Department of Anaesthesiology and Critical Care, Rouen University Hospital, Rouen, France

[7]Department of Anesthesiology and Critical Care, CHU Lille, Lille, France

Corresponding Author:
Antoine Lamer, PhD
Univ. Lille
CHU Lille
ULR 2694 - METRICS: Évaluation des technologies de santé et des pratiques médicales
2 place de Verdun
Lille, F-59000
France
Phone: 33 320626969
Email: antoine.lamer@chru-lille.fr

Abstract

Background: Electronic health records (EHRs, such as those created by an anesthesia management system) generate a large amount of data that can notably be reused for clinical audits and scientific research. The sharing of these data and tools is generally affected by the lack of system interoperability. To overcome these issues, Observational Health Data Sciences and Informatics (OHDSI) developed the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) to standardize EHR data and promote large-scale observational and longitudinal research. Anesthesia data have not previously been mapped into the OMOP CDM.

Objective: The primary objective was to transform anesthesia data into the OMOP CDM. The secondary objective was to provide vocabularies, queries, and dashboards that might promote the exploitation and sharing of anesthesia data through the CDM.

Methods: Using our local anesthesia data warehouse, a group of 5 experts from 5 different medical centers identified local concepts related to anesthesia. The concepts were then matched with standard concepts in the OHDSI vocabularies. We performed structural mapping between the design of our local anesthesia data warehouse and the OMOP CDM tables and fields. To validate the implementation of anesthesia data into the OMOP CDM, we developed a set of queries and dashboards.

Results: We identified 522 concepts related to anesthesia care. They were classified as demographics, units, measurements, operating room steps, drugs, periods of interest, and features. After semantic mapping, 353 (67.7%) of these anesthesia concepts were mapped to OHDSI concepts. Further, 169 (32.3%) concepts related to periods and features were added to the OHDSI vocabularies. Then, 8 OMOP CDM tables were implemented with anesthesia data and 2 new tables (EPISODE and FEATURE) were added to store secondarily computed data. We integrated data from 5,72,609 operations and provided the code for a set of 8 queries and 4 dashboards related to anesthesia care.

Conclusions: Generic data concerning demographics, drugs, units, measurements, and operating room steps were already available in OHDSI vocabularies. However, most of the intraoperative concepts (the duration of specific steps, an episode of hypotension, etc) were not present in OHDSI vocabularies. The OMOP mapping provided here enables anesthesia data reuse.

(J Med Internet Res 2021;23(10):e29259) doi:10.2196/29259

KEYWORDS

data reuse; common data model; Observational Medical Outcomes Partnership; anesthesia; data warehouse; reproducible research

Introduction

Observational health data collected from electronic health records (EHRs) can be valuable not only for direct health care delivery but also for secondary uses (ie, data reuse) in research, evaluating quality of care, and public health [1,2]. Concerns on data reuse include data validity and lack of reproducibility [3-5]. These concerns have driven the need for a framework to enhance the secondary use of health data [6]. To support reproducible research over a distributed research network, Observational Health Data Sciences and Informatics (OHDSI) provides the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) and a full range of open-source tools and methods [7-12]. OHDSI provides database scripts for implementing the CDM on various database systems, a terminology browser to navigate through vocabularies integrated into the OMOP CDM (Athena), a data quality tool used to characterize and visualize a database's conformity with the OMOP CDM (Achilles), methods for connecting to the OMOP CDM (DatabaseConnector), methods for the extract-transform-load process (WhiteRabbit, RabbitInAHat, and Usagi), methods for data extraction and transformation (OhdsiRTools and FeatureExtraction), and methods for statistical analyses and machine learning (PatientLevelPrediction, CohortMethod, CaseCrossover, and CaseControl) [13-15].

The OMOP CDM standardizes the vocabulary and structure of EHRs and medical claims data to promote interoperability and ensure that queries can be applied consistently to distributed databases. Integration of local data into the CDM involves conceptual mapping of local concepts into standard vocabulary concepts and structural mapping of local entities to standard entities in the OMOP CDM [8,16]. The essential conceptual and structural mapping of local data is time- and resource-consuming and may also result in the loss of information [11]. However, once mapped, the data offer new opportunities [8,11]. In 2020, more than 100 databases from 20 countries (corresponding to more than 0.5 billion patients) have been integrated into the OMOP CDM [12]. Most of the data come from claims databases studied for pharmacoepidemiological purposes [17-21] or from hospital clinical databases [22,23]. In the past decade, many studies have been carried out; they include patient-level predictions and estimations of the population-level effect [24-27]. Recently, Lane et al collected data on 9,00,000 patients in 15 centers using different software packages; this highlights opportunities for collaboration between centers and for increasing the power of such studies [28].

Even though many studies have been published, some aspects of integrating data into the OMOP CDM are still challenging. Cho et al showed that semantic mapping of concepts from organ transplantation registry forms was fastidious and that OMOP concepts covered only 55% of their vocabulary [29]. Michael et al mapped only 26% of local biospecimen records to the OMOP CDM owing to missing information [30]. Researchers have suggested adapting the CDM (by adding new concepts or new fields) to support the integration of biospecimen data. Warner et al added an extension to the OMOP CDM to support cancer treatments and handle episodes of care with a higher level of abstraction than that represented in the OMOP tables of low-level clinical events [31].

In the field of intraoperative management and anesthesiology, several retrospective studies have looked for links between hemodynamic variations (eg, hypotension) in the operating theater and negative postoperative outcomes (eg, death and acute kidney injury) [32-34]. Similar results were observed for the intraoperative tidal volume ventilation administered to patients [35]. In several cases, this work has made it possible to generate hypotheses for prospective studies, the results of which then validated the proposed hypotheses [36]. These studies were mainly performed with data automatically collected by anesthesia information management systems (AIMS) [37]. However, most of the studies were performed at a small number of centers, which reduced the results' external validity. The main specific features of data recorded in the operating room are their high frequency and high degree of precision, with 1 data point saved every 30 seconds for signals like the heart rate or the intra-arterial blood pressure. Another specific feature is the ability to transform raw data into more usable information or new variables that may better describe exposure to an insult. For example, the arterial pressure signal is computed into comprehensive hypotension events, including the number of episodes, area under the curve, and average time spent within or beyond a threshold [38,39]. In terms of anesthesia data, these data warehouse–based studies can be potentially extrapolated to an international dimension, with stronger evidence through data sharing. This sharing requires the prior homogenization of vocabularies, data formats, and data quality, as promoted by OMOP. However, anesthesia data have not previously been mapped into the OMOP CDM, and the proportion of the anesthesia vocabulary that has already been mapped has not been determined.

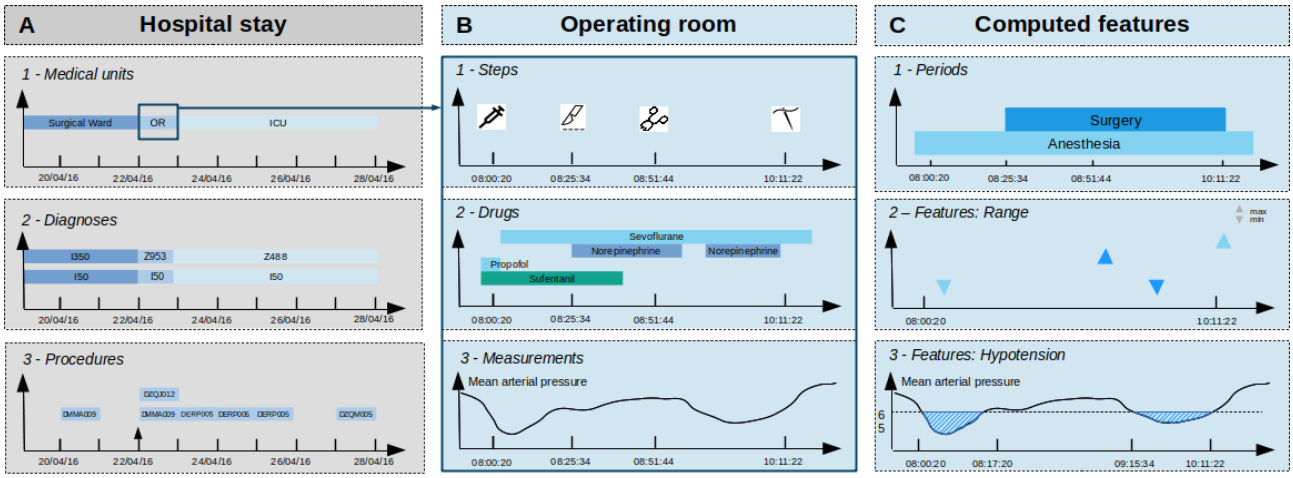
The primary objective of the present study was to standardize anesthesia data to the OMOP CDM. The secondary objectives were to provide vocabularies for the reuse of large-scale data and develop queries and dashboards related to the exploitation of anesthesia data using the OMOP CDM.

Methods

Study Data

Lille University Medical Center (Lille, France) has developed a clinical data warehouse with a local data model [40]. This data warehouse has been collecting data related to the hospital stay and operating room since 2010. Other features were also subsequently computed to facilitate data reuse. Hence, data were classified into three types, as shown in Figure 1: hospital stay data, operating room data, and computed features.

Figure 1. Example of local data organization for a cardiac surgery stay. A. hospital data from the Programme de Médicalisation des Systèmes d'Information database: medical units, diagnoses, and procedures. B. anesthesia information management systems data: steps in the procedures, drug administrations, and measurements. C. features computed from anesthesia information management systems data: periods of interest (anesthesia and surgery), features (range of mean arterial pressure during anesthesia and surgery, and the duration and number of episodes with a mean arterial pressure below 65 mm Hg).



Hospital Stay Data

Hospital stay data were extracted from the French national discharge database (Programme de Médicalisation des Systèmes d'Information [PMSI]) used by all hospitals in France and are presented in Figure 1A. The PMSI contains medical discharge reports entered after each hospital visit. The hospital stay data include all the characteristics of a patient's stay, such as the diagnosis (based on the International Classification of Diseases, 10th edition), medical procedures (based on the Classification Commune des Actes Médicaux), and admission and discharge dates. We have previously implemented the PMSI's administrative data into the OMOP CDM [20].

Operating Room Data

Operating room data were extracted from the hospital's dedicated AIMS [37] and are presented in Figure 1B. Various modules collect and centralize all the data referring to one case, from the preanesthetic evaluation to discharge from the postanesthesia care unit (PACU). These modules include continuously monitored parameters (eg, heart rate, blood pressure, respiratory rate, and tidal volume), drug administrations, and the main steps in anesthesia and surgery procedures.

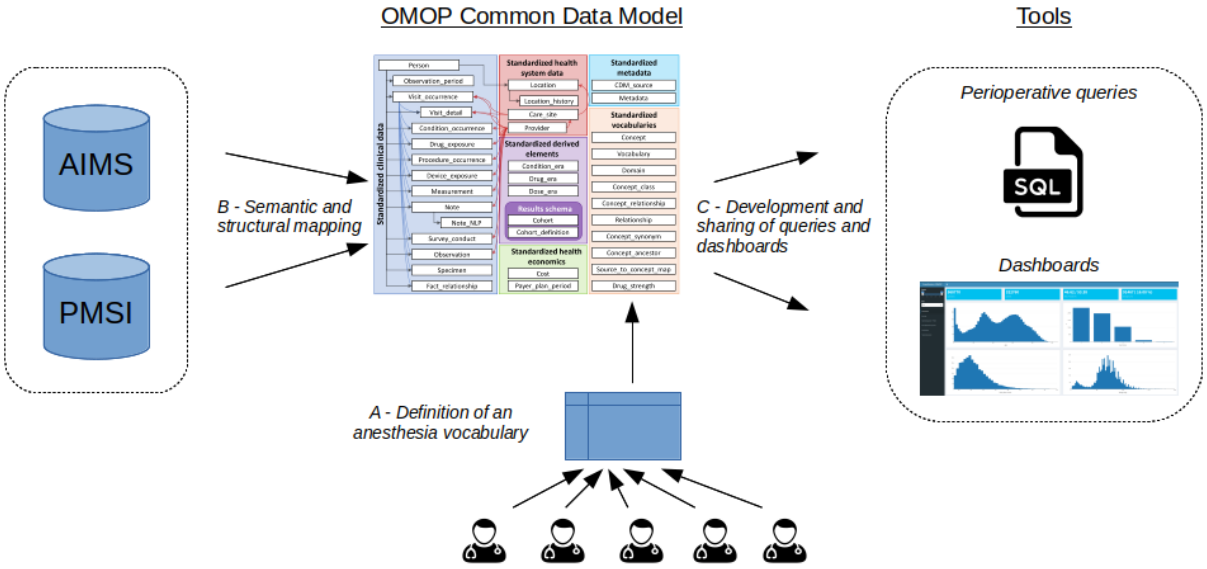
Computed Features

New features were computed to facilitate data reuse for research purposes [38,39]. First, we determined intraoperative periods of interest from events in time, as shown in Figure 1C-1. Second, we derived perioperative measurements and events from the periods of interest and then specified events (hypotension, tachycardia, and oxygen desaturation) as the ranges, medians, or means, indicated in Figures 1C-2 and 1C-3.

Semantic and Structural Mapping to the OMOP CDM

The vocabularies used to characterize the patients and anesthesia procedure were identified by 5 experts in anesthesia from 5 different centers (Lille, Amiens, APHP, Nancy, and Rouen) in France. The experts then selected the most relevant concepts for conducting care and research from within these vocabularies. Next, each local concept was mapped to a standard concept from the OHDSI vocabularies, as shown in Figure 2A. Figure 2B shows that structural mapping links the source data table to the OMOP data table and the source columns to the OMOP columns according to the OHDSI specifications [41]. The extract-transform-load process was implemented using a structured query language, and data were stored in a PostgreSQL 10.11 database (PostgreSQL Global Development Group) on Ubuntu 18.04.3.

Figure 2. Transformation of anesthesia data into the Observational Medical Outcomes Partnership common data model. A. selection of concepts related to anesthesia procedures by 5 anesthetists. B. semantic and structural mapping of anesthesia and Programme de Médicalisation des Systèmes d'Information data into the Observational Medical Outcomes Partnership common data model. C. development of shareable material for the exploitation of anesthesia data. AIMS: anesthesia information management systems; OMOP: Observational Medical Outcomes Partnership; PACU: postanesthesia care unit; PMSI: Programme de Médicalisation des Systèmes d'Information; SQL: structured query language.



Shareable Queries and Dashboards

To test the relevance of the OMOP CDM, we implemented 8 queries related to service audits and clinical research; these were based on the semantic and structural mapping implemented in our database. The queries were intended to provide the following information: (Q1) number of operations per year and per specialty department, (Q2) anesthesia procedures during an outpatient visit, (Q3) operations with fast-track surgery and no admission to the PACU, (Q4) operations with a mean arterial pressure below 65 mm Hg within 30 minutes of anesthesia induction, (Q5) administrations of norepinephrine, epinephrine, ephedrine, phenylephrine, dobutamine, or atropine received within 15 minutes of the first drop in the mean arterial pressure to below 65 mm Hg, (Q6) length of stay according to the score categories of the American society of anesthesiologists, (Q7) operations followed by a stay in the intensive care unit, and (Q8) characterization of the Mallampati grade.

In a previous work, we described the user-centered development, implementation, and preliminary evaluation of clinical dashboards related to anesthesia unit management and quality assessment in the Lille University Medical Center [42]. The user needs had been identified by conducting 21 end-user interviews. Several representations had been developed and submitted to end users for appraisal. After prioritization and feasibility assessment, 10 dashboards were ultimately implemented and deployed. Dashboards were evaluated by 20 end users (4 residents, 4 nurse anesthetists, and 12 anesthesiologists, including the head of the department and a unit manager). The mean (standard deviation) system usability score was 82.6 (11.5), which corresponded to excellent usability. As the dashboards were implemented from our data warehouse with local vocabulary and structured following a local data model, their codes could not be shared with other teams. In the current study, we selected 4 existing dashboards (population description, hemodynamic management, ventilation management, and postoperative outcome) and implemented them from the database now in the OMOP format, as shown in Figure 2C. The dashboards were implemented in R (The R Project for Statistical Computing) with the shiny, shinythemes, shinydashboard, and dplyr packages. The application was connected to the OMOP CDM via the DatabaseConnector package. We compared the new dashboards with the former versions to assess the possible loss of information.

Results

Semantic Mapping

The experts identified 8 types of vocabularies that had been custom-developed for the AIMS by software editors and anesthetists or that were used in the data warehouse: patient characteristics on the day of the procedure, types of visits, units, measurements, drugs, operation steps, periods, and features. Patient history–related vocabulary was not considered, as it was mainly documented manually, using synonyms, abbreviations, and negatives. From within the 8 mapped vocabularies, the experts selected the 522 concepts given in Table 1: 23 patient characteristics, 6 visits, 162 drugs, 45 measurement parameters, 67 units, 46 operation steps, 18 periods, and 155 features.

The experts looked for corresponding concepts in the OHDSI standardized vocabularies. Among the 522 concepts, 353 (67.7%) were successfully mapped to standard concepts for patient characteristics, visits, units, measurements, drugs, operation steps, and periods. All the concepts for patient characteristics, units, measurements, operation steps, and drugs were mapped. Further, 169 concepts (32.4%) in the visit, period, and feature vocabularies were not retrieved in the OHDSI standardized vocabularies and were thus added to the CONCEPT table. Specifically, the concept "operating room visit" was a new type of visit and helped distinguish visits to the operating

room from the other types of visits in care units (eg, intensive care and emergency units). The semantic mapping is described in Table 1. All the concepts are listed in Multimedia Appendix 1.

Drugs were mapped to standard concepts of the class "ingredient," as the clinical drug form is not correctly documented in the AIMS.

Table 1. Semantic mapping between anesthesia and Observational Health Data Sciences and Informatics vocabularies.

Source vocabularies	Concepts identified in source vocabularies, N	Corresponding standard OHDSI[a] vocabularies	Concepts mapped to standard OHDSI concepts, n (%)	New concepts added, n
Demographics	23	SNOMED[b]	23 (100)	0
Visits	6	Visit	5 (83.3)	1
Units	67	UCUM[c]/SNOMED	67 (100)	0
Measurements	45	LOINC[d]/SNOMED	45 (100)	0
Operation steps	46	SNOMED	46 (100)	0
Drugs	162	RxNorm	162 (100)	0
Period	18	—[e]	5 (28.8)	13
Feature	155	—	0 (0)	155

[a]OHDSI: Observational Health Data Sciences and Informatics.

[b]SNOMED: Systemized Nomenclature of Medicine.

[c]UCUM: Unified Code for Units of Measure.

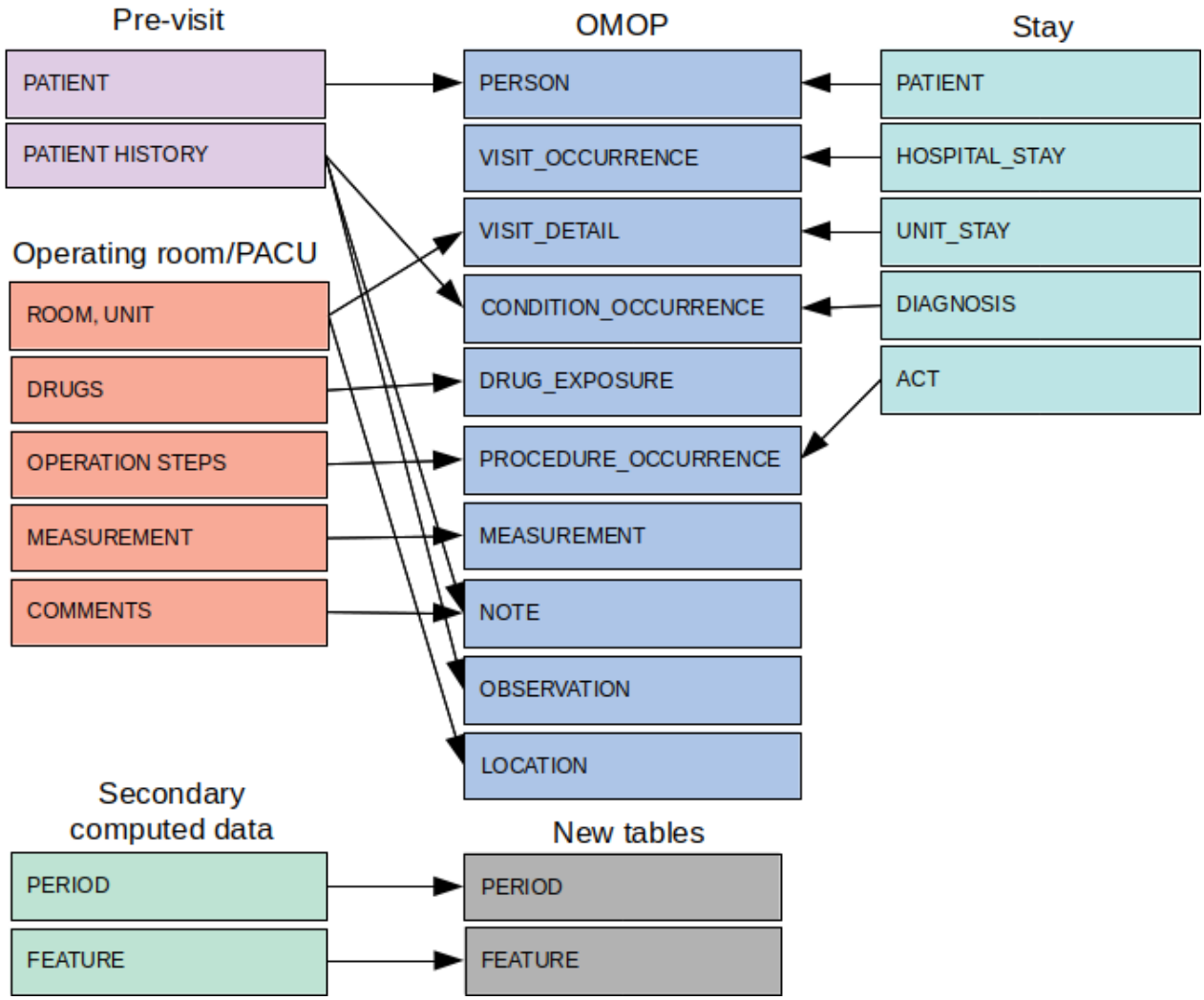
[d]LOINC: Logical Observation Identifiers Names and Codes.

[e]Not available.

Structural Mapping

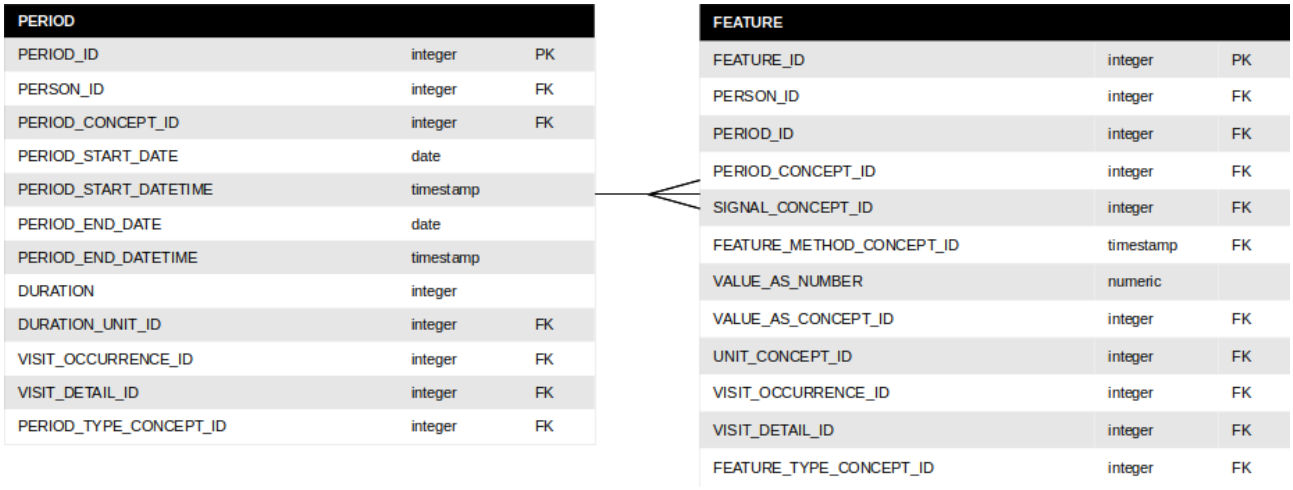
Each inpatient visit is defined a record in the VISIT_OCCURRENCE table. During a hospital stay, each move to a medical unit or an operating room for an operation is defined as a record in the VISIT_DETAIL table. Operating room visits were characterized with a new "operating room visit" concept, namely VISIT_DETAIL_CONCEPT_ID. This concept made it possible to differentiate between visits to care units and those to the operating room. Diagnoses and medical procedures documented in medical units were linked to the corresponding VISIT_DETAIL and VISIT_OCCURRENCE records. Measurements, drug administrations, and events documented in the operating room or PACU were linked to the corresponding operation by the VISIT_DETAIL_ID. Structural events were mapped to procedure_occurrence. Free-text entries from the preanesthesia consultation and those in the operating room were mapped to NOTE. Owing to the high volume, raw data for continuously monitored variables were not included in the measurement table but were kept aside in another schema. The RELATIONSHIP table was implemented with the relationships between the 214 anesthesia rooms (ie, preanesthesia consultation rooms, operating rooms, and the PACU) and the corresponding specialty departments. Structural mapping of the local clinical tables onto the OMOP tables is described in Figure 3.

Figure 3. Structural mapping of data related to the preanesthesia consultation and visits to the operating room, and the postanesthesia care unit in the Observational Medical Outcomes Partnership common data model. Integration of secondarily computed data necessitated the implementation of 2 new tables: PERIOD and FEATURE. OMOP: Observational Medical Outcomes Partnership; PACU: postanesthesia care unit.



We defined 2 new tables to store the computed data, namely period and feature. A period is defined by 2 milestones, a start event and an end event. The events may come from different sources: administration of a drug, a step in a procedure, consultation with a health care professional, or a visit to a health care unit. A period may be defined by an event date or time and a time interval, such as the start of a procedure and the next 30 minutes, or the administration of a drug and the last 10 minutes. A feature is defined by the combination of three concepts: a period (as defined above), a raw signal, and an aggregation method. The raw signal may include measurements of vital signs (eg, heart rate, arterial pressure, and oxygen saturation) or mechanical ventilation parameters (tidal volume, respiratory rate, and plateau pressure). The aggregation method may be a statistical indicator (eg, the mean, minimum, or maximum value) or an expert-driven rule [35]. The logical data model for these 2 tables is described in Figure 4.

Figure 4. Logical data model of PERIOD and FEATURE tables.



Integration

Records spanning 10 years were integrated into the OMOP CDM. It corresponded to 5,72,609 operations for 3,29,633 patients. The numbers of records per OMOP table are shown in Table 2, and the number of records per operation and those per hospital stay are given in Table 3.

Table 2. Number of records implemented in Observational Medical Outcomes Partnership tables for the 2 data sources.

OMOP[a] table	Number of records
PERSON	3,29,633
VISIT_OCCURRENCE	48,84,220
VISIT_DETAIL (from PMSI[b])	15,40,677
VISIT_DETAIL (from AIMS[c])	5,72,609
CONDITION_OCCURRENCE (from PMSI)	15,13,544
CONDITION_OCCURRENCE (from AIMS)	5,67,442
DRUG_EXPOSURE	86,12,045
PROCEDURE_OCCURRENCE (from PMSI)	11,66,227
PROCEDURE_OCCURRENCE (from AIMS)	5,58,734
OBSERVATION (from PMSI)	18,644
OBSERVATION (from AIMS)	49,45,451
NOTE	92,88,981
PERIOD	40,26,665
FEATURE	3,48,09,015
LOCATION	1,348

[a]OMOP: Observational Medical Outcomes Partnership.

[b]PMSI: Programme de Médicalisation des Systèmes d'Information.

[c]AIMS: anesthesia information management system.

Table 3. Median (IQR) number of records per operation and per hospital stay.

OMOP[a] table	Median (IQR) number of records per operation (AIMS[b])	Median (IQR) number of records per hospital stay (PMSI[c])
VISIT_DETAIL	1 (1-1)	1 (1-1)
CONDITION_OCCURRENCE	1 (1-1)	2 (2-5)
DRUG_EXPOSURE	10 (5-17)	__[d]
PROCEDURE_OCCURRENCE	9 (4-12)	2 (1-4)
OBSERVATION	10 (8-10)	—
NOTE	31 (12-40)	—
PERIOD	7 (5-10)	—
FEATURE	71 (42-84)	—

[a]OMOP: Observational Medical Outcomes Partnership.

[b]AIMS: anesthesia information management system.

[c]PMSI: Programme de Médicalisation des Systèmes d'Information.

[d]Not available.

Shareable Queries and Dashboards

Based on the anesthesia and hospital stay data, we developed 8 queries for application to the existing VISIT_OCCURRENCE, VISIT_DETAIL, CONDITION_OCCURRENCE, PROCEDURE_OCCURRENCE, DRUG_EXPOSURE, NOTE, CONCEPT, and RELATIONSHIP tables and the 2 new PERIOD and FEATURE tables. The query steps and queried tables are described in Table 4. All queries are detailed in Multimedia Appendix 2.

Population description, hemodynamic, ventilation, and postoperative outcome are the 4 dashboards available, as shown in Table 5 and Figure 5. They provide an overview of the population treated in the operating room, compliance with hemodynamic guidelines, compliance with ventilatory guidelines, and postoperative outcomes. Each dashboard can be configured through filtering by year and department. The tables PERSON, VISIT_DETAIL, OBSERVATION, and FEATURE were queried to feed the dashboards. Although the format of the data source differed between the 2 versions of the dashboards (local format vs OMOP format), the figures and results obtained were identical.

Table 4. List of queries in the context of the operating room visits and hospital stays.

Query ID	Query	Requirement	Query steps and queried tables
1	Number of operations per year and per specialty department	Identify the visit to the operating room and the corresponding department	• Identification of visits to the operating room with the new concept "operating room visit" (VISIT_DETAIL) • Relationship between care_site_id of the operating room and care_site_id of the department (CONCEPT_RELATIONSHIP, CONCEPT)
2	Anesthesia procedure during an outpatient visit	Cross-check data from two sources: operating room (AIMS[a]) and hospital stay (PMSI[b])	• Identification of visits to the operating room with the new concept "operating room visit" (VISIT_DETAIL, VISIT_OCCURRENCE)
3	Operations with fast-track surgery and no admission to the PACU[c]	Identify a specific period of the operation	• Identification of visits to the operating room with the new concept "operating room visit" (VISIT_DETAIL) • Joining with PACU periods (PERIOD)
4	Operations with an MAP[d]<65 mm Hg within 30 minutes of inducing anesthesia	Cross-check data from two secondarily computed, operation-specific periods	• Period P1 of hypotension with MAP<65 mm Hg (PERIOD) • Period P2 of anesthesia (PERIOD) • Joining of P1 and P2 with the start date of P1 in 30 minutes following the start date of P2
5	Administration of norepinephrine, epinephrine, ephedrine, phenylephrine, dobutamine, or atropine received within 15 minutes of the first drop in MAP to below 65 mm Hg	Cross-check data from a secondarily computed period and specific drug administrations	• First period P1 of MAP<65 mm Hg (PERIOD) • Administration A of norepinephrine, epinephrine, ephedrine, phenyleprine, dobutamine, or atropine (DRUG_EXPOSURE) • Linking P1 and A with the start date and time of A in the 15 minutes following the start date of P1 • Aggregation by drug
6	Length of stay by ASA[e] status	Cross-check data from two sources: the operating room (AIMS) and hospital stay (PMSI)	• Extraction of ASA status conditions (CONDITION_OCCURRENCE) • Linking of the operating room visit details to the visit occurrence (VISIT_OCCURRENCE) • Aggregation of the duration of visit occurrence by ASA status
7	Operations followed by a stay in the intensive care unit	Cross-check data from two sources: operating room (AIMS) and hospital stay (PMSI)	• Identification of visits to the operating room with the new concept "operating room visit" VD1[f] (VISIT_DETAIL) • Identification of visits to the intensive care unit VD2 (VISIT_DETAIL) • Linking VD1 to VD2 according to the visit_occurrence identifier and with VD2 start datetime>VD1 end datatime
8	Characterization of the Mallampati grade	Query the preanesthesia consultation	• Extraction of Mallampati scores (NOTE) • Aggregation by score

[a]AIMS: anesthesia information management system.

[b]PMSI: Programme de Médicalisation des Systèmes d'Information.

[c]PACU: postanesthesia care unit.

[d]MAP: mean arterial pressure.

[e]ASA: American Society of Anesthesiologists.

[f]VD: visit detail.

Table 5. Description of dashboards implemented with the Observational Medical Outcomes Partnership common data model.

Dashboard	Information/objective	Numeric indicators (number or percentage)	Graphics	OMOP[a] tables
Population description	Overview of the population cared for in the operating room	• Number of operations • Number of patients • Male/female ratio (%) • Number of urgent operations (%)	• Histogram of age • Bar plot of the ASA[b] • Status histogram of the BMI • Histogram of the weight	PERSON VISIT_DETAIL OBSERVATION
Hemodynamics	Compliance with hemodynamic guidelines	• Number of operations with MAP[c]<65 mm Hg • Number of operations with MAP>120 mm Hg • Number of operations with HR[d]<60 bpm[e] • Number of operations with SpO2[f]<90%	• Bar plot of the duration with MAP<65 mm Hg (min) • Bar plot of the duration with MAP>120 mm Hg • Bar plot of the duration with HR<60 bpm • Bar plot of the duration with SpO2< 90%	PERSON VISIT_DETAIL FEATURE
Ventilation	Compliance with ventilatory guidelines	• Number of operations with expiratory tidal volume>8 ml/kg IBW[g] • Number of operations with expiratory tidal volume>10 ml/kg IBW • Number of operations with expiratory tidal volume (ml) • Number of operations with expiratory tidal volume/IBW (ml/kg)	• Bar plot of the expiratory tidal volume>8 ml/kg IBW by sex and year • Line plot of the expiratory tidal volume/IBW	PERSON VISIT_DETAIL FEATURE
Postoperative outcome	Overview of postoperative outcome: mortality, duration of hospital stay, and intensive care unit stay	• Number of operations followed by a death during hospital stay (%) • Duration of hospital stay • Number of operations followed by a passage in intensive care (%)	• Bar plot of the number of deaths per year • Line plot of the number of passages in intensive care per year	PERSON VISIT_DETAIL OBSERVATION

[a]OMOP: Observational Medical Outcomes Partnership.

[b]ASA: American Society of Anesthesiologists.

[c]MAP: Mean arterial pressure.

[d]HR: heart rate.

[e]bpm: beats per minute.

[f]SpO_2: oxygen saturation.

[g]IBW: ideal body weight.

Figure 5. Clinical dashboard for the assessment of ventilatory guidelines. Number of operations with tidal volume>8 ml.kg-1 of ideal body weight, number of operations with tidal volume>10ml.kg-1 of ideal body weight, median (IQR) expiratory tidal volume, median (IQR) expiratory tidal volume/ ideal body weight change over time in the proportion of operations with tidal volume >8 ml.kg-1 of ideal body weight and change over time in expiratory tidal volume/ ideal body weight over the year.

The OMOP model has a row-oriented structure, with 1 data item per row. For example, each row of OBERVATION stores a single data item (ie, a weight or a BMI). In contrast, each query and dashboard must gather several data items (coming from a single table or several tables). Queries were developed with common table expressions, a syntax provided by PostgreSQL to write auxiliary statements for use in a larger query [43]. Dashboards needed to be implemented on top of the temporary tables gathering the results of a set of CTEs to reduce the response time of each query.

Discussion

Principal Results

In the present work, we integrated intraoperative anesthesia data into the OMOP CDM. To the best of our knowledge, this study is the first to have mapped intraoperative data into the OMOP CDM. First, experts from 5 French centers defined a list of concepts describing the anesthesia procedure and specific features. This list mainly comprised standardized concepts from the OHDSI vocabularies: patient history, patient characteristics on the day of the procedure, units, measurements, drugs, and procedure steps. When the corresponding concepts were missing, we added new concepts, particularly to characterize secondarily computed periods and features. Second, we implemented an extract-transform-load process to move perioperative data into the CDM. Third, we implemented common queries related to anesthesia procedures. As the OMOP CDM was initially developed for pharmacoepidemiology, we ensured that the mapping proposed for intraoperative data (and particularly the features specific to our work) could be easily queried. Finally, we developed shareable R scripts for the generation of anesthesia dashboards. These dashboards enabled us to ensure that hemodynamic and ventilatory guidelines were followed.

Limitations

First, we focused primarily on implementing the vocabulary related to the most common anesthetic procedures. Thus, it may not be sufficient to describe anesthetic management related to more specific procedures (obstetrics, ambulatory procedures, etc), but these could be added in the future. Second, the added concepts are not available in Athena at present and are therefore nonstandard concepts. While waiting for integration validation, the concepts are available on our git directory [44] and can be used and supplemented by other teams. Third, PERIOD and FEATURE are not supported by the OHDSI software stack. Further developments are needed to fully benefit from these new tables in the OHDSI tools and packages. Finally, CDMs may lose information owing to restrictions on the types of relationships proposed in relational models [45]. When integrating, care must be taken to ensure that the information realistically integrated is adequate to perform analyses afterward, and that any loss of information does not sanction the results and their interpretations.

Comparison With Prior Works

As observed in the studies that focused on specific data (apart from claims data) [29-31], we encountered difficulties with perioperative data. The main difficulty was using several local and custom vocabularies to document the intraoperative period; this contrasts with claims data, which are described according to terminologies. This problem required experts to define anesthesia-related concepts because the local concepts provided by the AIMS were not sufficient. Ryu et al have already reported that mapping by experts is an essential step [46]. Furthermore, the frequency of the recordings (every 30 seconds) in the operating theater produced a large volume of data. We decided not to retain the raw measurements in the measurement table so that the query response time remained acceptable. Raw measurements were stored in a similar measurement table on a twin schema. Finally, we had to compute new periods and

features that did not fit in the OMOP CDM tables. To achieve this, we developed 2 new period and feature tables.

Our present work might offer opportunities for research collaborations on intraoperative data with other centers. The material provided here could be used and enhanced by other centers. In combination with federated learning [47], the OMOP CDM provides tools needed for conducting reproducible research.

Conclusions

Generic data concerning demographics, drugs, units, measurements, and operating room steps were already available in OHDSI vocabularies. However, most of the intraoperative concepts (the duration of specific steps, episodes of hypotension, etc) were absent in the OHDSI vocabularies. We have performed OMOP mapping for reusing anesthesia data.

Acknowledgments

We thank the InterHop association and the OHDSI community for their support and advice.

Authors' Contributions

All authors contributed to the study conception and design. OAA, AB, AP, VP, and MDM contributed to the vocabulary mapping. AL implemented the extract-transform-load process. AL, OAA, and AB implemented the common queries and dashboards. AL drafted the manuscript as the first author. All authors approved the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Anesthesia-related concepts.
[XLSX File (Microsoft Excel File), 18 KB - jmir_v23i10e29259_app1.xlsx]

Multimedia Appendix 2
Common queries.
[DOCX File , 9 KB - jmir_v23i10e29259_app2.docx]

References

1. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol 2005 Apr;58(4):323-337. [doi: 10.1016/j.jclinepi.2004.10.012] [Medline: 15862718]
2. Brookhart M, Stürmer T, Glynn R, Rassen J, Schneeweiss S. Confounding control in healthcare database research: challenges and potential approaches. Med Care 2010 Jun;48(6):S114-S120 [FREE Full text] [doi: 10.1097/MLR.0b013e3181dbebe3] [Medline: 20473199]
3. Mayes LC, Horwitz RI, Feinstein AR. A collection of 56 topics with contradictory results in case-control research. Int J Epidemiol 1988 Sep;17(3):680-685. [doi: 10.1093/ije/17.3.680] [Medline: 3272133]
4. Collins GS, Mallett S, Omar O, Yu L. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. BMC Med 2011 Sep 08;9:103 [FREE Full text] [doi: 10.1186/1741-7015-9-103] [Medline: 21902820]
5. Collins GS, Omar O, Shanyinde M, Yu L. A systematic review finds prediction models for chronic kidney disease were poorly reported and often developed using inappropriate methods. J Clin Epidemiol 2013 Mar;66(3):268-277. [doi: 10.1016/j.jclinepi.2012.06.020] [Medline: 23116690]
6. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association white paper. J Am Med Inform Assoc 2007 Jan;14(1):1-9 [FREE Full text] [doi: 10.1197/jamia.M2273] [Medline: 17077452]
7. Stang PE, Ryan PB, Racoosin JA, Overhage JM, Hartzema AG, Reich C, et al. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. Ann Intern Med 2010 Nov 02;153(9):600-606. [doi: 10.7326/0003-4819-153-9-201011020-00010] [Medline: 21041580]
8. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc 2012 Jan;19(1):54-60 [FREE Full text] [doi: 10.1136/amiajnl-2011-000376] [Medline: 22037893]
9. Hripcsak G, Duke J, Shah N, Reich C, Huser V, Schuemie M, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. Stud Health Technol Inform 2015;216:574-578 [FREE Full text] [Medline: 26262116]

10. FitzHenry F, Resnic F, Robbins S, Denton J, Nookala L, Meeker D, et al. Creating a common Ddata model for comparative effectiveness with the Observational Medical Outcomes Partnership. Appl Clin Inform 2015;6(3):536-547 [FREE Full text] [doi: 10.4338/ACI-2014-12-CR-0121] [Medline: 26448797]
11. Voss E, Makadia R, Matcho A, Ma Q, Knoll C, Schuemie M, et al. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. J Am Med Inform Assoc 2015 May;22(3):553-564 [FREE Full text] [doi: 10.1093/jamia/ocu023] [Medline: 25670757]
12. Observational Health Data Sciences and Informatics. URL: https://www.ohdsi.org/ [accessed 2019-05-03]
13. Schuemie MJ, Ryan PB, DuMouchel W, Suchard MA, Madigan D. Interpreting observational studies: why empirical calibration is needed to correct p-values. Stat Med 2014 Jan;33(2):209-218 [FREE Full text] [doi: 10.1002/sim.5925] [Medline: 23900808]
14. Reps J, Schuemie M, Suchard M, Ryan P, Rijnbeek P. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. J Am Med Inform Assoc 2018 Aug;25(8):969-975 [FREE Full text] [doi: 10.1093/jamia/ocy032] [Medline: 29718407]
15. Johnston SS, Morton JM, Kalsekar I, Ammann EM, Hsiao C, Reps J. Using machine learning applied to real-world healthcare data for predictive analytics: an applied example in bariatric surgery. Value Health 2019 May;22(5):580-586. [doi: 10.1016/j.jval.2019.01.011] [Medline: 31104738]
16. Definition and DDLs for the OMOP common data model. OHDSI/CommonDataModel. URL: https://github.com/OHDSI/CommonDataModel [accessed 2019-05-03]
17. Reich C, Ryan PB, Stang PE, Rocca M. Evaluation of alternative standardized terminologies for medical conditions within a network of observational healthcare databases. J Biomed Inform 2012 Aug;45(4):689-696 [FREE Full text] [doi: 10.1016/j.jbi.2012.05.002] [Medline: 22683994]
18. Weinstein RB, Schuemie M, Ryan P, Stang P. Seasonality in acute liver injury? findings in two health care claims databases. Drug Healthc Patient Saf 2016 Mar;8:39-48 [FREE Full text] [doi: 10.2147/DHPS.S95399] [Medline: 27099532]
19. Haberson A, Rinner C, Schöberl A, Gall W. Feasibility of mapping Austrian health claims data to the OMOP common data model. J Med Syst 2019 Sep;43(10):314 [FREE Full text] [doi: 10.1007/s10916-019-1436-9] [Medline: 31494719]
20. Lamer A, Depas N, Doutreligne M, Parrot A, Verloop D, Defebvre M, et al. Transforming French electronic health records into the Observational Medical Outcome Partnership's common data model: a feasibility study. Appl Clin Inform 2020 Jan;11(1):13-22 [FREE Full text] [doi: 10.1055/s-0039-3402754] [Medline: 31914471]
21. Thurin NH, Lassalle R, Schuemie M, Pénichon M, Gagne JJ, Rassen JA, et al. Empirical assessment of case-based methods for drug safety alert identification in the French National Healthcare System database (SNDS): methodology of the ALCAPONE project. Pharmacoepidemiol Drug Saf 2020 Sep;29(9):993-1000. [doi: 10.1002/pds.4983] [Medline: 32133717]
22. Maier C, Lang L, Storf H, Vormstein P, Bieber R, Bernarding J, et al. Towards Implementation of OMOP in a German university hospital consortium. Appl Clin Inform 2018 Jan;9(1):54-61 [FREE Full text] [doi: 10.1055/s-0037-1617452] [Medline: 29365340]
23. Zhang X, Wang L, Miao S, Xu H, Yin Y, Zhu Y, et al. Analysis of treatment pathways for three chronic diseases using OMOP CDM. J Med Syst 2018 Nov;42(12):260 [FREE Full text] [doi: 10.1007/s10916-018-1076-5] [Medline: 30421323]
24. Yoon D, Ahn EK, Park MY, Cho SY, Ryan P, Schuemie MJ, et al. Conversion and data quality assessment of electronic health record data at a Korean tertiary teaching hospital to a common data model for distributed network research. Healthc Inform Res 2016 Jan;22(1):54-58 [FREE Full text] [doi: 10.4258/hir.2016.22.1.54] [Medline: 26893951]
25. Ryan PB, Madigan D, Stang PE, Overhage JM, Racoosin JA, Hartzema AG. Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. Stat Med 2012 Dec;31(30):4401-4415. [doi: 10.1002/sim.5620] [Medline: 23015364]
26. Ryan PB, Stang PE, Overhage JM, Suchard MA, Hartzema AG, DuMouchel W, et al. A comparison of the empirical performance of methods for a risk identification system. Drug Saf 2013 Oct;36 Suppl 1:S143-S158. [doi: 10.1007/s40264-013-0108-9] [Medline: 24166231]
27. Hripcsak G, Ryan PB, Duke JD, Shah NH, Park RW, Huser V, et al. Characterizing treatment pathways at scale using the OHDSI network. Proc Natl Acad Sci USA 2016 Jul;113(27):7329-7336 [FREE Full text] [doi: 10.1073/pnas.1510502113] [Medline: 27274072]
28. Lane JCE, Weaver J, Kostka K, Duarte-Salles T, Abrahao MTF, Alghoul H, OHDSI-COVID-19 consortium. Risk of hydroxychloroquine alone and in combination with azithromycin in the treatment of rheumatoid arthritis: a multinational, retrospective study. Lancet Rheumatol 2020 Nov;2(11):e698-e711 [FREE Full text] [doi: 10.1016/S2665-9913(20)30276-9] [Medline: 32864627]
29. Cho S, Sin M, Tsapepas D, Dale L, Husain SA, Mohan S, et al. Content coverage evaluation of the OMOP vocabulary on the transplant domain focusing on concepts relevant for kidney transplant outcomes analysis. Appl Clin Inform 2020 Aug;11(4):650-658. [doi: 10.1055/s-0040-1716528] [Medline: 33027834]
30. Michael C, Sholle E, Wulff R, Roboz G, Campion T. Mapping local biospecimen records to the OMOP common data model. AMIA Jt Summits Transl Sci Proc 2020 May;2020:422-429 [FREE Full text] [Medline: 32477663]

31. Warner JL, Dymshyts D, Reich CG, Gurley MJ, Hochheiser H, Moldwin ZH, et al. HemOnc: a new standard vocabulary for chemotherapy regimen representation in the OMOP common data model. J Biomed Inform 2019 Aug;96:103239 [FREE Full text] [doi: 10.1016/j.jbi.2019.103239] [Medline: 31238109]
32. Sessler D, Sigl J, Kelley S, Chamoun N, Manberg P, Saager L, et al. Hospital stay and mortality are increased in patients having a "triple low" of low blood pressure, low bispectral index, and low minimum alveolar concentration of volatile anesthesia. Anesthesiology 2012 Jun;116(6):1195-1203 [FREE Full text] [doi: 10.1097/ALN.0b013e31825683dc] [Medline: 22546967]
33. Walsh M, Devereaux P, Garg A, Kurz A, Turan A, Rodseth R, et al. Relationship between intraoperative mean arterial pressure and clinical outcomes after noncardiac surgery: toward an empirical definition of hypotension. Anesthesiology 2013 Sep;119(3):507-515 [FREE Full text] [doi: 10.1097/ALN.0b013e3182a10e26] [Medline: 23835589]
34. Komatsu R, You J, Mascha E, Sessler D, Kasuya Y, Turan A. Anesthetic induction with etomidate, rather than propofol, is associated with increased 30-day mortality and cardiovascular morbidity after noncardiac surgery. Anesth Analg 2013 Dec;117(6):1329-1337. [doi: 10.1213/ANE.0b013e318299a516] [Medline: 24257383]
35. Levin MA, McCormick PJ, Lin HM, Hosseinian L, Fischer GW. Low intraoperative tidal volume ventilation with minimal PEEP is associated with increased mortality. Br J Anaesth 2014 Jul;113(1):97-108 [FREE Full text] [doi: 10.1093/bja/aeu054] [Medline: 24623057]
36. Futier E, Constantin JM, Paugam-Burtz C, Pascal J, Eurin M, Neuschwander A, et al. IMPROVE Study Group. A trial of intraoperative low-tidal-volume ventilation in abdominal surgery. N Engl J Med 2013 Aug;369(4):428-437. [doi: 10.1056/NEJMoa1301082]
37. Eden A, Grach M, Schneider I, Lazarovici H, Goldik Z, Pizov R. Implementation of an anesthesia information management system (AIMS). Eur J Anaesthesiol 2005;22(Supplement 34):39-40. [doi: 10.1097/00003643-200505001-00138]
38. Lamer A, Jeanne M, Marcilly R, Kipnis E, Schiro J, Logier R, et al. Methodology to automatically detect abnormal values of vital parameters in anesthesia time-series: proposal for an adaptable algorithm. Comput Methods Programs Biomed 2016 Jun;129:160-171. [doi: 10.1016/j.cmpb.2016.01.004] [Medline: 26817405]
39. Lamer A, Jeanne M, Ficheur G, Marcilly R. Automated data aggregation for time-series analysis: study case on anaesthesia data warehouse. Stud Health Technol Inform 2016;221:102-106. [Medline: 27071886]
40. Lamer A, Jeanne M, Vallet B, Ditilyeu G, Delaby F, Tavernier B, et al. Development of an anesthesia data warehouse: preliminary results. IRBM 2013 Dec;34(6):376-378. [doi: 10.1016/j.irbm.2013.09.005]
41. Chapter 4 The common data model. The Book of OHDSI. URL: https://ohdsi.github.io/TheBookOfOhdsi/ [accessed 2021-09-27]
42. Laurent G, Moussa MD, Cirenei C, Tavernier B, Marcilly R, Lamer A. Development, implementation and preliminary evaluation of clinical dashboards in a department of anesthesia. J Clin Monit Comput 2020 May;35:617-626 [FREE Full text] [doi: 10.1007/s10877-020-00522-x] [Medline: 32418147]
43. WITH Queries (common table expressions). PostgreSQL Documentation. URL: https://www.postgresql.org/docs/13/queries-with.html [accessed 2021-09-27]
44. Larmer A. omop_anesthesia. 2021. URL: https://gitlab.com/antoinelamer/omop_anesthesia [accessed 2021-09-29]
45. Blaisure JC, Ceusters WM. Improving the 'fitness for purpose' of common data models through realism based ontology. AMIA Annu Symp Proc 2017;2017:440-447 [FREE Full text] [Medline: 29854108]
46. Ryu B, Yoon E, Kim S, Lee S, Baek H, Yi S, et al. Transformation of pathology reports into the common data model with oncology module: use case for colon cancer. J Med Internet Res 2020 Dec;22(12):e18526 [FREE Full text] [doi: 10.2196/18526] [Medline: 33295294]
47. Mangold P, Filiot A, Moussa M, Sobanski V, Ficheur G, Andrey P, et al. A decentralized framework for biostatistics and privacy concerns. Stud Health Technol Inform 2020 Nov;275:137-141. [doi: 10.3233/SHTI200710] [Medline: 33227756]

Abbreviations

AIMS: anesthesia information management systems
CDM: common data model
EHR: electronic health record
OHDSI: Observational Health Data Sciences and Informatics
OMOP: Observational Medical Outcomes Partnership
PACU: postanesthesia care unit
PMSI: Programme de Médicalisation des Systèmes d'Information

Edited by R Kukafka; submitted 26.04.21; peer-reviewed by M Sedlmayr, R Lee; comments to author 21.05.21; revised version received 14.06.21; accepted 05.07.21; published 29.10.21.

Please cite as:
Lamer A, Abou-Arab O, Bourgeois A, Parrot A, Popoff B, Beuscart JB, Tavernier B, Moussa MD
Transforming Anesthesia Data Into the Observational Medical Outcomes Partnership Common Data Model: Development and Usability Study
J Med Internet Res 2021;23(10):e29259
URL: https://www.jmir.org/2021/10/e29259
doi:10.2196/29259
PMID:

Original Paper

Public Attitudes to Digital Health Research Repositories: Cross-sectional International Survey

Giovanna Nunes Vilaza[1], BSc, MSc; David Coyle[2], BSc, MSc, PhD; Jakob Eyvind Bardram[1], BSc, MSc, PhD

[1]Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark

[2]School of Computer Science, University College Dublin, Dublin, Ireland

Corresponding Author:
Giovanna Nunes Vilaza, BSc, MSc
Department of Health Technology
Technical University of Denmark
Ørsteds Plads 345B
Kongens Lyngby, 2800
Denmark
Phone: 45 45253724
Email: gnvi@dtu.dk

Abstract

Background: Digital health research repositories propose sharing longitudinal streams of health records and personal sensing data between multiple projects and researchers. Motivated by the prospect of personalizing patient care (precision medicine), these initiatives demand broad public acceptance and large numbers of data contributors, both of which are challenging.

Objective: This study investigates public attitudes toward possibly contributing to digital health research repositories to identify factors for their acceptance and to inform future developments.

Methods: A cross-sectional online survey was conducted from March 2020 to December 2020. Because of the funded project scope and a multicenter collaboration, study recruitment targeted young adults in Denmark and Brazil, allowing an analysis of the differences between 2 very contrasting national contexts. Through closed-ended questions, the survey examined participants' willingness to share different data types, data access preferences, reasons for concern, and motivations to contribute. The survey also collected information about participants' demographics, level of interest in health topics, previous participation in health research, awareness of examples of existing research data repositories, and current attitudes about digital health research repositories. Data analysis consisted of descriptive frequency measures and statistical inferences (bivariate associations and logistic regressions).

Results: The sample comprises 1017 respondents living in Brazil (1017/1600, 63.56%) and 583 in Denmark (583/1600, 36.44%). The demographics do not differ substantially between participants of these countries. The majority is aged between 18 and 27 years (933/1600, 58.31%), is highly educated (992/1600, 62.00%), uses smartphones (1562/1600, 97.63%), and is in good health (1407/1600, 87.94%). The analysis shows a vast majority were very motivated by helping future patients (1366/1600, 85.38%) and researchers (1253/1600, 78.31%), yet very concerned about unethical projects (1219/1600, 76.19%), profit making without consent (1096/1600, 68.50%), and cyberattacks (1055/1600, 65.94%). Participants' willingness to share data is lower when sharing personal sensing data, such as the content of calls and texts (1206/1600, 75.38%), in contrast to more traditional health research information. Only 13.44% (215/1600) find it desirable to grant data access to private companies, and most would like to stay informed about which projects use their data (1334/1600, 83.38%) and control future data access (1181/1600, 73.81%). Findings indicate that favorable attitudes toward digital health research repositories are related to a personal interest in health topics (odds ratio [OR] 1.49, 95% CI 1.10-2.02; P=.01), previous participation in health research studies (OR 1.70, 95% CI 1.24-2.35; P=.001), and awareness of examples of research repositories (OR 2.78, 95% CI 1.83-4.38; P<.001).

Conclusions: This study reveals essential factors for acceptance and willingness to share personal data with digital health research repositories. Implications include the importance of being more transparent about the goals and beneficiaries of research projects using and re-using data from repositories, providing participants with greater autonomy for choosing who gets access to which parts of their data, and raising public awareness of the benefits of data sharing for research. In addition, future developments should engage with and reduce risks for those unwilling to participate.

(J Med Internet Res 2021;23(10):e31294) doi:10.2196/31294

KEYWORDS

digital medicine; health informatics; health data repositories; personal sensing; technology acceptance; willingness to share data; human-centered computing; ethics

Introduction

Background

Health research is increasingly adopting digital technologies to accelerate scientific discovery, as digital data sources increase scalability and predictive power for algorithmic inferences [1-3]. Novel data collection techniques include wearables and smartphone sensors to extract participants' behavioral features passively [4]. Records of calls and texts can flag social activity fluctuations; location tracking can reveal mobility patterns; heart rate measures can indicate sleep quality [5-7]. Ecological momentary assessments further complement such passive indicators by sampling individuals' health status in real time through questionnaires [8,9]. The motivation for considering pervasive and digital sources of health and behavioral information is related to the possibility of closely observing research patients' daily lives [10,11].

Intending to personalize future patient care, researchers search for scientific evidence by analyzing longitudinal streams of personal sensing data from large segments of the population [12,13]. Because of this expansion of personal sensing in the health domain, digital health research repositories are gaining momentum. An ambitious concept similar to biobanks [14], a digital health research repository allows multiple projects and researchers to share access to personal data streams beyond DNA and biosamples [15-17]. Although still in their initial steps, initiatives with this goal in mind include nationwide programs [18,19], university-led projects [20,21], and community-driven data platforms [22-24].

Despite promising benefits, barriers to public acceptance can hinder the successful implementation of digital health research repositories [25,26]. Without a diverse range of participants agreeing to contribute with their personal health data, repositories cannot accomplish their ambitious goal of providing reliable evidence for personalized medicine to the broad population [27]. Furthermore, a potential lack of acceptance is aggravated by ethical debates questioning which rights individuals should have following contribution of their data [28-30], especially if data are shared beyond a specific project's scope. Given such challenges, previous research has emphasized that in contrast to most current initiatives, which mainly cater to researchers' needs, health data repositories should attend more to participants' preferences to identify enablers for participation [31,32].

Previous Studies

Previous studies have investigated public attitudes toward biobanks [33-36] and digital health data [37,38] separately. Other past studies have examined motivations to contribute [39], privacy concerns [40], and access control preferences [41] for data sharing within health research in general, yet these studies consider only a few variables simultaneously and rarely inquire about the magnitude of specific attitudes [42,43]. To the best of our knowledge, published quantitative studies have not thoroughly examined how different factors can affect attitude and willingness to share in digital health research repositories' timely and emerging context. Thus, it remains unclear how the public perceives the risks and benefits of shared access of multiple sources of behavioral and health indicators, including digital sensing, for research repositories.

Study Goal

Given this research gap, an online cross-sectional survey was conducted examining public attitudes to research repositories storing health information, biosamples, personal sensing, and behavioral data. This survey study aims to identify implications for future developments by consulting those whose personal data are to be shared for research. The study took place in Denmark, where the project is funded, and the principal investigators are based. Furthermore, with the goal of investigating the potential contrast between 2 very different historical, social, and cultural contexts, we contacted a research group in Brazil to establish a partnership and conduct the study with a sample of Brazilian residents. This decision allowed a cross-country analysis that illuminated similarities and divergences between 2 very disparate contexts. The results contribute to substantial empirical evidence about enablers and barriers for participants' acceptance and discussions on how community engagement, technology design, and policymaking can lead to a stronger participant-centric development in this field.

Methods

Population, Sample, and Recruitment

Denmark and Brazil are very different in terms of population, geography, economy, and culture. Denmark is a small country in area and population (5.8 million inhabitants), with a robust economy and a strong focus on social welfare, which is evident given the country's investments in education, research, and health care. By contrast, Brazil, the fifth largest country globally (208 million inhabitants), has a diversified economy, rich biodiversity, and industrial potential but at the same is characterized by an unequal society. Most of the population still lacks access to high-quality education and health services given the vast disparities of wealth distribution across the country. These differences make the population of these 2 countries very contrasting.

This survey study was part of an academic consortium project, publicly funded, to develop a digital health research repository for youth mental health, in turn defining our main target population (young adults), but without excluding the possibility of collecting data from other, older groups, if those participants would be interested in the study. As the project investigators are in Denmark, participants were first recruited among young Danish residents. Later, to enable the comparison of findings with a divergent historical, cultural, and social context, we

sought to form a partnership with clinical researchers at a university in Northeast Brazil, the Universidade Federal do Rio Grande do Norte (UFRN). Therefore, recruitment efforts were also made to collect data from a sample of young adults in Brazil, but without being restrictive over the age groups that could respond. The online survey was deployed using LimeSurvey and hosted on a server at the Technical University of Denmark (DTU). The survey link was distributed across several channels in an extensive recruitment process that started on March 9, 2020, and ended on December 9, 2020. The link was made available in forums and newsletters at university portals, emailing lists, social networking groups, online chat platforms, and unpaid posts on Twitter. Recruitment in person happened once during an event hosted at DTU (prior to the surge of COVID-19). Participants were compensated with a cup of coffee at this event. Besides this one-time event, no other compensation was given to respondents to avoid providing incentives for repeated participation. Given the distribution of the survey in multiple channels without access restriction, a considerable number of responses came from participants beyond the target population (older than 27 years old). The research team decided not to exclude data from these respondents belonging to age groups above 27 years from analysis; instead, the data collected enabled another dimension for comparison (age). The sample, therefore, includes participants from all age groups residing in Brazil and Denmark.

Ethical and Legal Compliance

Following local jurisdictions, this survey study received ethical approval from the Institutional Review Board of the partner university in Northeast Brazil and was exempt from ethical approval in Denmark. As established by the European General Data Protection Regulation (GDPR), the first page of the survey included information about the study's purpose, which data were collected, measures to anonymity and confidentiality, and data handling processes. Participants were asked to provide their consent after reading this information and confirming that they were older than 18 years. Besides the consent question, none of the questions were mandatory, following standard ethical conduct principles for online research. In addition, the survey was anonymous (IP address and identifiable information were not collected).

Questionnaire Design

Overview

The instrument development was based on (1) several previous surveys and focus groups about public acceptance of biobanks, electronic health records, and clinical trial repositories [43-53]; (2) a previous qualitative study about enablers and barriers for participation in digital health research repositories [32]; and (3) the input from the research team, clinicians, statisticians, and participants of the target population (young adults). The instrument went through several iterations until the final version, which is the one available in Multimedia Appendix 1. First, questions were designed in English and this version was used to ask for the initial rounds of feedback from both experts and targeted participants. After each round of feedback, the questionnaire was incrementally modified. Once a final English version was agreed upon by the research team, the questionnaire was professionally translated to the official language of each country (Portuguese and Danish). The translations were then verified by native speakers from the research team (GV and JB) to ensure content validity. Using the translated versions and the original in English, pilot tests were conducted by the research team with small convenient samples of 5 young adults in Denmark (in person) and in Brazil (remotely). These pilot tests consisted of asking participants to fill the survey and provide feedback on the readability of the questions, comprehension of the vocabulary used for the answer options, navigation of the interface, and time taken to complete all questions. Participants unanimously expressed that the instrument was easy to use and understand and completion time was reasonable. All members of the research team then approved the distribution of the final versions of the questionnaire (in Portuguese, English, and Danish). The complete questionnaire is available in Multimedia Appendix 1. A summary of the survey questions and their rationale are described next. This study is the first to combine such a set of diverse factors to the best of our knowledge.

Demographics and Socioeconomics Questions

This first group of questions inquired about participant age group, gender, education level, country of residence, and usage and ownership of digital devices (computers, smartphones, smartwatches, smart home assistants, and tablets). Some individuals may have a gender that is neither male nor female. They may identify as both male and female at one time, different genders at different times, no gender at all, or dispute the very idea of only 2 genders. Therefore, the term "nonbinary" in this study refers to gender identities outside of the gender binary (male or female). The survey conducted in Brazil also contained 2 additional questions: race and household income (based on minimum salary). Minimum salary has been defined as the minimum amount of remuneration that an employer is required to pay for the work performed during a given period (usually per month), which cannot be reduced by collective agreement or an individual contract. In Brazil, at the time of the study, the minimum salary per month was 1040 Brazilian Reais (approximately US $188.45). Following recommendations by local Brazilian investigators, these questions were added to examine whether the sample reflected the Brazilian population's diversity, which, by contrast, was not considered a usual requirement by local investigators in Denmark.

Factors Related to Technology Acceptance

This group of questions asked participants about factors highlighted by previous research as essential for technology acceptance in data-sharing contexts: self-assessed health status [54], personal interest in health topics [34], previous participation in health research [33], and awareness of examples of data repositories [55]. This group of questions also asked participants about their current attitude toward digital health research repositories (positive, negative, or indifferent) [43,56] after being provided with the following short description of the concept: A research data repository is an online database containing data collected during research studies. In such repositories, deidentified data is to be re-used in the future by other research studies.

Motivations to Participate and Reasons for Concern

These questions asked participants how motivated they would feel by the following reasons to contribute to a research data repository: helping future patients, helping researchers, receiving results about themselves, knowing the research outcomes, getting financial compensation, and proposing questions to be investigated in future studies. Participants were also asked how concerned they would feel about the following risks if their data were stored in a health research repository: having their data used for profit without their knowledge, having data used for projects that they perceive as unethical, agreeing with terms and conditions that they do not fully understand, being socially discriminated against because of the information shared, becoming vulnerable to cyberattacks and blackmail, and being asked to provide more data in the future. Such questions about motivations and concerns were based on findings of a qualitative interview study [32] and previous research on motivations to contribute to research [57] and concerns related to data sharing in general [58,59]. The order of the answer options was randomized for each respondent to avoid order bias.

Access Control Preferences

This group of questions asked participants how desirable or undesirable different access control choices would be once they shared their data with a research platform (answers were not mutually exclusive). The listed answers were: to never be contacted after data are shared, to receive information about who is using the data, to decide who has access to which parts of the data, to have the repository managers decide who has access, to grant data access to public or academic institutions, and to give data access to private laboratories and companies. These questions were based on previous research about informed consent options in biobanks and health data–sharing contexts [48,51,58,60]. The order of the answer options was randomized for each respondent to avoid order bias.

Willingness to Share Data

Questions in this group concerned how comfortable or uncomfortable participants would feel about sharing different deidentified data sources for a research repository, as previous studies have shown that willingness to share personal health data varies according to the data source [61-63]. Data sources were grouped as (1) biospecimen samples and input data provided through health questionnaires (online or in-person); and (2) passive data collected through smartphone or wearable devices, without end user input. The first questions inquired about participants' willingness to share the following: clinical diagnosis (physical), clinical diagnosis (mental), family health status, DNA samples, food consumption, alcohol consumption, sleep patterns, and blood samples. These data types were based on previous studies of willingness to share clinical and health data for research [64]. The second group of questions inquired about participants' willingness to share frequency of social communication (calls/texts), the content of social communication (calls/texts), distances traveled per day, places visited, physical activity levels (heart rate), stress/emotional levels (heart rate), screen time, and apps used. The choice of data types to include in this second group was based on digital data sources previously identified as objective behavioral features for health research [5]. Based on previous studies that showed that different granularities might affect willingness to share, the options in this second group were purposely varied in terms of levels of detail provided by the sensor data (eg, frequency of calls/texts versus the content of calls/texts) [65]. The order of the answer options was randomized for each respondent to avoid order bias.

Statistical Analysis

Data were analyzed and visualized using the R Project for Statistical Computing (software environment for statistical computing and graphics). First, frequency distributions were used to characterize responses for each variable, and bivariate associations (odds ratio [OR]) examined relationships between variables. Following previously established reference values, an OR below 1.5 was considered weak and above 5.0 strong [66]. For a 95% CI, results were considered significant if $P<.05$. Then, a binary logistic regression was conducted to examine directional relationships between explanatory variables and participants' current attitudes toward digital health research repositories. Similarly, another binary logistic regression was conducted to examine directional relationships between explanatory variables and participants' willingness to share data types. Missing values from "prefer not to say" responses were removed before conducting these regression analyses and assumptions were verified beforehand.

Results

Survey Participants

A total of 2299 participants started answering the survey, of whom 1963 completed all questions (1963/2299, 85.38%). This paper includes only responses from participants living in Denmark (583/1600, 36.44%) and Brazil (1017/1600, 63.56%), thus excluding participants residing in other countries from the data analysis for this study (336/1963, 17.12%). The majority of the sample is aged between 18 and 27 years (933/1600, 58.31%); the second largest age group is between 28 and 37 years (459/1600, 28.69%). Only 12.56% (201/1600) were aged above 37 years. There are slightly more individuals who identify as females (891/1600, 55.69%) than males (682/1600, 42.63%). A majority of participants are educated, having at least a university degree (992/1600, 62.00%), own and use smartphones (1562/1600, 97.63%) and computers (1537/1600, 96.06%), but only 36.75% (588/1600) own and use more than 2 types of digital devices. The vast majority is currently in good, very good, or excellent health (1407/1600, 87.94%), while most are moderately, very, or extremely interested in health topics (1088/1600, 68.00%). Around half of the Brazilian participants (555/1017, 54.57%) are White and 43.17% (439/1017) are Black or Brown; most of the respondents living in Brazil have a monthly household income between 1 (1040 Brazilian Reais or US $190) and 5 (5200 Brazilian Reais or US $950) minimum salaries (739/1017, 72.66%). As explained in the previous section, information about race and income was not collected in the Danish survey. Further details on the sample characteristics are presented in Table 1.

Table 1. Participants' characteristics, awareness, past experiences, and attitudes.

Variables	All participants (N=1600), n (%)	Participants in Brazil (n=1017), n (%)	Participants in Denmark (n=583), n (%)
Age (years)			
18-27	933 (58.31)	613 (60.28)	320 (54.89)
28-37	459 (28.69)	273 (26.84)	186 (31.90)
38-47	105 (6.56)	80 (7.87)	25 (4.29)
48-57	64 (4.00)	38 (3.74)	26 (4.46)
>57	32 (2.00)	9 (0.88)	23 (3.95)
Prefer not to say	7 (0.44)	4 (0.39)	3 (0.51)
Gender			
Female	891 (55.69)	606 (59.59)	285 (48.89)
Male	682 (42.63)	399 (39.23)	283 (48.54)
Nonbinary	9 (0.56)	5 (0.49)	4 (0.69)
Prefer not to say	18 (1.13)	7 (0.69)	11 (1.89)
Self-reported race			
White	—[a]	555 (54.57)	—
Black or Brown	—	439 (43.17)	—
Yellow	—	3 (0.29)	—
Indigenous	—	2 (0.20)	—
Prefer not to say	—	18 (1.77)	—
Household income (monthly)[b]			
Less or equal to 1 minimum salary	—	114 (11.21)	—
Between 1 and 3 minimum salaries	—	340 (33.43)	—
Between 3 and 5 minimum salaries	—	399 (39.23)	—
Higher or equal to 5 minimum salaries	—	113 (11.11)	—
Prefer not to say	—	51 (5.01)	—
Education			
Less than secondary education	1 (0.06)	1 (0.10)	0 (0)
Currently on higher education	595 (37.19)	479 (47.10)	116 (19.90)
Higher education degree completed	992 (62.00)	527 (51.82)	465 (79.76)
Prefer not to say	12 (0.75)	10 (0.98)	2 (0.34)
Digital devices owned			
Smartphone(s)	1562 (97.63)	994 (97.74)	568 (97.43)
Computer(s)	1537 (96.06)	962 (94.59)	575 (98.63)
Tablet(s)	399 (24.94)	194 (19.08)	205 (35.16)
Smartwatch(es)	267 (16.69)	145 (14.26)	122 (20.93)
Smarthome assistant(s)	132 (8.25)	59 (5.80)	73 (12.52)
Number of digital device types owned			
0	5 (0.31)	5 (0.49)	0 (0)
1	64 (4.00)	50 (4.92)	14 (2.40)
2	943 (58.94)	666 (65.49)	277 (47.51)
3	437 (27.31)	227 (22.32)	210 (36.02)
4	119 (7.44)	54 (5.31)	65 (11.15)

Variables	All participants (N=1600), n (%)	Participants in Brazil (n=1017), n (%)	Participants in Denmark (n=583), n (%)
5 or more	32 (2.00)	15 (1.47)	17 (2.92)
Current health status			
Poor	30 (1.88)	16 (1.57)	14 (2.40)
Fair	157 (9.81)	119 (11.70)	38 (6.52)
Good	513 (32.06)	355 (34.91)	158 (27.10)
Very good	666 (41.63)	406 (39.92)	260 (44.60)
Excellent	228 (14.25)	117 (11.50)	111 (19.04)
Prefer not to say	6 (0.38)	4 (0.39)	2 (0.34)
Interest in health topics			
Not interested	39 (2.44)	21 (2.06)	18 (3.09)
Slightly interested	471 (29.44)	247 (24.29)	224 (38.42)
Moderately interested	124 (7.75)	72 (7.08)	52 (8.92)
Very interested	559 (34.94)	340 (33.43)	219 (37.56)
Extremely interested	405 (25.31)	336 (33.04)	69 (11.84)
Prefer not to say	2 (0.13)	1 (0.10)	1 (0.17)
Previous participation in a health research study			
No	815 (50.94)	440 (43.26)	375 (64.32)
Yes	763 (47.69)	567 (55.75)	196 (33.62)
Prefer not to say	22 (1.38)	10 (0.98)	12 (2.06)
Awareness of examples of research data repositories			
No	884 (55.25)	528 (51.92)	356 (61.06)
Yes	459 (28.69)	330 (32.45)	129 (22.13)
Not sure	245 (15.31)	152 (14.95)	93 (15.95)
Prefer not to say	12 (0.75)	7 (0.69)	5 (0.86)
Perception of digital health data repositories			
Positive	1339 (83.69)	927 (91.15)	412 (70.67)
Indifferent	188 (11.75)	53 (5.21)	135 (23.16)
Negative	45 (2.81)	19 (1.87)	26 (4.46)
Prefer not to say	28 (1.75)	18 (1.77)	10 (1.72)

[a]Data not collected.

[b]Ranges between 1 (1040 Brazilian Reais or US $190) and 5 (5200 Brazilian Reais or US $950).

Previous Participation, Awareness of Examples, and Current Attitude

Around half of the respondents participated in a health research study before (763/1600, 47.69%), and those who participated are more likely to have a moderate to high interest in health topics (OR 2.35, 95% CI 1.88-2.93; *P*<.001). By contrast, only a minority are aware of research data repository examples (459/1600, 28.69%). Those aware of examples are more likely to have a moderate to high interest in health topics (OR 3.02, 95% CI 2.30-3.96; *P*<.001) and to have been participants in previous health studies (OR 3.36, 95% CI 2.66-4.23; *P*<.001). In addition, most participants have a positive perception of health research data repositories (1339/1600, 83.69%), and those who have a positive perception are more likely to be aware of examples of research data repositories (OR 3.26, 95% CI 2.17-4.90; *P*<.001). Further details on the frequency distribution for these variables are shown in Table 1.

Results from a binary logistic regression show that interest in health topics (OR 1.49, 95% CI 1.10-2.02; *P*=.01), previous participation in health research studies (OR 1.70, 95% CI 1.24-2.35; *P*=.001), and awareness of examples of existing repositories (OR 2.78, 95% CI 1.83-4.38; *P*<.001) are significant factors influencing participants' current perception of digital health research repositories. See the results of the binary logistic regression in Table 2.

Table 2. Binary logistic regression model for the current perception of digital health data repositories (base: not positive perception).

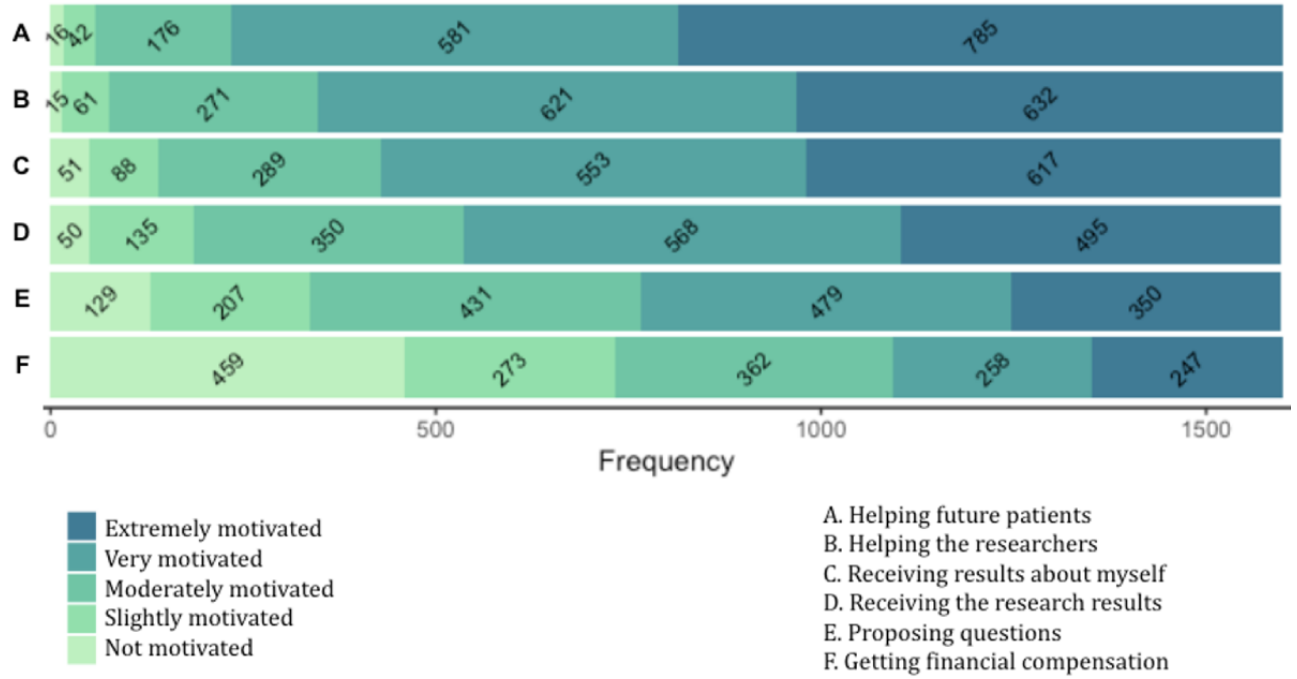
Factors for current perception digital health data repositories (base: not positive perception)	Estimate (B)	Standard error B	*P* value	Odds ratio (95% CI)
Age (base: above 27)				
Below 27 years	0.10	0.17	.56	1.11 (0.79-1.55)
Gender (base: not female)				
Female	−0.05	0.15	.73	0.95 (0.70-1.28)
Education (base: no university degree)				
With university degree	−0.29	0.17	.10	0.75 (0.52-1.06)
Device ownership (base: less than 2 device types)				
Owns more than 2 types	−0.26	0.15	.07	0.76 (0.56-1.04)
Health status (base: poor or fair health)				
Good, very good, or excellent health	0.06	0.23	.79	0.94 (0.57-1.74)
Interest in health (base: none or slight interest)				
Moderate to extreme interest	0.39	0.15	.01	1.49 (1.10-2.02)
Participation in health study (base: no past participation)				
Participated in a health study	0.53	0.16	.001	1.70 (1.24-2.35)
Awareness of an example (base: no awareness or not sure)				
Aware of an example of repository	1.02	0.22	<.001	2.78 (1.83-4.38)

Motivations to Participate

The majority of participants feel very or extremely motivated by helping future patients (1366/1600, 85.38%), helping researchers (1253/1600, 78.31%), receiving results about themselves (1170/1600, 73.13%), and receiving the results of the research (1063/1600, 66.44%). In addition, being provided with the possibility of suggesting research questions to be investigated is very or extremely motivating for more respondents (829/1600, 51.81%) than receiving financial compensation (505/1600, 31.56%), which is not motivating for 28.69% (459/1600). Multimedia Appendix 2 shows the entire distribution of responses, and Figure 1 displays this information as stacked bar charts.

Those who have a positive perception about health data repositories are more likely to be moderately, very, or extremely motivated by 5 out of 6 motivation sources: helping future patients (OR 9.44, 95% CI 5.43-16.40; *P*<.001), helping researchers (OR 5.74, 95% CI 3.56-9.25; *P*<.001), receiving results about themselves (OR 4.12, 95% CI 2.82-6.03; *P*<.001), receiving results of the research (OR 4.15, 95% CI 2.94-5.85; *P*<.001), and proposing questions to be investigated (OR 3.46, 95% CI 2.57-4.66; *P*<.001). Those moderately, very, or extremely interested in health topics are more likely to be moderately, very, or extremely motivated by receiving results of the research (OR 2.25, 95% CI 1.65-3.06; *P*<.001) and proposing questions to be investigated (OR 2.53, 95% CI 1.97-3.24; *P*<.001). The youngest segment (18-27 years old) is more likely to feel moderately, very, or extremely motivated to receive financial compensation (OR 1.92, 95% CI 1.57-2.35; *P*<.001).

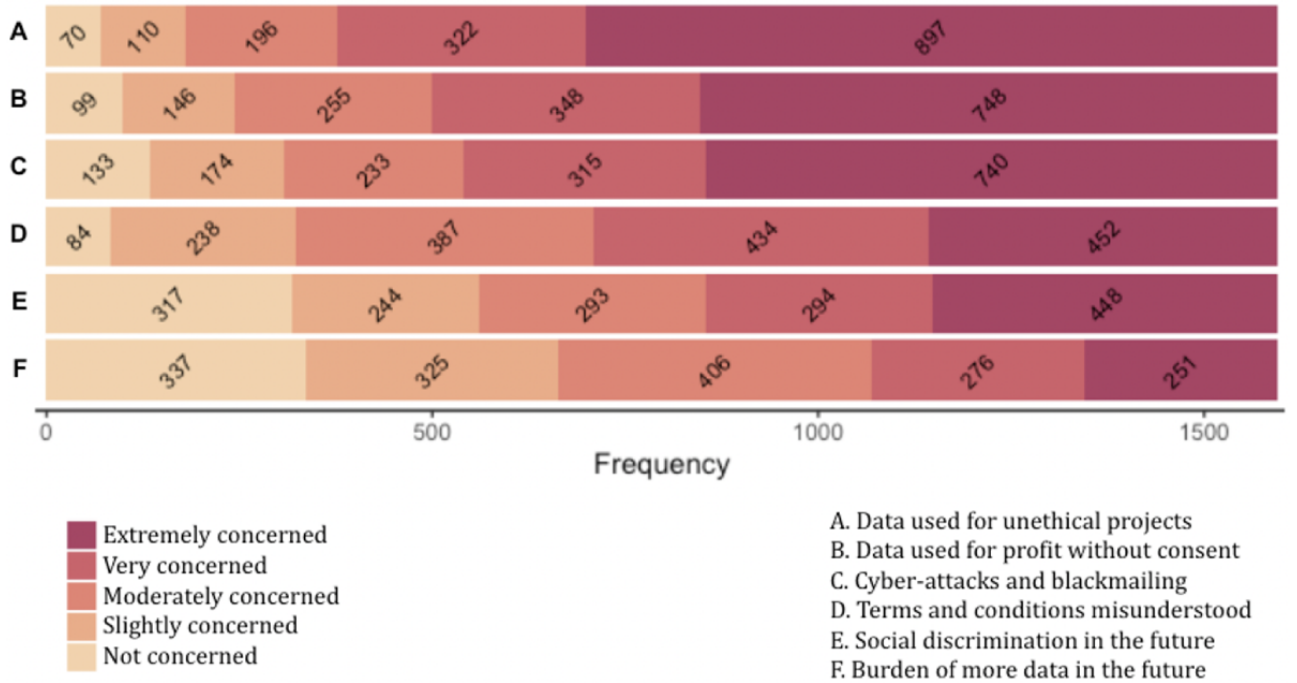
Figure 1. Bar chart displaying the distribution of answers for each motivation source.



Reasons for Concern

The vast majority of participants feel very or extremely concerned about having their data used for unethical projects (1219/1600, 76.19%) and profit making without their consent (1096/1600, 68.50%). In addition, the risk of becoming vulnerable to cyberattacks and blackmail is very or extremely concerning for 65.94% (1055/1600); the possibility of not understanding terms and conditions for 55.38% (886/1600); and the fear of being socially discriminated for 46.38% (742/1600). By contrast, not as many participants feel very or extremely concerned about the burden of being asked to share more data in the future (527/1600, 32.94%). Multimedia Appendix 3 shows the entire distribution of responses, and Figure 2 displays this information in the form of a stacked bar chart.

Figure 2. Bar chart displaying the distribution of answers for each reason for concern.



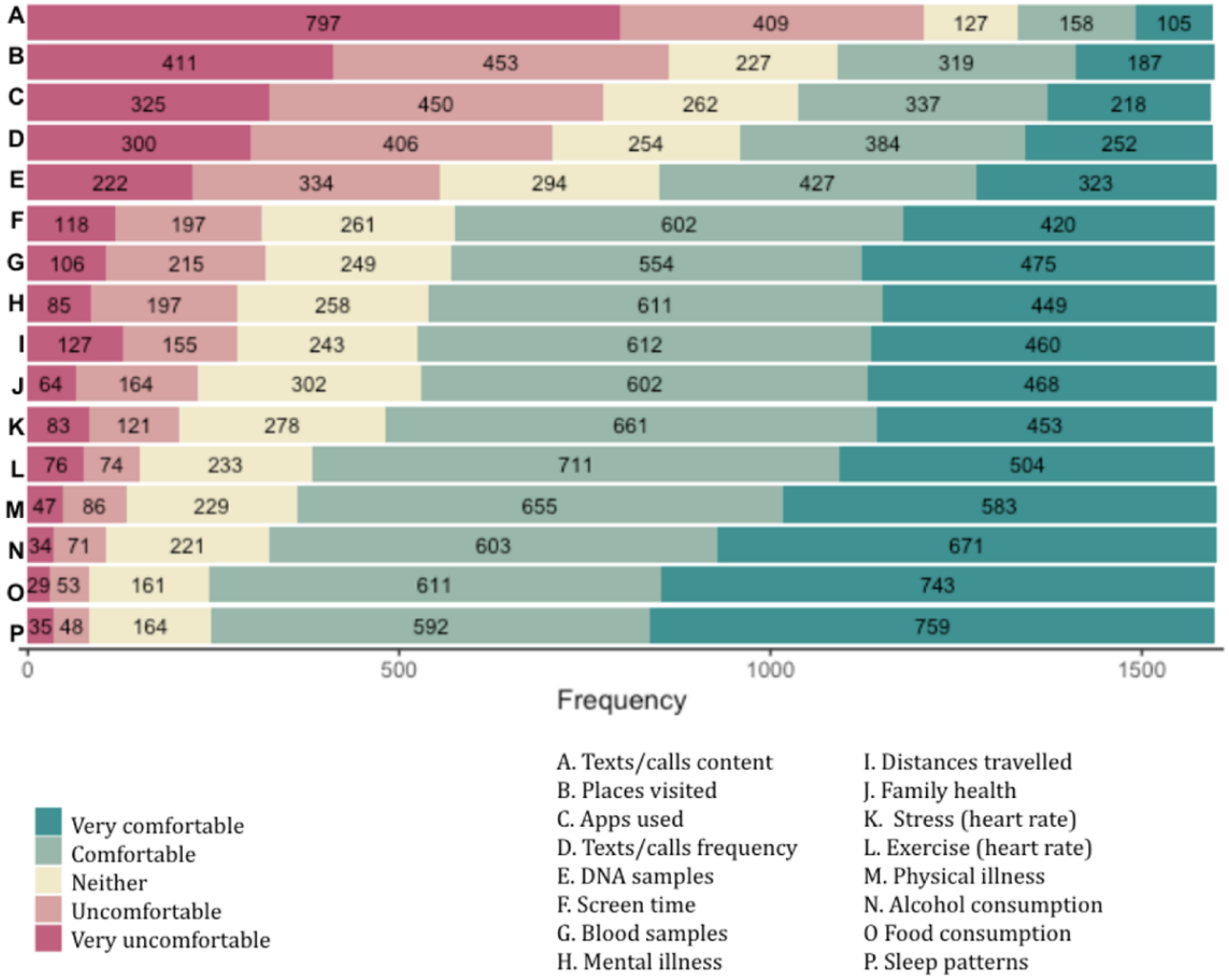
Willingness to Share Different Types of Data

Regarding the willingness to share specific data items, most participants feel uncomfortable or very uncomfortable sharing the content of texts and calls (1206/1600, 75.38%), while fewer participants feel uncomfortable or very uncomfortable sharing the frequency of texts and calls (706/1600, 44.13%). Places visited (864/1600, 54.00%) and apps used (775/1600, 48.44%)

are perceived as uncomfortable or very uncomfortable data to share by many.

By contrast, most participants feel comfortable or very comfortable sharing sleeping patterns (1351/1600, 84.44%), food consumption (1354/1600, 84.63%), alcohol consumption (1274/1600, 79.63%), physical illness diagnosis (1238/1600, 77.38%), physical activity levels (1215/1600, 75.94%), stress levels (1114/1600, 69.63%), family health history (1070/1600, 66.88%), distances traveled (1072/1600, 67.00%), mental illness diagnosis (1060/1600, 66.25%), blood samples (1029/1600, 64.31%), DNA samples (750/1600, 46.88%), and screen time (1022/1600, 63.88%). Multimedia Appendices 4 and 5 show the full distribution of responses, and Figure 3 displays this information as stacked bar charts.

Figure 3. Bar chart displaying the distribution of willingness to share across different data types.



Those who have a positive perception about health research repositories are more likely to feel comfortable or very comfortable sharing 15 out of 16 data items: physical illness diagnosis (OR 3.84, 95% CI 2.87-5.15; *P*<.001), mental illness diagnosis (OR 3.44, 95% CI 2.59-4.59; *P*<.001), family health history (OR 3.45, 95% CI 2.59-4.59; *P*<.001), DNA samples (OR 2.51, 95% CI 1.85-3.41; *P*<.001), blood samples (OR 2.31, 95% CI 1.74-3.06; *P*<.001), food consumption (OR 4.15, 95% CI 3.01-5.70; *P*<.001), alcohol consumption (OR 3.25, 95% CI 2.41-4.40; *P*<.001), sleep (OR 3.85, 95% CI 2.80-5.30; *P*<.001), screen time (OR 3.17, 95% CI 2.38-4.22; *P*<.001), apps used (OR 2.09, 95% CI 1.50-2.91; *P*<.001), frequency of calls or texts (OR 2.07, 95% CI 1.51-2.83; *P*<.001), distances traveled per day (OR 3.34, 95% CI 2.51-4.45; *P*<.001), places visited (OR 2.77, 95% CI 1.91-4.00; *P*<.001), physical activity levels (OR 3.43, 95% CI 2.56-4.59; *P*<.001), and stress levels (OR 3.60, 95% CI 2.70-4.79; *P*<.001). However, no significant association was found between having a positive perception of digital health research repositories and feeling comfortable with sharing the content of calls and texts (*P*=.03).

Those moderately, very, or extremely concerned about being discriminated against are more likely to feel uncomfortable or very uncomfortable sharing data about mental illness diagnosis (OR 2.26, 95% CI 1.66-3.07; *P*<.001). Those uncomfortable or very uncomfortable sharing information about app usage are more likely to be moderately, very, or extremely concerned about data being used for profit (OR 2.57, 95% CI 1.91-3.46; *P*<.001) and not understanding terms and conditions (OR 2.22, 95% CI 1.71-2.87; *P*<.001). Those not motivated or only slightly motivated by receiving results about themselves are more likely to feel uncomfortable or very uncomfortable with sharing information about alcohol consumption (OR 5.76, 95% CI 3.63-9.13; *P*<.001), distances traveled per day (OR 3.31, 95% CI 2.29-4.80; *P*<.001), stress levels (OR 6.46, 95% CI 4.43-9.44;

P<.001), and physical activity levels (OR 6.78, 95% CI 4.52-10.17; *P*<.001).

A small number of participants feel uncomfortable or very uncomfortable sharing any of the data items (94/1600, 5.88%). Those who feel uncomfortable or very uncomfortable sharing any data items are more likely to have a negative or indifferent perception about health research repositories (OR 3.91, 95% CI 2.49-6.14; *P*<.001). A binary logistic regression shows that age (OR 2.16, 95% CI 1.28-3.70; *P*=.004), digital device ownership (OR 1.90, 95% CI 1.14-3.26; *P*=.01), health status (OR 2.28, 95% CI 1.24-3.98; *P*=.01), and current attitude regarding digital health research repositories (OR 3.77, 95% CI 2.24-6.26; *P*<.001) are significant factors affecting participants' willingness to share data with a health research repository. Table 3 shows the results of the binary logistic regression.

Table 3. Binary logistic regression model for willingness to share data with repositories for health research (base: unwilling to share any data).

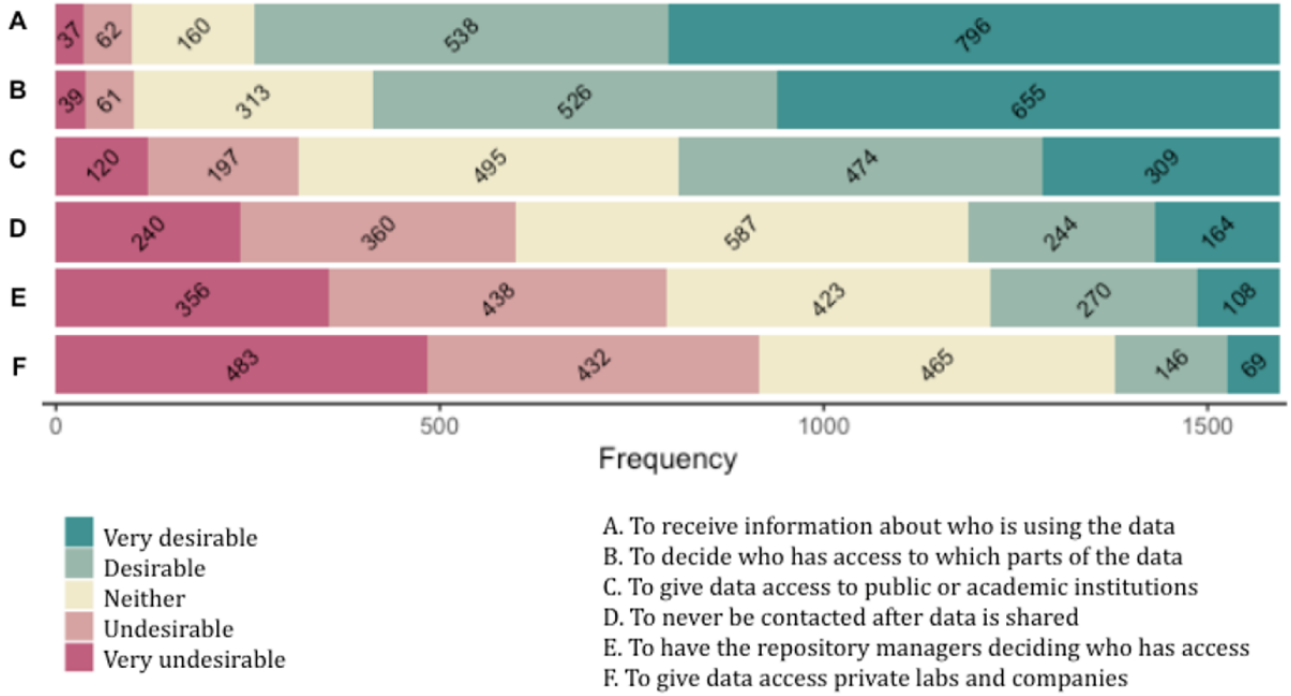
Factors for willingness to share data (base: unwilling to share any)	Estimate (B)	Standard error B	*P* value	Odds ratio (95% CI)
Age (base: above 27)				
Below 27 years	0.76	0.27	.004	2.16 (1.28-3.70)
Gender (base: not female)				
Female	0.06	0.23	.78	1.06 (0.66-1.70)
Education (base: no university degree)				
With university degree	0.21	0.28	.45	1.24 (0.70-2.16)
Device ownership (base: less than 2 device types)				
Owns more than 2 devices	0.64	0.26	.01	1.90 (1.14-3.26)
Health status (base: poor or fair health)				
Good, very good, or excellent health	−0.82	0.29	.005	2.28 (1.24-3.98)
Interest in health (base: no or slight interest)				
Moderate to extreme interest	−0.01	0.25	.95	0.99 (0.59-1.62)
Participation in health study (base: no past participation)				
Participated in a health study	0.14	0.24	.55	1.16 (0.71-1.90)
Awareness of an example (base: no awareness)				
Aware of an example of repository	−0.12	0.28	.65	0.88 (0.51-1.56)
Current perception (base: negative or indifferent)				
Positive current perception	1.32	0.26	<.001	3.77 (2.24-6.26)

Preferred Access Control Options

After collecting and sharing their data with a research platform, most participants find it desirable or very desirable to receive information about which projects access their data in the future (1334/1600, 83.38%). The majority also find it desirable or very desirable to decide who gets access to which parts of their data (1181/1600, 73.81%). By contrast, not being contacted is desirable or very desirable to only 25.50% of participants (408/1600), and the option to allow the owners of the repositories to decide who can access the data is desirable or very desirable only to 23.63% (378/1600). Finally, allowing public or academic institutions to access the data is desirable or very desirable for 48.94% (783/1600), while allowing private laboratories and companies to obtain access is desirable or very desirable to only 13.44% (215/1600). Multimedia Appendix 6 shows the entire distribution of responses, and Figure 4 displays this information as stacked bar charts.

Those who find it is desirable or very desirable to be informed about who is using their data are more likely to have a positive perception of health data repositories (OR 2.45, 95% CI 1.77-3.39; *P*<.001). Those moderately, very, or extremely concerned about data being used for unethical projects are more likely to find it desirable or very desirable to have control over how their data are used (OR 2.45, 95% CI 1.80-3.42; *P*<.001) and to be informed about it (OR 3.09, 95% CI 2.18-4.37; *P*<.001). Those moderately, very, or extremely concerned about data being used for profit are more likely to find it undesirable or very undesirable to have private laboratories and companies access their data (OR 2.24, 95% CI 1.69-2.96; *P*<.001).

Figure 4. Bar chart displaying the distribution of answers for each access control option.



Cross-country Analysis

The sample comprises 1017 respondents living in Brazil (1017/1600, 63.56%) and 583 living in Denmark (583/1600, 36.44%). The demographics of the participants residing in these 2 countries do not differ substantially, except for a higher percentage of female respondents and a lower percentage of respondents who completed a higher education degree within the Brazilian sample (Table 1). The Brazilian sample also has a higher percentage of extremely interested individuals in health topics than the Danish sample (Table 1).

Around half of the participants in Brazil participated in health research previously (567/1017, 55.75%), in contrast to a minority of the participants in Denmark (196/583, 33.62%). Similarly, the vast majority of participants from the Brazilian sample have a positive initial impression of health data repositories (927/1017, 91.15%), in contrast to a smaller majority of participants in Denmark (412/583, 70.67%). See Table 1 for complete information regarding these differences.

The majority of participants from both countries are highly motivated and concerned by similar sources of motivation and concerns; however, there are differences in the magnitude of the motivation and concern levels reported by those residing in Brazil and Denmark. The Brazilian sample is more likely to rate themselves as moderately, very, or extremely motivated by receiving results about themselves (OR 6.35, 95% CI 4.25-9.50; *P*<.001), proposing questions to be investigated (OR 6.08, 95% CI 4.67-7.91; *P*<.001), receiving results of the research (OR 4.13, 95% CI 2.98-5.72; *P*<.001), and helping the researchers (OR 3.36, 95% CI 2.07-5.44; *P*<.001). The Brazilian sample is also more likely to rate levels of concerns as moderately, very, or extremely concerning for all of the options listed: having data used for unethical projects (OR 5.44, 95% CI 3.86-7.66; *P*<.001), becoming vulnerable to cyberattacks and blackmail (OR 3.88, 95% CI 2.99-5.04; *P*<.001), having data used for profit without consent (OR 3.68, 95% CI 2.77-4.89; *P*<.001), being asked to provide more data (OR 3.28, 95% CI 2.65-4.06; *P*<.001), agreeing to terms without understanding them (OR 2.29, 95% CI 1.79-2.94; *P*<.001), and being socially discriminated against (OR 2.10, 95% CI 1.70-2.60; *P*<.001). Multimedia Appendices 2 and 3 show the frequency distribution of answers regarding motivations and concerns according to country of residence.

When it comes to access preferences, important differences arise between the 2 countries. The Brazilian sample is more likely to find it desirable or very desirable to receive information about who is using the data (OR 5.51, 95% CI 4.12-7.37; *P*<.001). By contrast, the Danish sample is more likely to find it desirable or very desirable never to be contacted (OR 3.63, 95% CI 2.87-4.60; *P*<.001), to have the repository managers decide who can obtain access (OR 2.84, 95% CI 2.24-3.60; *P*<.001), and to allow private organizations (OR 3.73, 95% CI 2.77-5.04; *P*<.001) and public institutions access the data (OR 4.51, 95% CI 3.61-5.63; *P*<.001). Multimedia Appendices 4-6 show the frequency distribution of answers regarding access control preferences and willingness to share according to country of residence.

In summary, the vast majority of the participants residing in Brazil have a positive attitude regarding the idea of health data repositories. These findings are further endorsed by the Brazilian sample reporting higher motivation to help the researchers and willingness to share several data types. However, those residing in Brazil are also more likely to be strongly concerned about all of the potential negative consequences. The Brazilian sample is also more likely to find it desirable to keep the control and be informed about the use of the shared data, rather than never being contacted, delegating control to repository owners, or allowing both private companies and public institutions to get access.

Discussion

Enablers for Acceptance

Our survey contributes novel empirical insights regarding an extensive set of factors contributing to the acceptance of repositories storing biosamples, health records, and digital data sources for observational research. Previous research suggests that individuals may view some loss of privacy as worthwhile to advance medical research and benefit future generations [4,33], with altruism being a strong incentive for participation in clinical studies [50,67]. Aligned with such previous research, we found that helping future patients and researchers is indeed a powerful source of motivation across our sample, with most participants also feeling very motivated by the prospect of being updated about research outcomes. Furthermore, our findings show that those who do not feel motivated by helping future patients and researchers are more likely to be unwilling to share data, highlighting the critical role of altruism in this context.

Participants are also motivated by learning about their health through the data they provide, aligning with past research [68]. By contrast, our findings indicate that financial compensation may not be a more decisive factor than other sources of motivation. For instance, being invited to suggest research questions for a project strongly motivates more participants than financial compensation. However, consistent with previous studies [41,50], the youngest participants in our sample are more likely to be motivated to share health data in exchange for financial benefits. Such observations reinforce the importance of providing both societal and individual benefits to accommodate different preferences.

Another essential enabler for acceptance is individuals' current perception of the idea of health research data repositories. In our survey, a positive perception appears to be associated with higher levels of motivation to help patients and researchers, and those who have a positive perception are also more likely to feel comfortable sharing 15 out of 16 data items. These results confirm past research highlighting that a positive opinion about biomedical research can predict willingness to participate [33], and attitudes about health care interventions can predict patient acceptance [56]. We also extend previous findings from other contexts [34,43,55] by providing evidence about key factors that can affect individuals' perceptions of digital health research repositories, emphasizing the critical role that positive past experiences and personal interests have in enabling favorable attitudes.

Regarding cross-country differences, the vast majority of our participants residing in Brazil have a positive perception of the idea of health data repositories, further demonstrated by their reported higher motivation to participate and higher willingness to share several data types. It could be speculated that such enthusiasm stems from the prospect of significantly improving an imperfect yet ubiquitous public health care system, which may become an essential enabler for acceptance as digital health emerges in Brazil [67]. By contrast, Denmark has a long history of using clinical databases and electronic health records for population-level clinical research [44]. This observation could explain why our sample residing in Denmark is more likely to find it desirable to allow repository owners to make decisions regarding access control, an arrangement already familiar to them, as the Danish public health authorities manage data use. These differences illustrate that acceptance depends not only on individual predispositions but also on broader sociocultural contexts [36].

Barriers for Acceptance

In contrast to such enablers, our findings show that even though participation in research repositories might occur under the promise of sharing deidentified data, participants still report concerns. Our sample's most substantial concern is the fear that their data will eventually be used for unethical research goals or profit without consent, which is a fear also reported by several previous studies [37,38,42]. Further aligned to previous research, the fear of cyberattacks or blackmail is considered very or extremely concerning to most of our participants [69-72]. Surprisingly, however, the fear of social discrimination is not as prevalent, contrasting a previous study's claim that this might be a core reason behind privacy concerns [58]. It is also surprising that the fear of not fully understanding terms and conditions was a more significant concern for participants than the burden of providing more data, which contradicts previous findings from another study [4].

The predominant concern of data misuse may explain the preference for more restrictive access control options. Many of our participants report feeling comfortable sharing their data if the purpose is to protect the common good, but the same does not apply to the prospect of supporting others' profit making, in alignment with previous research [36,44,73]. Related to this, the large majority of our sample want to receive information about the different projects using their data, and most also want to be deciding who can ultimately obtain access to their data, a finding which has been highlighted in other past studies [38,55,73-75]. By contrast, leaving this responsibility to repository owners is often not our participant' preferred option, especially within the Brazilian sample. Furthermore, approaches such as notification-only and opt-out options have been considered less acceptable than reconsent [73,76], showing the importance of reconsidering usual consent practices.

Another barrier to participation is that willingness to share data depends strongly on the data type [32,47,67], even though there are divergent findings in the literature about which data types people feel most uncomfortable sharing [37,38,42,61-65]. For example, previous studies with young adults have observed a high willingness to donate DNA samples [33,34], but 2 extensive worldwide surveys have observed the opposite [36,44]. Our analysis indicates that when compared with behavioral indicators such as food consumption and sleeping patterns, DNA and blood samples are among the data types most uncomfortable to be shared.

However, even more so than DNA, participants in our study feel uncomfortable sharing passive mobile and wearable sensing data. Interestingly, these are data with the less obvious connection to health in a traditional sense. While the relationship between health and food consumption or sleep might be apparent to many people, the relevance of app use or social communication data may be less noticeable. Such observation

is particularly relevant for behavioral health research contemplating passive data sources as a strategy to reduce the data collection burden for participants. Our results also add a more nuanced understanding of participants' willingness to share data. We empirically demonstrate that participants feel uncomfortable sharing more detailed and revealing data sources, such as apps used, frequency of texts and calls, and places visited, compared with broader and less granular information such as screen time, the content of texts and calls, and distances traveled. These findings have important implications for health research studies that consider collecting high granularity information, especially when it comes to location and social communication.

Furthermore, sociodemographic factors have been emphasized by several past studies as possible barriers to the willingness to share data [35,36,39,43,45,54,74]. Our analysis shows that participants' willingness to share data can be related to age, health status, and digital device ownership. However, contrary to previous studies, which observed that members of American ethnic groups other than White have higher odds of being unwilling to donate their DNA data [43,45], our study does not find a significant association between race and unwillingness to share. We also do not find significant associations between race and fear of discrimination [43,45] or desire to control data access [35]. However, our sample is in its vast majority young and educated, in contrast to these previous studies.

Research Implications

The empirical findings discussed above provide the basis for a series of implications for community engagement, technology design, and policymaking. First, we found evidence that a lack of knowledge about health research may be a challenge for public acceptance, which points to the importance of broadening public awareness. For instance, education and familiarity-increasing programs can be possible community engagement approaches and strengthened relationships between potential participants, clinicians, and health research experts may be helpful during recruitment and beyond [40]. Regardless of the medium, participant information could include explanations about the collaborative nature of contemporary health research and why digital data sources extracted passively may be necessary for answering specific research questions. Given the factors found to motivate and demotivate data sharing, it may be necessary to explain the benefits of sharing data types where the direct connection to health is not immediately visible. Additionally, appropriate communication may help to emphasize the importance of data collection compliance to participants, especially when it comes to experience sampling and the provision of frequent self-reports [77].

Personal health informatics could also be considered to increase the appeal of and the motivation for participating. Given that data collection may require interactions with mobile and wearable devices, it is a natural step to also provide participants with personalized data visualizations and, potentially, digital health interventions. However, digital tools for personal health must consider how existing health care practices complement (or hinder) novel approaches [78-80]. Interface design should focus on suitably informing patients about how their data relate to their health to facilitate rather than replace efficient clinician–patient relationships. Above all, risks to individual well-being should be avoided, as an intense "datafication" of personal health standards might prove to be more harmful than beneficial [81]. For instance, our analysis shows that those uncomfortable with sharing alcohol consumption, levels of stress, and physical exercise are less likely to feel motivated by receiving results about themselves. Thus, any consideration of adding personal health informatics features to health research systems should be mindful of the preferences of each individual.

Furthermore, our analysis makes it clear that broader acceptance will be challenging to achieve if contributing to health research repositories demands that participants share every digital source of data [38]. Health research projects might need to acknowledge that certain personal information is associated with social stigma [82], which may compromise willingness to participate in research as a whole. For instance, we observe a strong association between fear of discrimination and unwillingness to share mental illness diagnoses. For this reason, health research should consider personal boundaries by allowing participants to opt-out from specific data collection types and decide which level of details are to be shared. Even if individuals do not exercise this right to choose, the option to safely do that without negative consequences may still enhance trust [83].

When it comes to access control options, our results show that participants would like to be informed about the different projects which may access their data and customize their consent. Even though granular data control options may reduce privacy concerns [84], broad consent models are still the most used approach in current health research platforms [32], which means that once participants provide their consent, they are usually not consulted about data reuse in the future. The conception of digital systems for continuous communication with participants could transform consent practices. For instance, research participants could be consulted about whether they would like to receive a request each time a new project wants to use their data. Access requests could include details about who benefits from the research outcomes and how organizations use any profit. The possibility of opting out from data sharing could also be provided. Beyond allowing participants to make choices about data access, participants could further contribute with questions to a research project, which is an interest identified in our survey and explored in other research platforms [22]. Nondigital approaches could also be considered (eg, phone calls, letters) for those who prefer or do not have access to digital devices. However, a challenge is how to help participants stay informed and control their data without making them overwhelmed [42].

As pervasive sensing technologies become more refined and widespread in health research, those proposing shared-access repositories for collecting, sharing, and using such sensing data will need to take responsibility for identifying risks and be accountable for consequences against participants' best interests. Proactive legal and ethical guidelines are necessary, as current regulatory frameworks for digital health data sharing are relatively weak in some jurisdictions [44]. Likewise, regulatory board members and grant reviewers could evaluate how managers of digital health research repositories demonstrate

awareness of ethical considerations and strategies to mitigate possible negative consequences of participation. For instance, being transparent about the trustworthiness of the technical infrastructures and governance arrangements of the platforms hosting the data is essential, even if it means acknowledging challenges [50]. Clear and understandable evidence of compliance with regulations may help diminish individuals' reticence to share health data and increase public acceptance.

Finally, future developments should not ignore that without a diverse cohort providing data, research outcomes and benefits will be unevenly distributed [42,85]. Even though our sample, composed mainly of educated young individuals, does not show significant associations between race, income, and unwillingness to share data, other past studies have shown that these factors can be significant [43,45]. For this reason, communication efforts, interface design, and data sharing policies should be made accessible and inclusive by being mindful of language choices, cultural requirements, access costs, and participation demands (eg, owning and using smartphones and smartwatches). After all, strategies to increase acceptance should be motivated by research repository owners' genuine desire to make data sharing fairer and more ethical.

Limitations and Future Work

Based on our team experience, we suggest that similar surveys in the future should strive to focus recruitment efforts on racial, ethnic, gender, and disability minorities to achieve a higher representation from these groups. We also suggest that quantitative findings should be complemented with parallel qualitative investigations, to provide richer and subjective insight into justifications and reasonings behind responses. Another suggestion is to consider depicting data usage scenarios with illustrations, infographics, and narrative forms instead of purely descriptive texts.

In terms of methodological limitations, sampling bias is a common challenge of voluntary response samples, given that those who take the time to respond to online survey requests tend to have strong opinions compared with the rest of the population. Despite our efforts to use a diverse range of digital channels for recruitment, our sample is biased in terms of age and education level, partially due to the scope of the overarching project (mental health for young adults) and partially because the higher response rate stemmed from posts on university web portals. A constrained focus on specific population segments can be considered a strength, given that it allows the investigation of particular perspectives at a time; however, future work must seek to consider the perspectives of many other groups of individuals, especially underrepresented minorities.

Another methodological limitation is that, as an exploratory cross-sectional study, our results can only capture attitudes about hypothetical future participation. Thus, future work must consider evaluating participants' perspectives while taking part in an actual digital health research repository initiative. In addition, factors such as the burden of continuous data collection might be better examined throughout actual participation.

Finally, it could be speculated that public attitudes may shift following the global experience of a public health emergency (COVID-19 pandemic). Given that this survey was conducted during the first waves of the pandemic, future work is still needed to evaluate further consequences of this unprecedented crisis in the long term. In particular, the impact of contact tracing apps and vaccination passports may prove significant when it comes to the acceptance of digital health data storage on a population level.

Conclusion

This survey study reveals essential factors for potential acceptance and willingness to share personal data with a digital health research repository. In summary, most participants feel very motivated about helping future patients, helping researchers, and receiving results about their health; most also feel comfortable sharing data sources usually associated with health research, except DNA data. However, most respondents feel very concerned about the risk of cyberattacks, the possibility of data being used for unethical research goals or for-profit without consent, and the prospect of sharing personal sensing data, especially social communication and location. The majority of participants find it desirable to receive information about which projects access their data and would like to be able to decide who gets access to which parts of their data.

The analysis of such a large spectrum of variables and their relationships provides a strong foundation for suggesting implications for future developments. The implications discussed include to disseminate knowledge about health research; to value the role of transparency for trust development; to engage participants with the research process and their health management; to allow flexible and customizable data sharing; and to align policies and regulations with ethical considerations. Providing valuable benefits for individuals and reducing the risks involved in participation are essential requirements in this context, and by recognizing differences between groups, it is possible to better understand and respond to individual views and expectations.

Acknowledgments

This study is part of the TEAM (Technology Enabled Mental Health for Young People), a project funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 722561.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire applied to participants (survey questions).
[PDF File (Adobe PDF File), 447 KB - jmir_v23i10e31294_app1.pdf]

Multimedia Appendix 2
Distribution of answers for the question: “How motivated do you feel by the following?”.
[PDF File (Adobe PDF File), 43 KB - jmir_v23i10e31294_app2.pdf]

Multimedia Appendix 3
Distribution of answers for the question: “How concerned are you about the following?”.
[PDF File (Adobe PDF File), 43 KB - jmir_v23i10e31294_app3.pdf]

Multimedia Appendix 4
Distribution of answers for the question: “How comfortable do you feel about sharing the following data items (collected through questionnaires and surveys)?”.
[PDF File (Adobe PDF File), 53 KB - jmir_v23i10e31294_app4.pdf]

Multimedia Appendix 5
Distribution of answers for the question: “How comfortable do you feel about sharing the following data items (collected through mobile and wearable sensing)?”.
[PDF File (Adobe PDF File), 53 KB - jmir_v23i10e31294_app5.pdf]

Multimedia Appendix 6
Distribution of answers for the question: “How desirable are the following options?”.
[PDF File (Adobe PDF File), 46 KB - jmir_v23i10e31294_app6.pdf]

References

1. Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med 2015 Feb 26;372(9):793-795. [doi: 10.1056/NEJMp1500523] [Medline: 25635347]
2. Garcia-Ceja E, Riegler M, Nordgreen T, Jakobsen P, Oedegaard KJ, Tørresen J. Mental health monitoring with multimodal sensing and machine learning: A survey. Pervasive and Mobile Computing 2018 Dec;51:1-26. [doi: 10.1016/j.pmcj.2018.09.003]
3. Abdullah S, Choudhury T. Sensing Technologies for Monitoring Serious Mental Illnesses. IEEE MultiMedia 2018 Jan;25(1):61-75. [doi: 10.1109/mmul.2018.011921236]
4. Nebeker C, Lagare T, Takemoto M, Lewars B, Crist K, Bloss CS, et al. Engaging research participants to inform the ethical conduct of mobile imaging, pervasive sensing, and location tracking research. Transl Behav Med 2016 Dec;6(4):577-586. [doi: 10.1007/s13142-016-0426-4] [Medline: 27688250]
5. Rohani DA, Faurholt-Jepsen M, Kessing LV, Bardram JE. Correlations Between Objective Behavioral Features Collected From Mobile and Wearable Devices and Depressive Mood Symptoms in Patients With Affective Disorders: Systematic Review. JMIR Mhealth Uhealth 2018 Aug 13;6(8):e165 [FREE Full text] [doi: 10.2196/mhealth.9691] [Medline: 30104184]
6. Cornet VP, Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. J Biomed Inform 2018 Jan;77:120-132. [doi: 10.1016/j.jbi.2017.12.008] [Medline: 29248628]
7. Aledavood T, Triana HAM, Alakörkkö T, Kaski K, Saramäki J, Isometsä E, et al. Data Collection for Mental Health Studies Through Digital Platforms: Requirements and Design of a Prototype. JMIR Res Protoc 2017 Jun 09;6(6):e110 [FREE Full text] [doi: 10.2196/resprot.6919] [Medline: 28600276]
8. Harari GM, Lane ND, Wang R, Crosier BS, Campbell AT, Gosling SD. Using Smartphones to Collect Behavioral Data in Psychological Science: Opportunities, Practical Considerations, and Challenges. Perspect Psychol Sci 2016 Nov;11(6):838-854. [doi: 10.1177/1745691616650285] [Medline: 27899727]
9. Dogan E, Sander C, Wagner X, Hegerl U, Kohls E. Smartphone-Based Monitoring of Objective and Subjective Data in Affective Disorders: Where Are We and Where Are We Going? Systematic Review. J Med Internet Res 2017 Jul 24;19(7):e262 [FREE Full text] [doi: 10.2196/jmir.7006] [Medline: 28739561]
10. Broes S, Lacombe D, Verlinden M, Huys I. Sharing human samples and patient data: Opening Pandora’s box. Journal of Cancer Policy 2017 Sep;13:65-69. [doi: 10.1016/j.jcpo.2017.07.006]
11. Committee on Strategies for Responsible Sharing of Clinical Trial Data, Board on Health Sciences Policy, Institute of Medicine. Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk. Washington, DC: National Academies Press; 2015.

12. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2:3 [FREE Full text] [doi: 10.1186/2047-2501-2-3] [Medline: 25825667]
13. Bernardos A, Pires M, Ollé D, Casar J. Digital Phenotyping as a Tool for Personalized Mental Healthcare Internet. 2019 May Presented at: PervasiveHealth'19: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 20-23, 2019; Trento, Italy p. 403-408. [doi: 10.1145/3329189.3329240]
14. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015 Mar;12(3):e1001779 [FREE Full text] [doi: 10.1371/journal.pmed.1001779] [Medline: 25826379]
15. Hossain MA, Dwivedi YK, Rana NP. State-of-the-art in open data research: Insights from existing literature and a research agenda. Journal of Organizational Computing and Electronic Commerce 2015 Dec 02;26(1-2):14-40. [doi: 10.1080/10919392.2015.1124007]
16. Ferri M, Abdallah K. Preparing for responsible sharing of clinical trial data. N Engl J Med 2014 Jan 30;370(5):484-485. [doi: 10.1056/NEJMc1314515] [Medline: 24476448]
17. Broes S, Lacombe D, Verlinden M, Huys I. Toward a Tiered Model to Share Clinical Trial Data and Samples in Precision Oncology. Front Med (Lausanne) 2018 Jan 29;5:6 [FREE Full text] [doi: 10.3389/fmed.2018.00006] [Medline: 29435448]
18. All of Us Research Program Investigators T, Denny JC, Rutter JL, Goldstein DB, Philippakis A, Smoller JW, et al. The "All of Us" Research Program. N Engl J Med 2019 Aug 15;381(7):668-676 [FREE Full text] [doi: 10.1056/NEJMsr1809937] [Medline: 31412182]
19. Deutsch E, Duftschmid G, Dorda W. Critical areas of national electronic health record programs-is our focus correct? Int J Med Inform 2010 Mar;79(3):211-222. [doi: 10.1016/j.ijmedinf.2009.12.002] [Medline: 20079685]
20. Pencina MJ, Louzao DM, McCourt BJ, Adams MR, Tayyabkhan RH, Ronco P, et al. Supporting open access to clinical trial data for researchers: The Duke Clinical Research Institute-Bristol-Myers Squibb Supporting Open Access to Researchers Initiative. Am Heart J 2016 Feb;172:64-69. [doi: 10.1016/j.ahj.2015.11.002] [Medline: 26856217]
21. Bardram J. The CARP Mobile Sensing Framework -- A Cross-platform, Reactive, Programming Framework and Runtime Environment for Digital Phenotyping Internet. arXiv. Preprint posted online June 21, 2020 [FREE Full text]
22. Lunn MR, Lubensky M, Hunt C, Flentje A, Capriotti MR, Sooksaman C, et al. A digital health research platform for community engagement, recruitment, and retention of sexual and gender minority adults in a national longitudinal cohort study--The PRIDE Study. J Am Med Inform Assoc 2019 Aug 01;26(8-9):737-748 [FREE Full text] [doi: 10.1093/jamia/ocz082] [Medline: 31162545]
23. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]
24. Greshake Tzovaras B, Angrist M, Arvai K, Dulaney M, Estrada-Galiñanes V, Gunderson B, et al. Open Humans: A platform for participant-centered research and personal data exploration. Gigascience 2019 Jun 01;8(6):giz076 [FREE Full text] [doi: 10.1093/gigascience/giz076] [Medline: 31241153]
25. van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, et al. A systematic review of barriers to data sharing in public health. BMC Public Health 2014;14:1144 [FREE Full text] [doi: 10.1186/1471-2458-14-1144] [Medline: 25377061]
26. Stone MA, Redsell SA, Ling JT, Hay AD. Sharing patient data: competing demands of privacy, trust and research in primary care. Br J Gen Pract 2005 Oct;55(519):783-789 [FREE Full text] [Medline: 16212854]
27. Sankar PL, Parker LS. The Precision Medicine Initiative's All of Us Research Program: an agenda for research on its ethical, legal, and social issues. Genet Med 2017 Jul 8;19(7):743-750. [doi: 10.1038/gim.2016.183] [Medline: 27929525]
28. Kostkova P, Brewer H, de Lusignan S, Fottrell E, Goldacre B, Hart G, et al. Who Owns the Data? Open Data for Healthcare. Front Public Health 2016;4:7 [FREE Full text] [doi: 10.3389/fpubh.2016.00007] [Medline: 26925395]
29. Klingler C, Silva DS, Schuermann C, Reis AA, Saxena A, Strech D. Ethical issues in public health surveillance: a systematic qualitative review. BMC Public Health 2017 Apr 04;17(1):295 [FREE Full text] [doi: 10.1186/s12889-017-4200-4] [Medline: 28376752]
30. Clark K, Duckham M, Guillemin M, Hunter A, McVernon J, O'Keefe C, et al. Advancing the ethical use of digital data in human research: challenges and strategies to promote ethical practice. Ethics Inf Technol 2018 Nov 21;21(1):59-73. [doi: 10.1007/s10676-018-9490-4]
31. Vilaza G. What is the future of data sharing for research? Interactions 2020 Jul 09;27(4):54-56. [doi: 10.1145/3404211]
32. Vilaza G, Maharjan R, Coyle D, Bardram J. Futures for Health Research Data Platforms From the Participants' Perspectives. 2020 Oct Presented at: Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society; October 25-29, 2020; Tallinn, Estonia. [doi: 10.1145/3419249.3420110]
33. Merdad L, Aldakhil L, Gadi R, Assidi M, Saddick SY, Abuzenadah A, et al. Assessment of knowledge about biobanking among healthcare students and their willingness to donate biospecimens. BMC Med Ethics 2017 May 02;18(1):32 [FREE Full text] [doi: 10.1186/s12910-017-0195-8] [Medline: 28464877]
34. Tozzo P, Fassina A, Caenazzo L. Young people's awareness on biobanking and DNA profiling: results of a questionnaire administered to Italian university students. Life Sci Soc Policy 2017 Dec 10;13(1):9 [FREE Full text] [doi: 10.1186/s40504-017-0055-9] [Medline: 28601979]

35. Hong SJ, Drake B, Goodman M, Kaphingst KA. Race, Trust in Doctors, Privacy Concerns, and Consent Preferences for Biobanks. Health Commun 2020 Sep;35(10):1219-1228 [FREE Full text] [doi: 10.1080/10410236.2019.1623644] [Medline: 31167570]
36. Middleton A, Milne R, Almarri MA, Anwer S, Atutornu J, Baranova EE, Mendes, et al. Global Public Perceptions of Genomic Data Sharing: What Shapes the Willingness to Donate DNA and Health Data? Am J Hum Genet 2020 Oct 01;107(4):743-752 [FREE Full text] [doi: 10.1016/j.ajhg.2020.08.023] [Medline: 32946764]
37. Rieger A, Gaines A, Barnett I, Baldassano CF, Connolly Gibbons MB, Crits-Christoph P. Psychiatry Outpatients' Willingness to Share Social Media Posts and Smartphone Data for Research and Clinical Purposes: Survey Study. JMIR Form Res 2019 Aug 29;3(3):e14329 [FREE Full text] [doi: 10.2196/14329] [Medline: 31493326]
38. Rendina H, Mustanski B. Privacy, Trust, and Data Sharing in Web-Based and Mobile Research: Participant Perspectives in a Large Nationwide Sample of Men Who Have Sex With Men in the United States. J Med Internet Res 2018 Jul 04;20(7):e233 [FREE Full text] [doi: 10.2196/jmir.9019] [Medline: 29973332]
39. Karampela M, Ouhbi S, Isomursu M. Connected Health User Willingness to Share Personal Health Data: Questionnaire Study. J Med Internet Res 2019 Nov 27;21(11):e14537 [FREE Full text] [doi: 10.2196/14537] [Medline: 31774410]
40. Abdelhamid M, Gaia J, Sanders GL. Putting the Focus Back on the Patient: How Privacy Concerns Affect Personal Health Information Sharing Intentions. J Med Internet Res 2017 Sep 13;19(9):e169 [FREE Full text] [doi: 10.2196/jmir.6877] [Medline: 28903895]
41. Weitzman ER, Kaci L, Mandl KD. Sharing medical data for health research: the early personal health record experience. J Med Internet Res 2010 May 25;12(2):e14 [FREE Full text] [doi: 10.2196/jmir.1356] [Medline: 20501431]
42. Bouras A, Simoes EJ, Boren S, Hicks L, Zachary I, Buck C, et al. Non-Hispanic White Mothers' Willingness to Share Personal Health Data With Researchers: Survey Results From an Opt-in Panel. J Particip Med 2020 May 15;12(2):e14062 [FREE Full text] [doi: 10.2196/14062] [Medline: 33064096]
43. Middleton A, Milne R, Thorogood A, Kleiderman E, Niemiec E, Prainsack B, et al. Attitudes of publics who are unwilling to donate DNA data for research. Eur J Med Genet 2019 May;62(5):316-323 [FREE Full text] [doi: 10.1016/j.ejmg.2018.11.014] [Medline: 30476628]
44. Patil S, Lu H, Saunders CL, Potoglou D, Robinson N. Public preferences for electronic health data storage, access, and sharing - evidence from a pan-European survey. J Am Med Inform Assoc 2016 Nov 23;23(6):1096-1106 [FREE Full text] [doi: 10.1093/jamia/ocw012] [Medline: 27107445]
45. Sanderson SC, Brothers KB, Mercaldo ND, Clayton EW, Antommaria AHM, Aufox SA, et al. Public Attitudes toward Consent and Data Sharing in Biobank Research: A Large Multi-site Experimental Survey in the US. Am J Hum Genet 2017 Mar 02;100(3):414-427 [FREE Full text] [doi: 10.1016/j.ajhg.2017.01.021] [Medline: 28190457]
46. Lemke AA, Wolf WA, Hebert-Beirne J, Smith ME. Public and biobank participant attitudes toward genetic research participation and data sharing. Public Health Genomics 2010;13(6):368-377 [FREE Full text] [doi: 10.1159/000276767] [Medline: 20805700]
47. Ahram M, Othman A, Shahrouri M, Mustafa E. Factors influencing public participation in biobanking. Eur J Hum Genet 2014 Apr;22(4):445-451 [FREE Full text] [doi: 10.1038/ejhg.2013.174] [Medline: 23921537]
48. Trinidad SB, Fullerton SM, Bares JM, Jarvik GP, Larson EB, Burke W. Informed Consent in Genome-Scale Research: What Do Prospective Participants Think? AJOB Prim Res 2012 Jul 01;3(3):3-11 [FREE Full text] [doi: 10.1080/21507716.2012.662575] [Medline: 23493836]
49. Rathi V, Dzara K, Gross CP, Hrynaszkiewicz I, Joffe S, Krumholz HM, et al. Sharing of clinical trial data among trialists: a cross sectional survey. BMJ 2012 Nov 20;345:e7570 [FREE Full text] [doi: 10.1136/bmj.e7570] [Medline: 23169870]
50. Papoutsi C, Reed JE, Marston C, Lewis R, Majeed A, Bell D. Patient and public views about the security and privacy of Electronic Health Records (EHRs) in the UK: results from a mixed methods study. BMC Med Inform Decis Mak 2015 Oct 14;15:86 [FREE Full text] [doi: 10.1186/s12911-015-0202-2] [Medline: 26466787]
51. Luchenski S, Balasanthiran A, Marston C, Sasaki K, Majeed A, Bell D, et al. Survey of patient and public perceptions of electronic health records for healthcare, policy and research: study protocol. BMC Med Inform Decis Mak 2012 May 23;12:40 [FREE Full text] [doi: 10.1186/1472-6947-12-40] [Medline: 22621621]
52. Zanaboni P, Kummervold PE, Sørensen T, Johansen MA. Patient Use and Experience With Online Access to Electronic Health Records in Norway: Results From an Online Survey. J Med Internet Res 2020 Feb 07;22(2):e16144 [FREE Full text] [doi: 10.2196/16144] [Medline: 32031538]
53. Porteri C, Pasqualetti P, Togni E, Parker M. Public's attitudes on participation in a biobank for research: an Italian survey. BMC Med Ethics 2014 Nov 26;15(1):81 [FREE Full text] [doi: 10.1186/1472-6939-15-81] [Medline: 25425352]
54. Ivanov A, Sharman R, Rao HR. Exploring factors impacting sharing health-tracking records. Health Policy and Technology 2015 Sep;4(3):263-276. [doi: 10.1016/j.hlpt.2015.04.008]
55. Riordan F, Papoutsi C, Reed JE, Marston C, Bell D, Majeed A. Patient and public attitudes towards informed consent models and levels of awareness of Electronic Health Records in the UK. Int J Med Inform 2015 Apr;84(4):237-247 [FREE Full text] [doi: 10.1016/j.ijmedinf.2015.01.008] [Medline: 25649841]

56. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
57. Shah N, Coathup V, Teare H, Forgie I, Giordano GN, Hansen TH, et al. Motivations for data sharing-views of research participants from four European countries: A DIRECT study. Eur J Hum Genet 2019 May 30;27(5):721-729 [FREE Full text] [doi: 10.1038/s41431-019-0344-2] [Medline: 30700834]
58. Christin D, Reinhardt A, Kanhere SS, Hollick M. A survey on privacy in mobile participatory sensing applications. Journal of Systems and Software 2011 Nov;84(11):1928-1946. [doi: 10.1016/j.jss.2011.06.073]
59. Cilliers L. Wearable devices in healthcare: Privacy and information security issues. Health Inf Manag 2020;49(2-3):150-156. [doi: 10.1177/1833358319851684] [Medline: 31146589]
60. Nurgalieva L, Frik A, Ceschel F, Egelman S, Marchese M. Information Design in An Aged Care Context: Views of Older Adults on Information Sharing in a Care Triad. In: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2019 May Presented at: PervasiveHealth'19: The 13th International Conference on Pervasive Computing Technologies for Healthcare; May 20-23, 2019; Trento, Italy. [doi: 10.1145/3329189.3329211]
61. Rudnicka A, Cox A, Gould S. Why Do You Need This? In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI ?19. 2019 May Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland, UK. [doi: 10.1145/3290605.3300622]
62. Di Matteo D, Fine A, Fotinos K, Rose J, Katzman M. Patient Willingness to Consent to Mobile Phone Data Collection for Mental Health Apps: Structured Questionnaire. JMIR Ment Health 2018 Aug 29;5(3):e56 [FREE Full text] [doi: 10.2196/mental.9539] [Medline: 30158102]
63. Rooksby J, Morrison A, Murray-Rust D. Student Perspectives on Digital Phenotyping Internet. In: CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 May Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland, UK. [doi: 10.1145/3290605.3300655]
64. Weitzman ER, Kelemen S, Kaci L, Mandl KD. Willingness to share personal health record data for care improvement and public health: a survey of experienced personal health record users. BMC Med Inform Decis Mak 2012 May 22;12:39 [FREE Full text] [doi: 10.1186/1472-6947-12-39] [Medline: 22616619]
65. Schneegass S, Poguntke R, Machulla T. Understanding the Impact of Information Representation on Willingness to Share Information. In: CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 May Presented at: CHI '19: 2019 CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland, UK p. 1-6. [doi: 10.1145/3290605.3300753]
66. Chen H, Cohen P, Chen S. How Big is a Big Odds Ratio? Interpreting the Magnitudes of Odds Ratios in Epidemiological Studies. Communications in Statistics - Simulation and Computation 2010 Mar 31;39(4):860-864. [doi: 10.1080/03610911003650383]
67. Zammar G, Meister H, Shah J, Phadtare A, Cofiel L, Pietrobon R. So different, yet so similar: meta-analysis and policy modeling of willingness to participate in clinical trials among Brazilians and Indians. PLoS One 2010 Dec 16;5(12):e14368 [FREE Full text] [doi: 10.1371/journal.pone.0014368] [Medline: 21179556]
68. Nobile H, Vermeulen E, Thys K, Bergmann MM, Borry P. Why do participants enroll in population biobank studies? A systematic literature review. Expert Rev Mol Diagn 2013 Jan;13(1):35-47. [doi: 10.1586/erm.12.116] [Medline: 23256702]
69. Seltzer E, Goldshear J, Guntuku SC, Grande D, Asch DA, Klinger EV, et al. Patients' willingness to share digital health and non-health data for research: a cross-sectional study. BMC Med Inform Decis Mak 2019 Aug 08;19(1):157 [FREE Full text] [doi: 10.1186/s12911-019-0886-9] [Medline: 31395102]
70. Pratap A, Allred R, Duffy J, Rivera D, Lee HS, Renn BN, et al. Contemporary Views of Research Participant Willingness to Participate and Share Digital Data in Biomedical Research. JAMA Netw Open 2019 Nov 01;2(11):e1915717 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.15717] [Medline: 31747031]
71. Overby C, Maloney K, Alestock T, Chavez J, Berman D, Sharaf R, et al. Prioritizing Approaches to Engage Community Members and Build Trust in Biobanks: A Survey of Attitudes and Opinions of Adults within Outpatient Practices at the University of Maryland. J Pers Med 2015 Jul 28;5(3):264-279 [FREE Full text] [doi: 10.3390/jpm5030264] [Medline: 26226006]
72. Ma Y, Dai H, Wang L, Zhu L, Zou H, Kong X. Consent for use of clinical leftover biosample: a survey among Chinese patients and the general public. PLoS One 2012;7(4):e36050 [FREE Full text] [doi: 10.1371/journal.pone.0036050] [Medline: 22558323]
73. Kim KK, Joseph JG, Ohno-Machado L. Comparison of consumers' views on electronic data sharing for healthcare and research. J Am Med Inform Assoc 2015 Jul;22(4):821-830 [FREE Full text] [doi: 10.1093/jamia/ocv014] [Medline: 25829461]
74. Haeusermann T, Fadda M, Blasimme A, Tzovaras BG, Vayena E. Genes wide open: Data sharing and the social gradient of genomic privacy. AJOB Empir Bioeth 2018;9(4):207-221. [doi: 10.1080/23294515.2018.1550123] [Medline: 30596357]
75. Ostherr K, Borodina S, Bracken RC, Lotterman C, Storer E, Williams B. Trust and privacy in the context of user-generated health data. Big Data & Society 2017 Apr 17;4(1):205395171770467. [doi: 10.1177/2053951717704673]

76. Ludman EJ, Fullerton SM, Spangler L, Trinidad SB, Fujii MM, Jarvik GP, et al. Glad you asked: participants' opinions of re-consent for dbGap data submission. J Empir Res Hum Res Ethics 2010 Sep;5(3):9-16 [FREE Full text] [doi: 10.1525/jer.2010.5.3.9] [Medline: 20831417]
77. van Berkel N, Ferreira D, Kostakos V. The Experience Sampling Method on Mobile Devices. ACM Comput. Surv 2018 Jan 12;50(6):1-40. [doi: 10.1145/3123988]
78. Epstein DA, Caraway M, Johnston C, Ping A, Fogarty J, Munson SA. Beyond Abandonment to Next Steps: Understanding and Designing for Life after Personal Informatics Tool Use. Proc SIGCHI Conf Hum Factor Comput Syst 2016 May;2016:1109-1113 [FREE Full text] [doi: 10.1145/2858036.2858045] [Medline: 28503678]
79. Bhat KS, Kumar N. Sociocultural Dimensions of Tracking Health and Taking Care. Proc. ACM Hum.-Comput. Interact 2020 Oct 14;4(CSCW2):1-24. [doi: 10.1145/3415200]
80. Chandwani R, Kulkarni V. Who's the doctor?: Physicians' perception of internet informed patients in India. In: CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 May Presented at: CHI '16: The 2016 CHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose, CA p. 3091-3102. [doi: 10.1145/2858036.2858500]
81. Luo Y, Oh CY, Jean BS, Choe EK. Interrelationships Between Patients' Data Tracking Practices, Data Sharing Practices, and Health Literacy: Onsite Survey Study. J Med Internet Res 2020 Dec 22;22(12):e18937 [FREE Full text] [doi: 10.2196/18937] [Medline: 33350960]
82. Warner M, Gutmann A, Sasse MA, Blandford A. Privacy Unraveling Around Explicit HIV Status Disclosure Fields in the Online Geosocial Hookup App Grindr. Proc. ACM Hum.-Comput. Interact 2018 Nov 01;2(CSCW):1-22. [doi: 10.1145/3274450]
83. Kuntsman A, Miyake E, Martin S. Re-thinking Digital Health: Data, Appisation and the (im)possibility of 'Opting out'. Digit Health 2019 Oct 09;5:2055207619880671 [FREE Full text] [doi: 10.1177/2055207619880671] [Medline: 31636917]
84. Abdelhamid M. Greater patient health information control to improve the sustainability of health information exchanges. J Biomed Inform 2018 Jul;83:150-158 [FREE Full text] [doi: 10.1016/j.jbi.2018.06.002] [Medline: 29894811]
85. Ibrahim H, Liu X, Zariffa N, Morris AD, Denniston AK. Health data poverty: an assailable barrier to equitable digital health care. Lancet Digit Health 2021 Apr;3(4):e260-e265 [FREE Full text] [doi: 10.1016/S2589-7500(20)30317-4] [Medline: 33678589]

Abbreviations

DTU: Technical University of Denmark
GDPR: General Data Protection Regulation

Edited by G Eysenbach; submitted 17.06.21; peer-reviewed by H Pratomo; comments to author 23.07.21; revised version received 16.09.21; accepted 20.09.21; published 29.10.21.

Please cite as:
Nunes Vilaza G, Coyle D, Bardram JE
Public Attitudes to Digital Health Research Repositories: Cross-sectional International Survey
J Med Internet Res 2021;23(10):e31294
URL: https://www.jmir.org/2021/10/e31294
doi:10.2196/31294
PMID:

Original Paper

Questioning the Yelp Effect: Mixed Methods Analysis of Web-Based Reviews of Urgent Cares

Dian Hu[1], BSc; Cindy Meng-Hsin Liu[2], BA, MPH, MD, PhD; Rana Hamdy[3,4], BA, MPH, MSc, MD; Michael Cziner[2], BA, MPH; Melody Fung[2], BSc, MPH; Samuel Dobbs[5], BSc, MPH; Laura Rogers[2], BSc; Monique Mitchell Turner[6], BA, MA, PhD; David André Broniatowski[1], BSc, MSc, PhD

[1]Department of Engineering Management and Systems Engineering, School of Engineering and Applied Science, George Washington University, Washington, DC, United States

[2]Department of Environmental and Occupational Health, Milken Institute School of Public Health, George Washington University School of Medicine and Health Sciences, Washington, DC, United States

[3]Division of Infectious Diseases, Children's National Hospital, Washington, DC, United States

[4]Department of Pediatrics, George Washington University, Washington, DC, United States

[5]Department of Health Policy and Management, Milken Institute School of Public Health, George Washington University School of Medicine and Health Sciences, Washington, DC, United States

[6]Department of Communication, College of Communication, Arts, and Sciences, Michigan State University, East Lansing, MI, United States

Corresponding Author:
Dian Hu, BSc
Department of Engineering Management and Systems Engineering
School of Engineering and Applied Science
George Washington University
B1800, Science and Engineering Hall 2700
800 22nd St NW
Washington, DC, 20052
United States
Phone: 1 2027251564
Email: hudian@gwmail.gwu.edu

Abstract

Background: Providers of on-demand care, such as those in urgent care centers, may prescribe antibiotics unnecessarily because they fear receiving negative reviews on web-based platforms from unsatisfied patients—the so-called *Yelp effect*. This effect is hypothesized to be a significant driver of inappropriate antibiotic prescribing, which exacerbates antibiotic resistance.

Objective: In this study, we aimed to determine the frequency with which patients left negative reviews on web-based platforms after they expected to receive antibiotics in an urgent care setting but did not.

Methods: We obtained a list of 8662 urgent care facilities from the Yelp application programming interface. By using this list, we automatically collected 481,825 web-based reviews from Google Maps between January 21 and February 10, 2019. We used machine learning algorithms to summarize the contents of these reviews. Additionally, 200 randomly sampled reviews were analyzed by 4 annotators to verify the types of messages present and whether they were consistent with the Yelp effect.

Results: We collected 481,825 reviews, of which 1696 (95% CI 1240-2152) exhibited the Yelp effect. Negative reviews primarily identified operations issues regarding wait times, rude staff, billing, and communication.

Conclusions: Urgent care patients rarely express expectations for antibiotics in negative web-based reviews. Thus, our findings do not support an association between a lack of antibiotic prescriptions and negative web-based reviews. Rather, patients' dissatisfaction with urgent care was most strongly linked to operations issues that were not related to the clinical management plan.

(J Med Internet Res 2021;23(10):e29406) doi:10.2196/29406

KEYWORDS

urgent care; doctor-patient communication; doctor web-based review; review websites

Introduction

The World Health Organization has deemed antibiotic resistance, which is primarily caused by antibiotic overuse, to be one of the world's most pressing health problems [1]. Antibiotic overuse is widespread; approximately one-third of all antibiotics prescribed in US outpatient settings are unnecessary [2]. Care providers have admitted to prescribing antibiotics—even when antibiotics are unnecessary—when they assume that patients will be unsatisfied without an antibiotic prescription [3,4]. However, care providers' assumptions about a patient's expectations frequently do not match patients' actual expectations for an antibiotic [5]. Furthermore, prior literature on patient satisfaction and provider-patient communication have suggested that there are other factors that drive patient satisfaction [6]. For example, Welschen et al [7] found that receiving information or reassurance was more strongly associated with satisfaction than receiving an antibiotic prescription in primary care. Ong et al [8] found that patient satisfaction was not related to the receipt of antibiotics but was related to the belief that patients had a better understanding of their illness. Stearns et al [9] found that patients generally had equal levels of visit satisfaction regardless of their antibiotic treatment status.

Despite findings that do not support a link between patient satisfaction and antibiotics, care providers still report prescribing antibiotics unnecessarily because they fear that dissatisfied patients will leave negative reviews [10]. Authors in several media outlets [11,12] have coined the term *Yelp effect*; they propose that providers of on-demand care, such as those in urgent care centers, may prescribe antibiotics unnecessarily to prevent patients from leaving negative web-based reviews. There is a perception among urgent care providers that many of their patients expect to receive antibiotics, even when they are clinically unnecessary, and that patients will leave negative reviews on web-based platforms if these expectations are not met. To avoid negative web-based reviews, which could impact care providers' pay and the performance of urgent care centers, care providers are driven to prescribe antibiotics, even when they are clinically unnecessary.

Concerns about the Yelp effect are magnified by the surging popularity of review websites, such as reviews on Google Pages, Yelp, and Healthgrades [11]. Even though most physician ratings on web-based platforms are positive [13,14], at least 1 study has demonstrated that in web-based physician reviews, the words *medication* and *prescription* are mentioned in more negative contexts [15]. This perceived connection between negative web-based reviews and antibiotics is hypothesized to be a significant driver of inappropriate antibiotic prescribing and is evident in at least 1 petition that has received 40,000 signatures supporting the removal of web-based doctor reviews [16].

Concerns about the impact of negative reviews appear to be valid; patients often use web-based reviews when deciding whether and where to seek treatment [17]. Furthermore, these reviews may not accurately reflect the quality of medical care. For example, Daskivich et al [18] conducted an analysis of 5 popular web-based platforms and showed that there was no significant association between web-based consumer review scores and standard quality guidelines. Likewise, Yaraghi et al [19] found that consumers tend to perceive care provider ratings from nonclinical websites to be as important as the ratings from government websites. Thus, web-based reviews can drive care providers' behaviors in ways that negatively impact public health.

At present, the extent to which patients' expectations for antibiotic treatment drive negative web-based reviews is unknown. Even though web-based reviews have been mined for other health topics [20,21], to our knowledge, ours is the first study to evaluate web-based reviews and antibiotic prescribing. Thus, we sought to determine whether there truly is a Yelp effect by evaluating how frequently patients leave negative web-based reviews regarding a lack of antibiotic prescriptions.

To determine the prevalence of the Yelp effect, we analyzed a large sample of web-based reviews of urgent care centers in the US by calculating the proportion of negative reviews that exhibit a message regarding a lack of antibiotic prescriptions. Specifically, we sought to (1) quantify the proportion of negative reviews that were posted due to patients (reviewers) not receiving an antibiotic and (2) evaluate the content of negative reviews of urgent care centers.

Methods

Data Collection

We used the Yelp application programming interface (API) between October 1 and October 20, 2018, to obtain a list of facilities that were tagged as "urgent care" in the United States. We also retrieved the star ratings for all reviews that were associated with these facilities. We removed all facilities that did not have a US zip code. We were unable to obtain the full texts of these reviews—Yelp's terms of service prohibit web scraping, and Yelp declined to provide us with permission to use the text. Thus, by using the list of facilities that was obtained from the Yelp API, we ran the data collection algorithm from January 21 and February 10, 2019, and obtained all of the Google Maps reviews of each urgent care facility that were posted before January 21, 2019. Unlike Yelp, Google's terms of service permit the collection of their reviews as long as it does not put undue burdens on Google's servers. To collect data from Google Maps reviews, we designed a computer program (Multimedia Appendix 1) for automatically collecting these reviews based on their Google Maps URLs. The Google Maps URLs of the urgent care facilities were collected after we presented their titles and addresses (obtained from the Yelp API) to 20 workers on Amazon Mechanical Turk, a crowdsourcing marketplace. These 20 workers were selected based on their prior demonstrated ability to successfully collect a set of 100 known URLs. Workers received US $0.09 for each URL collected. To summarize the contents of web-based reviews, we used a machine learning algorithm that was designed to summarize text—the latent Dirichlet allocation (LDA) topic model [22]. This algorithm was implemented in the LDA python package [23] (default settings were used) to

the fit the model to the Google Maps data set. Topic models identify review topics automatically without human intervention by examining the word co-occurrence statistics within each review [24]. For each topic, we calculated the total number of word tokens that were found in positive reviews (4 or 5 stars) and negative reviews (1 or 2 stars). We fitted 3 topic models (one model with 10 topics, another with 20 topics, and another with 50 topics) to the data. We selected the model that generated the most coherent topics without a large increase in perplexity (a measure of model goodness of fit that is commonly used in natural language processing; Multimedia Appendix 2). Afterward, we extracted antibiotic-related reviews by using a list of keywords that was generated by one of the authors (RH)—a pediatrician who specializes in antibiotic stewardship (see Multimedia Appendix 3 for a keyword list). For each review, we examined the proportion of words in each topic. We applied the same procedure to the subset of reviews containing antibiotic-related keywords. We then developed a qualitative codebook to determine the content and sentiment of 200 reviews that were sampled at random from all reviews containing these keywords. This study was approved by The George Washington University Committee on Human Research Institutional Review Board (Federal Wide Assurance number: FWA00005945; institutional review board registration number: 180804).

Data Annotation

Four authors (LR, MF, MC, and SD) affiliated with The George Washington University Antibiotic Resistance Action Center collectively reviewed a subset of the 200 randomly sampled reviews to determine the types of messages that were present in the reviews [25]. After this initial review and the development of inductive codes, the reviews were categorized into one of the following categories in the codebook: (1) the *Yelp effect* category (the patient wanted antibiotics but did not receive them); (2) the *opposite of the Yelp effect* category (the patient received antibiotics but did not want them); (3) the *convenience, inconvenience, and wait times* category; (4) the *staff competence or incompetence, courtesy and attitude, and satisfaction of care* category; (5) the *cost and price of drugs per visit (including sticker shock)* category; (6) the *other prescription-related complaints* category; and (7) the *other or none of the above* category. Additionally, all reviews were annotated as "positive" (eg, the patient was satisfied with their care, and the review had 4 or 5 stars) or "negative" (eg, the patient was dissatisfied with care, and the review had fewer than 4 stars).

The four annotators then independently reviewed the same 200 randomly sampled reviews to assign them to 1 of the 7 categories. The final categories for the reviews were assigned based on the majority category among annotators. If there was no majority category, disagreements were resolved discursively until a consensus category was agreed upon.

Some reviews mentioned that the reviewer did not receive an antibiotic when it was expected, even if that was not the main message. Thus, after assigning each message a primary category, the same four annotators revisited all 200 reviews to determine if they mentioned the Yelp effect in passing (ie, whether the review mentioned an unfulfilled expectation for antibiotics). A code for mentioning the presence or absence of the Yelp effect—even if it was mentioned in passing—was then assigned to each review as a secondary code. By using these 200 annotated samples, we inferred population proportions for each category in our codebook and calculated 95% CIs.

Results

Distribution of Google and Yelp Data

By using the Yelp API, we identified 8662 unique urgent care facilities that had 84,127 unique reviews. We collected 481,825 US-based reviews from Google Maps. Of these, 340,328 (70.63%) contained some text. The average star rating in Yelp reviews was significantly lower than the average star rating in Google Map reviews ($t_{565,950}$=82.38; P<.001).

Figures 1–4 display the distributions of the number of reviews and mean reviews stars from both Yelp and Google Maps on a 2-by-2 map. The distributions show an apparent difference between various geographical regions of the United States. The biased distribution is intriguing and can prompt many other research questions. Therefore, all state-by-state maps featuring the same information are being hosted on The George Washington University cloud, which can be made available to other researchers upon request.

Figure 1. Total number of reviews in US counties (Yelp data).

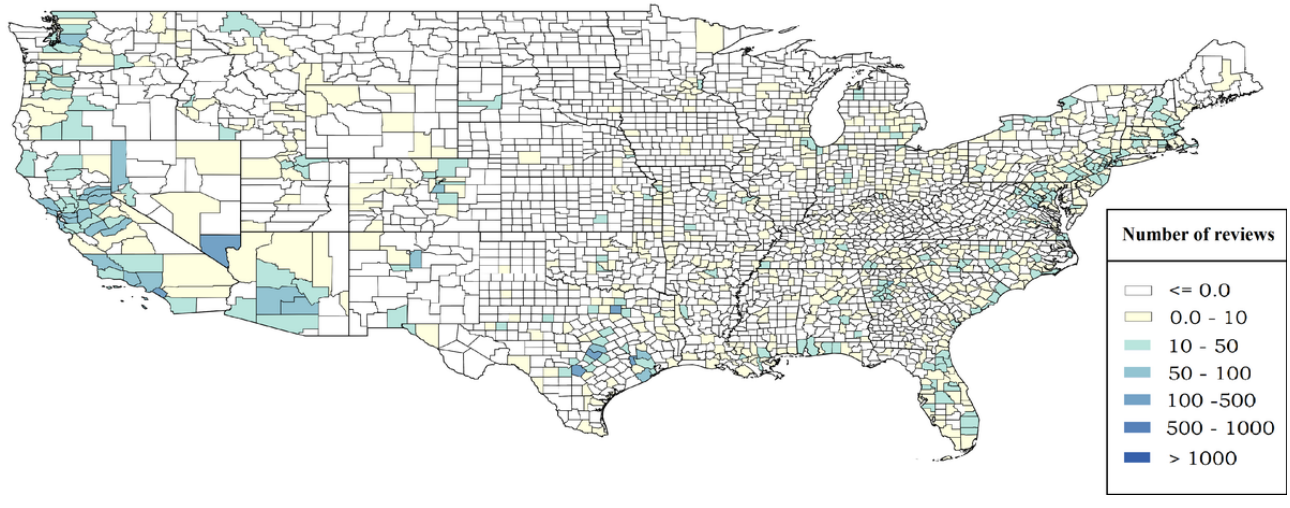


Figure 2. Mean review stars in US counties (Yelp data).

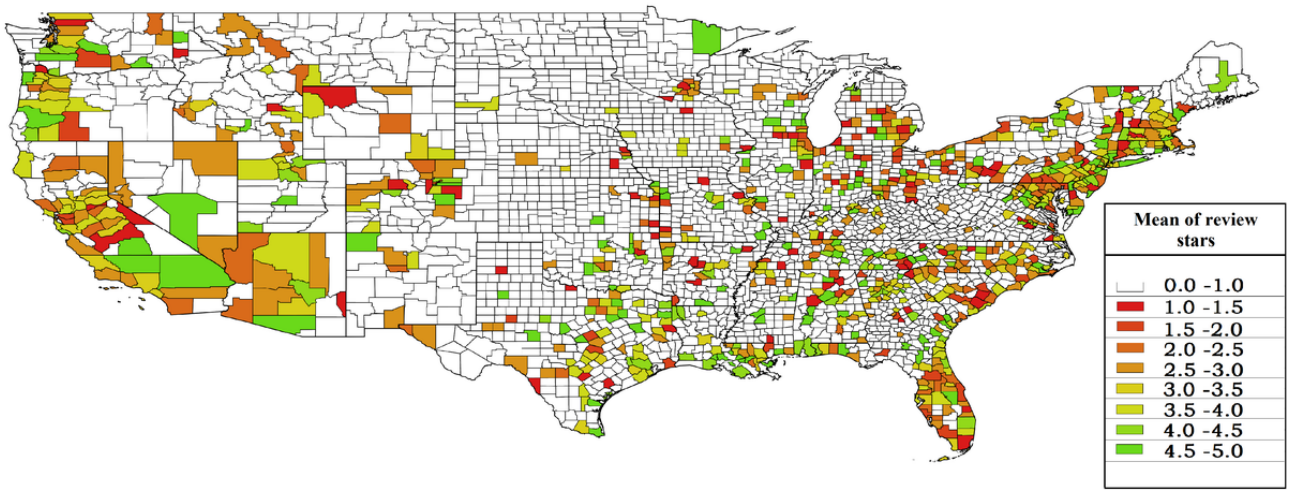


Figure 3. Total number of reviews in US counties (Google Maps data).

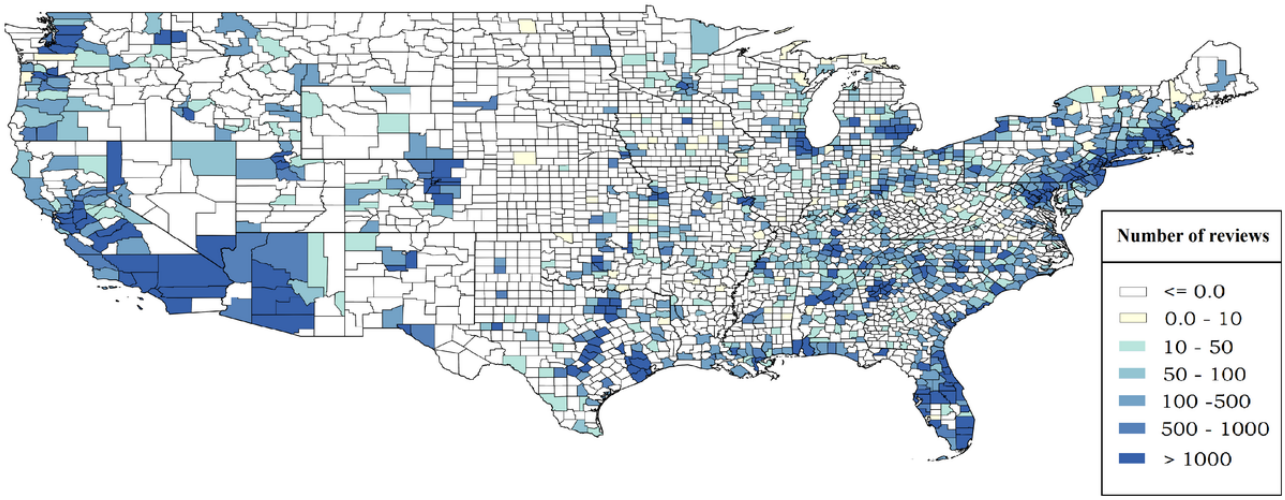
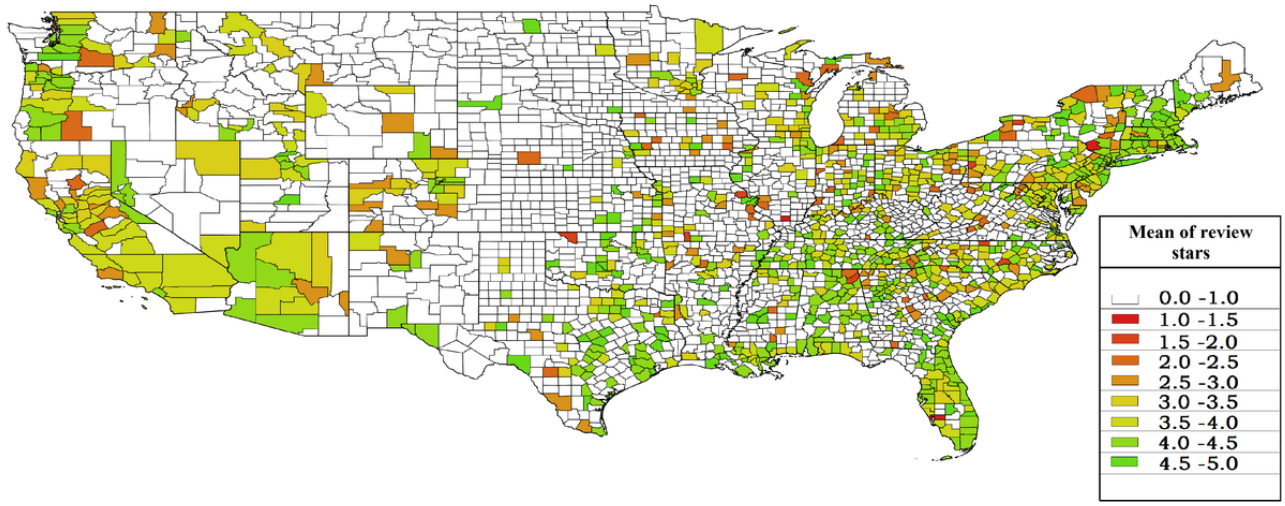


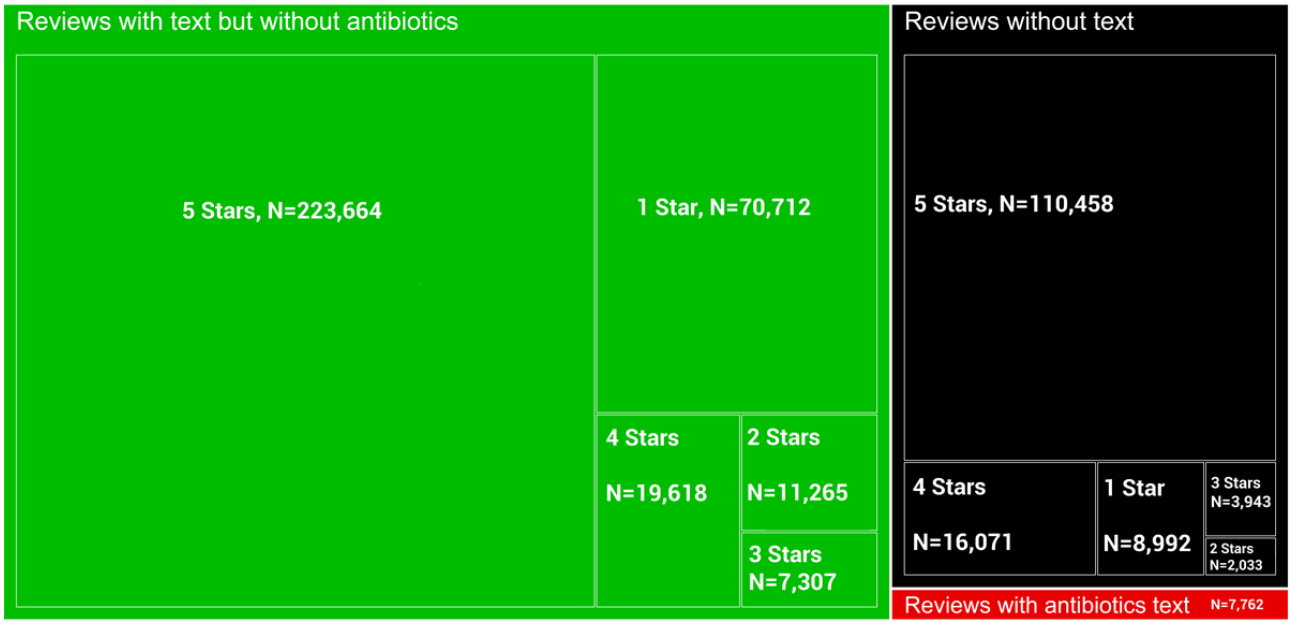
Figure 4. Mean review stars in US counties (Google Maps data).



We categorized the 481,825 reviews with text from Google Maps into the following three groups: (1) reviews without any text (n=141,497; star rating: mean 4.53); (2) reviews with text that did not mention antibiotic-related keywords (n=332,566; star rating: mean 3.94); and (3) reviews with text that mentioned antibiotic-related keywords (n=7762; star rating: mean 2.40; Figure 5). We found significant differences in the average star ratings across these three groups; a posthoc Tukey honestly significant difference test for multiple comparisons showed that reviews with antibiotic-related keywords had the lowest ratings compared to those in reviews without text (mean difference=−2.13; P<.001) and reviews with text but no

keywords (mean difference=−1.54; *P*<.001). Reviews without text were significantly more positive than reviews with text but no keywords (mean difference=0.59; *P*<.001; all descriptive statistics are in Multimedia Appendix 4).

Figure 5. Distribution of Google reviews. Reviews with text but no antibiotic-related keywords are shown in green. Reviews with antibiotic-related keywords are shown in red. Reviews without text are shown in black.



Topic Modeling Results of Google Data

The LDA topic modeling analysis (Figure 6; Multimedia Appendix 2) yielded 20 topics. The most negative topics pertained to rude staff, wait times, billing, callbacks, and other aspects of customer experience. Although the “infections and symptoms” topic was also predominantly negative, this topic did not make up a plurality of the reviews (481,760/6,795,468, 7.09%). With regard to topics with antibiotic-related keywords, Figure 7 shows that the most common topic pertained to infections and symptoms and was predominantly negative; however, several other topics pertaining to customer experience were also predominantly negative.

Figure 6. Topics generated by the latent Dirichlet allocation algorithm based on the text of all Google reviews in our data set. The size of each topic is proportional to the number of words that were assigned to each topic by the algorithm. Words are further segmented according to the sentiment of each review. Reviews with 4-5 stars are positive, reviews with 1-2 stars are negative, and reviews with 3 stars are neutral.

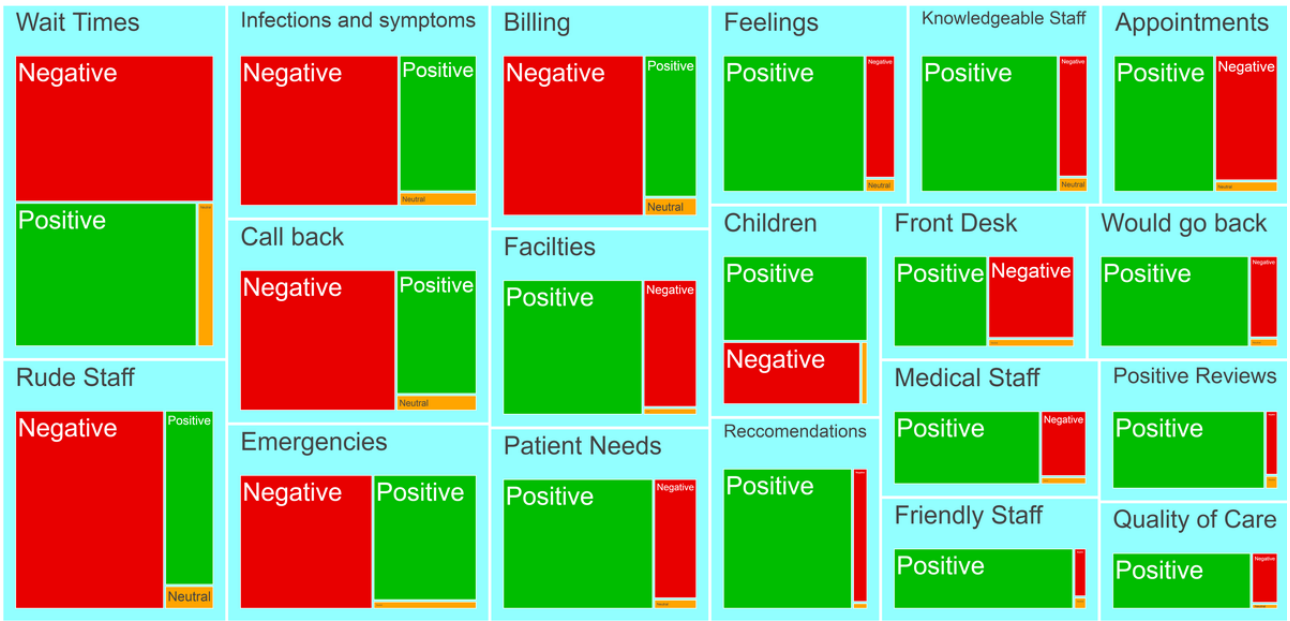
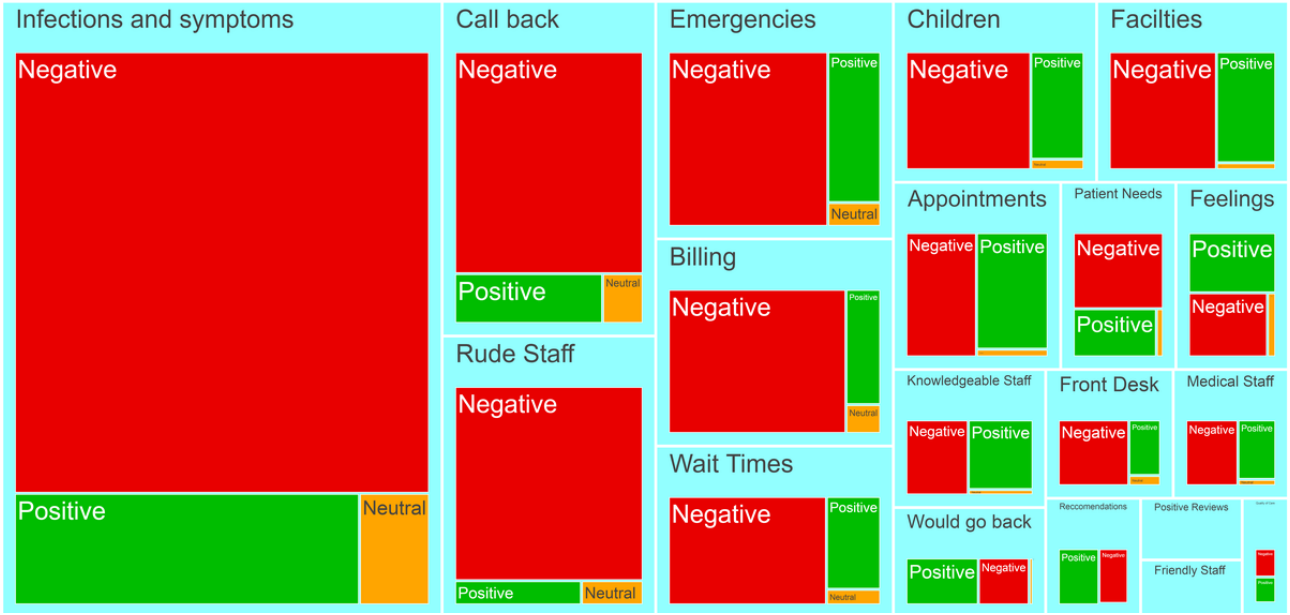


Figure 7. Subset of topic modeling results for reviews containing antibiotic-related keywords. The size of each topic is proportional to the number of words that were assigned to each topic by the algorithm. Words are further segmented according to the sentiment of each review. Reviews with 4-5 stars are positive, reviews with 1-2 stars are negative, and reviews with 3 stars are neutral.



Annotation Results

The annotators who labeled the seven inductive content categories achieved moderate reliability (Fleiss κ=0.42) with the first set of 100 reviews. All annotators agreed on sentiment. After disagreements were resolved, a second round of annotation for the next set of 100 reviews yielded substantial agreement (Fleiss κ=0.65), and disagreements were again resolved by consensus among reviewers. Table 1 summarizes the results of these annotations. Of a total of 200 reviews, we found that only 5 reviews (2.5%; 95% CI 0.3%-4.7%) exhibited the Yelp effect as the primary message category. By applying CIs to the full set of 8078 reviews containing antibiotic-related keywords, we expected that between 27 and 377 reviews would exhibit the Yelp effect as the primary message with 95% confidence. Thus, in our data set of 481,825 reviews, at most, 377 (0.08%) were expected to exhibit the Yelp effect.

Table 1. The seven primary message categories of the 200 messages.

Primary message	Reviews (N=200), n (%)
Yelp effect (the patient expected to receive antibiotics but did not receive an antibiotic)	5 (2.5)
Counter Yelp effect (the patient received antibiotics but did not want them)	1 (0.5)
Convenience, inconvenience, and wait times (positive and negative sentiment)	36 (18)
Staff competence or incompetence, courtesy and attitude, and satisfaction of care (positive and negative sentiment)	138 (69)
Cost and price of drugs per visit (including sticker shock; positive and negative sentiment)	13 (6.5)
Other complaints (positive and negative sentiment)	3 (1.5)
Other or none of the above	4 (2)

The annotators also reexamined all reviews to determine if they mentioned a Yelp effect in passing. We found that of a total of 200 reviews, 42 (21%) had some mention of reviewers not having received antibiotics when they were expected (Fleiss κ=0.75). Thus, with 95% confidence, between 1240 and 2152 of our 8078 reviews with antibiotic-related keywords exhibited the Yelp effect, even if it was only mentioned in passing. In our data set of 481,825 reviews, at most, 2152 (0.45%) were expected to exhibit the Yelp effect.

Discussion

Principal Findings

Our data suggest that the Yelp effect is quite rare. Out of a set of almost half a million reviews (N=481,824), fewer than 1 in 1250 (0.08%) seemed to contain the Yelp effect as the primary message (with 95% confidence). Furthermore, with 95% confidence, fewer than 1 in 225 (0.45%) reviews seemed to contain the Yelp effect as the primary or secondary message.

In contrast, we found that (138/200) 69% of the reviews in our annotated data set focused primarily on assessments of staff competence and the quality of personal interactions. This suggests that in terms of the extent that a Yelp effect exists,

patients express this effect by questioning the expertise or personal qualities of urgent care staff. This may put urgent care providers in a bind; although they should not prescribe antibiotics inappropriately, a failure to explain to patients why the patients' preferred treatment is ineffective may lead to reviews that are designed to undermine care providers' credibility, expertise, and personal qualities. Thus, it is of paramount importance that both care providers and urgent care staff provide high-quality care and leave patients with a meaningful understanding of why they received the treatment that they did. For example, prior work has shown that patients' expectations for antibiotics are associated with categorical gist perceptions of the risks and benefits of antibiotics [26-28] and that patients more likely to be satisfied when they understand the gist of appropriate prescribing. This promotes the need to better communicate rationales for prescribing antibiotics in a manner that enhances patients' insights into why decisions are made and, by extension, their assessments of care providers' competence. Naturally, care providers' attitudes toward patient care are also important.

We aimed to answer the following question: is there a Yelp effect? The 2.28% (7762/340,328) of Google Maps reviews that mentioned antibiotics were indeed significantly more negative than those without antibiotic-related keywords ($P<.001$). Furthermore, our results show that reviews of urgent care centers on Yelp are significantly more negative compared to those on Google Maps ($P<.001$). Thus, we cannot rule out the existence of a Yelp effect on either Yelp or Google Maps. However, our results show that antibiotic prescription is merely one of the many potentially addressable issues in doctor-patient communication and may not be the primary source of negative web-based reviews. Indeed, patient satisfaction seems to have been most strongly linked to customer service issues (eg, wait times, rude staff, billing practices, etc). Thus, we must question whether claims regarding the impact of antibiotic prescriptions on negative reviews of urgent care centers are exaggerated. In recent years, some authors have suggested the presence of an effect that is similar to the Yelp effect in the context of opioid prescription [29,30]. However, similar to our findings, other studies have shown that these negative reviews are primarily comments on physicians' attributes or administrative attributes [31].

Limitations

The limitations of our work include our inability to hand-annotate all of the 481,825 reviews in our data set. Instead, we annotated 200 of the 7762 (2.58%) messages that were identified to have antibiotic-related keywords. This limitation was mitigated by the fact that these 200 messages were selected uniformly at random, meaning that they are likely to be representative of messages with antibiotic-related keywords. It is possible that our choice of keywords might have introduced selection bias; specifically, we assumed that patients who expected to (but did not) receive antibiotics would have said so in their reviews. Thus, we cannot rule out the possibility that patients were insincere when providing their reasons for negative reviews. However, our findings clearly indicate that patients were willing to express dissatisfaction with several other topics that do not directly pertain to antibiotics. Web-based reviews also often lack key patient information (eg, visit reason, medical history, and demographics). Finally, we do not claim that our results generalize beyond urgent care settings.

Conclusion

Our analysis shows that the Yelp effect may not be a major driver of negative sentiments in web-based reviews. Rather than compromise medical and public health recommendations by acceding to the potentially faulty perceptions resulting from patients' desires, urgent care facilities should instead invest in efforts for improving patients' overall experience, such as reducing wait times, making billing practices transparent, and investing in training staff members to adhere to the best standards of customer service. Although these steps may not prevent all negative reviews, our analysis suggests that antibiotic prescribing need not be the focal point for patient satisfaction in urgent care settings.

Acknowledgments

This work was supported by a grant from The George Washington University's Cross-Disciplinary Research Fund.

Conflicts of Interest

None declared

Multimedia Appendix 1
Google Maps review collection algorithm.
[DOCX File , 28 KB - jmir_v23i10e29406_app1.docx]

Multimedia Appendix 2
Results and interpretations of 3 latent Dirichlet allocation models.
[XLSX File (Microsoft Excel File), 49 KB - jmir_v23i10e29406_app2.xlsx]

Multimedia Appendix 3
Antibiotic-related keywords.

[DOCX File , 17 KB - jmir_v23i10e29406_app3.docx]

Multimedia Appendix 4
Descriptive statistics.
[XLSX File (Microsoft Excel File), 29 KB - jmir_v23i10e29406_app4.xlsx]

References

1. Antibiotic resistance: Multi-country public awareness survey. World Health Organization. 2015. URL: https://apps.who.int/iris/bitstream/handle/10665/194460/9789241509817_eng.pdf?sequence=1&isAllowed=y [accessed 2021-09-01]
2. Fleming-Dutra KE, Hersh AL, Shapiro DJ, Bartoces M, Enns EA, File TM, et al. Prevalence of inappropriate antibiotic prescriptions among US ambulatory care visits, 2010-2011. JAMA 2016 May 03;315(17):1864-1873. [doi: 10.1001/jama.2016.4151] [Medline: 27139059]
3. Mangione-Smith R, McGlynn EA, Elliott MN, Krogstad P, Brook RH. The relationship between perceived parental expectations and pediatrician antimicrobial prescribing behavior. Pediatrics 1999 Apr;103(4 Pt 1):711-718. [doi: 10.1542/peds.103.4.711] [Medline: 10103291]
4. May L, Gudger G, Armstrong P, Brooks G, Hinds P, Bhat R, et al. Multisite exploration of clinical decision making for antibiotic use by emergency medicine providers using quantitative and qualitative methods. Infect Control Hosp Epidemiol 2014 Sep;35(9):1114-1125 [FREE Full text] [doi: 10.1086/677637] [Medline: 25111919]
5. Stivers T, Mangione-Smith R, Elliott MN, McDonald L, Heritage J. Why do physicians think parents expect antibiotics? What parents report vs what physicians believe. J Fam Pract 2003 Feb;52(2):140-148. [Medline: 12585992]
6. What drives inappropriate antibiotic use in outpatient care? The Pew Charitable Trusts. 2017 Jun 28. URL: https://www.pewtrusts.org/en/research-and-analysis/issue-briefs/2017/06/what-drives-inappropriate-antibiotic-use-in-outpatient-care/ [accessed 2021-09-01]
7. Welschen I, Kuyvenhoven M, Hoes A, Verheij T. Antibiotics for acute respiratory tract symptoms: patients' expectations, GPs' management and patient satisfaction. Fam Pract 2004 Jun;21(3):234-237. [doi: 10.1093/fampra/cmh303] [Medline: 15128681]
8. Ong S, Nakase J, Moran GJ, Karras DJ, Kuehnert MJ, Talan DA, EMERGEncy ID NET Study Group. Antibiotic use for emergency department patients with upper respiratory infections: prescribing practices, patient expectations, and patient satisfaction. Ann Emerg Med 2007 Sep;50(3):213-220. [doi: 10.1016/j.annemergmed.2007.03.026] [Medline: 17467120]
9. Stearns CR, Gonzales R, Camargo Jr CA, Maselli J, Metlay JP. Antibiotic prescriptions are associated with increased patient satisfaction with emergency department visits for acute respiratory tract infections. Acad Emerg Med 2009 Oct;16(10):934-941 [FREE Full text] [doi: 10.1111/j.1553-2712.2009.00522.x] [Medline: 19799568]
10. Zgierska A, Rabago D, Miller MM. Impact of patient satisfaction ratings on physicians and clinical care. Patient Prefer Adherence 2014 Apr 03;8:437-446. [doi: 10.2147/PPA.S59077] [Medline: 24729691]
11. Mckenna M. The Yelping of the American doctor. Wired. 2018 Mar 22. URL: https://www.wired.com/story/the-yelping-of-the-american-doctor/ [accessed 2021-09-01]
12. Lagasse J. Doctors' online reviews show more polarization, have more impact than those of other industries. Healthcare Finance. 2019 Dec 20. URL: https://www.healthcarefinancenews.com/news/doctors-online-reviews-show-more-polarization-have-more-impact-those-other-industries/ [accessed 2021-09-01]
13. López L, Weissman JS, Schneider EC, Weingart SN, Cohen AP, Epstein AM. Disclosure of hospital adverse events and its association with patients' ratings of the quality of care. Arch Intern Med 2009 Nov 09;169(20):1888-1894. [doi: 10.1001/archinternmed.2009.387] [Medline: 19901141]
14. Kadry B, Chu LF, Kadry B, Gammas D, Macario A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. J Med Internet Res 2011 Nov 16;13(4):e95 [FREE Full text] [doi: 10.2196/jmir.1960] [Medline: 22088924]
15. Paul MJ, Wallace BC, Dredze M. What affects patient (dis) satisfaction? Analyzing online doctor ratings with a joint topic-sentiment model. In: Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence. 2013 Presented at: AAAI-13: Expanding the Boundaries of Health Informatics Using AI; July 14-18, 2013; Phoenix, Arizona, USA.
16. Remove online reviews of doctors!. Change.org. URL: https://www.change.org/p/yelp-remove-online-reviews-of-doctors/ [accessed 2021-09-01]
17. Zhang KZK, Zhao SJ, Cheung CMK, Lee MKO. Examining the influence of online reviews on consumers' decision-making: A heuristic–systematic model. Decis Support Syst 2014 Nov;67:78-89. [doi: 10.1016/j.dss.2014.08.005]
18. Daskivich TJ, Houman J, Fuller G, Black JT, Kim HL, Spiegel B. Online physician ratings fail to predict actual performance on measures of quality, value, and peer review. J Am Med Inform Assoc 2018 Apr 01;25(4):401-407 [FREE Full text] [doi: 10.1093/jamia/ocx083] [Medline: 29025145]
19. Yaraghi N, Wang W, Gao GG, Agarwal R. How online quality ratings influence patients' choice of medical providers: Controlled experimental survey study. J Med Internet Res 2018 Mar 26;20(3):e99 [FREE Full text] [doi: 10.2196/jmir.8986] [Medline: 29581091]

20. Cawkwell PB, Lee L, Weitzman M, Sherman SE. Tracking hookah bars in New York: Utilizing Yelp as a powerful public health tool. JMIR Public Health Surveill 2015 Nov 20;1(2):e19 [FREE Full text] [doi: 10.2196/publichealth.4809] [Medline: 27227137]
21. Ranard BL, Werner RM, Antanavicius T, Schwartz HA, Smith RJ, Meisel ZF, et al. Yelp reviews of hospital care can supplement and inform traditional surveys of the patient experience of care. Health Aff (Millwood) 2016 Apr;35(4):697-705 [FREE Full text] [doi: 10.1377/hlthaff.2015.1030] [Medline: 27044971]
22. Blei DM, Ng AY, Jordan MI. Latent dirichllocation. J Mach Learn Res 2003 Mar 01;3:993-1022 [FREE Full text] [doi: 10.5555/944919.944937]
23. Riddell A, Hopper T, Luo S, Leinweber K, Grivas A. lda-project/lda: 1.1.0. Zenodo. URL: https://zenodo.org/record/1412135 [accessed 2021-09-01]
24. Blei DM. Probabilistic topic models. Commun ACM 2012 Apr 01;55(4):77-84. [doi: 10.1145/2133806.2133826]
25. Krippendorff K. Content Analysis: An Introduction to Its Methodology. Thousand Oaks, California: SAGE Publications Inc; 2013.
26. Broniatowski DA, Klein EY, Reyna VF. Germs are germs, and why not take a risk? Patients' expectations for prescribing antibiotics in an inner-city emergency department. Med Decis Making 2015 Jan;35(1):60-67 [FREE Full text] [doi: 10.1177/0272989X14553472] [Medline: 25331913]
27. Klein EY, Martinez EM, May L, Saheed M, Reyna V, Broniatowski DA. Categorical risk perception drives variability in antibiotic prescribing in the emergency department: A mixed methods observational study. J Gen Intern Med 2017 Oct;32(10):1083-1089 [FREE Full text] [doi: 10.1007/s11606-017-4099-6] [Medline: 28634909]
28. Broniatowski DA, Klein EY, May L, Martinez EM, Ware C, Reyna VF. Patients' and clinicians' perceptions of antibiotic prescribing for upper respiratory infections in the acute care setting. Med Decis Making 2018 Jul;38(5):547-561 [FREE Full text] [doi: 10.1177/0272989X18770664] [Medline: 29847253]
29. Wenghofer EF, Wilson L, Kahan M, Sheehan C, Srivastava A, Rubin A, et al. Survey of Ontario primary care physicians' experiences with opioid prescribing. Can Fam Physician 2011 Mar;57(3):324-332 [FREE Full text] [Medline: 21402971]
30. Sundling RA, Logan DB, Tawancy CH, So E, Lee J, Logan K. Opioid prescribing habits of podiatric surgeons following elective foot and ankle surgery. Foot (Edinb) 2020 Dec;45:101710. [doi: 10.1016/j.foot.2020.101710] [Medline: 33137545]
31. Orhurhu MS, Salisu B, Sottosanti E, Abimbola N, Urits I, Jones M, et al. Chronic pain practices: An evaluation of positive and negative online patient reviews. Pain Physician 2019 Sep;22(5):E477-E486 [FREE Full text] [Medline: 31561660]

Abbreviations

API: application programming interface
LDA: latent Dirichlet allocation

Edited by T Kool; submitted 06.04.21; peer-reviewed by N Yaraghi, DK Yon; comments to author 14.05.21; revised version received 12.08.21; accepted 26.08.21; published 08.10.21.

Please cite as:
Hu D, Liu CMH, Hamdy R, Cziner M, Fung M, Dobbs S, Rogers L, Turner MM, Broniatowski DA
Questioning the Yelp Effect: Mixed Methods Analysis of Web-Based Reviews of Urgent Cares
J Med Internet Res 2021;23(10):e29406
URL: https://www.jmir.org/2021/10/e29406
doi:10.2196/29406
PMID:34623316

Original Paper

Comparing the Impact of Online Ratings and Report Cards on Patient Choice of Cardiac Surgeon: Large Observational Study

Xuan Li[1], PhD; Shin-Yi Chou[2], PhD; Mary E Deily[2], PhD; Mengcen Qian[3], PhD

[1]Capital One Financial Corporation, McLean, VA, United States

[2]Department of Economics, Lehigh University, Bethlehem, PA, United States

[3]School of Public Health, Fudan University, Key Laboratory of Health Technology Assessment, Ministry of Health, Shanghai, China

Corresponding Author:
Mengcen Qian, PhD
School of Public Health, Fudan University, Key Laboratory of Health Technology Assessment, Ministry of Health
130 Dong'an Road
Shanghai, 200032
China
Phone: 86 13524622077
Email: qianmengcen@fudan.edu.cn

Abstract

Background: Patients may use two information sources about a health care provider's quality: online physician reviews, which are written by patients to reflect their subjective experience, and report cards, which are based on objective health outcomes.

Objective: The aim of this study was to examine the impact of online ratings on patient choice of cardiac surgeon compared to that of report cards.

Methods: We obtained ratings from a leading physician review platform, Vitals; report card scores from Pennsylvania Cardiac Surgery Reports; and information about patients' choices of surgeons from inpatient records on coronary artery bypass graft (CABG) surgeries done in Pennsylvania from 2008 to 2017. We scraped all reviews posted on Vitals for surgeons who performed CABG surgeries in Pennsylvania during our study period. We linked the average overall rating and the most recent report card score at the time of a patient's surgery to the patient's record based on the surgeon's name, focusing on fee-for-service patients to avoid impacts of insurance networks on patient choices. We used random coefficient logit models with surgeon fixed effects to examine the impact of receiving a high online rating and a high report card score on patient choice of surgeon for CABG surgeries.

Results: We found that a high online rating had positive and significant effects on patient utility, with limited variation in preferences across individuals, while the impact of a high report card score on patient choice was trivial and insignificant. About 70.13% of patients considered no information on Vitals better than a low rating; the corresponding figure was 26.66% for report card scores. The findings were robust to alternative choice set definitions and were not explained by surgeon attrition, referral effect, or admission status. Our results also show that the interaction effect of rating information and a time trend was positive and significant for online ratings, but small and insignificant for report cards.

Conclusions: A patient's choice of surgeon is affected by both types of rating information; however, over the past decade, online ratings have become more influential, while the effect of report cards has remained trivial. Our findings call for information provision strategies that incorporate the advantages of both online ratings and report cards.

(J Med Internet Res 2021;23(10):e28098) doi:10.2196/28098

KEYWORDS

online physician reviews; report cards; cardiac surgeons; patient choice

Introduction

Since the early 1990s, federal and state agencies in the United States have published hospital report cards that measure the risk-adjusted performance of physicians so as to encourage quality-informed decisions by patients. Generally, complex mathematical tools are applied to inpatient records and clinical data to correct for patient characteristics and severity of conditions to produce arguably reliable and objective quality information, a process that is costly in time and resources.

However, studies have found that patient use of this information is below expectation: although reductions in volume for poor-performing providers is observed, the size of the effect is small, and highly rated providers are not always rewarded with increased market share [1-4]. Possible reasons include low awareness [5], high cognitive burden of understanding and interpreting quality metrics [6], and the absence of information more interesting to patients, such as health gains, waiting time, or bedside manner [7].

Fortunately, the spread over the last decade of physician review platforms provides a new information source that patients can use when choosing a physician. These platforms (ie, Vitals, ZocDocs, and RateMDs) allow patients to assign a general rating, indicate detailed ratings of specific aspects of their experience, and write comments, which are then disclosed to the public (see Panel A of Figure S1 in Multimedia Appendix 1). In contrast to report cards, online reviews are prompt, straightforward, and more subjective.

Recent studies of the reliability of physician online ratings, however, show mixed results [8-11]. Some researchers argue that online ratings may not reflect actual quality information due to "fake" reviews and a lower response rate among patients with less favorable outcomes [12,13]. Further, while some researchers found that patients treated by cardiac surgeons with lower online ratings had lower odds of surviving [14], others found no correlation between online ratings and clinical quality measures [15,16].

However, despite being controversial as quality measures, online physician reviews may still affect patient choice. First, such ratings are popular: surveys suggest that 50% to 75% of US consumers choosing a physician consult physician review platforms [17,18]. Second, online reviews serve as a form of word-of-mouth. Previous studies suggest that the impact of market-based learning on patient choice is greater than public report cards, and that narratives from nonprofessionals have a powerful impact [19,20]. Unfortunately, few studies have directly examined the impact of online ratings on patient demand. Early attempts used the volume of appointments scheduled via review platforms as the outcome measure [21-24]. However, previous studies are subject to two major concerns: first, online appointments do not necessarily result in an actual office visit, which introduces measurement errors; second, good online ratings and the decision of providing online scheduling may be correlated with unobserved surgeon characteristics, resulting in selection of physicians [25].

In this study, we investigated patient responses to different quality information sources using both online ratings and report card scores. We took advantage of the quality reporting based on risk-adjusted mortality rates achieved by individual cardiac surgeons in Pennsylvania, United States, which the state publishes every 1 to 2 years. We compared the impact of online ratings on patient choice with that of report card scores by linking both types of rating information to inpatient discharges in Pennsylvania over a 10-year period.

Methods

Data and Variables

Inpatient Records

We used inpatient discharge records collected by the Pennsylvania Health Care Cost Containment Council (PHC4) to identify coronary artery bypass graft (CABG) surgeries (ICD-9-CM [International Classification of Diseases, Ninth Revision, Clinical Modification] procedure codes 36.10-36.19) performed from January 2008 through December 2017. Each record included the quarter of admission, patient characteristics (eg, age, sex, and home zip code), the operating physician's license number, and a hospital identifier. We used the operating physician's license number and the Pennsylvania State Licensing System to identify the physician's full name and practice location, which were then used to identify the online ratings of the physician. In this step, we matched information for 99.4% of surgeons. We focused on surgeons who received at least one set of report card scores during our study time frame and obtained a sample of 37,354 CABG surgeries performed by 184 surgeons.

Online Physician Ratings

Our online ratings were from the website Vitals. Vitals was launched in 2008 and is a leading physician review platform in the United States. We chose Vitals for two important reasons. First, it is free for both patients and providers to use and, thus, has become very popular, accumulating reviews for over 1 million doctors and 165,000 facilities. In the first half of 2018, for example, over 940,000 people in United States visited it each month. The usage of Vitals in Pennsylvania ranked in the top quartile among all the states (Figure S2 and Table S1 in Multimedia Appendix 1 give more information about Vitals, showing variations in Vitals usage across states and specialties, respectively). Second, Vitals provides the date that each review was posted, allowing us to identify the average ratings that a patient would see at the time of his or her surgery.

Each review on Vitals includes a required overall rating, optional ratings of eight specific aspects of the patient's experience (ie, ease of making appointments, promptness, friendliness of the staff, accuracy of diagnosis, bedside manner, time spent with the patient, whether there was appropriate follow-up, and wait time), and optional written comments. All ratings range from one to five stars. For each provider, Vitals displays all received reviews and provides a summary with the average and the distribution of the overall rating earned by the physician across all of his or her practice hospitals (see Panel B of Figure S1 in Multimedia Appendix 1). We identified surgeons who performed CABG in Pennsylvania during our study period and who were reviewed on Vitals using surgeons' full names, specialties, and locations. All reviews posted before January 1, 2018, were scraped. A total of 1096 reviews for 132 out of 184 surgeons (71.7%) were obtained (see Figure S3 in Multimedia Appendix 1 for variations in rating volumes across years).

We calculated the average overall ratings at the time of a patient's surgery at the surgeon-by-quarter levels and matched this information to inpatient records based on the surgeon's

license number and the patient's quarter of admission. In a given quarter, ratings were the same for the same surgeon who practiced at multiple hospitals. Following the literature [14], we used a categorical form of the rating information to address the potential underrepresentation of low ratings on Vitals (see Figure S4 in Multimedia Appendix 1) while also taking into account surgeons with no ratings. We gave a surgeon a "high online rating" each quarter when the surgeon's average rating, based on reviews posted up to the most recent quarter, was at least four stars. We also created an indicator that equals 1 if the surgeon had not yet been reviewed on Vitals.

Report Card Scores

We obtained report card scores for providers from the seven issues of PHC4's Cardiac Surgery Report published during our study period. Table S2 in Multimedia Appendix 1 lists the publication dates and data collection periods for each issue. Report cards provide quality scores for each surgeon who performed at least 30 CABG surgeries during the data collection period, indicating whether they had mortality rates that were greater than expected, as expected, or lower than expected based on risk-adjusted health outcomes of their patients (see Figure S5 in Multimedia Appendix 1). We used the reported summary measure for each surgeon across his or her practice hospitals. Following the literature [3], we assumed that patients used the most recent report card to inform their decisions and, therefore, that the effective period of each report card was from its publication date to the publication date of the next report card. Accordingly, we linked the most recent report card information to inpatient records, based on the surgeon's name. We created an indicator, "high report card score," which equals 1 if the surgeon had lower than expected mortality rates on the most recent PHC4 report card. We generated a dummy variable for surgeons with no score in the most recent report card.

Control Variables

We obtained hospital characteristics from the 2010 American Hospital Association's Annual Survey of Hospitals. These were the number of beds, whether the hospital was a member of the Council of Teaching Hospitals, whether it had a cardiology intensive care unit (CICU), and the hospital's zip code. We calculated the distance between a patient's zip code and the zip code of the admitting hospital to capture the proximity of surgeons to patients.

The Final Inpatient Sample

We restricted our sample to patients with fee-for-service plans since their choices are expected to be less restricted by insurance networks. The final inpatient sample with full information consisted of 12,521 CABG discharges performed by 184 surgeons.

Empirical Strategy

We estimated a random coefficient logit model, also known as the mixed logit model, to characterize a patient's choice of surgeon [26]. The model has become the preferred method in hospital and physician choice analyses for two reasons [3,27]. First, the model estimates random coefficients on choice characteristics, thus eliminating the need to assume independence of irrelevant alternatives, which is likely to be violated in a model of surgeon choice. Second, the model allows for random taste variation among patients so that we can better incorporate consideration of patient heterogeneity. This feature is particularly important in our study because patients may differ in their knowledge of and willingness to use information from online ratings and report cards.

We assumed that each patient chooses among a set of possible surgeon-hospital pairs (ie, the surgeon's rating is the same across the different hospitals where he or she works), and the observed portion of utility is additively separable in hospital and surgeon characteristics. The utility of patient i for choosing hospital j and surgeon k at time t can be specified as follows:

where Q_{ijkt} is a vector of quality information variables and X_{ijk} is a vector of observed hospital characteristics, including the patient's distance from the hospital, and surgeon fixed effects, which capture the effect of time-invariant unobserved heterogeneity across physicians. ε_{ijkt} is an idiosyncratic error term, which is assumed to be random with an independent and identically distributed extreme value distribution. Coefficients of the quality information and the distance variable were modeled as random variables with normal distributions to represent heterogeneity in patient tastes. The preference parameter, β_i, is the parameter of interest, as it captures the marginal utilities of the different sources of quality information.

The probability that patient i chooses surgeon k conditional on α_i and β_i is as follows:

where y_i^S represents patient i's choice of surgeon, g indexes the hospitals in patient i's choice set, and B_g^t represents the set of surgeons available at hospital set g at time t. The unconditional choice probability can be then obtained by integrating the conditional choice probability from equation 2 over the probability density functions of α and β as follows:

where p_{ijkt} represents the probability of patient i choosing hospital j and surgeon k at time t, and $f(\,)$ is the density function of a normal distribution. Through numeric integration, the log likelihood function of equation 3 can be maximized to yield estimates of the means and SDs of the preference parameters [28,29] .

To define the choice set, we drew a circle with its radius centered on a patient's zip code and considered all surgeon-hospital pairs located within that circle as possible alternatives for the patient. In our main analysis, we followed the literature and used a radius of 50 miles, excluding discharges from patients who traveled beyond this cutoff [3], reducing the sample by 14.21% (1779/12,521). The estimation sample for the mixed logit models includes a total of 973,953 patient-alternative pairs.

Our variables of interest were the dummy variables for good quality information. We used two approaches to evaluate the impact of good ratings on patient choice. First, we calculated the average marginal effect (AME) of a surgeon receiving good ratings by doing the following: randomly selecting one surgeon from the choice set of each patient and bootstrapping 1000 times to calculate changes in the predicted probability of the surgeon being chosen when receiving a high rating versus not, then taking the average over all randomly selected surgeons.

Second, we calculated the impact of good ratings on patients' willingness to travel (WTT). Because patients generally preferred a hospital closer to home, we used a patient's WTT for a surgeon-hospital pair as an analog of the willingness to pay reported in preference studies involving prices. The WTT was calculated as the negative ratio of the estimated coefficient for high ratings to the estimated coefficient for distance.

Robustness Checks

We performed four robustness checks. First, attrition of low-performing surgeons might lead to an overestimation of the effect of favorable quality information. We assume a surgeon has exited a patient's choice set if he or she performed no CABG surgeries during the quarter that patient was admitted and no additional surgeries throughout the rest of our study period.

Second, we examined the effect of replacing the choice set based on a fixed 50-mile radius centered on the patient's residence to a variable radius determined by the distance necessary to encompass 90% of patients who received CABG at a hospital (ie, the practice location of the surgeon). The variable radius captures the actual market of each practice hospital. Accordingly, we then redefined the choice set of a patient as surgeons whose practice hospitals had an actual market that covered the patient's residence.

Third, we excluded patients with urgent and emergency conditions because these patients are less likely to have adequate time to gather information before making a choice. Finally, we considered the possibility that surgeons with good ratings also have better personal connections with primary care doctors or cardiologists and, thus, receive more referrals. As the inpatient records provide the license number of the doctors who made referrals, we restricted our sample to patients without this information in their records to address this concern.

Results

Descriptive Statistics

Table 1 reports descriptive statistics for key variables, first for all fee-for-service CABG discharges in Pennsylvania during our study time frame, and then for a sample restricted by travel distance. The average distance to the hospital dropped by more than half in the restricted sample due to the exclusion of patients who traveled more than 50 miles. All other observed characteristics were similar across the samples. In the restricted sample, at the time of a surgery, on average, 25.66% of surgeries were performed by surgeons with a high online rating, 68.43% by surgeons with no rating on Vitals, 6.04% by surgeons who received a high report card score, and 18.45% by surgeons with no report card scores.

Table 1. Descriptive statistics of the inpatient sample.

Characteristic	Inpatient sample without restrictions[a] (N=12,521), mean (SD)	Inpatient sample restricted to discharges within 50 miles[b] (N=10,742), mean (SD)
Panel A: patient characteristics		
Age (years)	70.09 (9.39)	70.23 (9.32)
Male (%)	69.79 (45.92)	69.66 (45.97)
White (%)	86.64 (34.03)	85.88 (34.83)
Nonemergency admission (%)	68.48 (46.46)	67.28 (46.92)
Distance to hospital (miles)	32.94 (106.58)	14.17 (11.87)
Panel B: surgeon characteristics (%)		
High online rating	25.57 (43.62)	25.66 (43.68)
No online rating	68.60 (46.41)	68.43 (46.48)
High report card score	5.85 (23.46)	6.04 (23.83)
No report card score	18.34 (38.70)	18.45 (38.79)
Panel C: hospital characteristics		
Teaching hospitals (%)	56.28 (49.61)	51.04 (49.99)
Bed size, n	543.04 (359.77)	531.06 (355.82)
Contains a cardiology intensive care unit (%)	93.09 (25.36)	92.09 (27.00)

[a]Inpatient sample consisting of fee-for-service patients who received a coronary artery bypass graft in Pennsylvania from January 2008 to December 2017.

[b]Subset of the above sample who traveled less than 50 miles.

Main Results

Table 2 reports estimates of the random coefficient logit model. For each specification, we reported the estimated mean and SD of the effect of the quality information and travel distance. The estimated means represented the average responses of the patients in our sample, and the estimated SDs captured the degree of heterogeneity in such responses among patients. We found that the estimated mean of "high online rating" was positive and significant, suggesting that the average patient was more likely to choose a surgeon who had a higher online rating compared with lower-rated surgeons. In contrast, the estimated mean of "high report card score" was small and insignificant, suggesting no return in terms of patient preference for surgeons whose patients were less likely to die than expected.

Table 2. Random coefficient logit estimates: fee-for-service patients' choices in Pennsylvania, United States[a].

Variable	Mean (SE)	*P* value	SD (SE)	*P* value
High online rating	0.317 (0.058)	<.001	0.0643 (0.216)	.77
No online rating	0.536 (0.064)	<.001	1.015 (0.117)	<.001
High report card score	–0.057 (0.093)	.54	0.326 (0.381)	.39
No report card score	–0.516 (0.091)	<.001	0.828 (0.243)	<.001
Distance to hospital	–0.278 (0.005)	<.001	0.122 (0.004)	<.001

[a]The table presents random coefficient logit estimates from the estimation sample, which consists of within-50-mile alternatives for patients who received a fee-for-service coronary artery bypass graft in Pennsylvania, United States. The number of observations (n=973,953) is the number of patient-alternative pairs. Other control variables included hospital characteristics (ie, bed size, indicator for member of the Council of Teaching Hospitals, and indicator for having a cardiology intensive care unit) and surgeon fixed effects.

Our results suggest that patients responded differently to surgeons with no online ratings in Vitals and surgeons with no report card scores. The positive estimated mean of "no online rating" indicates that the average patient was more likely to choose a surgeon who had not been reviewed on Vitals compared to a surgeon who had been reviewed and received a low rating. The findings suggested that it was worse for a surgeon to appear on Vitals with a low rating than to not appear at all. In contrast, the estimated mean of "no report card score" was negative. Since the cardiac report cards assigned a rating to all surgeons performing at least 30 surgeries during the data collection period, the finding implied a desire by patients to avoid surgeons that performed fewer operations.

We found little heterogeneity among patients with respect to a surgeon having a high online rating or a high report card score: all patients viewed these positively. Patient heterogeneity was statistically significant, however, for missing online ratings or report card scores. The significant estimated SDs for "no report card score" and "no online rating" suggested that some patients may treat the absence of such information as a signal of lower quality, while others may not. Based on the cumulative standard normal distribution, we found that 70.13% (φ [0.536/1.015]) of patients considered no information better than low ratings on Vitals, while the other 29.87% thought the opposite. The corresponding figures were 26.66% and 73.34% (φ [0.516/0.828]) for "no report card score," suggesting a slightly more unanimous opinion that a missing report card score was a negative signal.

We calculated AMEs and WTT for a high online rating to further understand the impact of this information. We found that holding other observables constant, a surgeon's probability of being chosen would increase by 0.69 percentage points, on average, if the surgeon's online rating changed from low to high. We defined the baseline probability of being chosen as the average ratio of the total number of selected surgeons over the total number of surgeons within patients' choice sets. The above AME estimates of "high online rating" corresponded to a 22.47% increase in choice probability, compared with the baseline. In terms of WTT, in our sample an average patient was willing to travel 0.59 miles to reach a surgeon with high ratings on Vitals.

Additional Analyses

Table 3 presents results of the robustness checks. The first column reports estimates from a sample restricted to surgeons who remained in the market for the entire 10-year period. The second column presents results using the variable radius approach to define the choice sets. The third column reports results from a sample excluding urgent and emergency admissions. The fourth column displays results for nonreferral admissions only. All of these results were qualitatively similar with those presented in Table 2, suggesting that our main results were robust to surgeon attrition, alternative choice set definitions, and emergency status of admissions, and that our findings were not driven by a referral effect.

Although the marginal effect of a high online rating is more influential than a high report card score on patient utilities on average, the relative importance of the two information sources to patient choices may have changed over time. The number of online reviews on Vitals for cardiac surgeons operating in Pennsylvania was relatively small until 2014; since then, they have become increasingly popular (Figure S3 in Multimedia Appendix 1). We explore this further by adding a set of interactions between each rating variable and a linear year trend to equation 1 to capture any temporal pattern in the main effects of the rating information.

Table 3. Robustness checks[a].

Variable	Excluding exited surgeons[b]	*P* value	Variable radius[c]	*P* value	Nonemergency[d]	*P* value	Nonreferral[e]	*P* value
High online rating								
Mean (SE)	0.167 (0.067)	.01	0.142 (0.072)	.050	0.164 (0.070)	.02	0.172 (0.068)	.01
SD (SE)	0.826 (0.278)	.003	0.721 (0.398)	.07	0.004 (0.322)	.99	0.368 (0.399)	.36
No online rating								
Mean (SE)	0.505 (0.067)	<.001	0.543 (0.075)	<.001	0.458 (0.079)	<.001	0.447 (0.074)	<.001
SD (SE)	1.089 (0.195)	<.001	1.247 (0.218)	<.001	1.154 (0.126)	<.001	1.076 (0.160)	<.001
High report card score								
Mean (SE)	0.016 (0.096)	.87	–0.060 (0.124)	.63	–0.073 (0.085)	.39	0.036 (0.138)	.79
SD (SE)	0.324 (0.396)	.41	0.650 (0.305)	.03	0.190 (0.429)	.66	0.445 (0.482)	.36
No report card score								
Mean (SE)	–0.273 (0.045)	<.001	–0.278 (0.047)	<.001	–0.282 (0.055)	<.001	–0.316 (0.055)	<.001
SD (SE)	0.109 (0.273)	.69	0.053 (0.268)	.84	0.169 (0.252)	.50	0.199 (0.344)	.56
Distance to hospital								
Mean (SE)	–0.282 (0.005)	<.001	–0.327 (0.008)	<.001	–0.245 (0.005)	<.001	–0.293 (0.006)	<.001
SD (SE)	0.126 (0.005)	<.001	0.189 (0.009)	<.001	0.102 (0.005)	<.001	0.134 (0.005)	<.001

[a]All random coefficient logit models also included hospital characteristics (ie, bed size, indicator for member of the Council of Teaching Hospitals, and indicator for having a cardiology intensive care unit) and surgeon fixed effects.

[b]Each column is a separate regression. This column reports estimates from a sample restricted to surgeons who did not exit the market. The number of observations (ie, the number of patient-alternative pairs) is 798,150.

[c]This column reports results using the variable radius approach to define the choice sets. The number of observations is 354,226.

[d]This column reports results from a sample excluding urgent and emergency admissions. The number of observations is 498,903.

[e]This column reports results for nonreferral admissions only. The number of observations is 682,852.

Table 4 reports the results. The estimated means of the interaction of "high online rating" and a linear year trend were positive, while those of the interaction of "high report card score" and a linear year trend were around four times smaller and insignificant. Figure 1 depicts the marginal utilities of receiving a high online rating and a high report card score over time. The marginal effects of a high online rating were imprecisely estimated in the early years and then turned positive and significant after 2014. In contrast, the estimated marginal effects of a high report card score did not strengthen, remaining insignificant over our study period. Our results suggest that online ratings have become more influential for patient choice over time, while report card scores have not.

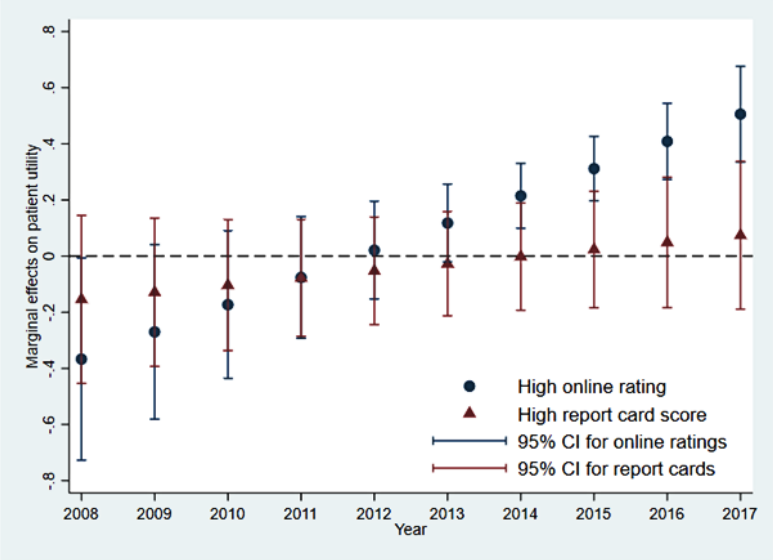
Table 4. Temporal variations in the main effects of quality information[a].

Variable	Mean (SE)	*P* value	SD (SE)	*P* value
High online rating	–0.367 (0.184)	.046	0.106 (0.335)	.75
No online rating	–0.134 (0.213)	.53	0.840 (0.188)	<.001
High report card score	–0.154 (0.153)	.31	0.354 (0.357)	.32
No report card score	–0.907 (0.056)	<.001	0.052 (0.144)	.72
Distance to hospital	–0.277 (0.005)	<.001	0.122 (0.004)	<.001
High online rating × year	0.097 (0.027)	<.001	N/A[b]	N/A
No online rating × year	0.091 (0.033)	.005	N/A	N/A
High report card score × year	0.025 (0.024)	.29	N/A	N/A
No report card score × year	0.165 (0.012)	<.001	N/A	N/A

[a]The mean and SD columns present results from equation 1 with the interaction of quality variables and a linear time trend controlled. Estimation is based on discharges of fee-for-service patients who traveled no more than 50 miles. The number of observations (n=973,953) is the number of patient-alternative pairs. Other control variables included hospital characteristics (ie, bed size, indicator for member of the Council of Teaching Hospitals, and indicator for having a cardiology intensive care unit) and surgeon fixed effects.

[b]N/A: not applicable; these coefficients were not modeled as random variables.

Figure 1. Changes in marginal effects of a high online rating and a high report card score on patient utility from 2008 to 2017. The marginal effects of a high online rating and a high report card score on patient utility are plotted with their 95% CIs by year. Calculations were based on estimates reported in Table 4.



Discussion

This paper investigated the impact of online physician ratings on patients' choice of CABG surgeon from 2008 through 2017 in Pennsylvania and compared it to the influence of report card scores. We estimated random coefficient logit models to allow for patient heterogeneity in preferences for surgeons with different ratings using a sample of fee-for-service patients to avoid impacts of insurance networks on patient choices. Our results show that positive assessments on the online physician rating platform had a significantly positive impact on a patient's choice of surgeon, and that the impact increased over our study period. In contrast, the effect of a good report card score was trivial and did not change much over these years.

Two reasons may explain the increasing influence of online ratings. First, online ratings are much easier to use. If a patient searches for a surgeon, most of the returned links are to physician review websites. In contrast, to access report card scores, patients must know the scores exist, go to the correct website of the state agency, download the report, understand the meaning of the scores, and read through the report to find the physician that they are interested in. Second, online ratings can provide information on things like communication skills, friendliness of the staff, or ease of making an appointment, attributes that patients care about, but which are not available from report cards.

However, one concern about the increasing influence of online ratings is that the information provided on these review websites may steer patients away from surgeons who provide the best quality in terms of health outcomes. To shed light on this issue, we used text mining and machine learning techniques on the written comments on Vitals to identify attributes more likely

to result in a higher overall rating. In our scraped reviews, about 13% of the ratings included written comments.

First, we calculated frequency of words and phrases. We found that phrases related to doctors' attitudes toward patients, such as "bedside manner," "doctor caring compassionate," and "office staff," were more frequently mentioned. In comments accompanying low ratings, patients often complained about scheduling issues and the time that the doctor spent with them. We then employed the classification and regression trees method to identify the strongest predictors of ratings among the frequently mentioned words [30]. Our results showed that the words "insurance," "wait," "appointment," and "questions" were the strongest predictors of ratings. About 50% of the comments that contained "insurance" gave one star; around 79% of the comments that did not contain either "insurance" or "wait" gave five stars. This suggested that insurance was a purely negative factor: when insurance worked well, it did not contribute to high ratings, but when patients raised concerns about insurance-related issues with their surgeons, they were very likely to give them a low rating. Our findings implied that online ratings reflected more nonquality attributes of services than report cards.

Our study extends previous research regarding impacts of information on patient behaviors in several dimensions. First, this is the first study to compare the impacts of online ratings and public report cards on patient choice, providing insight into patient behavior when different types of quality information are available. Second, we presented new evidence on the impact of online ratings by examining inpatient volumes. Consistent with previous studies [21], we showed that online ratings exerted a positive effect on the popularity of physicians. However, previous literature generally relied on changes in volumes of virtual appointments to examine this issue. We, instead, linked the ratings directly to actual admissions and included in the analysis surgeons that had not been reviewed. Our approach helped avoid measurement error and selection of surgeons because online appointments do not necessarily end up with office visits and the absence of quality information may correlate with other surgeon characteristics. Third, we constructed a large sample that spans a 10-year period, allowing us to investigate changes over time in the relative importance of online ratings and report card scores for patient choice.

Our results yield important implications for both policy makers and online review platforms. Online ratings provide richer information about health services that patients seem to value and are becoming increasingly influential on patient choice. Nonetheless, such ratings are not driven by traditional quality indicators, and the report cards that do provide objective quality metrics have little impact. It is important to find ways to provide the health service market with information that incorporates the advantages of both online ratings and report cards. For policy makers, it is urgent to improve report card systems, for example, by making them easier to access and interpret, by working to increase media coverage of report card information, and perhaps by adding attribute measures that patients care about [1,27]. For online review platforms, it is important to work out ways to increase the importance of self-reported quality attributes in overall ratings.

Our study has several caveats. We discussed our results in terms of influences on patients' choices, but we were not able to determine whether a patient had read online ratings or report cards before selecting a surgeon. Our results are, thus, better interpreted as an intent-to-treat effect. Moreover, although we found that a high online rating had a significant impact, we were not able to explore the mechanisms of the impact with our current data. For example, the subpopulation who are responsive to the online ratings information could be different, or online ratings may affect patient choices through younger members of a family. Understanding the underlying mechanisms is important and could be a focus of future research, because such knowledge would facilitate the design of systems that more effectively guide patients toward higher-quality providers.

Acknowledgments

MQ gratefully acknowledges financial supports from the National Natural Science Foundation of China (grant 71704027) and the "Chenguang Program" supported by the Shanghai Municipal Education Commission (grant 17CG03). SC and XL are grateful for the support provided by Lehigh University (Data X Innovation Grant). We also thank PK Kannan, Seth Richards-Shubik, Chad D Meyerhoefer, Adam Biener, and Reagan A Baughman for their valuable comments and suggestions. An earlier version of this article was presented at the 7th Conference of the American Society of Health Economists. PHC4, its agents, and its staff bear no responsibility or liability for the results of the analysis. The results are solely the views of the authors.

Authors' Contributions

XL, MQ and S-YC initiated the concept and research design. All the authors contributed to acquisition and interpretation of data. XL and MQ conducted statistical analyses. XL drafted the manuscript. MED, MQ and S-YC reviewed and revised the manuscript. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials.

[DOCX File , 348 KB - jmir_v23i10e28098_app1.docx]

References

1. Bhandari N, Scanlon D, Shi Y, Smith R. Consumer use of provider quality report cards: The role of dissemination and media coverage. Med Care 2020 Apr;58(4):368-375. [doi: 10.1097/MLR.0000000000001279] [Medline: 31876660]
2. Dranove D, Sfekas A. Start spreading the news: A structural estimate of the effects of New York hospital report cards. J Health Econ 2008 Sep;27(5):1201-1207. [doi: 10.1016/j.jhealeco.2008.03.001] [Medline: 18420293]
3. Wang J, Hockenberry J, Chou S, Yang M. Do bad report cards have consequences? Impacts of publicly reported provider quality information on the CABG market in Pennsylvania. J Health Econ 2011 Mar;30(2):392-407. [doi: 10.1016/j.jhealeco.2010.11.006] [Medline: 21195494]
4. Werner RM, Norton EC, Konetzka RT, Polsky D. Do consumers respond to publicly reported quality information? Evidence from nursing homes. J Health Econ 2012 Jan;31(1):50-61. [doi: 10.1016/j.jhealeco.2012.01.001] [Medline: 22307033]
5. Marshall MN, Shekelle PG, Leatherman S, Brook RH. The public release of performance data: What do we expect to gain? A review of the evidence. JAMA 2000 Apr 12;283(14):1866-1874. [doi: 10.1001/jama.283.14.1866] [Medline: 10770149]
6. Werner RM, Konetzka RT, Polsky D. Changes in consumer demand following public reporting of summary quality ratings: An evaluation in nursing homes. Health Serv Res 2016 Jun;51 Suppl 2:1291-1309 [FREE Full text] [doi: 10.1111/1475-6773.12459] [Medline: 26868034]
7. Gutacker N, Siciliani L, Moscelli G, Gravelle H. Choice of hospital: Which type of quality matters? J Health Econ 2016 Dec;50:230-246 [FREE Full text] [doi: 10.1016/j.jhealeco.2016.08.001] [Medline: 27590088]
8. Bardach NS, Asteria-Peñaloza R, Boscardin WJ, Dudley RA. The relationship between commercial website ratings and traditional hospital performance measures in the USA. BMJ Qual Saf 2013 Mar;22(3):194-202 [FREE Full text] [doi: 10.1136/bmjqs-2012-001360] [Medline: 23178860]
9. Schlesinger M, Grob R, Shaller D. Using patient-reported information to improve clinical practice. Health Serv Res 2015 Dec;50 Suppl 2:2116-2154 [FREE Full text] [doi: 10.1111/1475-6773.12420] [Medline: 26573890]
10. Daskivich TJ, Houman J, Fuller G, Black JT, Kim HL, Spiegel B. Online physician ratings fail to predict actual performance on measures of quality, value, and peer review. J Am Med Inform Assoc 2018 Apr 01;25(4):401-407 [FREE Full text] [doi: 10.1093/jamia/ocx083] [Medline: 29025145]
11. Lantzy S, Anderson D. Can consumers use online reviews to avoid unsuitable doctors? Evidence from RateMDs.com and the Federation of State Medical Boards. Decis Sci 2019 Jun 26;51(4):962-984. [doi: 10.1111/deci.12398]
12. Mayzlin D, Dover Y, Chevalier J. Promotional reviews: An empirical investigation of online review manipulation. Am Econ Rev 2014 Aug 01;104(8):2421-2455. [doi: 10.1257/aer.104.8.2421]
13. de Barra M, Eriksson K, Strimling P. How feedback biases give ineffective medical treatments a good reputation. J Med Internet Res 2014 Aug 21;16(8):e193 [FREE Full text] [doi: 10.2196/jmir.3214] [Medline: 25147101]
14. Lu SF, Rui H. Can we trust online physician ratings? Evidence from cardiac surgeons in Florida. Manage Sci 2018 Jun;64(6):2557-2573. [doi: 10.1287/mnsc.2017.2741]
15. Gray BM, Vandergrift JL, Gao GG, McCullough JS, Lipner RS. Website ratings of physicians and their quality of care. JAMA Intern Med 2015 Feb;175(2):291-293. [doi: 10.1001/jamainternmed.2014.6291] [Medline: 25437252]
16. Okike K, Peter-Bibb TK, Xie KC, Okike ON. Association between physician online rating and quality of care. J Med Internet Res 2016 Dec 13;18(12):e324 [FREE Full text] [doi: 10.2196/jmir.6612] [Medline: 27965191]
17. Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Public awareness, perception, and use of online physician rating sites. JAMA 2014 Feb 19;311(7):734-735. [doi: 10.1001/jama.2013.283194] [Medline: 24549555]
18. Hedges L, Couey C. How patients use online reviews. Software Advice. 2020 Apr 03. URL: https://www.softwareadvice.com/resources/how-patients-use-online-reviews/ [accessed 2019-07-02]
19. Dafny L, Dranove D. Do report cards tell consumers anything they don't know already? The case of Medicare HMOs. Rand J Econ 2008;39(3):790-821. [doi: 10.1111/j.1756-2171.2008.00039.x] [Medline: 19013864]
20. Huppertz JW, Carlson JP. Consumers' use of HCAHPS ratings and word-of-mouth in hospital choice. Health Serv Res 2010 Dec;45(6 Pt 1):1602-1613 [FREE Full text] [doi: 10.1111/j.1475-6773.2010.01153.x] [Medline: 20698896]
21. Wu H, Lu N. Online written consultation, telephone consultation and offline appointment: An examination of the channel effect in online health communities. Int J Med Inform 2017 Nov;107:107-119. [doi: 10.1016/j.ijmedinf.2017.08.009] [Medline: 29029686]
22. Lu W, Wu H. How online reviews and services affect physician outpatient visits: Content analysis of evidence from two online health care communities. JMIR Med Inform 2019 Dec 02;7(4):e16185 [FREE Full text] [doi: 10.2196/16185] [Medline: 31789597]
23. Xu Y, Armony M, Ghose A. The interplay between online reviews and physician demand: An empirical investigation. Manage Sci 2021 Mar 02. [doi: 10.1287/mnsc.2020.3879]
24. Shukla A, Gao G, Agarwal R. How digital word-of-mouth affects consumer decision making: Evidence from doctor appointment booking. Manage Sci 2021 Mar;67(3):1546-1568. [doi: 10.1287/mnsc.2020.3604]
25. Gao G, Greenwood BN, Agarwal R, McCullough JS. Vocal minority and silent majority: How do online ratings reflect population perceptions of quality. MIS Q 2015;39(3):565-590 [FREE Full text]

26. McFadden D, Train K. Mixed MNL models for discrete response. J Appl Econ 2000 Sep;15(5):447-470. [doi: 10.1002/1099-1255(200009/10)15:5<447::aid-jae570>3.0.co;2-1]
27. Pope DG. Reacting to rankings: Evidence from "America's Best Hospitals". J Health Econ 2009 Dec;28(6):1154-1165. [doi: 10.1016/j.jhealeco.2009.08.006] [Medline: 19818518]
28. Hole AR. Fitting mixed logit models by using maximum simulated likelihood. Stata J 2007 Sep 01;7(3):388-401. [doi: 10.1177/1536867x0700700306]
29. Train KE. Discrete Choice Methods With Simulation. 2nd edition. New York, NY: Cambridge University Press; 2009.
30. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Boca Raton, FL: Chapman & Hall/CRC; 1984.

Abbreviations

AME: average marginal effect
CABG: coronary artery bypass graft
CICU: cardiology intensive care unit
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
PHC4: Pennsylvania Health Care Cost Containment Council
WTT: willingness to travel

Edited by G Eysenbach; submitted 21.02.21; peer-reviewed by IV George, A Sinaiko; comments to author 17.03.21; revised version received 11.08.21; accepted 03.10.21; published 28.10.21.

Please cite as:
Li X, Chou SY, Deily ME, Qian M
Comparing the Impact of Online Ratings and Report Cards on Patient Choice of Cardiac Surgeon: Large Observational Study
J Med Internet Res 2021;23(10):e28098
URL: https://www.jmir.org/2021/10/e28098
doi:10.2196/28098
PMID:34709192

Original Paper

Willingness to Share Wearable Device Data for Research Among Mechanical Turk Workers: Web-Based Survey Study

Casey Overby Taylor[1,2], PhD; Natalie Flaks-Manov[2,3], PhD; Shankar Ramesh[4], MSc; Eun Kyoung Choe[4], PhD

[1]Departments of Medicine and Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, MD, United States

[2]Institute for Computational Medicine, Johns Hopkins University, Baltimore, MD, United States

[3]Department of Medicine, Division of General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, United States

[4]College of Information Studies, University of Maryland, College Park, MD, United States

Corresponding Author:
Casey Overby Taylor, PhD
Departments of Medicine and Biomedical Engineering
Johns Hopkins University School of Medicine
217D Hackerman Hall
3101 Wyman Park Dr
Baltimore, MD, 21218
United States
Phone: 1 4432876657
Email: cot@jhu.edu

Abstract

Background: Wearable devices that are used for observational research and clinical trials hold promise for collecting data from study participants in a convenient, scalable way that is more likely to reach a broad and diverse population than traditional research approaches. Amazon Mechanical Turk (MTurk) is a potential resource that researchers can use to recruit individuals into studies that use data from wearable devices.

Objective: This study aimed to explore the characteristics of wearable device users on MTurk that are associated with a willingness to share wearable device data for research. We also aimed to determine whether compensation was a factor that influenced the willingness to share such data.

Methods: This was a secondary analysis of a cross-sectional survey study of MTurk workers who use wearable devices for health monitoring. A 19-question web-based survey was administered from March 1 to April 5, 2018, to participants aged ≥18 years by using the MTurk platform. In order to identify characteristics that were associated with a willingness to share wearable device data, we performed logistic regression and decision tree analyses.

Results: A total of 935 MTurk workers who use wearable devices completed the survey. The majority of respondents indicated a willingness to share their wearable device data (615/935, 65.8%), and the majority of these respondents were willing to share their data if they received compensation (518/615, 84.2%). The findings from our logistic regression analyses indicated that Indian nationality (odds ratio [OR] 2.74, 95% CI 1.48-4.01, *P*=.007), higher annual income (OR 2.46, 95% CI 1.26-3.67, *P*=.02), over 6 months of using a wearable device (OR 1.75, 95% CI 1.21-2.29, *P*=.006), and the use of heartbeat and pulse tracking monitoring devices (OR 1.60, 95% CI 0.14-2.07, *P*=.01) are significant parameters that influence the willingness to share data. The only factor associated with a willingness to share data if compensation is provided was Indian nationality (OR 0.47, 95% CI 0.24-0.9, *P*=.02). The findings from our decision tree analyses indicated that the three leading parameters associated with a willingness to share data were the duration of wearable device use, nationality, and income.

Conclusions: Most wearable device users indicated a willingness to share their data for research use (with or without compensation; 615/935, 65.8%). The probability of having a willingness to share these data was higher among individuals who had used a wearable for more than 6 months, were of Indian nationality, or were of American (United States of America) nationality and had an annual income of more than US $20,000. Individuals of Indian nationality who were willing to share their data expected compensation significantly less often than individuals of American nationality (*P*=.02).

(J Med Internet Res 2021;23(10):e19789) doi:10.2196/19789

KEYWORDS

wearables; personal data; research participation; crowdsourcing

Introduction

A wearable device is a small hardware technology that people wear on various parts of their bodies. These devices' functions may include tracking and monitoring health, fitness, food, and aging-related metrics [1]. Using wearables has become very popular in part because data collection and presentation occur in real time [2] and due to their potential for having a positive effect on health and fitness [3]. In addition, regular feedback provided by wearables shows promise for positively influencing physical activity and weight loss outcomes [4]. From a research perspective, wearables hold promise for collecting data from study participants in a convenient, scalable way that is more likely to reach a broad and diverse population than traditional research approaches [5,6].

Previous research on wearable device data sharing for research has yielded mixed results regarding the willingness to share such data [7-9]. Amazon Mechanical Turk (MTurk) is a potential resource that researchers can use to recruit individuals into studies that use data from wearables. MTurk is a crowdsourcing platform that allows researchers to post human intelligence tasks (HITs) that "workers" can choose to complete for compensation [10]. Previous studies have assessed the opinions of individuals from targeted populations such as older adults [7], health app users [8], and patients [9]. Although there have been studies that have recruited wearable users from MTurk [11,12], to our knowledge, these participants' characteristics have not been examined. The objective of this study was to explore the characteristics of wearable device users on MTurk that are associated with a willingness to share data for research use. We also aimed to determine whether compensation was a factor that influenced the willingness to share such data. This study will help researchers who are considering MTurk as a research avenue to recruit wearable device users from MTurk and to understand the potential benefits and limitations.

Methods

Study Design and Population

This was a secondary analysis of a cross-sectional survey study of MTurk workers who use wearable devices for health monitoring. These workers participated in the survey from March 1 to April 5, 2018. Only adults over the age of 18 years were eligible for this study. Details on the recruitment strategy and survey instrument were reported in the primary publication [13] and are summarized in the *Recruitment Strategy and Survey Instrument* section.

Recruitment Strategy and Survey Instrument

We developed a 19-question web-based survey to explore the characteristics that are associated with a willingness to share wearable device data for research use. This survey was prepared by using Qualtrics (Qualtrics International Inc). After publishing a short description of the survey to the MTurk interface (ie, a HIT), MTurk workers who met the eligibility requirement were able to review the HIT, and if they chose to undertake the survey task, they were referred to the external Qualtrics survey website (Multimedia Appendix 1).

On the Qualtrics survey website, the web-based recruitment strategy involved four parts—an introduction to this study, screening questions, survey questions, and postsurvey steps. The introduction provided an overview of this study, which included details about the time limit, compensation, risks, benefits, and the contact details of the investigators. The screening portion was included to make sure that the respondents were qualified to take the survey and that they understood important concepts. Comprehension was assessed by using multiple-choice questions on the meanings of *health monitoring technology* and *health monitoring data* after providing descriptions. In order to assess whether a respondent was qualified to take the survey, they were asked to indicate if they use a wearable device to monitor their health (Multimedia Appendix 1). Only participants who answered the comprehension questions correctly and indicated that they use a wearable to monitor their health could move on to complete the survey. The survey included questions about demographics, experience with MTurk, motivations for participating in MTurk HITs, experience with using wearables, and interest in submitting wearable device data for research purposes. For the postsurvey steps, upon the successful completion of the survey, respondents were given a short message of appreciation for their participation. They were also provided with a random, automatically generated validation code. In order to receive payment for completing the survey, the validation code had to be entered on MTurk. Each participant was paid US $0.40 for their responses.

Pilot tests for this process were conducted to determine the time required for completing the survey. Additionally, to make sure that the efficiency of the data did not decrease, we ensured that respondents from all of the previous batches were unable to respond to the latest batch of HITs.

This study was judged as one that imposed only minimal risks on participants and was determined to be exempt from institutional review by both the Johns Hopkins University (protocol code: IRB00158371) and University of Maryland (protocol code: IRB 1165377) institutional review boards.

Outcomes and Associated Variables

The primary outcome was a willingness to share wearable device data for research purposes. The secondary outome was a willingness to share wearable device data if compensation is provided. The focus of the analysis was to understand the characteristics associated with a willingness to share wearable device data and those associated with a willingness to share such data if compensation is provided. The surveyed demographic characteristics included age (18-24 years, 25-34 years, 35-44 years, and ≥45 years), sex (male and female), nationality (American [United States of America], Indian, and other), annual income in US dollars (<US $20,000; US $20,000-US $39,999; US $40,000-US $69,999; and ≥US

$70,000), and education (high school diploma or other, some college, a bachelor's degree, and a graduate degree or work). In addition, the surveyed wearable device characteristics included the purpose of use (heartbeat and pulse tracking; sleep tracking; step tracking; and diet-related tracking, that is, the tracking of calories, body fat, or nutrients) and the duration of use (<6 months and ≥6 months). We also tested for a possible association between the willingness to share wearable device data and the average time spent on doing HITs per week (0 to <2 hours, 2 to <4 hours, 4 to <8 hours, 8 to <20 hours, and ≥20 hours).

Statistical Analysis

Bivariate analyses via chi-square tests were used for categorical variables to estimate differences in characteristics between workers with and without a willingness to share wearable device data for research use. Similar analyses were performed for workers who were willing to share such data to estimate differences in characteristics between those who were expecting compensation and those who were not. With regard to multivariable analyses, we performed a logistic regression to evaluate the relationship between a willingness to share wearable device data and the individual characteristics of MTurk workers. A similar model was used to analyze the willingness to share such data if compensation is provided. In addition, we performed decision tree analyses to identify subgroups that have a higher probability to share wearable data for research. All analyses were performed by using R version 3.6.2 (R Foundation for Statistical Computing).

Results

A total of 935 MTurk workers who use wearable devices completed the survey. Almost 90% (827/935, 88.4%) of participants were aged 18 to 44 years, 58.9% (551/935) were male, 64.7% (605/935) were of American nationality, and about 60% (578/935, 61.8%) had been using wearables for more than 6 months (Table 1). Two-thirds of all respondents (615/935, 65.8%) indicated a willingness to share their wearable device data. Nationality, annual income, education, the average time spent on completing HITs per week, the duration of using wearable devices, and the use of wearables that track heartbeat and pulse had a statistically significant association with a willingness to share data without expecting compensation in the univariate analysis (Table 1).

Table 1. Univariate analysis of the association between the willingness to share wearable device data for research and Amazon Mechanical Turk workers' characteristics and between the willingness to share data if compensation is provided and workers' characteristics.

Characteristics	Total (N=935), n (%)	Willingness to share wearable device data			Willingness to share wearable device data if compensation is provided[a]		
		No (n=320), n (%)	Yes (n=615), n (%)	*P* value	No (n=97), n (%)	Yes (n=518), n (%)	*P* value
Sex							
Male	551 (58.9)	188 (58.8)	363 (59)	N/A[b]	62 (63.9)	301 (58.1)	N/A
Female	375 (40.1)	129 (40.3)	246 (40)	.97	33 (34)	213 (41.1)	.27
Other or prefer not to say	9 (1)	3 (0.9)	6 (1)	N/A	2 (2.1)	4 (0.8)	N/A
Age (years)							
18-24	130 (13.9)	52 (16.3)	78 (12.7)	N/A	14 (14.4)	64 (12.4)	N/A
25-34	490 (52.4)	159 (49.7)	331 (53.8)	.59	49 (50.5)	282 (54.4)	.83
35-44	207 (22.1)	70 (21.9)	137 (22.3)	N/A	24 (24.7)	113 (21.8)	N/A
≥45	100 (10.7)	37 (11.6)	63 (10.2)	N/A	10 (10.3)	53 (10.2)	N/A
Prefer not to say	8 (0.9)	2 (0.6)	6 (1)	N/A	0 (0)	6 (1.2)	N/A
Nationality							
American (United States of America)	605 (64.7)	222 (69.4)	383 (62.3)	N/A	45 (46.4)	338 (65.3)	N/A
Indian	210 (22.5)	53 (16.6)	157 (25.5)	.008	43 (44.3)	114 (22)	<.001
Other	120 (12.8)	45 (14.1)	75 (12.2)	N/A	9 (9.3)	66 (12.7)	N/A
Annual income (US $)							
<20,000	226 (24.2)	92 (28.7)	134 (21.8)	N/A	35 (36.1)	99 (19.1)	N/A
20,000-39,999	236 (25.2)	82 (25.6)	154 (25)	.03	20 (20.6)	134 (25.9)	.01
40,000-69,999	276 (29.5)	94 (29.4)	182 (29.6)	N/A	26 (26.8)	156 (30.1)	N/A
≥70,000	194 (20.7)	52 (16.2)	142 (23.1)	N/A	16 (16.5)	126 (24.3)	N/A
Prefer not to say	3 (0.3)	0 (0)	3 (0.5)	N/A	0 (0)	3 (0.6)	N/A
Education							
Graduate degree or work	192 (20.5)	62 (19.4)	130 (21.1)	N/A	19 (19.6)	111 (21.4)	N/A
Bachelor's degree	472 (50.5)	147 (45.9)	325 (52.8)	.04	53 (54.6)	272 (52.5)	.90
Some college	210 (22.5)	83 (25.9)	127 (20.7)	N/A	21 (21.6)	106 (20.5)	N/A
High school diploma or other	61 (6.5)	28 (8.8)	33 (5.4)	N/A	4 (4.1)	29 (5.6)	N/A
Average time spent on doing human intelligence tasks per week (hours)							
0 to <2	120 (12.8)	46 (14.4)	74 (12)	N/A	12 (12.4)	62 (12)	N/A
2 to <4	183 (19.6)	77 (24.1)	106 (17.2)	.04	15 (15.5)	91 (17.6)	.45
4 to <8	225 (24.1)	63 (19.7)	162 (26.3)	N/A	31 (32)	131 (25.3)	N/A
8 to <20	228 (24.4)	75 (23.4)	153 (24.9)	N/A	18 (18.6)	135 (26.1)	N/A
≥20	179 (19.1)	59 (18.4)	120 (19.5)	N/A	21 (21.6)	99 (19.1)	N/A
Duration of using wearable devices (months)							
≤6	357 (38.2)	150 (46.9)	207 (33.7)	N/A	48 (49.5)	159 (30.7)	N/A
>6	578 (61.8)	170 (53.1)	408 (66.3)	<.001	49 (50.5)	359 (69.3)	.001
Purpose of wearable device[c]							
Heartbeat and pulse tracking	446 (47.7)	131 (40.9)	315 (51.2)	.004	49 (50.5)	266 (51.4)	.97

Characteristics	Total (N=935), n (%)	Willingness to share wearable device data			Willingness to share wearable device data if compensation is provided[a]		
		No (n=320), n (%)	Yes (n=615), n (%)	*P* value	No (n=97), n (%)	Yes (n=518), n (%)	*P* value
Sleep tracking	323 (34.5)	103 (32.2)	220 (35.8)	.31	32 (33)	188 (36.3)	.61
Step tracking	648 (69.3)	218 (68.1)	430 (69.9)	.62	54 (55.7)	376 (72.6)	.001
Diet (calories, body fat, and nutrition)	552 (59)	180 (56.2)	372 (60.5)	.24	61 (62.9)	311 (60)	.68

[a]Data are from those who are willing to share data.

[b]N/A: not applicable.

[c]More than 1 option could be selected by the same person.

Among those who were willing to share their wearable device data, 84.2% (518/615) indicated that they were willing to do so if compensation was provided. Nationality, annual income, the duration of using wearable devices, and the use of wearables that track steps had a significant univariate association with a willingness to share data if compensation is provided. A more granular breakdown of the duration of using wearable devices among our study population is shown in Multimedia Appendix 2.

The findings from our logistic regression analyses indicated adjusted associations between the willingness of MTurk workers to share wearable device data and their characteristics. In particular, Indian nationality (odds ratio [OR] 3.09, 95% CI 1.92-5.02); annual incomes of US $20,000-US $39,999, US $40,000-US $69,999, and ≥US $70,000 (OR 1.61, 95% CI 1.04-2.52; OR 1.73, 95% CI 1.12-2.7; OR 2.32, 95% CI 1.4-3.87, respectively); over 6 months of using a wearable (OR 1.74, 95% CI 1.26-2.4); and the use of heartbeat and pulse tracking wearables (OR 1.58, 95% CI 1.17-2.14) were associated with a higher willingness to share data for research use (Table 2).

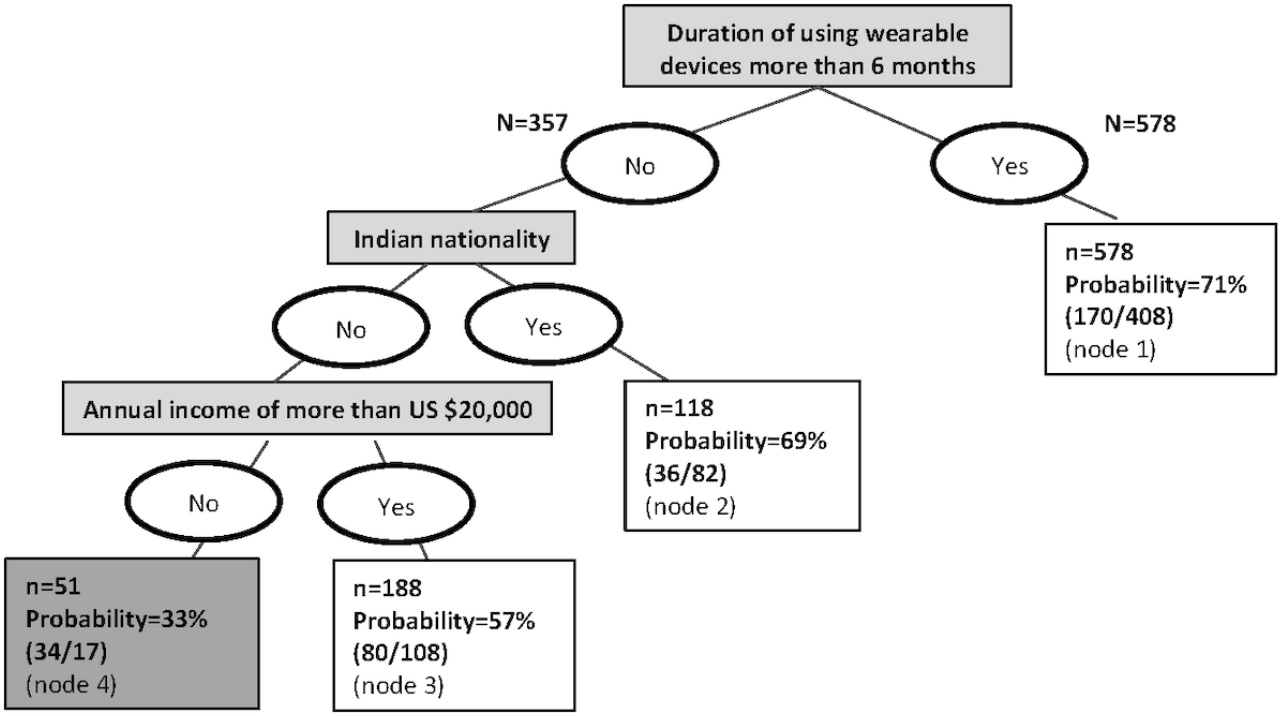
Table 2. Multivariable analysis (logistic regression) of the association between the willingness of Amazon Mechanical Turk workers to share wearable device data for research and their characteristics and between their willingness to receive compensation for donating wearable device data and their characteristics.

Characteristics	Willingness to share wearable device data		Willingness to share wearable device data if compensation is provided	
	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
MTurk workers' characteristics				
Sex				
Male	1.00 (reference)	Reference	1.00 (reference)	Reference
Female	1.10 (0.78-1.43)	.53	1.17 (0.72-1.92)	.53
Age (years)				
18-24	1.00 (reference)	Reference	1.00 (reference)	Reference
25-34	0.99 (0.56-1.42)	.96	1.12 (0.54-2.24)	.75
35-44	0.97 (0.50-1.45)	.91	0.74 (0.33-1.6)	.45
≥45	0.88 (0.37-1.39)	.64	0.78 (0.3-2.1)	.62
Nationality				
American	1.00 (reference)	Reference	1.00 (reference)	Reference
Indian	2.74 (1.48-4.01)	.007	0.47 (0.24-0.9)	.02
Other	1.13 (0.64-1.62)	.61	1.16 (0.54-2.74)	.71
Annual income (US $)				
<20,000	1.00 (reference)	Reference	1.00 (reference)	Reference
20,000-39,999	1.61 (0.93-2.29)	.08	1.72 (0.88-3.43)	.11
40,000-69,999	1.76 (1.02-2.51)	.045	1.31 (0.68-2.55)	.42
≥70,000	2.46 (1.26-3.67)	.02	1.52 (0.71-3.32)	.29
Education				
Graduate degree or work	1.00 (reference)	Reference	1.00 (reference)	Reference
Bachelor's degree	1.31 (0.81-1.81)	.23	0.97 (0.52-1.78)	.93
Some college	1.10 (0.60-1.60)	.69	0.64 (0.29-1.39)	.27
High School diploma or other	0.93 (0.32-1.54)	.82	1.03 (0.31-4.13)	.96
Average time spent on doing human intelligence tasks per week (hours)				
0 to <2	1.00 (reference)	Reference	1.00 (reference)	Reference
2 to <4	0.73 (0.37-1.10)	.15	1.22 (0.51-2.88)	.65
4 to <8	1.41 (0.71-2.11)	.25	0.89 (0.4-1.88)	.76
8 to <20	1.16 (0.59-1.73)	.58	1.61 (0.69-3.66)	.26
≥20	1.16 (0.56-1.76)	.60	1.25 (0.54-2.83)	.59
Health monitoring technologies				
Duration of using wearable devices (months)				
≤6	1.00 (reference)	Reference	1.00 (reference)	Reference
>6	1.75 (1.21-2.29)	.006	1.60 (0.99-2.6)	.06
Purpose of wearable device[a]				
Heartbeat and pulse tracking	1.60 (1.14-2.07)	.01	1.07 (0.67-1.71)	.77
Sleep tracking	1.11 (0.77-1.44)	.54	1.00 (0.62-1.65)	.99
Step tracking	1.17 (0.77-1.57)	.41	1.45 (0.85-2.44)	.17
Diet (calories, body fat, and nutrition)	1.09 (0.77-1.41)	.60	0.96 (0.59-1.57)	.88

[a]More than 1 option could be selected by the same person.

The decision tree analysis identified three characteristics that were significantly associated with the willingness to share wearable data for research and subdivided all participants into 4 segments (nodes). The first characteristic in the tree (highest importance) was the use of health monitoring wearables for more than 6 months. If the response was positive (ie, "yes"; node 1; n=578), these participants had a 71% probability of having a willingness to share wearable data. For participants who indicated that they have been using wearables for 6 months or less, the second question was "what is your nationality?" If they were of Indian nationality (node 2; n=118), then the probability of having a willingness to share wearable data was 69%. If the participant was not Indian, the next question was "what is your annual income?" Individuals with an annual income of more than US $20,000 (node 3; n=188) had a 57% probability of having a willingness to share wearable data. The last segment of participants, who indicated a wearable device use duration of less than 6 months, were not Indian, and had an annual income of less than US $20,000 (node 4; n=51), had a low probability of being willing to share wearable data (33%). The accuracy of the decision tree was 0.684 (Figure 1).

Figure 1. Decision tree analysis for the classification of Amazon Mechanical Turk workers who are willing or not willing to share wearable device data for research purposes (N=935; Accuracy=0.684).



Discussion

Principal Findings

In this study, we explored the characteristics of wearable device users on MTurk that are associated with a willingness to share wearable device data for research use. Our findings show that about two-thirds of individuals (615/935, 65.8%) who used wearables indicated a willingness to share their data for research use. The probability of having a willingness to share wearable device data was higher among individuals who had been using wearables for more than 6 months, individuals with higher incomes, individuals of Indian nationality, and individuals who use wearables for heartbeat and pulse tracking. The majority of those who were willing to share wearable data (518/615, 84.2%) preferred to receive compensation for sharing such data. Further, MTurk workers from India expected compensation significantly less often than American workers (*P*=.02). No other factors were associated with the willingness to share wearable device data if compensation is provided.

Our principal finding is that most MTurk workers (615/935, 65.8%) were willing to share their wearable data for research; however, the results of similar studies based on different populations are mixed. One study of older adults [7] and another study of health and fitness app users [8] found that a majority of participants were willing to share their digital device data. The study with older adults also found an association between income and a willingness to share such data [7]. In another study of patients who were seeking care in an academic emergency department, a minority (40%) of patients indicated a willingness to share wearable device data with researchers [9]. More patients were willing to share other types of data (eg, music streaming data), and a majority were willing to share all surveyed digital data types after death, including wearable device data.

Our finding that a willingness to share wearable device data may differ due to experience with using such devices is similar to the findings of others. For example, a survey found that those with a self-rated low or medium level of expertise with wearables were less willing to share such data [14]. Another survey of health and fitness app users found that "quantified-selfers" were significantly more willing to share their personal data on a public scientific database compared to "non–quantified-selfers" [8]. The relationship between experience and the willingness to share data may be influenced by factors such as a knowledge of data protection laws and

people's comfort with the privacy implications of sharing data. Furthermore, individuals who have more experience with engaging in self-tracking or data sharing activities may be different from individuals with less experience (eg, being more accustomed to sharing data due to previously participating in self-tracking studies or finding value in sharing and discussing their data with others as part of the quantified self movement).

This study is among the few investigating the willingness to share wearable device data if financial compensation is provided. In a review of motivating factors that influence participation in genomic studies, the authors found compensation to be the least important factor [15]. Genomic studies however largely do not involve wearable data donation. In another study that had a sample of individuals that was more similar to ours (ie, a population in which self-tracking was common), the authors found that a majority "probably would" or "definitely would" be willing to share their data and that a majority would be "more" or "much more" willing to share data if they were compensated [16]. Different from that study, we also investigated characteristics that are associated with the willingness to share if compensation is provided and found nationality to be the only statistically significant factor (P=.02).

We also found that people who collect heartbeat and pulse tracking data are more willing to share wearable device data. When considering the two data representation levels (ie, sensor data and derived information [14]) that were present among the data types covered in our survey, heartbeat and pulse tracking data were the only data type at the sensor data level. The other data types were derived information from accelerometors (ie, those for sleep tracking and step tracking) or other sensors (ie, those for tracking calories, body fat, nutrients). This finding might be suggestive of a lower concern with data privacy, given that derived information only makes use of a subset of the available raw data. However, given that we did not make the distinction between raw and derived data in our survey, it is very possible that survey respondents were unaware of the kinds of derived information that can be collected from heartbeat and pulse tracking devices (eg, health status and life expectations). According to the work of Schneegass et al [14], wearable device users' understanding of the relationship between sensor data and the derived information from sensor data is still limited. Thus, when considering mechanisms for increasing people's willingness to share wearable device data, requesting access to derived information may be the most appropriate approach for maximizing the transparency on data that would be studied in the research.

The demographic characteristics of the MTurk workers who enrolled in this study were in close accord with those in another study that was conducted at around the same time [17]. Most of the workers were from the United States (605/935, 64.7% in this study vs the 75% in the work of Difallah et al [17]), the second largest group was from India (210/935, 22.5% in our survey vs the 16% in the work of Difallah et al [17]), and the remaining respondents were from other countries. Males constituted 58.9% (551/935) of the respondents in our study. This is similar to what we have observed among MTurk workers for most countries except the United States [17,18]. The population of MTurk workers who use wearable devices is slightly younger than the general population of MTurk workers. Almost 90% (827/935, 88.4%) of our population were aged 18 to 45 years, whereas 70% of general MTurk workers are aged 18 to 50 years [17]. The age distribution in our study however is similar to that of wearable device users in the United States [19]. The annual income of our population was lower than that of the general US population [20] but was similar to the income of MTurk workers [17]. The median income of the study population was below US $40,000, whereas the median income of MTurk workers is about US $47,000 and the median income of the US general population is about US $60,000.

Limitations

First, the generalizability of MTurk survey responses is a common concern for researchers. Our study found some similarities to and some differences from the general population of wearable device users. Our sample may also have been more driven by compensation than a general study population due to our use of MTurk. Future work might compare the MTurk sample with other types of samples, such as web panels [21] and social media forum users who are not offered compensation. Second, data quality concerns are also common with crowdsourcing studies such as ours. In order to improve the quality of our data, we built a comprehension screening test into our survey. Although it was not explored in this study, high data quality may also be better ensured through the use of sampling strategies that take the experience and reputation of workers into consideration [22]. Third, opinions about a willingness to share wearable device data may differ from actual decisions. This has been observed in a study of the willingness to participate in biobank research [23]. Similar factors that influence participation in biobank research that are less influential in a hypothetical context, such as trust, may play a part in whether data are actually shared for research. Others have found that a willingness to share digital data, for example, may be a function of people's trust in scientific teams [24]. Additionally, informed consent for data donation in situations where individuals may no longer have control over their data after the donation has been completed was not explicitly described to the participants in this study, and this may have influenced their actual willingness. Fourth, the complexity and the multiplicity of influences on wearable data donation could not be fully captured by this study. For example, the demographic characteristics of individuals that were related to a higher probability of having a willingness to share wearable data were high income and Indian nationality. There are likely cultural contexts that are relevant to these factors, such as social norms and the data protection policies of different countries, which influence attitudes toward data donation. Future works that recruit participants by using the MTurk platform should explore other factors that may influence the decision to share data, such as trust; control over what and how data are used; informed consent considerations (eg, broad consent, dynamic consent, etc); and interactions between individual factors and possible broader factors, such as social norms and data protection policies. Furthermore, a better understanding of the extent to which the openness to data sharing translates into actual behaviors exhibited by the broad population of wearable device users will help to guide researchers seeking to recruit

these individuals. A person's comfort with sharing data prior to a data request, for example, may influence the extent to which the hypothetical willingness to share data translates into an actual data donation.

Implications

Crowdsourcing Wearable Device Data Donation

This study summarizes the characteristics of potential wearable users who were recruited from MTurk. Missing descriptions of crowd workers' characteristics in study reports are a frequent issue in health research studies [25]. Although there are other studies that have recruited wearable users from MTurk [11,12], to our knowledge these workers' characteristics have not been examined. Once participants give approval for accessing their wearable device data, mechanisms for granting access, such as on-device permissions (eg, as with Apple HealthKit) and the OAuth 2.0 protocol [26], can be used to allow researchers to remotely access wearable device users' data. Therefore, the crowdsourcing aspect of MTurk can allow researchers to quickly reach a large number and a diversity of study participants [10]. As MTurk is considered further as a source for recruiting study participants who are willing to share wearable device data, the secondary use of such data for research will require the careful consideration of current data protection laws, such as the European Union General Data Protection Regulation and US Health Insurance Portability and Accountability Act.

Enabling Wearable Data Donation

In order to enable wearable device data donation by MTurk workers, technical approaches are needed to obtain permissions for the secondary use of data and for transferring data to research platforms that have the appropriate levels of privacy and security. There are platforms, such as mCerebrum DataKit [27] and Open Humans [28], that have been designed to collect mobile sensor data from multiple sources, including wearable devices. However, there are few devices that support the transfer of data [29]. A recent review of iOS personal health apps showed that only 33% of apps support application programming interfaces—a data access method that allows for the most fine-grained data access [26]. Given the high willingness of MTurk workers to share wearable device data that was identified in our study, future work might enable wearable data access for devices that are commonly used by MTurk workers and establish opt-in data collection services that are compatible with existing platforms.

Conclusions

In this study, we examined MTurk workers' willingness to share wearable device data for research use and the characteristics associated with a willingness to share such data. We found that about two-thirds of wearable device users on MTurk (615/935, 65.8%) indicated a willingness to share their data for research use. Among those who were willing to share such data, most (518/615, 84.2%) indicated that they were willing to share their data if compensation is provided. The probability of having a willingness to share wearable device data was higher among individuals who had used a wearable for more than 6 months, were of Indian nationality, or were of American nationality and had an annual income of more than US $20,000. Overall, our findings are encouraging and should be considered by crowdsourcing research studies that involve wearable device data sharing. Existing platforms for opt-in data collection should be used to achieve this goal.

Acknowledgments

This work was completed at Johns Hopkins University and University of Maryland, College Park. This study is a secondary analysis of the data presented in the master's thesis of SR, which is titled *Survey on Health Device Use by MTurk Participants* [13].

Conflicts of Interest

None declared.

Multimedia Appendix 1
Human intelligence task description and comprehension check question.
[DOCX File , 14 KB - jmir_v23i10e19789_app1.docx]

Multimedia Appendix 2
The duration of wearable device use among Amazon Mechanical Turk survey participants.
[DOCX File , 22 KB - jmir_v23i10e19789_app2.docx]

References

1. Kaewkannate K, Kim S. A comparison of wearable fitness devices. BMC Public Health 2016 May 24;16:433 [FREE Full text] [doi: 10.1186/s12889-016-3059-0] [Medline: 27220855]
2. Casselman J, Onopa N, Khansa L. Wearable healthcare: Lessons from the past and a peek into the future. Telematics and Informatics 2017 Nov;34(7):1011-1023. [doi: 10.1016/j.tele.2017.04.011]
3. Higgins JP. Smartphone applications for patients' health and fitness. Am J Med 2016 Jan 17;129(1):11-19. [doi: 10.1016/j.amjmed.2015.05.038] [Medline: 26091764]

4. Fawcett E, Van Velthoven MH, Meinert E. Long-term weight management using wearable technology in overweight and obese adults: Systematic review. JMIR Mhealth Uhealth 2020 Mar 10;8(3):e13461 [FREE Full text] [doi: 10.2196/13461] [Medline: 32154788]
5. Steinhubl SR, McGovern P, Dylan J, Topol EJ. The digitised clinical trial. Lancet 2017 Nov 11;390(10108):2135. [doi: 10.1016/S0140-6736(17)32741-1] [Medline: 29143747]
6. Arvidsson D, Fridolfsson J, Börjesson M. Measurement of physical activity in clinical practice using accelerometers. J Intern Med 2019 Aug;286(2):137-153 [FREE Full text] [doi: 10.1111/joim.12908] [Medline: 30993807]
7. Seifert A, Christen M, Martin M. Willingness of older adults to share mobile health data with researchers. GeroPsych (Bern) 2018 Mar;31(1):41-49. [doi: 10.1024/1662-9647/a000181]
8. Chen J, Bauman A, Allman-Farinelli M. A study to determine the most popular lifestyle smartphone applications and willingness of the public to share their personal data for health research. Telemed J E Health 2016 Aug;22(8):655-665. [doi: 10.1089/tmj.2015.0159] [Medline: 26958742]
9. Seltzer E, Goldshear J, Guntuku SC, Grande D, Asch DA, Klinger EV, et al. Patients' willingness to share digital health and non-health data for research: a cross-sectional study. BMC Med Inform Decis Mak 2019 Aug 08;19(1):157 [FREE Full text] [doi: 10.1186/s12911-019-0886-9] [Medline: 31395102]
10. Mason W, Suri S. Conducting behavioral research on Amazon's Mechanical Turk. Behav Res Methods 2012 Mar;44(1):1-23. [doi: 10.3758/s13428-011-0124-6] [Medline: 21717266]
11. Epstein DA, Kang JH, Pina LR, Fogarty J, Munson SA. Reconsidering the device in the drawer: Lapses as a design opportunity in personal informatics. Proc ACM Int Conf Ubiquitous Comput 2016 Sep 12;2016:829-840 [FREE Full text] [doi: 10.1145/2971648.2971656] [Medline: 28516173]
12. Epstein DA, Caraway M, Johnston C, Ping A, Fogarty J, Munson SA. Beyond abandonment to next steps: Understanding and designing for life after personal informatics tool use. Proc SIGCHI Conf Hum Factor Comput Syst 2016 May;2016:1109-1113 [FREE Full text] [doi: 10.1145/2858036.2858045] [Medline: 28503678]
13. Ramesh S. Survey on health device use by MTurk participants. Digital Repository at the University of Maryland. 2018. URL: http://hdl.handle.net/1903/21073 [accessed 2021-09-23]
14. Schneegass S, Poguntke R, Machulla T. Understanding the impact of information representation on willingness to share information. 2019 May Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland, United Kingdom. [doi: 10.1145/3290605.3300753]
15. Goodman D, Bowen D, Wenzel L, Tehrani P, Fernando F, Khacheryan A, et al. The research participant perspective related to the conduct of genomic cohort studies: A systematic review of the quantitative literature. Transl Behav Med 2018 Jan 29;8(1):119-129 [FREE Full text] [doi: 10.1093/tbm/ibx056] [Medline: 29385589]
16. Bietz M, Bloss C, Calvert S, Godino J, Gregory J, Claffey M, et al. Opportunities and challenges in the use of personal health data for health research. J Am Med Inform Assoc 2016 Apr;23(e1):e42-e48 [FREE Full text] [doi: 10.1093/jamia/ocv118] [Medline: 26335984]
17. Difallah D, Filatova E, Ipeirotis P. Demographics and dynamics of Mechanical Turk workers. 2018 Feb Presented at: WSDM 2018: The Eleventh ACM International Conference on Web Search and Data Mining; February 5-9, 2018; Marina del Rey, California, USA. [doi: 10.1145/3159652.3159661]
18. Mechanical Turk surveys. MTurk Tracker. URL: http://demographics.mturk-tracker.com/#/gender/us [accessed 2021-09-17]
19. Forrester Research (Analytics). Smart wearable users in the United States 2020-2024, by generation (in millions). Statista. URL: https://www.statista.com/statistics/1202075/smart-wearable-users-generation-united-states/ [accessed 2021-09-23]
20. U.S. Census Bureau QuickFacts: United States. United States Census Bureau. URL: https://www.census.gov/quickfacts/fact/table/US/PST045218 [accessed 2021-09-23]
21. Baker R, Blumberg SJ, Brick JM, Couper MP, Courtright M, Dennis JM, et al. Research Synthesis: AAPOR Report on Online Panels. Public Opin Q 2010 Oct 20;74(4):711-781 [FREE Full text] [doi: 10.1093/poq/nfq048]
22. Robinson J, Rosenzweig C, Moss AJ, Litman L. Tapped out or barely tapped? Recommendations for how to harness the vast and largely unused potential of the Mechanical Turk participant pool. PLoS One 2019 Dec 16;14(12):e0226394. [doi: 10.1371/journal.pone.0226394] [Medline: 31841534]
23. Johnsson L, Helgesson G, Rafnar T, Halldorsdottir I, Chia K, Eriksson S, et al. Hypothetical and factual willingness to participate in biobank research. Eur J Hum Genet 2010 Nov;18(11):1261-1264 [FREE Full text] [doi: 10.1038/ejhg.2010.106] [Medline: 20648060]
24. Pratap A, Allred R, Duffy J, Rivera D, Lee HS, Renn BN, et al. Contemporary views of research participant willingness to participate and share digital data in biomedical research. JAMA Netw Open 2019 Nov 01;2(11):e1915717 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.15717] [Medline: 31747031]
25. Créquit P, Mansouri G, Benchoufi M, Vivot A, Ravaud P. Mapping of crowdsourcing in health: Systematic review. J Med Internet Res 2018 May 15;20(5):e187 [FREE Full text] [doi: 10.2196/jmir.9330] [Medline: 29764795]
26. Kim Y, Lee B, Choe EK. Investigating data accessibility of personal health apps. J Am Med Inform Assoc 2019 May 01;26(5):412-419 [FREE Full text] [doi: 10.1093/jamia/ocz003] [Medline: 30861531]

27. Hossain SM, Hnat T, Saleheen N, Nasrin NJ, Noor J, Ho BJ, et al. mCerebrum: A mobile sensing software platform for development and validation of digital biomarkers and interventions. Proc Int Conf Embed Netw Sens Syst 2017 Nov;2017:7 [FREE Full text] [doi: 10.1145/3131672.3131694] [Medline: 30288504]
28. Tzovaras BG, Angrist M, Arvai K, Dulaney M, Estrada-Galiñanes V, Gunderson B, et al. Open Humans: A platform for participant-centered research and personal data exploration. Gigascience 2019 Jun 01;8(6):giz076 [FREE Full text] [doi: 10.1093/gigascience/giz076] [Medline: 31241153]
29. Muzny M, Henriksen A, Giordanengo A, Muzik J, Grøttland A, Blixgård H, et al. Wearable sensors with possibilities for data exchange: Analyzing status and needs of different actors in mobile health monitoring systems. Int J Med Inform 2020 Jan;133:104017. [doi: 10.1016/j.ijmedinf.2019.104017] [Medline: 31778885]

Abbreviations

HIT: human intelligence task
MTurk: Amazon Mechanical Turk
OR: odds ratio

Edited by R Kukafka; submitted 01.05.20; peer-reviewed by S Munson, S Park; comments to author 21.07.20; revised version received 22.02.21; accepted 12.09.21; published 21.10.21.

Please cite as:
Taylor CO, Flaks-Manov N, Ramesh S, Choe EK
Willingness to Share Wearable Device Data for Research Among Mechanical Turk Workers: Web-Based Survey Study
J Med Internet Res 2021;23(10):e19789
URL: https://www.jmir.org/2021/10/e19789
doi:10.2196/19789
PMID:34673528

Original Paper

A Human, Organization, and Technology Perspective on Patients' Experiences of a Chat-Based and Automated Medical History–Taking Service in Primary Health Care: Interview Study Among Primary Care Patients

Evalill Nilsson[1*], PhD; Annette Sverker[2*], PhD; Preben Bendtsen[3*], MD, PhD; Ann Catrine Eldh[4,5*], PhD, RN

[1]eHealth Institute, Department of Medicine and Optometry, Linnaeus University, Kalmar, Sweden

[2]Department of Activity and Health and Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden

[3]Department of Medical Specialists in Motala and Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden

[4]Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden

[5]Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden

[*]all authors contributed equally

Corresponding Author:
Evalill Nilsson, PhD
eHealth Institute
Department of Medicine and Optometry
Linnaeus University
Hus Vita, Universitetskajen
Kalmar, 391 82
Sweden
Phone: 46 730377913
Email: evalill.nilsson@lnu.se

Abstract

Background: The use of e-visits in health care is progressing rapidly worldwide. To date, studies on the advantages and disadvantages of e-consultations in the form of chat services for all inquiries in primary care have focused on the perspective of health care professionals (HCPs) rather than those of end users (patients).

Objective: This study aims to explore patients' experiences using a chat-based and automated medical history–taking service in regular, tax-based, not-for-profit primary care in Sweden.

Methods: Overall, 25 individual interviews were conducted with patients in the catchment areas of 5 primary care centers (PCCs) in Sweden that tested a chat-based and automated medical history–taking service for all types of patient inquiries. The semistructured interviews were transcribed verbatim before content analysis using inductive and deductive strategies, the latter including an unconstrained matrix of human, organization, and technology perspectives.

Results: The service provided an easily managed way for patients to make written contact with HCPs, which was considered beneficial for some patients and issues but less suitable for others (acute or more complex cases). The automated medical history–taking service was perceived as having potential but still derived from what HCPs need to know and how they address and communicate health and health care issues. Technical skills were not considered as necessary for a mobile phone chat as for handling a computer; however, patients still expressed concern for people with less digital literacy. The opportunity to take one's time and reflect on one's situation before answering questions from the HCPs was found to reduce stress and prevent errors, and patients speculated that it might be the same for the HCPs on the other end of the system. Patients appreciated the ability to have a conversation from almost anywhere, even from places not suitable for telephone calls. The asynchronicity of the chat service allowed the patients to take more control of the conversation and initiate a chat at any time at their own convenience; however, it could also lead to lengthy conversations where a single issue in the worst cases could take days to close. The opportunity to upload photographs made some visits to the PCC redundant, which would otherwise have been necessary if the ordinary telephone service had been used, saving patients both time and money.

Conclusions: Patients generally had a positive attitude toward e-visits in primary care and were generally pleased with the prospects of the digital tool tested, somewhat more with the actual chat than with the automated history-taking system preceding

the chat. Although patients expect their PCC to offer a range of different means of communication, the human, organization, and technology analysis revealed a need for more extensive (end) user experience design in the further development of the chat service.

(J Med Internet Res 2021;23(10):e29868) doi:10.2196/29868

KEYWORDS

digital encounter; digital healthcare; e-consultation; e-health; interview; patient perspective; primary healthcare; qualitative study; telemedicine; telehealth

Introduction

e-Visits in Swedish Primary Care

The digitalization of health care is now progressing rapidly, not least because of the current COVID-19 pandemic, when client-to-provider telemedicine (also known as digital, virtual, or e-consultations; digital encounters; or e-visits) may ensure access to health care at times of social distancing [1,2]. Such e-visits may include videoconferencing; emailing; text messaging via web portals; SMS text messaging; and, more recently, chat services, predominantly designed for mobile devices (especially mobile phones) with the intention of a quick reply.

Following a change in the legislation to increase access to primary care in Sweden, for-profit medical companies with web-based services only have offered chat services since 2014 for most patient inquiries in primary care for the entire population of Sweden, regardless of which primary care center (PCC) the patient actually belongs to. Although controversial from a reimbursement perspective (as an invoice will be sent to the patient's regular PCC for each of such e-visits) and criticized for initiating certain treatments without a proper physical examination, these services have become increasingly popular for patients over the years. Subsequently, chat services are emerging also in regular tax-based, not-for-profit primary care, and contrary to the web-based medical services described earlier, the regular primary care offers the opportunity to convert chat conversations into clinical visits [3]. Despite the growing use of these chat services by patients, earlier interview studies have focused on the perspectives of health care professionals (HCPs) [4-7]. To our knowledge, before our study, no reports on patients' experiences of chat services for all kinds of inquiries in regular primary care had been published, from neither Sweden nor elsewhere.

Automated Medical History–Taking

e-Visit services often include automated medical history taking, as do the aforementioned chat services in Swedish primary care. Automated medical history taking can improve the quality of care and increase patient satisfaction; patients can document their medical history without any interference or time constraints, and HCPs can prepare for consultations in a time-efficient manner [8,9]. Earlier studies from primary care have suggested many benefits of using automated medical history taking, such as enabling triage prioritization and increasing diagnostic precision [10]. Potential risks include erroneous information, loss of nonverbal cues, and low digital literacy. As long as the questions are carefully designed to describe their health status accurately, patients using automated medical history taking in emergency care have reported that it helps them to organize their thoughts and thus enables better dialog with their physician [11].

Human, Organization, and Technology Perspectives on e-Visits

The rapid introduction of new and innovative digital communication tools in health care, often without taking enough time to consider implementation strategies from a theoretical perspective, has highlighted the need for thorough evaluations. Studying the implementation of e-visits from the three perspectives of human, organization, and technology (HOT) in parallel and given equal attention can render a more holistic evaluation [12-15]. Human factors include aspects such as user satisfaction; organizational factors cover aspects such as structure; and technological factors include aspects such as information and system quality. The more these aspects fit with each other, the greater the potential. The greatest risks are found in the intersections and interactions among them, and therefore, it is important for successful implementation to identify and rectify any gaps.

The aim of this study is to explore patients' experiences of using a chat-based and automated medical history–taking service in Swedish, regular, tax-based, not-for-profit primary care.

Methods

Design

An explorative design formed this study, with data collected through semistructured individual interviews and analyzed using content analysis.

Setting

Health care in Sweden is regionalized and managed by 21 independent county councils. In the county of Östergötland in southeastern Sweden, a chat-based service (provided by a commercial company) within primary care was tested by 5 PCCs for about a year, starting from April and May 2019. These PCCs were representative of the region, and Sweden, including both urban and rural areas. Normally, patients who need to consult their PCC for various health-related requests would contact their PCC through a telephone triage system, where they register their telephone number and are assigned a specific point in time later in the day when they will be contacted by a nurse at the PCC for triage, that is, Swedish patients are not accustomed to being put in touch directly with their personal physician, nurse, or physiotherapist, when contacting their PCC. When all available time slots for talking to the triage nurse are taken for

a particular day, patient callings are advised automatically by the system to try again the following day.

During the test period, adult patients calling their PCC between 8 AM and 3 PM, Monday to Friday (weekdays), were offered the use of a digital chat-based communication system as an alternative to the telephone triage. A voice message specified the alternative, offering the choice of receiving a link to the digital service via a text message on their mobile phone, along with a guaranteed response in the chat from the PCC within 2 hours. It was also possible to access the service through the websites of the PCCs. To log in to the system, the patient used a common secure personal identification system. Before the actual chat, an automated medical history–taking service was offered, where the patient responded to a battery of automatically generated questions about their chief complaint and current health status. The questions were rule-based, although response adaptive, and included both fixed and free response alternatives.

Sample

Patients logging in to the chat service in September to November 2019 were asked if they were willing to participate in the evaluation study. If so, they ticked a box giving the researchers permission to contact them with more information about the study. During this period, nearly 600 patients ticked the box and 25 of them were selected (5 from each PCC) in stratified purposeful sampling, to ensure participation of both men and women of all ages. In addition, age- and sex-matching reserves were selected if patients had changed their minds or otherwise were unable to participate in the study. In total, 35 patients had to be invited as 10 declined for various or unknown reasons. The demographics of the participants are presented in Table 1.

Table 1. Demographics of the participants (N=25).

Demographics	Total (N=25)	Women (n=13)	Men (n=12)
Age (years), mean (SD; range)	51 (19.0; 21-81)	48 (19.3; 21-81)	53 (18.4; 22-76)
Living in urban areas, n (%)	19 (76)	9 (69)	10 (83)
Manage your own electronic devices,[a] n (%)			
Always or almost always	23 (92)	11 (84)	12 (100)
Most of the time	1 (4)	1 (8)	0 (0)
Some of the time	1 (4)	1 (8)	0 (0)

[a]Nobody chose the response alternative *never or almost never.*

Procedure

Data Collection

The selected patients were contacted via the mobile phone number registered while using the chat service. Information about the study was given orally and also sent by mail or email to those who agreed to participate in the study. A suitable time for the telephone interview was agreed upon, according to the availability of the patient.

The interviews were conducted by 2 of the researchers in the team: author AS conducted 12 interviews and a trained research associate (Catharina Linderoth) conducted the other 13 (as further noted in *Acknowledgments*), supported by an agreed semistructured interview guide compiled for this study. The guide comprised three areas: (1) digital communication (in general), (2) experience of the digital tool in question, and (3) digital patient-professional relationship. Probes were available but were used only if the respondents did not naturally expand on the subjects. Four final demographic questions were asked: (1) age, (2) sex, (3) whether the respondents considered themselves living in a rural or urban area, and (4) how confident the respondent was in managing problems with their electronic devices (Table 1). The interview guide was tested and validated in a pilot study at the beginning of the test period (April-May 2019), indicating that it was comprehensive and corresponded to the study purpose. Before each interview began, the participant received a recap of the study information before the participant's informed consent, which was recorded. All interview recordings were transcribed verbatim by a skilled secretarial service. The interviews lasted between 8 and 31 minutes, with a mean of 21 (SD 5) minutes. The transcripts rendered 153 pages of one-and-a-half-spaced text. No data were reported back to the participants or PCCs.

Data Analysis

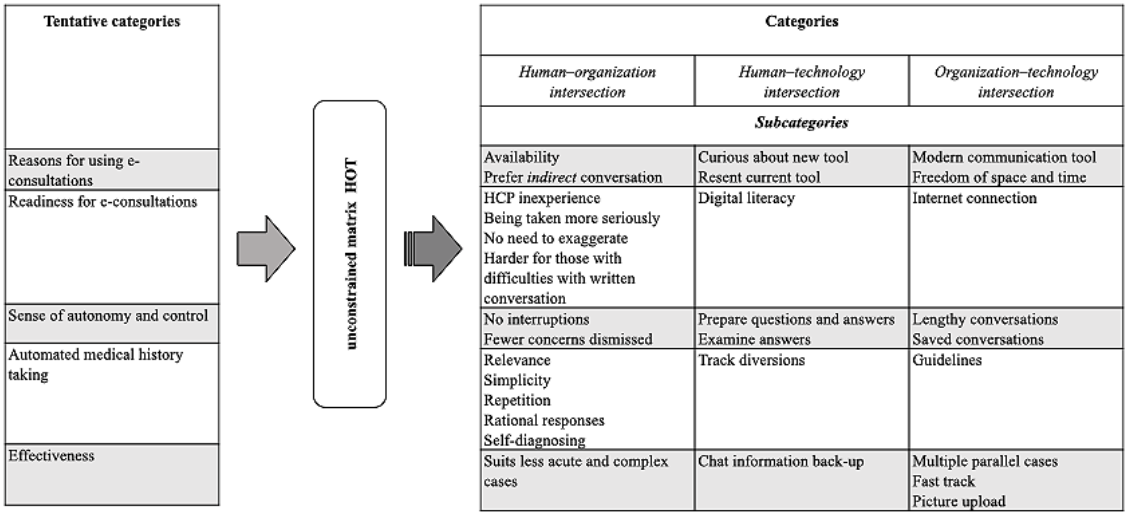
All transcripts were read separately by all authors in an inductive manner and everyone then presented, in writing, their individual general understanding of the data set, per interview and as a whole. A subsequent team discussion informed a common understanding of tentative categories. To advance the structured analysis, the elements of HOT were considered by the team, generating an unconstrained matrix, as described by Elo and Kyngäs [16]. To conclude, the relationships among the elements were mastered and quotations illuminating the findings were identified across the data set [17].

Results

Overview

Patients' experiences of the chat-based and automated medical history–taking service signify intersections among HOT aspects. An overview of the categories identified in correspondence with all three HOT elements is presented in Figure 1, along with the intersections among the three elements, which are described in more detail later.

Figure 1. The forming of categories by means of the unconstrained matrix: tentative categories analyzed from an HOT perspective and subcategories informing the intersection among the three elements (human, organization, and technology). HCP: health care professional; HOT: human, organization, and technology.



The Human-Organization Intersection

Patients trying out this new digital communication tool thought it would be a quicker way to get in contact with their PCC. A more specific reason was the preference for written communication, for example, because talking to other people induces anxiety, especially if the subject is health related, the issue is sensitive or embarrassing to raise face-to-face, or simply because it suits them better than oral communication:

> *When I need to get in contact with healthcare it induces anxiety, which makes it very difficult for me to pick up the phone and make the call. This makes me postpone it for far too long and therefore this chat service has made things much easier for me.* [Interview #13]

However, the patients raised concerns for other patients with reading and writing difficulties or vision impairment. Furthermore, digital chatting was not considered an optimal way to communicate regarding emergencies and severe illnesses or when a deeper and more intimate conversation was required.

Regarding the automated medical history–taking service, patients found the questions fairly easy to understand, but they were uncertain about the accuracy of their own responses, wondering which information would be vital and which would be redundant. In the absence of clear guidelines, patients began diagnosing themselves by trying to answer the questions. Furthermore, from their point of view, patients were both lacking relevant and receiving irrelevant and repetitive questions, some of which pertained to issues that the patients felt sure were already documented in their electronic health record (EHR):

> *When you talk [on the phone] to the nurse or caregiver they will ask follow-up questions, so that you understand better what it's all about. To express yourself verbally and not have the right knowledge behind it could easily become difficult.* [Interview #04]

In the subsequent chat, patients were never interrupted, and their concerns were less easily dismissed than in telephone calls; rather, they sensed that they were taken more seriously in the chat. Therefore, they did not feel the need to exaggerate their symptoms to get an appointment. The medical assessment was mainly regarded as fair and the advice as constructive. However, as some personnel seemed to be more skilled than others, the patients made suggestions regarding further training in written communication and digital chatting for HCPs:

> *In the chat you can say all that you want to say without anyone hanging up or interrupting or making something up just to dismiss you. So, all information will be put forward and they will read it and have to take it more seriously than over the phone.* [Interview #06]

Overall, data show that, regarding the human-organization interaction, patients expected the organization to be more prepared for chatting than turned out to be the case.

The Human-Technology Intersection

The technical standard of digital services is important to patients. As the patients expressed some negativity regarding the current telephone triage system, there was an openness to new tools. Common curiosity with new technical solutions was a reason for trying out the new service. Although all participants considered themselves able to manage most everyday problems with their electronic devices, technical skills were not considered necessary for the chat in the same way as for handling, for example, one's computer. Yet, patients conveyed concerns for people with less digital literacy, such as equity concerns about the service not aiding older people properly, as older people as a group are the least digitally literate group in society.

Although patients appreciated the potential benefits of an automated history-taking system, they expressed a sense of unease in answering the questions in certain ways, leading to undesired consequences, such as selecting a response alternative that guided them away from the line of questioning that they felt was relevant, with no opportunity to retract. The *back-up*, that is the possibility in the subsequent chat for correcting any mistakes from the automated history-taking procedure, was considered reassuring but ineffective, duplicating efforts.

The opportunity for patients to take their time and reflect on the situation before answering questions from HCPs was stress reducing, error preventing, and facilitated follow-ups. Besides being less stressful for themselves, patients contemplated how chatting would be less stressful also for the HCPs for the same reasons:

> *Writing everything down means I can check if I have forgotten anything. I can list things and add text before I send it, but when you talk to people you might forget things, important things, and when we are done talking, I remember, oh, I should have mentioned this and that!* [Interview #09]

The Organization-Technology Intersection

As web-based chatting in general nowadays is a common service in Sweden, in commercial and public sectors, patients suggested that people have become accustomed to this kind of customer-centric digital service. The patients anticipated that their usual HCPs would offer similar services. The chat permitted patients to have a conversation from almost anywhere, even from places not suitable for telephone calls, such as at work or on a bus. However, although most other chat services in society are synchronous, the chat in this study was asynchronous, which was disappointing to some patients. Although asynchronicity allows patients to take more control over the conversation and initiate a chat at any time at their own convenience, it could also lead to lengthy conversations where a single concern in the worst case could take days to close. This meant that health care personnel at times had been replaced by others in the chat, leaving the patient with a sense of having to start all over again. However, at the same time, patients reported that an asynchronous chat service supposedly provided efficiency, with personnel being able to handle several cases simultaneously. A chat service is expected to be a faster route of communication than, for example, a telephone service; however, this was not always the case, which patients attributed partly to the asynchronous structure of the chat and partly to staff's inexperience with the system, anticipating that the latter would change and improve over time:

> *I think the difficulty lies in the inactivity in the chat. Because when they write a question and I respond and it takes them hours to answer they will have lost the flow, because they have been doing many other things in-between. It felt like they neither had the energy nor the time to go back in the text flow, which meant they had to start all over again, to some extent.* [Interview #19]

The opportunity to upload photographs made some visits to the PCC redundant that would otherwise have been necessary if a telephone service had been used, saving both time and money of patients. However, an adequate internet connection is required, not least when uploading photographs, which was pointed out as an obstacle:

> *To be able to upload pictures of, for example, wounds and eczemas and stuff like that, that makes it [the chat service] nothing but positive.* [Interview #16]

The fact that technology allows for the (whole) conversation to be saved word for word, also in the EHR, for future reference was described as a new aspect to take into account when expressing oneself in the chat contacts; the patients were used to the HCPs summarizing the conversation in the EHR by extracting only vital aspects and using concise and professional language. Furthermore, improvements such as a fast track for simple cases and guidelines in the system for answering the questions were found to be desirable:

> *It is possible to follow up from both sides, they can see what I have written and they can see what others that work there have answered as well. The follow-up is better than with telephone conversations.* [Interview #18]

In summary, patients were predominantly positive about the chat service and felt it added value, even emotionally:

> *I felt welcomed, and that is not always the case.* [Interview #23]

Discussion

Principal Findings

The results from this study indicate that patients would welcome the chat service becoming permanent, as one of the several ways of contacting their PCC. However, there are several aspects from the HOT perspective that must be considered before implementation across primary care.

A health care organization will have to balance the desire to provide access and simplicity for patients with the obligation to provide healthy and safe working conditions for staff and uphold data privacy, security, and high-quality care (*human-organization*). Compared with the earlier study from our research group regarding the perspectives of HCPs, patients in this study seemed to be somewhat more satisfied with the chat service than the HCPs were [4]. For example, although anxious to offer patients ample availability, HCPs reported that patients used the chat service to gain quicker access to health care regardless of the degree of urgency of their health problems and sometimes initiated contact for the same complaint through several channels in parallel, which may add to the workload of primary care. As the chat was asynchronous, it was not always the fast route that patients anticipated, and perhaps the service should be renamed and the term *chat* should be reserved for synchronous services.

Both patients in this study and HCPs in the earlier study [4] regarded the chat as being more on the patients' terms than traditional face-to-face or telephone communication, although

patients, by comparison, seemed to find that as more positive than did the HCPs. Although no earlier interview studies were found with patients regarding modern and fast, primarily mobile phone based, chat services for all inquiries in primary care, there are several studies (both interview and questionnaire studies) regarding patients' views on other text messaging services in primary care, for inquiries in general, such as in this study, and for the self-management of chronic diseases or health-related interventions, where the patient-professional contact has already been established. The results of these studies are consistent with the findings of this study regarding patients' views on the appropriateness and availability of e-visits and the pros and cons of written communication [18-23].

How patients are received by HCPs is the most common grievance when assessing health care from a patient safety perspective [24]. Although chatting may diminish problems such as interruptions and not getting enough time to express yourself, new problems seem to emerge instead, such as HCPs inexperience with professional chat communication about health and health care issues. Interestingly, regarding the chat service in this study, this issue was only raised by patients and not by the HCPs themselves [4]. One might expect patients' opinions to change when the service has settled after having been operational for some time; however, in fact, patients' views in this study were unexpectedly similar to those in interviews conducted during the pilot study in May 2019 (data not yet published). Further studies are needed to examine if patient safety is affected by this line of communication and if training of HCPs in written communication and chatting would be beneficial [25,26].

The automated medical history–taking service is meant to provide health care personnel decision support and faster triage and management. This requires that the questions are relevant and easy both to understand and respond to. Verbal anamnesis, or medical history taking, is usually documented by HCPs asking the patient questions and follow-up questions, guiding the story, and immediately correcting mutual errors and misunderstandings. The automated medical history–taking service in this study seemed to be instigated by the needs of the organization rather than by the needs of the patient. Naturally, the choice of questions must have a medical basis, but patients would have much to add, for example, in which cases questions are perceived as hard to respond to, irrelevant, or inexplicably repetitive and when questions of relevance to them seem to be missing. The health care organization cannot simply assume that questions that are obvious to trained professionals will be automatically understood by patients and should not underestimate the importance of patients having the opportunity to convey all their concerns as they see fit. Patients' suggestions about a more pronounced user experience design approach when developing automated medical history–taking systems should be given consideration (*human-technology)* [18].

An adequate internet connection was considered an important prerequisite for using the new digital communication tool, but it is a societal rather than merely a health care concern. On the other hand, patient privacy and data security are very much health care concerns (*technology-organization*) [27]. Contrary to early studies from the beginning of the 21st century about patient-provider contacts via email in primary care, for example, where patients raised concerns about security and lost messages [28], patients in this study did not raise any cybersecurity or legal issues during the interviews. This may be interpreted as a solid trust in the system or a lack of awareness of the risks of digital communication, either in general or at the time of the interview. Trust has been shown to have a positive correlation with the intention to use e-consultation services [29]; if so, respondents lacking trust most likely would not consider trying the service in the first place and, therefore, were not available for this study with its current design.

Earlier studies have shown that organizational elements are crucial for the successful implementation of health information technology [30]. In this case, the chat service was trialed following only a limited training session for the staff engaged (organized by the commercial company providing the digital tool and only regarding the technical and administrative aspects of the tool, not regarding how to create a therapeutic alliance with patients using digital tools) [4] and very limited information to the patients. Furthermore, aspects such as cost-effectiveness will impact the implementation process. For example, the intention was to staff the service so that the person responsible for the chat could concentrate on that task and not have to alternate with other tasks. However, because of limited resources, this was not possible at all PCCs, as the influx of cases was not enough to warrant the higher cost of assigning a nurse solely to the chat service for all workdays (8 hours a day) [4]. Thus, the stakeholder's perspective was, to a certain extent, missing in both the creation and implementation of the chat service, and therefore evaluation from this perspective is even more important. Findings from this study indicate that patients provide perspectives of interest to further planning and implementation, confirming that the organization, technology, and human and social aspects need to be considered throughout the implementation of health technology innovations [12,30].

Methodological Discussion

Engaging patients in quality improvement is a suggested standard procedure, and the end user perspective is vital for any health care development [31]. However, a limitation of this study is that the participants may constitute a sample of people more used to computers and mobile electronic devices than others; having a mobile phone, tablet, or computer was a technical prerequisite for using the chat service. To explore the technical interface in a wider context, respondents should represent those less used to electronic devices as well. However, digital illiteracy may not be the major reason for not choosing to engage in a digital encounter; nonusers often report simply preferring to meet and speak with their health care providers [28,32]. Nevertheless, the novelty of this study is its inclusion of a variety of people with firsthand experience as patients using a digital chat with their primary care provider regarding all possible inquiries; thus, it may be of interest for similar contexts and further investigations of tools enabling digital communication [33].

Conclusions

Patients in general had a positive attitude toward e-visits in primary care and were generally pleased with the prospects of

the digital tool, somewhat more with the chat service than with the automated history-taking system preceding the chat. Although patients expect their PCC to offer a range of different means of communication, the HOT analysis revealed a need for more extensive (end) user experience design in the further development of the PCC chat service.

Acknowledgments

This study was funded by Region Östergötland and approved by the Swedish Ethical Review Authority (2019-04123). The authors are very grateful to project manager Emma Stockman (Region Östergötland) for assistance with details regarding the implementation of the primary care center chat service.

Authors' Contributions

EN, AS, PB, and ACE developed the study plan; EN initiated the interviews, and AS conducted them together with research associate Catharina Linderoth (Linköping University), to whom the authors would like to extend their sincere gratitude. EN led the analysis, and all authors contributed to the naïve analysis and repeated discussions. EN drafted the manuscript. All coauthors made substantial contributions in discrete sections and to the overall text, including agreeing to the final version.

Conflicts of Interest

None declared.

References

1. Mold F, Cooke D, Ip A, Roy P, Denton S, Armes J. COVID-19 and beyond: virtual consultations in primary care-reflecting on the evidence base for implementation and ensuring reach: commentary article. BMJ Health Care Inform 2021 Jan;28(1):e100256 [FREE Full text] [doi: 10.1136/bmjhci-2020-100256] [Medline: 33436372]
2. Doraiswamy S, Abraham A, Mamtani R, Cheema S. Use of telehealth during the COVID-19 pandemic: scoping review. J Med Internet Res 2020 Dec 01;22(12):e24087 [FREE Full text] [doi: 10.2196/24087] [Medline: 33147166]
3. Ekman B, Thulesius H, Wilkens J, Lindgren A, Cronberg O, Arvidsson E. Utilization of digital primary care in Sweden: descriptive analysis of claims data on demographics, socioeconomics, and diagnoses. Int J Med Inform 2019 Jul;127:134-140. [doi: 10.1016/j.ijmedinf.2019.04.016] [Medline: 31128825]
4. Eldh AC, Sverker A, Bendtsen P, Nilsson E. Health care professionals' experience of a digital tool for patient exchange, anamnesis, and triage in primary care: qualitative study. JMIR Hum Factors 2020 Dec 14;7(4):e21698 [FREE Full text] [doi: 10.2196/21698] [Medline: 33315014]
5. Cajander Å, Larusdottir M, Hedström G. The effects of automation of a patient-centric service in primary care on the work engagement and exhaustion of nurses. Qual User Exp 2020 Sep 19;5:9. [doi: 10.1007/s41233-020-00038-x]
6. Entezarjou A, Bolmsjö BB, Calling S, Midlöv P, Milos Nymberg V. Experiences of digital communication with automated patient interviews and asynchronous chat in Swedish primary care: a qualitative study. BMJ Open 2020 Jul 23;10(7):e036585 [FREE Full text] [doi: 10.1136/bmjopen-2019-036585] [Medline: 32709650]
7. Johansson A, Larsson M, Ivarsson B. General practitioners' experiences of digital written patient dialogues: a pilot study using a mixed method. J Prim Care Community Health 2020;11:2150132720909656 [FREE Full text] [doi: 10.1177/2150132720909656] [Medline: 32133905]
8. Zakim D. Development and significance of automated history-taking software for clinical medicine, clinical research and basic medical science. J Intern Med 2016 Sep 12;280(3):287-299 [FREE Full text] [doi: 10.1111/joim.12509] [Medline: 27071980]
9. Banks J, Farr M, Salisbury C, Bernard E, Northstone K, Edwards H, et al. Use of an electronic consultation system in primary care: a qualitative interview study. Br J Gen Pract 2018 Jan;68(666):e1-e8 [FREE Full text] [doi: 10.3399/bjgp17X693509] [Medline: 29109115]
10. Jimenez G, Tyagi S, Osman T, Spinazze P, van der Kleij R, Chavannes NH, et al. Improving the primary care consultation for diabetes and depression through digital medical interview assistant systems: narrative review. J Med Internet Res 2020 Aug 28;22(8):e18109 [FREE Full text] [doi: 10.2196/18109] [Medline: 32663144]
11. Arora S, Goldberg AD, Menchine M. Patient impression and satisfaction of a self-administered, automated medical history-taking device in the Emergency Department. West J Emerg Med 2014 Feb;15(1):35-40 [FREE Full text] [doi: 10.5811/westjem.2013.2.11498] [Medline: 24695871]
12. Yusof MM, Kuljis J, Papazafeiropoulou A, Stergioulas LK. An evaluation framework for Health Information Systems: human, organization and technology-fit factors (HOT-fit). Int J Med Inform 2008 Jul;77(6):386-398. [doi: 10.1016/j.ijmedinf.2007.08.011] [Medline: 17964851]
13. Ericson L, Hammar T, Schönström N, Petersson G. Stakeholder consensus on the purpose of clinical evaluation of electronic health records is required. Health Policy Technol 2017 Jun;6(2):152-160. [doi: 10.1016/j.hlpt.2017.02.005]

14. Saparova D, Belden J, Williams J, Richardson B, Schuster K. Evaluating a federated medical search engine: tailoring the methodology and reporting the evaluation outcomes. Appl Clin Inform 2014 Aug 13;5(3):731-745 [FREE Full text] [doi: 10.4338/ACI-2014-03-RA-0021] [Medline: 25298813]
15. Kilsdonk E, Peute LW, Jaspers MW. Factors influencing implementation success of guideline-based clinical decision support systems: a systematic review and gaps analysis. Int J Med Inform 2017 Feb;98:56-64. [doi: 10.1016/j.ijmedinf.2016.12.001] [Medline: 28034413]
16. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
17. Eldh AC, Årestedt L, Berterö C. Quotations in qualitative studies: reflections on constituents, custom, and purpose. Int J Qual Methods 2020 Jan 01;19. [doi: 10.1177/1609406920969268]
18. Farr M, Banks J, Edwards HB, Northstone K, Bernard E, Salisbury C, et al. Implementing online consultations in primary care: a mixed-method evaluation extending normalisation process theory through service co-production. BMJ Open 2018 Mar 19;8(3):e019966 [FREE Full text] [doi: 10.1136/bmjopen-2017-019966] [Medline: 29555817]
19. Cowie J, Calveley E, Bowers G, Bowers J. Evaluation of a digital consultation and self-care advice tool in primary care: a multi-methods study. Int J Environ Res Public Health 2018 May 02;15(5):896 [FREE Full text] [doi: 10.3390/ijerph15050896] [Medline: 29724040]
20. Katz SJ, Nissan N, Moyer CA. Crossing the digital divide: evaluating online communication between patients and their providers. Am J Manag Care 2004 Sep;10(9):593-598 [FREE Full text] [Medline: 15515991]
21. Grønning A, Assing Hvidt E, Nisbeth Brøgger M, Fage-Butler A. How do patients and general practitioners in Denmark perceive the communicative advantages and disadvantages of access via email consultations? A media-theoretical qualitative study. BMJ Open 2020 Oct 26;10(10):e039442 [FREE Full text] [doi: 10.1136/bmjopen-2020-039442] [Medline: 33109668]
22. Edwards HB, Marques E, Hollingworth W, Horwood J, Farr M, Bernard E, et al. Use of a primary care online consultation system, by whom, when and why: evaluation of a pilot observational study in 36 general practices in South West England. BMJ Open 2017 Dec 22;7(11):e016901 [FREE Full text] [doi: 10.1136/bmjopen-2017-016901] [Medline: 29167106]
23. Chang F, Paramsothy T, Roche M, Gupta NS. Patient, staff, and clinician perspectives on implementing electronic communications in an interdisciplinary rural family health practice. Prim Health Care Res Dev 2017 Mar;18(2):149-160. [doi: 10.1017/S1463423616000414] [Medline: 27995826]
24. Reader TW, Gillespie A, Roberts J. Patient complaints in healthcare systems: a systematic review and coding taxonomy. BMJ Qual Saf 2014 Aug;23(8):678-689 [FREE Full text] [doi: 10.1136/bmjqs-2013-002437] [Medline: 24876289]
25. Konttila J, Siira H, Kyngäs H, Lahtinen M, Elo S, Kääriäinen M, et al. Healthcare professionals' competence in digitalisation: a systematic review. J Clin Nurs 2019 Mar;28(5-6):745-761. [doi: 10.1111/jocn.14710] [Medline: 30376199]
26. Alpert JM, Dyer KE, Lafata JE. Patient-centered communication in digital medical encounters. Patient Educ Couns 2017 Oct;100(10):1852-1858 [FREE Full text] [doi: 10.1016/j.pec.2017.04.019] [Medline: 28522229]
27. Elwyn G, Barr PJ, Piper S. Digital clinical encounters. BMJ 2018 May 14;361:k2061. [doi: 10.1136/bmj.k2061] [Medline: 29760149]
28. Moyer CA, Stern DT, Dobias KS, Cox DT, Katz SJ. Bridging the electronic divide: patient and provider perspectives on e-mail communication in primary care. Am J Manag Care 2002 May;8(5):427-433 [FREE Full text] [Medline: 12019595]
29. Li J, Liu M, Liu X, Ma L. Why and when do patients use e-consultation services? The trust and resource supplementary perspectives. Telemed J E Health 2018 Jan;24(1):77-85. [doi: 10.1089/tmj.2016.0268] [Medline: 28686084]
30. Cresswell K, Sheikh A. Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review. Int J Med Inform 2013 May;82(5):e73-e86. [doi: 10.1016/j.ijmedinf.2012.10.007] [Medline: 23146626]
31. Bombard Y, Baker GR, Orlando E, Fancott C, Bhatia P, Casalino S, et al. Engaging patients to improve quality of care: a systematic review. Implement Sci 2018 Jul 26;13(1):98 [FREE Full text] [doi: 10.1186/s13012-018-0784-z] [Medline: 30045735]
32. Nijland N, van Gemert-Pijnen JE, Boer H, Steehouder MF, Seydel ER. Increasing the use of e-consultation in primary care: results of an online survey among non-users of e-consultation. Int J Med Inform 2009 Oct;78(10):688-703. [doi: 10.1016/j.ijmedinf.2009.06.002] [Medline: 19625210]
33. Polit DF, Beck CT. Generalization in quantitative and qualitative research: myths and strategies. Int J Nurs Stud 2010 Nov;47(11):1451-1458. [doi: 10.1016/j.ijnurstu.2010.06.004] [Medline: 20598692]

Abbreviations

EHR: electronic health record
HCP: health care professional
HOT: human, organization, and technology
PCC: primary care center

Edited by R Kukafka; submitted 23.04.21; peer-reviewed by A Entezarjou, R Ciorap; comments to author 28.06.21; revised version received 02.07.21; accepted 27.07.21; published 18.10.21.

Please cite as:
Nilsson E, Sverker A, Bendtsen P, Eldh AC
A Human, Organization, and Technology Perspective on Patients' Experiences of a Chat-Based and Automated Medical History–Taking Service in Primary Health Care: Interview Study Among Primary Care Patients
J Med Internet Res 2021;23(10):e29868
URL: https://www.jmir.org/2021/10/e29868
doi:10.2196/29868
PMID:34661544

Original Paper

Adoption of Machine Learning Systems for Medical Diagnostics in Clinics: Qualitative Interview Study

Luisa Pumplun[1], MSc; Mariska Fecho[1], MSc; Nihal Wahl[1], Dr rer pol; Felix Peters[1], MSc; Peter Buxmann[1], Prof Dr

Software & Digital Business Group, Technical University of Darmstadt, Darmstadt, Germany

Corresponding Author:
Luisa Pumplun, MSc
Software & Digital Business Group
Technical University of Darmstadt
Hochschulstraße 1
Darmstadt, 64289
Germany
Phone: 49 6151 16 24221
Email: luisa.pumplun@tu-darmstadt.de

Abstract

Background: Recently, machine learning (ML) has been transforming our daily lives by enabling intelligent voice assistants, personalized support for purchase decisions, and efficient credit card fraud detection. In addition to its everyday applications, ML holds the potential to improve medicine as well, especially with regard to diagnostics in clinics. In a world characterized by population growth, demographic change, and the global COVID-19 pandemic, ML systems offer the opportunity to make diagnostics more effective and efficient, leading to a high interest of clinics in such systems. However, despite the high potential of ML, only a few ML systems have been deployed in clinics yet, as their adoption process differs significantly from the integration of prior health information technologies given the specific characteristics of ML.

Objective: This study aims to explore the factors that influence the adoption process of ML systems for medical diagnostics in clinics to foster the adoption of these systems in clinics. Furthermore, this study provides insight into how these factors can be used to determine the ML maturity score of clinics, which can be applied by practitioners to measure the clinic status quo in the adoption process of ML systems.

Methods: To gain more insight into the adoption process of ML systems for medical diagnostics in clinics, we conducted a qualitative study by interviewing 22 selected medical experts from clinics and their suppliers with profound knowledge in the field of ML. We used a semistructured interview guideline, asked open-ended questions, and transcribed the interviews verbatim. To analyze the transcripts, we first used a content analysis approach based on the health care–specific framework of nonadoption, abandonment, scale-up, spread, and sustainability. Then, we drew on the results of the content analysis to create a maturity model for ML adoption in clinics according to an established development process.

Results: With the help of the interviews, we were able to identify 13 ML-specific factors that influence the adoption process of ML systems in clinics. We categorized these factors according to 7 domains that form a holistic ML adoption framework for clinics. In addition, we created an applicable maturity model that could help practitioners assess their current state in the ML adoption process.

Conclusions: Many clinics still face major problems in adopting ML systems for medical diagnostics; thus, they do not benefit from the potential of these systems. Therefore, both the ML adoption framework and the maturity model for ML systems in clinics can not only guide future research that seeks to explore the promises and challenges associated with ML systems in a medical setting but also be a practical reference point for clinicians.

(J Med Internet Res 2021;23(10):e29301) doi:10.2196/29301

KEYWORDS

machine learning; clinics; diagnostics; adoption; maturity model

Introduction

Machine Learning Systems for Medical Diagnostics

The ongoing digitalization is influencing the everyday activities of almost every individual, both in their private and professional lives. This transformation is particularly evident in health care, where the integration of health information technologies (HITs), such as electronic health records or clinical decision support systems, enables significant improvements in processes such as emergency medical care, diagnostics, and therapy [1-3]. However, the integration of HITs is not a panacea but leads to major challenges in clinics as, fueled by these technologies, physicians have to handle an ever-growing volume of patient data and complexity of interacting systems [4]. Moreover, societal problems further complicate the provision of health services to the population, as age-related diseases are on the rise because of demographic shifts and global pandemics such as the COVID-19 crisis are overburdening clinics, pushing medical personnel to the limits of their capacity [5,6].

Artificial intelligence (AI) as the "science and engineering of making intelligent machines, especially intelligent computer programs" [7] could help relieve this burden on physicians as AI is capable of solving tasks previously reserved for human intelligence [8]. In particular, machine learning (ML), as a subfield of AI, is currently one of the fastest growing technological approaches, opening up a wide range of possibilities for medicine [9,10]. Therefore, in the remainder of this research work, we focus on ML systems, that is, information systems (IS) that learn to perform certain tasks autonomously through experience without receiving explicit human instructions. Instead, ML systems use algorithms to search large amounts of data for patterns to create their own rules and strategies on how to deal with a particular problem. The identified rules can then be applied to solve a task [9,11-13]. ML systems can be particularly useful in solving problems for which the rules are difficult to derive and express. This is the case, for example, in image recognition; for instance, how can the image of a cat be explained in terms of pixels, what shapes of ears are allowed, and how can they be recognized in a picture [13]. From the prediction of patient admissions in clinics to therapy support, ML systems can help solve various problems in medicine [10,14]. However, one application area of particular value to researchers and practitioners in which ML systems could have a major impact on the overall well-being of the population is medical diagnostics [15,16]. In this context, ML systems can help identify patterns in medical data (eg, in medical scans, pathology slides, electrocardiograms, and written diagnoses) and sort possible conditions according to their likelihood [17,18]. A distinction can be made between ML serving to take over entire areas of responsibility from physicians and supporting them in their decision-making process. In the near future, ML systems will mainly be used as intelligent decision support rather than to automate medical diagnostics fully [10,17,19,20]. Thus, current cases in research and practice show that an increasing number of such assistive ML systems are presently finding their way into medical workflows. For example, ML systems are being developed, refined, and deployed to help in the early diagnosis of COVID-19 based on entered symptoms or medical images such as computed tomography scans and algorithms such as deep convolutional neural networks [21]. These systems raise the hope of making medical diagnostics of COVID-19 and also other diseases faster, more efficient, and consistent, and thus more valuable as they are able to compare patient data with a database that is larger than any physician's experience. Consequently, applying ML systems in patient care could make the difference between life and death by enabling more effective and efficient diagnostics [10,17].

Challenges of Adopting Machine Learning Systems in Clinics

However, despite this enormous promise, the integration of ML systems also poses challenges that have prevented the widespread adoption of these systems in clinics to date [22]. More specifically, clinics cannot draw on their experience from adopting other HITs, as ML differs substantially from prior technologies. Specifically, ML systems learn from high volumes of data instead of being explicitly programmed [12]. Although traditional clinical decision support systems rely on rule-based systems that produce deterministic outputs, ML systems derive their solutions based on complex statistical methods, leading to several consequences. First, ML systems are becoming increasingly complex and commonly resemble black boxes; that is, their mechanisms for generating predictions are opaque to humans. For example, ML systems based on deep neural networks make predictions using millions of parameters, and humans cannot comprehend each and every calculation. Second, ML systems that learn from data will almost never be able to perform tasks perfectly, for example, make classifications with 100% accuracy [11,19]. This is mainly because of the ML system reliance on statistical patterns, which will never be able to cover all edge cases. Third, the operationalization of ML systems in practice is challenging, largely because complex relationships between different types of artifacts (eg, data sets, models, and source codes) have to be managed [23]. Whereas traditional clinical decision support systems rely on human-defined rules that are instantiated in software code, ML systems are a result of applying algorithms to data, thus creating an additional dependency. All artifacts have to be versioned, and their dependencies must be tracked to comply with regulations and ensure reproducibility. Owing to these complicating factors, organizations in various industries struggle to integrate ML systems into their processes. Therefore, initial research is looking at the challenges that ML systems pose in terms of organizational adoption [24-27]. However, clinics differ considerably from other organizations, as they not only possess unique structures, management processes, and requirements for HIT adoption but are also responsible for their patients' lives [28]. In these medical settings, the characteristics of ML systems are particularly problematic as physicians and patients rely on profound diagnoses and the correct functionality of ML systems at any time [19]. Consistent with the call of Davison and Martinson [29] for more context-specific research, studies regarding the adoption of ML systems in clinics must, therefore, reflect on both, the specific characteristics of ML systems and clinics. Such context-specific research on the organizational adoption of ML systems in clinics is becoming more prevalent

in recent times [10,30,31]. Thematically, researchers mainly investigate the individual acceptance of physicians [19,31] and the technical specifics of ML systems, such as their lack of transparency [32,33]. However, the problem with existing research is that most of these publications are merely reviews and rely on the personal understanding and experience of the authors. Rare exceptions are, for example, Hofmann et al [34], Sandhu et al [31], and Sun and Medaglia 35], who made use of qualitative research methods. Hofmann et al [34] examined the opportunities and challenges of ML systems in radiology, whereas Sandhu et al [31] and Sun and Medaglia [35] studied the introduction of 2 specific ML-based diagnostic decision support systems in clinics. Although these publications already offer a first insight into the possible factors along the adoption process of ML systems, they are not sufficient to understand the process in its entirety.

Objectives and Research Approach

In particular, to our knowledge, no work exists that theoretically embeds the organizational adoption process of ML systems in clinics and presents it based on empirical evidence. Rather, current research focuses on individual acceptance criteria instead of taking a holistic, organizational perspective [19,31]. Therefore, clinics lack an integral overview of the requirements that ML systems imply and that they need to address to harness the potential of these systems for their diagnostic processes. Guided by the call of Shaw et al [10] for more research on the adoption of ML systems in clinics and the lack of prior integral research, our study thus aims to answer the following first research question: which specific factors influence the adoption process of ML systems in medical diagnostics?

Moreover, previous research does not elaborate on how these factors may manifest in a range of different stages and how these stages determine an overarching maturity score. However, such a maturity model could shed further light on the adoption process of ML systems in clinics by providing an empirically grounded and operationalized construct to measure adoption progress [36,37]. Therefore, the maturity model could not only be applied in future empirical research but also allow clinics to assess their as-is situation and evaluate potential courses of action for ML adoption. Therefore, our research sets out to investigate the following second research question: how can the identified factors be used to establish a maturity model for the adoption process of ML systems in clinics?

To answer these research questions, we conducted a qualitative study based on explorative interviews (N=22) with experts working for clinics or suppliers of clinics. To structure the key findings of our empirical investigation, we referred to the health care–specific framework of nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) for a conceptual basis [38]. Although this adoption framework provides a foundation, it is not sufficient to represent the full adoption process of ML systems in clinics, given the particular characteristics of ML systems. To provide a more context-specific framework [29], we drew on qualitative data to gradually adapt and expand the existing framework by several factors specific to the adoption process of ML systems for clinical diagnostics. Moreover, we used qualitative data to develop a maturity model that can help researchers and clinicians understand the possible range of ML adoption stages in clinics and determine an overarching maturity score. Overall, we aim to provide a practical reference point for clinicians to integrate ML systems more effectively into their diagnostic processes.

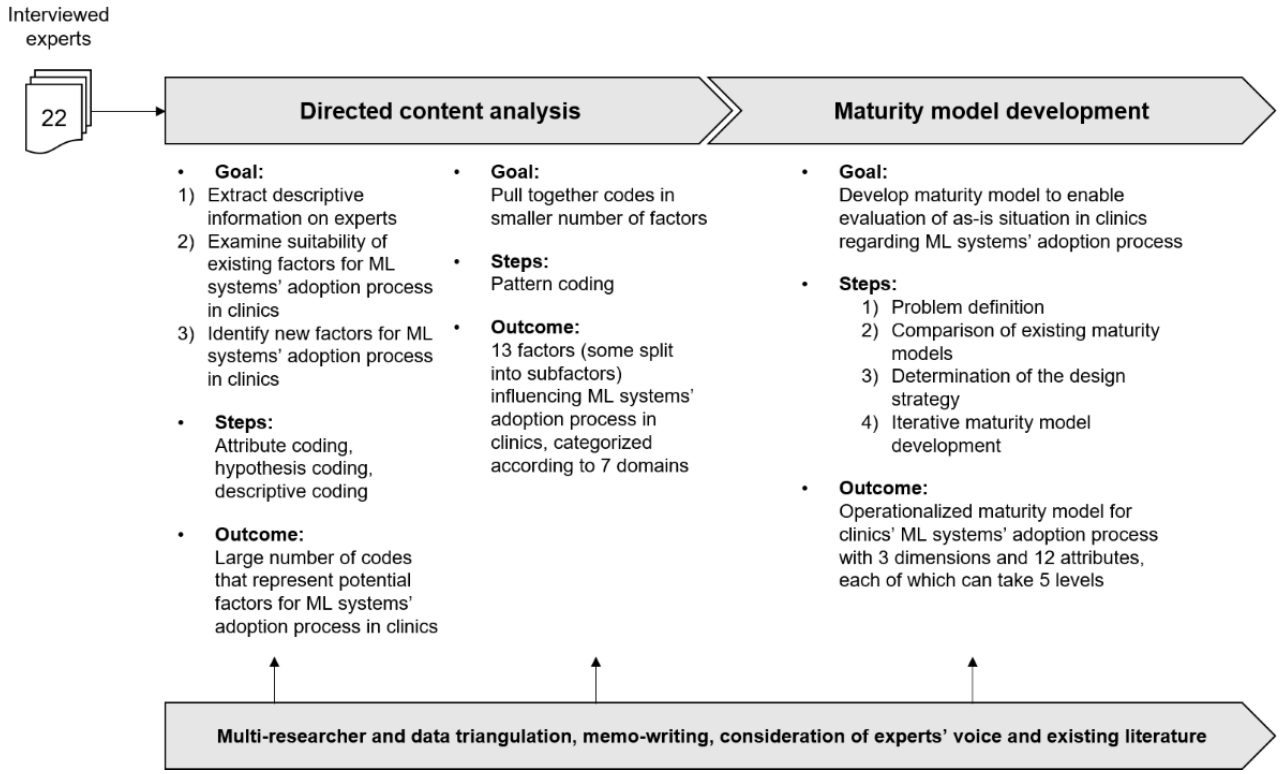
In the next section, we describe our qualitative research design, introduce directed content analysis as our basic data analysis methodology, and explain the development process of the ML maturity model in detail. We then present the empirical results of our study to provide a valuable basis for further research and guidance to clinics aiming to integrate ML systems within their diagnostic processes. Finally, we conclude by discussing the theoretical and practical implications of our study and showing perspectives for future research.

Methods

Overview

Qualitative data provide a rich source of information that can help to better understand emerging, highly complex research subjects [39]. Therefore, to understand the complex adoption process of ML systems and derive a maturity model, we used a qualitative approach to “see the world through the eyes of the people being studied” [39]. In this regard, we applied the key informant method and conducted in-depth interviews with experts (N=22) who have particular qualifications and specialized knowledge on the topic investigated [40]. We led these interviews according to a semistructured interview guideline to ensure that all relevant questions were posed. The questionnaire included general questions about the person, questions about previous knowledge in the field of ML systems, the assessment of potentials and challenges of ML systems for medicine, and further, more detailed questions about the prerequisites in clinics to adopt ML systems for diagnostics. Owing to the qualitative approach, we kept the guideline open and flexible to allow adaptations to the respective interviewed expert, their position, and knowledge base [41]. We analyzed the qualitative data with the help of *directed content analysis* [42] and the methodological approach for *maturity model development* [36]. For an overview of the research procedure, please refer to Figure 1.

Figure 1. Overview of research procedure, illustration based on Jöhnk et al [25]. ML: machine learning.



During the research process, we used several practices to obtain rigor and trustworthiness. To begin with, we defined 2 clear research questions and a conceptual framework that we used as input for our research design. Furthermore, we followed a theoretical sampling approach by iterating between data collection and analysis until we reached theoretical saturation [43]. In this way, we drew on the results from preceding interviews to select further experts and, for example, interviewed not only physicians and managers from clinics but also managers from HIT suppliers to obtain a more holistic perspective. In this regard, considering suppliers allowed us to gain an external, less biased perspective on the adoption of ML systems in clinics. Therefore, we found the additional supplier perspective to be particularly useful in triangulating the data and increasing the validity of our findings [44]. Moreover, different medical disciplines were considered in the interviews (eg, radiology, pathology, and internal medicine) to allow for different perspectives on medical diagnostic processes (eg, interpretation of medical scans, pathology slides, and electrocardiograms) and obtain more generalizable results [45]. The resulting number of interviews is comparable with those of other qualitative studies in IS health care research [31,34,46,47]. With regard to data analysis, we followed a structured and reproducible approach to evaluate the qualitative data [36,42]. During this whole process, a multiresearcher triangulation took place to include different perspectives on the research topic [44]. In that sense, we discussed all data analysis steps and results intensively with the authors and with further qualified researchers from the fields of IS, computer science, and medicine. We recorded the results of these discussions in the form of memos to make them available in the following analysis stages [48]. For later documentation of the results, we decided to include "the voice of participants" [49] and thus quote directly from the interviews while presenting our findings. Where possible, we have additionally incorporated existing—so far scattered—literature that backs up and contextualizes particular statements made by interviewed experts, thus demonstrating the relevance of the findings from the interviews [25].

Data Collection and Sample Selection

Qualitative data were collected in 2 rounds. We conducted a first round of in-depth interviews from the second to the last quarter of 2019. This round of interviews included most participants (15/22, 68% of experts) and formed the basis for content analysis and maturity model development. However, the adoption of ML systems in clinics has progressed significantly in recent times. Therefore, we conducted a further round of interviews (7/22, 32% of experts) in the first quarter of 2021 to capture potential new insights from clinics on the research subject. Moreover, we shared the identified factors and the complete operationalized maturity model with the second-round interview participants to verify and refine the findings from the first panel. All the interviews were conducted in 2 European countries (Germany and Switzerland).

To identify suitable participants for both rounds of interviews, we searched for experts in professional networks, clinic websites, and at relevant conferences on ML in medicine. We interviewed qualified experts, who had detailed knowledge of clinical processes, had profound experience with ML systems, and were involved in the respective decision-making processes [50]. Of the 22 interviewed experts, 5 (23%) were physicians, 8 (36%) held a hybrid position (ie, physicians with additional leadership responsibilities), and 9 (41) worked as full-time managers or information technology staff in the medical field. The participants worked for 11 different clinics and 5 HIT suppliers. Four clinics are privately financed, and the others are

public, providing a view of both privately and publicly funded clinics. All clinics and suppliers are currently running projects related to ML. On average, each expert interview lasted 48 minutes and took place in a private space. The interviews were audio recorded and transcribed after mutual agreement. In 3 interviews, we only took notes as the participants did not consent to recording. For an overview of the experts, see Table 1.

Table 1. Overview of interviewed experts.

ID	Position	Specialty	Expertise (years)
Clinics: key informants of clinics			
C-01	Physician	Radiology	3
C-02	Physician	Radiology	15
C-03	Physician	Radiology	8
C-04	Physician	Cardiology	3
C-05	Physician	Neuroradiology	3
C-06	Physician[a]	Neuroradiology	9
C-07	Physician[a]	Internal medicine	19
C-08	Physician[a]	Internal medicine	35
C-09	Physician[a]	Pathology	18
C-10	Physician[a]	Radiology	37
C-11	Physician[a]	Gynecology	40
C-12	Physician[a]	Otolaryngology	25
C-13	Physician[a]	Cardiology	12
C-14	Chief technology officer	Cardiology	8
C-15	Chief technology officer	Biomedicine	20
C-16	Director	Internal medicine	12
Health information technology (HIT) suppliers: key informants of clinics' HIT suppliers			
S-01	Director	Nephrology	20
S-02	Director	Biomedicine	22
S-03	Director	Genetics	10
S-04	Head of research and development	Radiology	2
S-05	System-engineer	Pathology	3
S-06	Innovation project lead	Surgery	3

[a]Physician with leadership responsibilities.

Directed Content Analysis

Our first goal was to identify the factors that are specific to the adoption process of ML systems in clinics and are not yet sufficiently covered by existing theories. As ML systems have an innovative character because of their novel, complex technical characteristics, we followed the steps of directed content analysis to extend existing theory on the adoption of innovations [42].

The process of adopting innovations in organizations is an overarching process that evolves from initial awareness of technology to a solidified interest and a subsequent adoption decision, to its implementation in the organization, and finally to continued adoption [51]. Presently, adoption research regarding HITs has started to look beyond the mere awareness of a technology to include the later stages of the adoption process [38]. In this context, ML systems own highly specific characteristics that will necessitate a significant change in the organization structure and working routines eventually [11,19]. Therefore, the whole adoption process of ML systems should be considered thoroughly. To capture this, we used the NASSS framework as a conceptual basis. NASSS has primarily been developed for the health care context by combining established health and social care frameworks and can be used to analyze the full adoption process of an HIT, including the implementation phase and continued adoption of the technology. It includes several *domains*, namely *technology* and its features, the *organization* that aims to adopt the *technology*, the *wider system* of an *organization*, the *condition* to be diagnosed and treated, the demand and supply side *value proposition* associated with HIT, and the *adopter system* consisting of patients, their

relatives, and medical staff. Furthermore, it explicitly conceptualizes the *embedding* and *adaptation* of the HIT within a clinic over time [38]. Each domain, in turn, comprises several *factors* that specify the domain considered. These are, for example, the regulatory issues related to a technology (wider system) or the value a technology can have for a patient (value proposition). The suitability of the NASSS framework for the topic under study is evidenced by recent research calling for the use of the framework for empirical work on the adoption process of ML systems in clinics [30]. The NASSS framework forms the basis for our research but is insufficient to explain the specific adoption process of *ML systems* in clinics and, therefore, needs to be reconsidered. In this regard, we used the framework as a starting point, and it was adapted and expanded, taking into account the qualitative data [42].

Specifically, we applied an iterative multicycle coding process that is in line with directed content analysis, which consists of 2 coding cycles, between which we moved back and forth [52]. The first cycle comprised 3 different types of coding. Using *attribute coding* enabled us to receive descriptive information concerning the participant. *Hypothesis coding* was used to consider the prespecified conceptual framework (ie, NASSS) and to examine the suitability of existing domains and factors regarding the adoption process (eg, domain: value proposition; factor: patients' value through ML). In contrast, the *descriptive coding* approach allowed us to identify new aspects that go beyond the conceptual framework by disregarding formerly identified domains and factors. As the coding procedure during the first cycle has led to a large number of constructs, we used *pattern coding* within the second coding cycle to pull together the codes into a smaller number of factors [52]. We performed the analysis using the NVivo 12 (QSR International) software. The result of the analysis is a holistic overview of domains, factors, and subfactors that influence the adoption process of ML systems for diagnostics (see section *Factors Influencing the Adoption Process of ML Systems in Clinics*).

Maturity Model Development

In a further step of our data analysis, we aimed to use (a subset of) the factors identified during content analysis to create a maturity model that can help clinics to assess their current state in the ML system adoption process. Organizations can have different maturities with regard to the management of technologies. To determine the maturity score of an organization regarding a certain type of technology, specified maturity (assessment) models can be used [36]. These models constitute an instrument for organizations to "measure and assess domain capabilities at a given point in time" [53]. In this context, maturity models are valuable tools for organizations to assess and document their as-is state and, based on this, achieve directions for transformation and prioritization of potential investments [36,54]. Therefore, a maturity model comprises different *dimensions* that are subdivided according to specific *attributes*, each of which can take different *maturity levels*. Dimensions represent capability areas, for example, in the field of technology management, that should be exhaustive and distinct from each other. Attributes further specify these dimensions and represent practices, activities, or measures that can be taken by the organization and contribute to an organization's maturity. Levels, on the other hand, are archetypal degrees of maturity which are often represented as a 5-step sequence of stages expressed by different labels [36,55-57]. Becker et al [36] differentiated 5 levels, namely, (1) *initial*, (2) *assessing*, (3) *determined*, (4) *managed*, and (5) *optimizing*. The descriptions characterizing these levels may vary depending on the level definitions and the subject of investigation. However, in general, an attribute is considered to be at an *initial* (1) level if the processes investigated are still in their infancy, chaotic, and not consciously controlled by the organization, whereas the most advanced level *optimized* (5) stands for those attributes whose processes are already actively and continuously improved with the help of standardized feedback mechanisms [55,58]. The overall maturity score of the organization, which can take one of the 5 levels described, results from the compilation of the individual attribute levels.

In recent years, maturity models have made their way into the health care sector. A literature review conducted by Carvalho et al [59] showed that clinical researchers and practitioners have established and applied various specified maturity models to understand and evaluate the integration of different HITs. However, there are no studies in the existing literature or insights from practice on a specific maturity model related to ML systems in clinics. To create a new maturity model for the ML adoption process in clinics, we followed the systematic development process outlined by Becker et al [36], which is loosely based on the design science methodology of Hevner et al [60]. This methodological approach includes 4 steps that structure the development of maturity models and 4 more that accompany the application of maturity models in practice. As our primary goal was to create a maturity model for the adoption process of ML systems in clinics rather than the subsequent application of the model in clinical practice, we focused primarily on the first 4 steps.

The first step of the maturity model development process by Becker et al [36] is to define the problem underlying maturity development. The aim of this study was to provide researchers and clinics with the opportunity to evaluate the clinic status quo in the adoption process of ML systems. As clinics still struggle to integrate ML systems into their processes, we consider this problem particularly relevant and topical [22]. After defining the problem domain and the target group, we searched for existing maturity models from adjacent research fields. In particular, we identified 3 maturity models that, although not specific to clinics, are drawn from the field of AI: the *artificial intelligence maturity model* by Alsheibani et al [61], the *five maturity levels of managing AI* by Lichtenthaler [62], and the *machine learning maturity framework* established by Akkiraju et al [63]. All of them use a 5-level maturity scale ranging from an *initial* (1) level to *optimized* or *integrated* (5). Although the framework by Akkiraju et al [63] was strongly technically oriented, Alsheibani et al [61] and Lichtenthaler [62] incorporated a management perspective as well. Although the identified maturity models helped provide a structure for the model to be built (eg, levels and potential attributes) and specific wordings that could be used (eg, "no data exist to train AI" [61]), no model is complete in itself or tailored to clinics. As clinics are highly specific in their structures and processes [28],

we took initial ideas from the existing models but widely supplemented and concretized these ideas with the help of the content analysis results. In particular, we designed a new maturity model that is specific to ML adoption in clinics, but which incorporates some basic structures and descriptions from existing models. In the following core step, the actual development of the maturity model takes place. We adopted an iterative approach that included 4 substeps: design-level selection, approach selection, model design, and testing. In total, 3 iterations were performed to develop the maturity model. In the first iteration, the existing maturity models and the results of the directed content analysis were considered to build a basic concept. In the second iteration, additional researchers from the field of IS and computer science were brought in to discuss and optimize the maturity model. In the third round, the maturity model was shared, discussed, and tested with 8 of the medical experts [36]. Within these iterations, we decided to adopt a multidimensional maturity assessment based on the results of the previously conducted content analysis. In particular, a subset of 3 domains was used for the dimensions of the maturity model; the corresponding factors or subfactors form 12 attributes that further specify these dimensions. Thereby, only those domains and factors were selected that clinics can modify themselves and are not set by external forces that are beyond the clinics' reach (eg, from the wider system). The resulting attributes were then populated with individual-level descriptions using the qualitative interview data. Therefore, we started with the 2 extreme levels *initial* (1) and *optimized* (5) for each attribute, and the formulations for the levels in between were derived from the interview data, the existing maturity models and literature, or logical inference. The complete maturity model, including dimensions, attributes, and levels, was then discussed with 8 of the medical experts, who confirmed its comprehensiveness, consistency, and adequacy. Following Joachim et al [64], the maturity model was mathematically operationalized to enable clinics to calculate an overall maturity score. In addition, we have developed a web application for using the maturity model that clinicians can apply to calculate their maturity level in the process of ML system adoption. The result of these iterative development steps is an evaluated applicable maturity model that can help researchers and clinics assess the current state of clinics in adopting ML systems (see section *A Maturity Model for ML Systems in Clinics*).

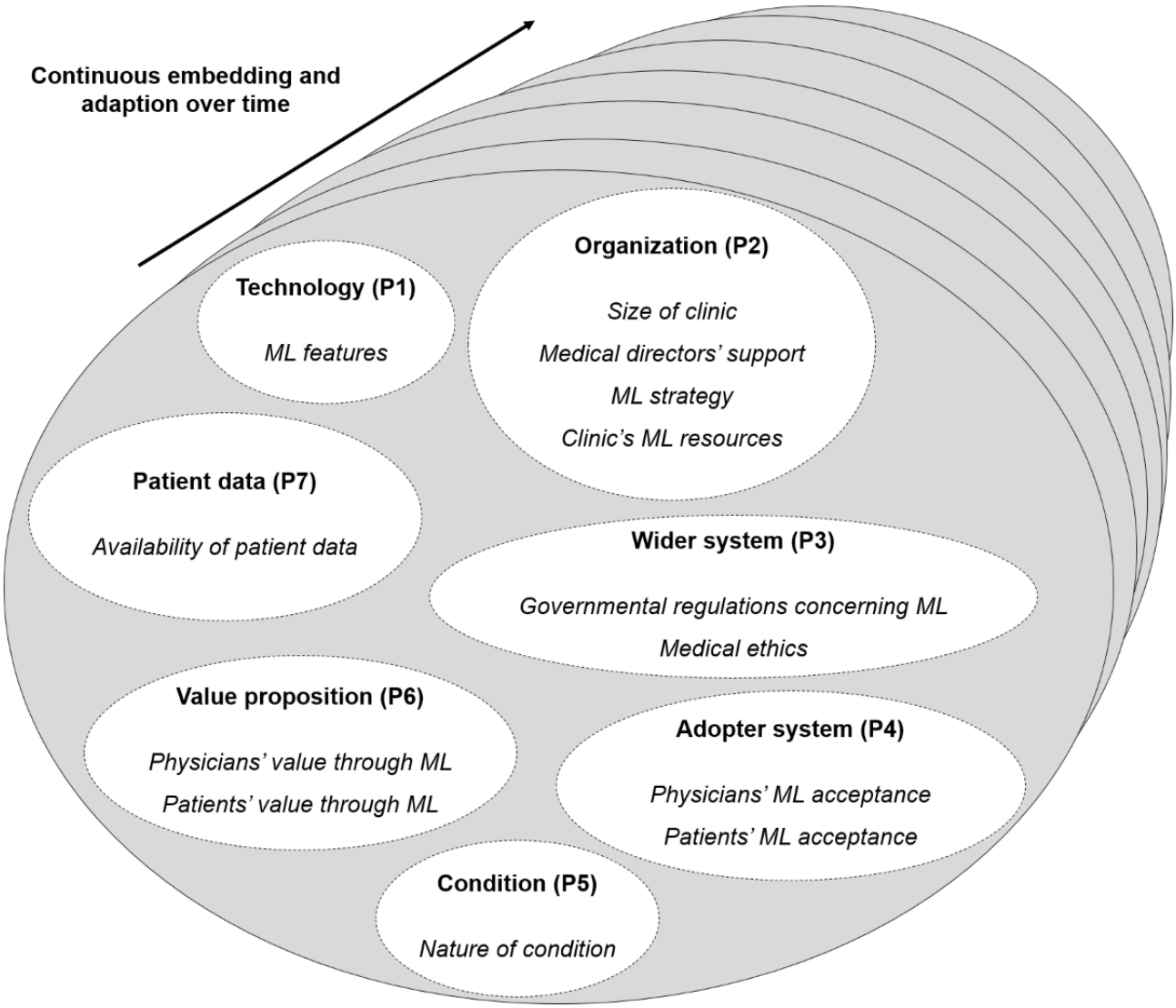
Results

Factors Influencing the Adoption Process of ML Systems in Clinics

Overview

As diagnostic procedures can differ within different medical specialties, the data analysis focuses on common factors that affect the adoption process of ML systems for diagnostics in clinics and can be derived across all disciplines. An integrative overview of these factors is shown in Figure 2. In the following section, we present and discuss the results of our directed content analysis. For this purpose, we structured our findings according to the domains: technology, organization, wider system, adopter system, condition, value proposition, and the new domain patient data. The aforementioned domains interact with each other to enable the continuous embedding and adaptation of ML systems in clinics over time [38,65,66]. In line with the existing literature, we thus did not formulate a separate domain to address the deep integration of ML systems across time. Rather, we assumed the embedding and adaptation over time to be a dynamic process in which, depending on the phase in the adoption process, specific domains and associated challenges are particularly relevant.

Figure 2. Integrative framework for the adoption process of machine learning systems in clinics. ML: machine learning.



Technology

The features of technology are factors that are already considered within the original NASSS framework [38]. Nevertheless, as outlined earlier, ML systems encompass several highly specific characteristics that cannot be compared with those of other HITs. Therefore, the existing general technical features factor is not sufficient to capture the properties of ML and has to be specified further.

As one subfactor of *ML features*, the interviewees pointed out the *lack of transparency* of ML systems as a major obstacle for the clinic's adoption of ML systems. ML systems based on neural networks can consist of multiple processing layers and up to billions of numerical weights, hampering the comprehensibility of ML systems to humans [11,32,33]. Especially in high-stakes decision-making processes such as medical diagnostics, this can lead to major issues, as ML systems do not always provide correct suggestions (S-05). As a result, the experts state that physicians need to know exactly what the critical features considered by ML systems are and how identified patterns lead to conclusions. This is required so that physicians can assess the ML system's recommendations and suggest an appropriate diagnosis and therapy. One of the experts underlines this aspect:

> *You will never make these existential decisions dependent on a black box, where it is not possible to understand what led to the recommendation.* [C-08]

Another subfactor of ML features is the *ability to adapt* their functioning if being retrained on novel data. This can become relevant either when the ML system is transferred to another context (eg, another clinic) or needs to be retrained after some time; for example, new medical research results are gained or the patient demographic structure shifts. Clinics thus have to deal with an opaque system that is able to change its reasoning, making the outcome of an ML system unpredictable. Accordingly, experts see the adaptability of ML systems as another factor that has to be addressed by clinics (C-08, S-01, S-03, and S-05). To adopt ML systems, clinics need to have a clear strategy in place on how to cope with the opacity and adaptability of self-learning ML systems. Thus, we state our first proposition:

- P1: The features of ML systems (ie, lack of transparency and adaptability) will impede their adoption in clinics.

Organization

Looking at the organization domain, 4 factors emerged during the interviews. These are the *size of a clinic*, *medical directors' ML support*, *ML strategy*, and *clinic's resources for ML*.

The size of a clinic is a newly identified factor that was not specifically considered in the original NASSS framework. However, the interviewed experts emphasize that small clinics usually have fewer resources than large clinics, which could hamper the adoption of ML systems (C-15). In the specific context of ML systems, larger clinics further care for a higher number of different patients and thus have access to more patient data, which are needed to train ML systems appropriately (S-01).

Furthermore, experts state that clinic medical directors need to support the adoption of ML systems for diagnostic processes to guarantee financial and nonfinancial support for the new technology (C-03). In this regard, ML systems for medical diagnostics affect the core business of clinics and thus have strategic relevance [67]. As medical directors develop the clinic's strategy, they are responsible for paving the way for the readiness of clinics to adopt ML systems. This is in line with prior research that states the significance of medical directors' support regarding the adoption of strategically relevant HITs in clinics [68,69].

As ML systems are a strategically relevant innovation, not only is the support of the directors necessary but also the establishment of an overarching, long-term ML strategy. The importance of an innovation strategy is also confirmed by an expert who emphasizes its relevance, especially against the background of the adoption of ML systems in a hospital network:

> *When I want to launch it to the 1900 other hospitals, I have to think about a classic transformation strategy.* [C-16]

Such a strategy should include a plan of structured activities that contribute to the successful adoption of ML systems over time and should be supported by the clinic's medical directors (C-03).

One of the most frequently stated factors within the domain organization is the clinic's resource. This factor is similar to the factor capacity to innovate already included in the original NASSS framework but is subdivided into novel subfactors (ie, *clinic's technical infrastructure*, *clinic's financing structure*, and *clinic's medical and ML methods expertise*). In line with existing literature [10,70], some of the experts report that clinics frequently rely on a wide range of clinical legacy systems, which are often proprietary to the suppliers, not connected, and based on outdated software and hardware:

> *The primary challenge [...] is that the clinic usually consists of [...] million proprietary systems that are not connected.* [C-01]

This difficulty is not only present within the clinic itself but also translates to the interorganizational level. Although some experts state that their clinics already have some special data networks in place, almost half of the experts stress that health care organizations have not yet connected their data to systems in and outside the clinic (C-01, C-03, C-04, C-05, C-06, C-08, C-09, C-13, C-15, and S-04). However, experts emphasize the importance of having a high-performance technical infrastructure that can efficiently access data from multiple sources, for example, via secure internal (within clinic) and external data networks (eg, clinic-to-primary care), which has the computing capacity needed to train ML systems (C-01, C-03, C-04, C-05, C-09, C-13, and S-04). Therefore, a clinic's existing technical infrastructure could pose a major challenge to the adoption of ML systems.

Furthermore, the interviewed experts pointed out the problem of the current financing structure of clinics, which leads to strict budgetary constraints, especially in publicly funded institutions (C-04, C-05, C-11, C-12, and C-13). In this regard, an interviewee states that one part of their budget is assigned to daily costs, such as medication. The other part of the budget can be used to purchase large-scale medical equipment, such as x-ray systems. Thus, the development and setup of ML systems are not covered by either of the 2 parts, and often, no specific ML budget can be claimed (C-08).

Beyond that, there is a lack of personnel in clinics having expertise in both medicine and ML methods such as data science or data engineering:

> *The shortage of medical specialists hits us twice as hard. We feel this at the medical professional side [...], but it is also very apparent at the technical side.* [C-14]

Both fields of knowledge are regarded as highly important for the adoption of ML systems by many experts (C-01, C-04, C-05, C-14, and S-02). Although a medical background can help identify relevant training data or assess the functionality of the ML system, ML method expertise is needed to train, integrate, and operate ML systems as presently, only scattered out-of-the-box ML systems exist for application in medicine, requiring clinics to develop and maintain ML systems by themselves (C-01, C-14, and S-02). Therefore, clinics need specific expertise in the field of ML methods in addition to their medical understanding to develop, set up, and run ML systems in clinics. In sum, we propose the following:

- P2: A larger clinic size, medical directors' ML support, formulation of an ML strategy, and availability of resources for ML (ie, technical infrastructure, ML budget, expertise in the field of medicine, and ML methods) will facilitate the adoption of ML systems in clinics.

Wider System

With regard to the wider system, there are 2 relevant factors influencing the adoption of ML systems: *governmental regulations concerning ML* and *medical ethics*. Governmental regulations are a factor already known from the original NASSS framework. Nevertheless, the interviews revealed some particularities that were not covered by the general concept and are described below. Medical ethics is a factor that has not been captured by the NASSS so far but has been identified through our study.

In the field of medicine, there are several governmental regulations that must be taken into consideration when adopting ML systems. The following subfactors could be identified: *medical approval of ML systems*, *accountability*, and the *protection of sensitive personal data*.

The experts drew attention to the fact that HIT offered in the market and used in clinics is subject to several laws. This includes the need for medical approval conducted by legal authorities or HIT suppliers themselves (C-03, C-05, and C-12). In the United States, the Food and Drug Administration is responsible for the admission of medical products. In Europe, the HIT suppliers themselves need to perform a conformity assessment procedure, for example, based on the Medical Device Regulation [71,72]. As mentioned before, most ML systems are currently being developed by the clinics themselves and have not undergone any approval process (C-03). However, legal approval of ML systems is not trivial, as the systems can learn from new experiences and adapt themselves as described above:

> *It is not obvious how evidence can be obtained for an [ML] model that differs significantly at the beginning, middle, and end of the study. If you want to approve a medical device today, you have to describe the intended use in detail.* [S-01]

The Food and Drug Administration addresses this legal uncertainty in an official statement that proposes an action plan for innovative approaches to more effectively approve adaptive ML systems [72]. The European Medicines Agency is also still in the early stages of defining and establishing an approval process for ML systems [73]. Therefore, legal ambiguities could represent a hurdle for clinics to adopt ML systems for diagnostics.

In addition to the medical approval of an ML system, there is the question of accountability for diagnoses. The experts interviewed indicated that it is questionable who takes over responsibility if the diagnosis prepared by an ML system is inaccurate (C-06, C-14, and S-05). It is also unclear who can be held liable—the HIT provider, the clinic, or the physician who is providing the medical diagnosis. An expert underlines this aspect with the following words:

> *Then there are certainly [...] legal problems, for example: who is responsible for the interpretation and possibly wrong results of the ML model?* [C-14]

According to the current state of the art, ML systems cannot be held responsible for their output, as a registered physician is always obliged to validate and interpret the system's results and perform the final diagnosis (C-16). However, it would ease the decision of clinics to opt for ML systems if there were a legal specification, especially if ML systems are increasingly able to automate steps of sensitive processes such as diagnostics (C-14 and C-15).

Another subfactor of governmental regulations, which could be identified as relevant for the adoption process of ML systems for diagnostics, is the protection of sensitive personal patient data. Patient data are widely considered as highly sensitive [74] and are under special protection by national and international laws (C-02, C-04, C-13, S-02, and S-05). For example, the General Data Protection Regulation in Europe only permits the processing of health data if the patient explicitly accepts or if the clinic can provide particular reasons for the use of the data [75]. Thus, the respondents emphasized the clinics' concerns in obtaining the necessary patient data to train the ML system (C-02, C-10, and S-06).

Using ML systems for diagnostic processes fueled medical ethics concerns among interviewees. On the one hand, ML systems are able to improve the efficiency and effectiveness of diagnostics (C-15, C-16, and S-02) and, on the other hand, the suggestions provided by ML systems are deduced based on statistical methods that recognize patterns in patient data that can be biased (C-15). Furthermore, the experts claimed that ML systems that are fed with patient data could determine whether a patient tends to develop a disease. This type of medical application would contradict the "patient's right not to know" (C-15). Summarizing these remarks, we set up the proposition:

- P3: Uncertainties in governmental regulations, strict requirements for the protection of sensitive patient data, and existing medical ethics will impede the adoption of ML systems in clinics.

Adopter System

The NASSS framework suggests that the successful adoption of ML systems is strongly influenced by individuals who are supposed to use the system or are affected by their suggestions. In this context, 2 ML-specific factors turned out to be relevant according to the interviews, which further specify the domain: *physician* and *patient ML acceptance*.

More than half of the interviewed experts stated that physicians' acceptance is essential for the adoption of ML systems in clinics (C-01, C-02, C-03, C-05, C-06, C-08, C-09, C-12, C-14, C-15, S-03, and S-06). As ML systems have the ability to solve tasks that were previously performed by humans, physicians might feel interchangeable in their job (C-03, C-07, S-03, and S-05). ML systems are trained on large sets of data that exceed the experience of any single physician, setting new standards for medical diagnostics. In this regard, most experts are concerned that physicians could reject ML systems for their daily work:

> *As a doctor who may have ten or 20 years of experience [...], would I like to be taught by a machine [...]?* [S-03]

These concerns have recently found their way into pertinent research, demonstrating the relevance of the topic [19,30,31,34]. However, it is also evident that the acceptance of ML systems differs among different age groups. In particular, physicians who belong to the group of digital natives are more willing to understand and ultimately use ML systems (S-04 and S-06).

Most interviewees stated the importance of patients' views on the use of ML systems for medical diagnostics. Although a physician is still involved in the decision-making process, patients might refuse the use of an ML system as the physician may be influenced by suggestions for possible conditions that are derived statistically and could be affected by biases. Furthermore, personal, sensitive patient data have to be

processed to gain results. Therefore, experts state that patient acceptance of ML systems is highly relevant for the adoption of ML systems for diagnostics (C-02, C-06, and C-14). We thus conclude the following:

- P4: Physicians' and patients' acceptance of ML systems will facilitate the adoption of ML systems in clinics.

Condition

As specified within the NASSS framework, patient condition affects the applicability of a technology. This is not only the case for conventional HITs but also holds true for ML systems, as stated by the interviewed experts (C-02 and C-09). ML systems have a narrow focus and can only deal with specific delimited problems [11,12]. However, the human body is a highly complex and not fully understood system that can hardly be delineated. Medical conditions can be complex, poorly understood, or even unpredictable, for example, when multiple comorbidities are involved, making it difficult for ML systems to provide a clear diagnostic recommendation (C-02 and S-02). Therefore, the nature of the condition affects the applicability of ML systems, which can only handle delimited problems in the diagnostic process. Thus, the use of ML systems will be limited to the diagnosis of certain conditions:

- P5: The limited applicability of ML systems for the diagnosis of specific conditions will impede the adoption of ML systems in clinics.

Value Proposition

The value proposition is another domain of the NASSS framework that we were able to concretize by analyzing the interviews. According to the experts, the adoption of ML systems could result in the creation of *value for* both *physicians* and *patients* (C-03, C-10, and C-14).

Integrating ML systems in their daily work enables physicians to improve the effectiveness and efficiency of their diagnostics as they can base their decisions on a broad database that is evaluated within a few seconds (C-16):

> *If you have the choice among a pathologist who has already looked at 10,000 cuts [...] compared to one who has created only 500 findings, whom would you chose? But [...] AI has not only 10,000 but 500,000 findings in its memory.* [C-08]

In this regard, ML systems that are, for example, based on image recognition algorithms can surpass the ability of the human eye to capture details and patterns in x-rays [76]. If used for a second opinion, ML systems thus increase the quality of physicians' work (C-02 and C-11).

In addition, patients could directly benefit from a decision that is faster and more informed if physicians use ML systems for diagnostics as a supportive tool (C-10 and C-16). We thus propose the following:

- P6: The additional value for physicians and patients created through ML systems will facilitate the adoption of ML systems in clinics.

Patient Data

During the interviews, nearly all experts stated the *availability of patient data* as crucial for the adoption of ML systems for diagnostics. In this regard, patient data have to be available to develop and train the ML system in the first place and subsequently retrain it during use. This factor comprises various subfactors (ie, *digitization of patient data*, *unified data formats*, *data quality standards*, *data anonymization*, and *representativeness of training data*) which are described in the following section.

According to the experts, most clinics generate high volumes of patient data through their daily diagnostic processes (C-03, C-05, S-01, S-04, S-05, and S-06), which is basically a positive feature as an appropriate amount of data is needed to train ML systems [11,20,35,77]. However, although high volumes of data are generated, many processes in clinics are still paper-based, which lowers the proportion of patient data available in digitized form:

> *Data are often not digitized, much is still in paper files, not structured, which means that the data availability is really extremely [...] poor.* [C-03]

This observation is in line with prior research concerning clinics that are lagging behind at using digitized technologies and digitizing patient data [1]. As a consequence, the interviewed experts see the integration of an electronic medical record system as a prerequisite for the application of ML systems (C-16, C-03, C-04, and C-13).

Furthermore, interviewing the experts revealed that medical patient data, if available in digitized form, are usually provided in a variety of proprietary data formats as many disparate clinical legacy systems from different suppliers have to interact to enable physicians to provide laboratory tests, diagnostic images, or clinical notes. These proprietary data formats are often difficult or impossible to convert, making the generation of consistent formats highly problematic (C-03, C-04, and S-04). The problem of differing data formats in clinics has already been recognized outside the ML context, for example, in research on the adoption of cloud solutions in health care environments [78]. Nevertheless, it is particularly critical for the introduction and use of ML systems that the patient data be processed for training and retraining the system. Although the first research has been conducted to allow for the transformation of different medical data types in one format [79], most clinics have not yet been able to implement unified standards for patient data to enable processing and analysis by ML systems.

Furthermore, digitized patient data are often stored in unstructured file types, such as images, texts, or videos (C-01, C-03, C-07, C-13, C-15, and S-04). The experts cautioned that the quality of unstructured data is highly dependent on the particular clinic where the data are generated and their clinical staff (C-06, C-07, and S-04). For instance, physician letters are frequently written in free text formats, which are filled with synonyms and can be interpreted individually. More specifically, personal formulations are used, such as the description of a tumor size as compared with that of a walnut (C-07). Thus, patient data are not only hard to harness and have to be

transferred to a machine-readable format first (C-03 and C-04) but also lack common quality standards (S-04), impeding the extraction of generalizable patterns through ML. Clinics aiming to adopt ML systems to support their diagnostics should therefore set standards for data creation, for example, by establishing a common language that physicians use when creating free texts. Such efforts are already being driven by some in-clinic as well as national initiatives (C-12 and C-16). In addition, other primary structured data sources could be connected, such as data from laboratory findings, to complement the unstructured data [80].

Moreover, the experts strongly emphasize that clinics that want to use patient data to train ML systems need to anonymize the sensitive data before processing them through an ML system (C-15 and S-06). However, anonymizing data might remove valuable information, which could be important for obtaining a diagnosis. For instance, information about a person's residence could facilitate a diagnosis if a disease is more prevalent regionally (C-15). Therefore, it is necessary for clinics to find the right balance of anonymization and information value to be able to use the data despite data protection regulations and still preserve all the information necessary to find meaningful correlations through ML systems. The first steps are already being taken in technical research to balance protection and the quality of sensitive data effectively [81,82].

According to the experts, the selection of the right training data is especially important in a health care context, as wrong diagnoses may have an impact on patients' lives. This leads to another aspect of patient data to be considered: the representativeness of training data. Patients in clinics vary in many aspects, from an outer perspective (eg, age, gender, and hair color) as well as from inner functioning (eg, size of organs and blood values; C-01 and S-05). If ML systems are trained based on an external database (eg, collected via data exchange) that is demographically or regionally skewed compared with the clinic's conditions, false conclusions could be drawn by the system. In this context, an expert raised the example of an ML system supporting the detection of skin melanomas, which is mainly trained on a sample of patients with a similar phenotype. Therefore, this pretrained ML system cannot be easily transferred to patients of other ages or with other skin pigmentations (C-01). In addition, the representativeness of the data is affected when different clinical systems, such as different radiographic systems, collect data as the resolution of the medical equipment may vary from provider to provider (S-04). As training data for supervised learning need to be labeled by humans, the same could be said regarding the expertise and working philosophy of physicians, which could be highly heterogeneous depending on the physician's knowledge state and working environment (C-09, C-14, and S-05).

The availability of patient data is a factor that is decisive for the adoption process of ML systems that need to be fed and retrained:

- P7: The availability and exchange of a large amount of digitized patient data for training (that are uniformly formatted, of high quality, anonymized but informative, and representative of the clinic) will facilitate the adoption of ML systems in clinics.

A Maturity Model for ML Systems in Clinics

Overview

Against the background that no maturity model for the adoption process of ML systems in clinics could be found in research and practice, we created a concept for a maturity model and present the model below. On the basis of our empirical results, the model is intended to enable researchers and clinics to quantify the overall maturity of clinics within the adoption process of ML systems. We followed the design process of Becker et al [36] to conceptualize a maturity model that comprises 3 dimensions and 12 attributes, each of which is operationalized by 5 corresponding levels (Multimedia Appendix 1). The dimensions and attributes are derived from a subset of the results presented in the previous section, whereby the dimensions were inferred from the domains and the attributes from the factors or subfactors that can be modified by the respective clinic itself. Specifically, the dimensions organization (P2), adopter system (P4), and patient data (P7) and their respective subfactors were taken into account, as these can be controlled by the clinic itself, whereas the technology (P1), the wider system (P3), the condition (P5), and the value proposition (P6) are influenced by factors that are not in the hands of a single organization.

It is necessary to operationalize the model mathematically to render the maturity model applicable for research and practice. To this end, we followed the approach of Joachim et al [64], which has already been used for the operationalization of other maturity models (eg, in the area of business intelligence [83]). We assume that maturity evolves linearly in 5 levels $l \in L$ with L={1,2,3,4,5}, starting with *initial* (1) and ending with *optimized* (5) [83]. The maturity model for the adoption of ML systems in clinics consists of 3 dimensions, d, each of which consists of a set of attributes I_d in turn. Therefore, the overall maturity score of a clinic is composed of the maturity score of all dimensions, whereby the maturity of each dimension d depends on the maturity within the corresponding attributes $a \in I_d$. As a clinic can have different maturities in the different dimensions and attributes of a dimension, a stepwise estimation of the overall maturity score must be made. Therefore, a two-step process is followed in which (1) the *maturity score of the dimensions* (ie, Mat_a) is determined first based on the respective attributes, followed by (2) the calculation of the *overall maturity score of a clinic* (ie, Mat).

Maturity Score of the Dimensions

At the lowest layer, each attribute a can take a value $x_a \in A$ with A={1,2,3,4,5} depending on the actual maturity of the clinic regarding the attribute, ranging from initial (1) to optimized (5). To determine the actual maturity value of each attribute in a dimension, a clinic must assess its own as-is situation by comparing the level descriptions (within each attribute) with their current adoption state in the clinic (Multimedia Appendix 1). For example, a clinic has a maturity value of x_a=1 for the attribute *digitization of patient data* if it has nearly no digitized data available for training ML systems

and is thus at an initial level of maturity. In the next step, all maturity values x_a of the attributes within a dimension d are compared with all possible maturity levels l to determine the level with the smallest distance to the set of attributes of a dimension. To operationalize the comparison, a weighted Euclidean distance metric $Dist_d(l)$ is used in line with prior research [64,83]:

where n_d represents the total number of dimensions and n_l is the total number of levels. As a result, each clinic receives 5 distance values (for 5 levels, l) per dimension. To obtain the maturity score of a dimension Mat_d, the level m associated with the minimum of these distance values needs to be selected per dimension:

Overall Maturity Score of the Clinic

On the basis of the distinct maturity scores Mat_d of the 3 dimensions, the overall maturity score Mat can be calculated in the second step. Again, we use a Euclidean distance metric $Dist(l)$ to compare the maturity scores of the dimensions with levels l (Equation 3). The final overall maturity score of a clinic striving to adopt ML systems is determined by the minimum distance (Equation 4):

To make the maturity model easily applicable for practitioners from clinics and researchers in the field of adoption science, we have developed a free-access web application based on the described mathematical operationalization, which calculates the maturity level of a clinic based on a questionnaire (Figures 3 and 4). This questionnaire includes the attributes as well as their level descriptions and is provided on the web [84].

Figure 3. Determine your clinic's readiness for machine learning–supported diagnostics (screenshot 1 of the web application). ML: machine learning.

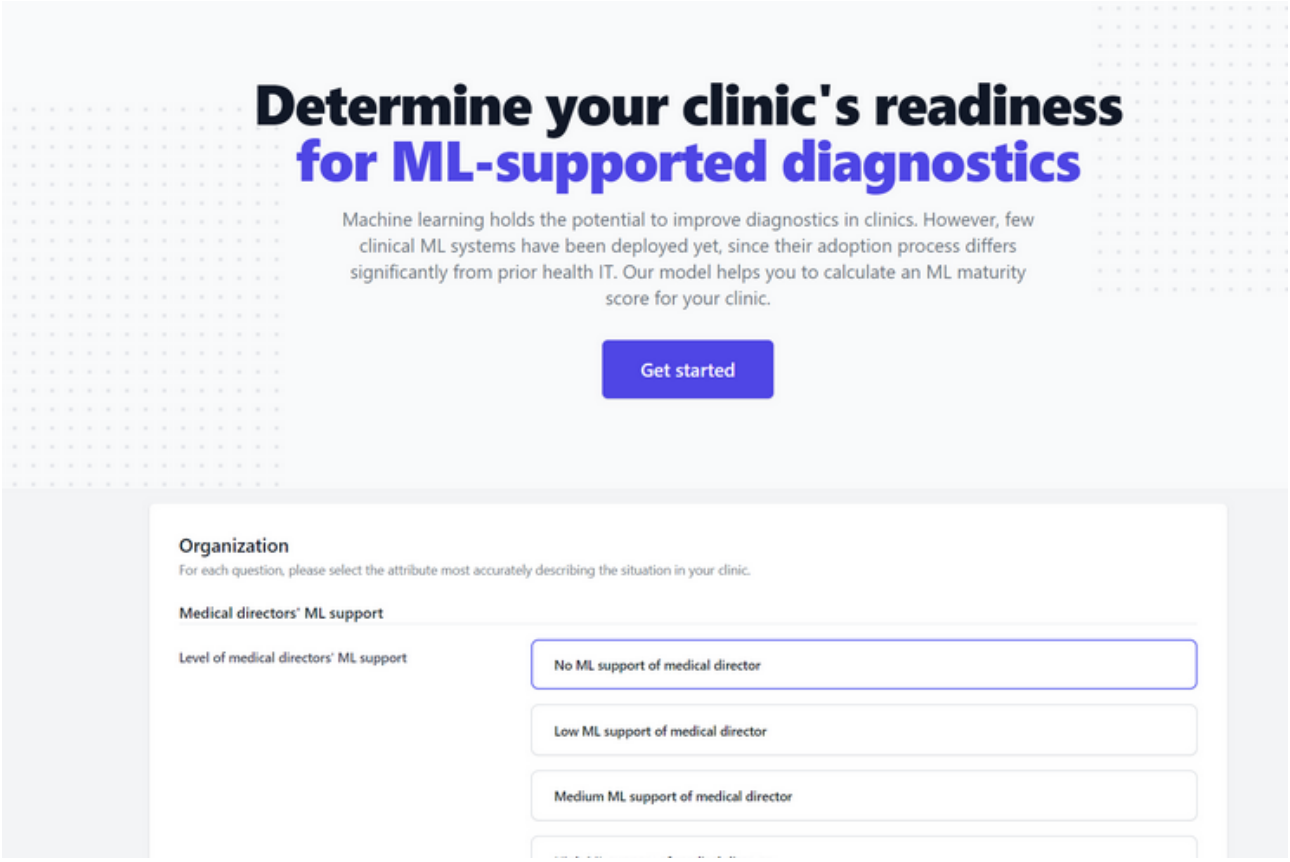
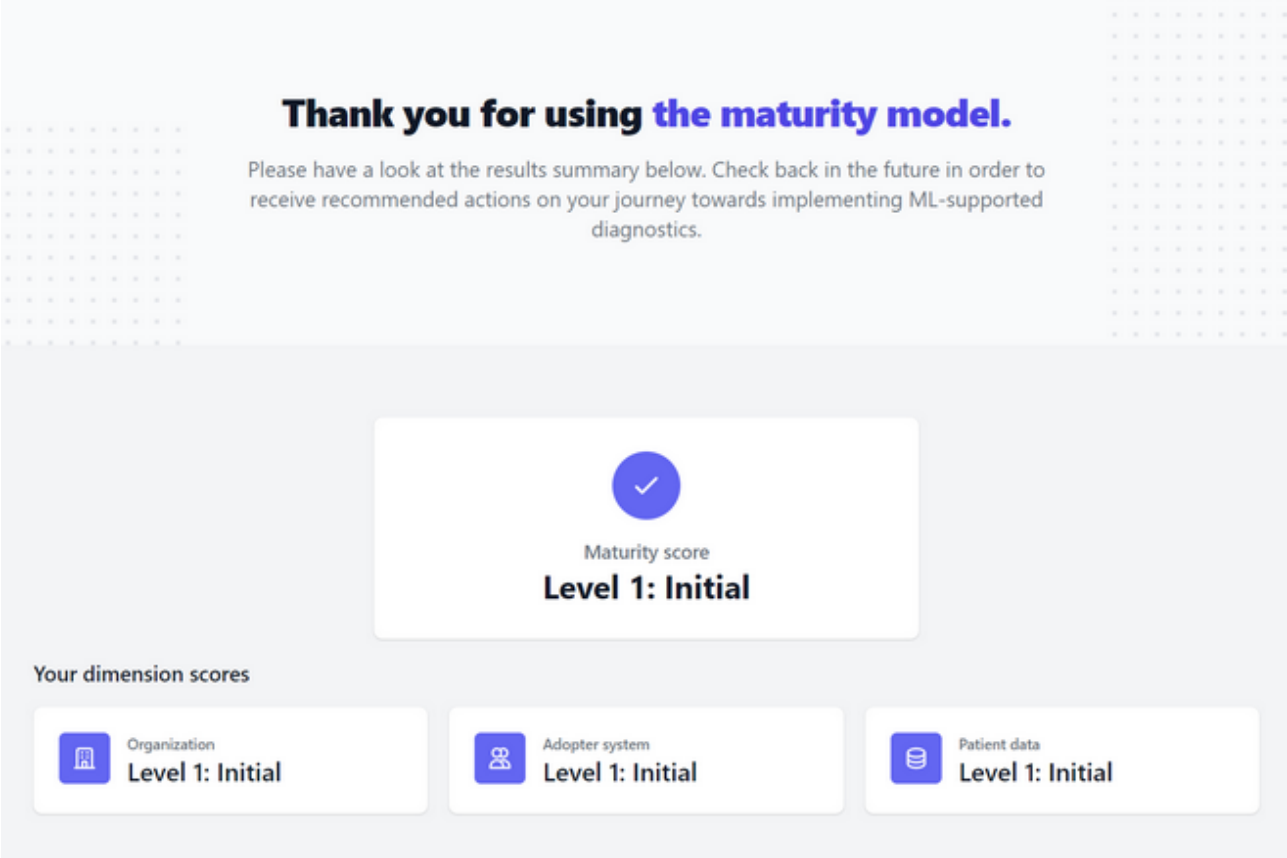


Figure 4. Thank you for using the maturity model (screenshot 2 of the web application). ML: machine learning.



Discussion

Principal Findings

ML has an impact on all areas of human life, including the health care system. In this regard, ML systems offer the opportunity to make diagnostics more efficient and informed. However, to harness ML for such an application, clinics need to deeply integrate ML systems into their clinical practice, a challenge that most clinics have not yet been able to overcome [20]. As clinics own highly individual, patient-oriented processes, it is crucial for researchers to reflect on this specific context [28,29]. However, prior research is lagging behind to provide empirically proven factors that influence the adoption process of ML systems in clinics for diagnostic processes. To address this shortcoming, we set up a qualitative study to (1) establish an integrated overview of factors specific to an ML system adoption process in clinics based on the NASSS framework and (2) create an operationalized maturity model that clinics can apply to assess their as-is state of ML adoption progress to decide on further actions and prioritize investments.

Limitations and Future Research Opportunities

Before we discuss our contributions to theory and practice in detail, it is necessary to clarify the limitations of this study and show room for further research. As we pursued a qualitative approach, our results are based on the expertise of the 22 interviewees. To counteract potential problems of generalizability, we have not only applied various criteria to ensure rigor and trustworthiness of our study (eg, theoretical saturation, multiresearcher and data triangulation, and inclusion of multiple medical disciplines) but also carefully selected only highly involved experts. Nevertheless, it might be interesting for further research to perform a follow-up study to validate the proposed framework and maturity model quantitatively. In this regard, it might be informative to evaluate the derived maturity model by applying it in clinics. In doing so, it could also be investigated whether practitioners attach different importance to attributes and dimensions. On the basis of these findings, the maturity calculation could be adjusted by introducing weights for attributes and dimensions.

Moreover, we conducted the interviews in only 2 European countries. As health care systems vary across nations, interviewing experts from other regions with different economic and cultural prerequisites could lead to differing results. Nevertheless, the relevance of the findings for the international context was substantiated with the help of existing literature and practice contributions from international authorities, which are cited in the *Results* section. For example, the report of the Food and Drug Administration shows that the issue of medical approval of ML systems is also being discussed in the United States [72]. However, replication of this study in other countries would be useful to highlight possible differences within the adoption process of ML systems in clinics.

In addition, the rapid development of increasingly advanced ML algorithms could lead to systems that can not only augment but also automate diagnostic processes. Investigating automated diagnostics, which has not yet been applied in clinics, could produce different findings, although the results obtained in this study could provide first indications.

Theoretical Contributions

Despite the limitations discussed, our study makes several important contributions to research. To begin with, we demonstrated that the NASSS framework can be applied but has to be adapted and expanded to explain the full adoption process of ML systems for diagnostics in clinics. To the best of our knowledge, this is the first study to provide an empirically proven and integrative overview of the factors determining the adoption of ML systems for clinical diagnostics and thus show what clinics need to consider to effectively integrate ML systems into their processes. Therefore, we contribute to and extend prior adoption research in health informatics, which has recently called for looking at the entire adoption process of HITs rather than just the initial awareness of the technology [38]. Although the identified factors are specific to diagnostic processes, it is conceivable that they may be applicable to other scenarios in which the cost of errors is high, such as ML-based treatment recommendations or medical prognoses in clinics.

Moreover, we have developed the first maturity model for ML system adoption in clinics, which contributes to the IS and medical body of knowledge by providing an empirically grounded and strategically derived artifact that depicts medical and ML-specific attributes and their level descriptions in detail. More specifically, the maturity model shows which attributes determine the status quo of clinics in adopting ML systems, how these attributes may manifest in descriptors according to 5 different maturity levels, and how clinics can evaluate their as-is state in the adoption process of ML systems. Researchers can apply the developed maturity model, for example, as an instrument in statistical studies investigating the adoption of ML systems in clinics. More specifically, the model can be used to operationalize the dependent variable in structural equation models or as a variable for multigroup comparisons [85], for example, to study the antecedents of clinical adoption of ML systems. Therefore, both the adoption framework and the maturity model for ML systems in clinics can guide future health care–centric research that seeks to explore the promises and challenges associated with ML systems in a medical setting.

Practical Contributions

In addition, the empirically based results hold relevant findings for practitioners, who are increasingly facing rising health care costs, demographic changes, and overcrowding of the clinics, and thus need to improve the efficiency and effectiveness of their clinical processes. ML systems could be a solution to these problems but have so far only been sporadically integrated into clinics [22]. In fact, our qualitative study shows that most clinics still have major problems integrating ML systems into their diagnostics. In this regard, the derived framework provides medical directors with a holistic overview of potential enablers and inhibitors during the adoption process of ML systems in clinics and could provide a roadmap for practitioners.

Moreover, the developed maturity model can be used by clinics to obtain the first impression of their as-is situation in the adoption process of ML systems and to quantify it in an overall maturity score (see the website [84] to easily apply the model). Assessing the maturity score with the help of the model not only helps to make external comparisons between clinics but also to identify internal deviations of certain attributes from the overall status. This allows clinics to invest especially in these attributes that are far from the present overall performance and lower the clinic's maturity score significantly to date. Thereby, the maturity model allows practitioners working for clinics to analyze their clinic's current status quo, identify shortcomings, prioritize possible courses of action, and efficiently allocate scarce resources depending on the respective degree of maturity. In this way, our research can help practitioners identify tailored requirements for the successful adoption of ML systems in clinics and build relevant capabilities and resources needed in the age of AI.

Acknowledgments

This research was funded by the German Federal Ministry of Education and Research and the Hessen State Ministry for Higher Education, Research, and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

Conflicts of Interest

This paper builds on a conference paper [86]. This paper extends the earlier work, in particular by expanding the underlying sample size and developing a maturity model.

Multimedia Appendix 1
Maturity model for machine learning systems in clinics.
[PNG File , 366 KB - jmir_v23i10e29301_app1.png]

References

1. Hufnagl C, Doctor E, Behrens L, Buck C, Eymann T. Digitisation along the patient pathway in hospitals. In: Proceedings of the 27th European Conference on Information Systems - ECIS. 2019 Presented at: 27th European Conference on Information Systems - ECIS; Jun 8-14, 2019; Stockholm & Uppsala, Sweden URL: https://eprints.qut.edu.au/204984/
2. Wang X, Sun J, Wang Y, Liu Y. Deepen electronic health record diffusion beyond breadth: game changers and decision drivers. Inf Syst Front 2021 Jan 07:1-12. [doi: 10.1007/s10796-020-10093-6]
3. Sun J, Qu Z. Understanding health information technology adoption: A synthesis of literature from an activity perspective. Inf Syst Front 2014 Apr 29;17(5):1177-1190. [doi: 10.1007/s10796-014-9497-2]

4. Bardhan I, Chen H, Karahanna E. Connecting systems, data, and people: A multidisciplinary research roadmap for chronic disease management. MIS Q 2020;44(1):185-200. [doi: 10.25300/MISQ/2020/14644]
5. Li T, Zhang Y, Gong C, Wang J, Liu B, Shi L, et al. Prevalence of malnutrition and analysis of related factors in elderly patients with COVID-19 in Wuhan, China. Eur J Clin Nutr 2020;74(6):871-875 [FREE Full text] [doi: 10.1038/s41430-020-0642-3] [Medline: 32322046]
6. Kankanhalli A, Hahn J, Tan S, Gao G. Big data and analytics in healthcare: Introduction to the special section. Inf Syst Front 2016;18(2):233-235. [doi: 10.1007/s10796-016-9641-2]
7. McCarthy J. What is artificial intelligence? Stanford University. 2007. URL: http://jmc.stanford.edu/articles/whatisai/whatisai.pdf [accessed 2021-04-01]
8. Rai A, Constantinides P, Sarker S. Editor's comments: Next-generation digital platforms: Toward human-AI hybrids. MIS Q 2019;43(1):iii-iix [FREE Full text]
9. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science 2015 Jul 17;349(6245):255-260. [doi: 10.1126/science.aaa8415] [Medline: 26185243]
10. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial intelligence and the implementation challenge. J Med Internet Res 2019 Jul 10;21(7):e13659 [FREE Full text] [doi: 10.2196/13659] [Medline: 31293245]
11. Brynjolfsson E, Mitchell T. What can machine learning do? Workforce implications. Science 2017 Dec 22;358(6370):1530-1534. [doi: 10.1126/science.aap8062] [Medline: 29269459]
12. Russell SJ, Norvig P. Artificial Intelligence: A Modern Approach. 3rd Ed. Boston, MA: Addison Wesley; 2016.
13. Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. NPJ Digit Med 2020 Sep 24;3(1):1-8 [FREE Full text] [doi: 10.1038/s41746-020-00333-z] [Medline: 33043150]
14. Thrall JH, Li X, Li Q, Cruz C, Do S, Dreyer K, et al. Artificial intelligence and machine learning in radiology: Opportunities, challenges, pitfalls, and criteria for success. J Am Coll Radiol 2018 Mar;15(3 Pt B):504-508. [doi: 10.1016/j.jacr.2017.12.026] [Medline: 29402533]
15. Paton C, Kobayashi S. An open science approach to artificial intelligence in healthcare. Yearb Med Inform 2019 Aug;28(1):47-51 [FREE Full text] [doi: 10.1055/s-0039-1677898] [Medline: 31022753]
16. Shahid N, Rappon T, Berta W. Applications of artificial neural networks in health care organizational decision-making: A scoping review. PLoS One 2019;14(2):1-22 [FREE Full text] [doi: 10.1371/journal.pone.0212356] [Medline: 30779785]
17. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019;25(1):30-36. [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
18. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
19. Lebovitz S. Diagnostic doubt and artificial intelligence: An inductive field study of radiology work. In: Proceedings of the 40th International Conference on Information Systems (ICIS). 2019 Presented at: 40th International Conference on Information Systems (ICIS); December 15-18, 2019; Munich, Germany URL: https://aisel.aisnet.org/icis2019/future_of_work/future_work/11/
20. Matheny ME, Whicher D, Israni ST. Artificial intelligence in health care: A report from the national academy of medicine. J Am Med Assoc 2020 Feb 11;323(6):509-510. [doi: 10.1001/jama.2019.21579] [Medline: 31845963]
21. Jin C, Chen W, Cao Y, Xu Z, Tan Z, Zhang X, et al. Development and evaluation of an artificial intelligence system for COVID-19 diagnosis. Nat Commun 2020 Oct 09;11(1):5088-5014 [FREE Full text] [doi: 10.1038/s41467-020-18685-1] [Medline: 33037212]
22. Kuan R. Adopting AI in health care will be slow and difficult. Harvard Bus Rev. 2019. URL: https://hbr.org/2019/10/adopting-ai-in-health-care-will-be-slow-and-difficult [accessed 2021-04-01]
23. Amershi S, Begel A, Bird C, DeLine R, Gall H, Kamar E, et al. Software engineering for machine learning: A case study. In: Proceedings of the IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP).: IEEE; 2019 Presented at: IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP); May 25-31, 2019; Montréal, QC, Canada. [doi: 10.1109/icse-seip.2019.00042]
24. Pumplun L, Tauchert C, Heidt M. A new organizational chassis for artificial intelligence—Exploring organizational readiness factors. In: Proceedings of the 27th European Conference on Information Systems (ECIS). 2019 Presented at: 27th European Conference on Information Systems (ECIS); June 8-14, 2019; Stockholm & Uppsala, Sweden URL: https://aisel.aisnet.org/ecis2019_rp/106/
25. Jöhnk J, Weißert M, Wyrtki K. Ready or not, AI comes— an interview study of organizational ai readiness factors. Bus Inf Syst Eng 2020;63(1):5-20. [doi: 10.1007/s12599-020-00676-7]
26. Alsheibani S, Cheung Y, Messom C. Factors inhibiting the adoption of artificial intelligence at organizational-level: A preliminary investigation. In: Proceedings of the 25th Americas Conference on Information Systems (AMCIS). 2019 Presented at: 25th Americas Conference on Information Systems (AMCIS); August 15-17, 2019; Cancun, Mexico URL: https://dblp.org/db/conf/amcis/amcis2019.html
27. Kruse L, Wunderlich N, Beck R. Artificial intelligence for the financial services industry: What challenges organizations to succeed. In: Proceedings of the Hawaii International Conference on System Sciences. 2019 Presented at: 52nd Hawaii International Conference on System Sciences; January 8 - 11, 2019; Wailea, Hawaii. [doi: 10.24251/hicss.2019.770]

28. Poba-Nzaou P, Uwizeyemungu S, Raymond L, Paré G. Motivations underlying the adoption of ERP systems in healthcare organizations: Insights from online stories. Inf Syst Front 2012;16(4):591-605. [doi: 10.1007/s10796-012-9361-1]
29. Davison RM, Martinsons MG. Context is king! Considering particularism in research design and reporting. J Inf Technol 2016;31(3):241-249. [doi: 10.1057/jit.2015.19]
30. Alami H, Lehoux P, Denis JL, Motulsky A, Petitgand C, Savoldelli M, et al. Organizational readiness for artificial intelligence in health care: insights for decision-making and practice. J Health Organ Manag 2020;35(1):106-114 [FREE Full text] [doi: 10.1108/JHOM-03-2020-0074] [Medline: 33258359]
31. Sandhu S, Lin AL, Brajer N, Sperling J, Ratliff W, Bedoya AD, et al. Integrating a machine learning system into clinical workflows: qualitative study. J Med Internet Res 2020;22(11):e22421 [FREE Full text] [doi: 10.2196/22421] [Medline: 33211015]
32. Arora A. Conceptualising artificial intelligence as a digital healthcare innovation: an introductory review. Med Devices (Auckl) 2020;13:223-230 [FREE Full text] [doi: 10.2147/MDER.S262590] [Medline: 32904333]
33. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2019;17(1):1-9 [FREE Full text] [doi: 10.1186/s12916-019-1426-2] [Medline: 31665002]
34. Hofmann P, Oesterle S, Rust P, Urbach N. Machine learning approaches along the radiology value chain – Rethinking value propositions. In: Proceedings of the Twenty-Seventh European Conference on Information Systems (ECIS). 2019 Presented at: Twenty-Seventh European Conference on Information Systems (ECIS); June 8-14, 2019; Stockholm & Uppsala, Sweden URL: http://publica.fraunhofer.de/dokumente/N-559057.html
35. Sun TQ, Medaglia R. Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. Gov Inf Q 2019;36(2):368-383. [doi: 10.1016/j.giq.2018.09.008]
36. Becker J, Knackstedt R, Pöppelbuß J. Developing maturity models for IT management. Bus Inf Syst Eng 2009;1(3):213-222. [doi: 10.1007/s12599-009-0044-5]
37. Poeppelbuss J, Niehaves B, Simons A, Becker J. Maturity models in information systems research: Literature search and analysis. Commun Assoc Inf Syst 2011;29:505-532. [doi: 10.17705/1cais.02927]
38. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: A new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
39. Greener S. Business Research Methods. London, UK: Ventus Publishing; 2008.
40. Bagozzi RP, Yi Y, Phillips LW. Assessing construct validity in organizational research. Adm Sci Q 1991;36(3):421-458. [doi: 10.2307/2393203]
41. Myers MD, Newman M. The qualitative interview in IS research: Examining the craft. Inf Organ 2007;17(1):2-26. [doi: 10.1016/j.infoandorg.2006.11.001]
42. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
43. Flick U. Triangulation in qualitative research. In: Flick U, von Kardoff E, Steinke I, editors. A Companion to Qualitative Research. London, UK: SAGE Publications; 2004:178-183.
44. Carter N, Bryant-Lukosius D, DiCenso A, Blythe J, Neville AJ. The use of triangulation in qualitative research. Oncol Nurs Forum 2014;41(5):545-547. [doi: 10.1188/14.ONF.545-547] [Medline: 25158659]
45. Benbasat I, Goldstein DK, Mead M. The case research strategy in studies of information systems. MIS Q 1987 Sep;11(3):369-386. [doi: 10.2307/248684]
46. Horan TA, Schooley B. Inter-organizational emergency medical services: Case study of rural wireless deployment and management. Inf Syst Front 2005 May;7(2):155-173. [doi: 10.1007/s10796-005-1476-1]
47. Whittaker R. Issues in mHealth: findings from key informant interviews. J Med Internet Res 2012;14(5):e129 [FREE Full text] [doi: 10.2196/jmir.1989] [Medline: 23032424]
48. Urquhart C, Lehmann H, Myers MD. Putting the “Theory” back into grounded theory: Guidelines for grounded theory studies in information systems. Inf Syst J 2010;20(4):357-381. [doi: 10.1111/j.1365-2575.2009.00328.x]
49. Creswell JW. Qualitative Inquiry and Research Design: Choosing Among Five Traditions. 2nd Ed. California, USA: SAGE Publications; 2007.
50. Bogner A, Littig B, Menz W. Introduction: Expert interviews – an introduction to a new methodological debate. In: Bogner A, Littig B, Menz W, editors. Interviewing Experts. London, UK: Palgrave Macmillan; 2009:1-13.
51. Rogers EM. Diffusion of Innovations, 5th Edition. New York, USA: The Free Press; 2003.
52. Saldaña J. The Coding Manual for Qualitative Researchers. London, UK: SAGE Publications; 2009.
53. De Bruin T, Rosemann M, Freeze R, Kulkarni U. Understanding the main phases of developing a maturity assessment model. In: Proceedings of the 16th Australasian Conference on Information Systems (ACIS). 2005 Presented at: 16th Australasian Conference on Information Systems (ACIS); Nov 29 - Dec 2, 2005; Sydney, Australia URL: https://asu.pure.elsevier.com/en/publications/understanding-the-main-phases-of-developing-a-maturity-assessment
54. Paulk M, Curtis B, Chrissis M, Weber C. Capability maturity model, version 1.1. IEEE Softw 1993 Jul;10(4):18-27. [doi: 10.1109/52.219617]

55. Fraser P, Moultrie J, Gregory M. The use of maturity models/grids as a tool in assessing product development capability. In: Proceedings of the IEEE International Engineering Management Conference. 2002 Presented at: IEEE International Engineering Management Conference; Aug. 18-20, 2002; Cambridge, UK. [doi: 10.1109/iemc.2002.1038431]
56. Mettler T, Rohner P. Situational maturity models as instrumental artifacts for organizational design. In: Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology. 2009 Presented at: DESRIST '09: Design Science Research in Information Systems and Technologies; May 7 - 8, 2009; Philadelphia, USA p. 1-9. [doi: 10.1145/1555619.1555649]
57. Raber D, Winter R, Wortmann F. Using quantitative analyses to construct a capability maturity model for business intelligence. In: Proceedings of the 45th Hawaii International Conference on System Sciences. 2012 Presented at: 45th Hawaii International Conference on System Sciences; Jan. 4-7, 2012; Maui, Hawaii, USA. [doi: 10.1109/hicss.2012.630]
58. Humphrey W. Characterizing the software process: a maturity framework. IEEE Softw 1988 Mar;5(2):73-79. [doi: 10.1109/52.2014]
59. Carvalho JV, Rocha Á, Abreu A. Maturity models of healthcare information systems and technologies: A literature review. J Med Syst 2016;40(6):131. [doi: 10.1007/s10916-016-0486-5] [Medline: 27083575]
60. Hevner AR, March ST, Park J, Ram S. Design science in information systems research. MIS Q 2004;28(1):75-105. [doi: 10.2307/25148625]
61. Alsheibani S, Cheung Y, Messom C. Towards an artifificial intelligence maturity model: From science fiction to business facts. In: Proceedings of the 23rd Pacific Asia Conference on Information Systems. 2019 Presented at: 23rd Pacific Asia Conference on Information Systems; July 8-12, 2019; Xi'an, China URL: https://aisel.aisnet.org/pacis2019/46/
62. Lichtenthaler U. Five maturity levels of managing AI: From isolated ignorance to integrated intelligence. J Innov Manag 2020 Mar 30;8(1):39-50. [doi: 10.24840/2183-0606_008.001_0005]
63. Akkiraju R, Sinha V, Xu A, Mahmud J, Gundecha P, Liu Z, et al. Characterizing machine learning processes: A maturity framework. In: Proceedings of the 18th International Conference on Business Process Management. 2020 Presented at: 18th International Conference on Business Process Management; September 13-18, 2020; Seville, Spain p. 17-31. [doi: 10.1007/978-3-030-58666-9_2]
64. Joachim N, Beimborn D, Weitzel T. An instrument for measuring SOA maturity. In: Proceedings of the 32nd International Conference on Information Systems. 2011 Presented at: 32nd International Conference on Information Systems; December 4-7, 2011; Shanghai, China URL: https://dblp.org/db/conf/icis/icis2011.html
65. James HM, Papoutsi C, Wherton J, Greenhalgh T, Shaw SE. Spread, scale-up, and sustainability of video consulting in health care: Systematic review and synthesis guided by the NASSS framework. J Med Internet Res 2021;23(1):e23775 [FREE Full text] [doi: 10.2196/23775] [Medline: 33434141]
66. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-world implementation of video outpatient consultations at macro, meso, and micro levels: mixed-method study. J Med Internet Res 2018;20(4):e150 [FREE Full text] [doi: 10.2196/jmir.9897] [Medline: 29625956]
67. Zhu K, Kraemer KL. Post-adoption variations in usage and value of e-business by organizations: cross-country evidence from the retail industry. Inf Syst Res 2005 Mar;16(1):61-84. [doi: 10.1287/isre.1050.0045]
68. Yang Z, Sun J, Zhang Y, Wang Y. Understanding SaaS adoption from the perspective of organizational users: A tripod readiness model. Comput Hum Behav 2015 Apr;45:254-264. [doi: 10.1016/j.chb.2014.12.022]
69. Lian J, Yen DC, Wang Y. An exploratory study to understand the critical factors affecting the decision to adopt cloud computing in Taiwan hospital. Int J Inf Manag 2014 Feb;34(1):28-36. [doi: 10.1016/j.ijinfomgt.2013.09.004]
70. Panch T, Mattie H, Celi LA. The "inconvenient truth" about AI in healthcare. NPJ Digit Med 2019;77:1-3 [FREE Full text] [doi: 10.1038/s41746-019-0155-4] [Medline: 31453372]
71. Migliore A. On the new regulation of medical devices in Europe. Expert Rev Med Devices 2017 Dec;14(12):921-923. [doi: 10.1080/17434440.2017.1407648] [Medline: 29157016]
72. Artificial Intelligence/Machine Learning (AI/ML)-based Software as a Medical Device (SaMD) action plan. Food and Drug Administration. 2021. URL: https://www.fda.gov/media/145022/download [accessed 2021-06-20]
73. EMA Regulatory Science to 2025: Strategic reflection. European Medicines Agency. 2020. URL: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/ema-regulatory-science-2025-strategic-reflection_en.pdf [accessed 2021-06-20]
74. Fox G, James TL. Toward an understanding of the antecedents to health information privacy concern: A mixed methods study. Inf Syst Front 2020 Aug 21:1-26. [doi: 10.1007/s10796-020-10053-0]
75. Regulation (EU) 2016/679 of the European Parliament and of the Council. European Parliament. 2016. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj [accessed 2021-06-20]
76. Akcay S, Kundegorski ME, Willcocks CG, Breckon TP. Using deep convolutional neural network architectures for object classification and detection within x-ray baggage security imagery. IEEE Trans Inform Forensic Secur 2018 Sep;13(9):2203-2215. [doi: 10.1109/tifs.2018.2812196]
77. He M, Li Z, Liu C, Shi D, Tan Z. Deployment of artificial intelligence in real-world practice: Opportunity and challenge. Asia Pac J Ophthalmol (Phila) 2020;9(4):299-307. [doi: 10.1097/APO.0000000000000301] [Medline: 32694344]
78. Zhang Y, Qiu M, Tsai C, Hassan MM, Alamri A. Health-CPS: Healthcare cyber-physical system assisted by cloud and big data. IEEE Syst J 2017 Mar;11(1):88-95. [doi: 10.1109/Jsyst.2015.2460747]

79. Lee J, Liu C, Shang N, Jiang X, Chen K, Kalluri KS, et al. Generate the concept representation using OMOP ontology graph. Observational Health Data Sciences and Informatics. 2019. URL: https://www.ohdsi.org/2019-us-symposium-showcase-52/ [accessed 2021-04-02]
80. Varghese J. Artificial intelligence in medicine: chances and challenges for wide clinical adoption. Visc Med 2020 Dec;36(6):443-449 [FREE Full text] [doi: 10.1159/000511930] [Medline: 33442551]
81. Prasser F, Kohlmayer F, Lautenschläger R, Kuhn KA. ARX--A comprehensive tool for anonymizing biomedical data. AMIA Annu Symp Proc 2014;2014:984-993 [FREE Full text] [Medline: 25954407]
82. Schneider T, Tkachenko O. Towards efficient privacy-preserving similar sequence queries on outsourced genomic databases. In: Proceedings of the 2018 Workshop on Privacy in the Electronic Society. 2018 Presented at: CCS '18: 2018 ACM SIGSAC Conference on Computer and Communications Security; October 15, 2018; Toronto, Canada p. 71-75. [doi: 10.1145/3267323.3268956]
83. Raber D, Wortmann F, Winter R. Towards the measurement of business intelligence maturity. In: Proceedings of the 21st European Conference on Information Systems. 2013 Presented at: 21st European Conference on Information Systems; Jun 6 - 8, 2013; Utrecht, Netherlands URL: https://aisel.aisnet.org/ecis2013_cr/95/
84. Maturity Model. 2021. URL: https://ml-maturity-clinics.netlify.app/ [accessed 2021-09-09]
85. Ply JK, Moore JE, Williams CK, Thatcher JB. IS Employee attitudes and perceptions at varying levels of software process maturity. MIS Q 2012;36(2):601-624. [doi: 10.2307/41703469]
86. Pumplun L, Fecho M, Wahl-Islam N, Buxmann P. Machine learning systems in clinics–how mature is the adoption process in medical diagnostics? In: Proceedings of the 54th Hawaii International Conference on System Sciences. 2021 Presented at: 54th Hawaii International Conference on System Sciences; January 5, 2021; Kauai, Hawaii, USA. [doi: 10.24251/HICSS.2021.762]

Abbreviations

AI: artificial intelligence
HIT: health information technology
IS: information systems
ML: machine learning
NASSS: nonadoption, abandonment, scale-up, spread, and sustainability

Edited by R Kukafka; submitted 02.04.21; peer-reviewed by D Valdes, S Pandey; comments to author 11.05.21; revised version received 25.06.21; accepted 27.07.21; published 15.10.21.

Please cite as:
Pumplun L, Fecho M, Wahl N, Peters F, Buxmann P
Adoption of Machine Learning Systems for Medical Diagnostics in Clinics: Qualitative Interview Study
J Med Internet Res 2021;23(10):e29301
URL: https://www.jmir.org/2021/10/e29301
doi:10.2196/29301
PMID:34652275

Viewpoint

Harnessing Machine Learning to Personalize Web-Based Health Care Content

Ahmad Guni[1,2], BSc, MBBS; Pasha Normahani[1,2], BSc, MBBS, MSc, MRCS; Alun Davies[1,2], MA, DM, DSc, FRCS, FHEA, FACPh, FLSW; Usman Jaffer[1,2], BSc, MSc, PhD, FRCS

[1]Department of Surgery and Cancer, Imperial College London, London, United Kingdom

[2]Imperial Vascular Unit, Imperial College Healthcare NHS Trust, London, United Kingdom

Corresponding Author:
Ahmad Guni, BSc, MBBS
Department of Surgery and Cancer
Imperial College London
Exhibition Road
London, SW7 2AZ
United Kingdom
Phone: 44 7803434969
Email: ahmad.guni@nhs.net

Abstract

Web-based health care content has emerged as a primary source for patients to access health information without direct guidance from health care providers. The benefit of this approach is dependent on the ability of patients to access engaging high-quality information, but significant variability in the quality of web-based information often forces patients to navigate large quantities of inaccurate, incomplete, irrelevant, or inaccessible content. Personalization positions the patient at the center of health care models by considering their needs, preferences, goals, and values. However, the traditional methods used thus far in health care to determine the factors of high-quality content for a particular user are insufficient. Machine learning (ML) uses algorithms to process and uncover patterns within large volumes of data to develop predictive models that automatically improve over time. The health care sector has lagged behind other industries in implementing ML to analyze user and content features, which can automate personalized content recommendations on a mass scale. With the advent of *big data* in health care, which builds comprehensive patient profiles drawn from several disparate sources, ML can be used to integrate structured and unstructured data from users and content to deliver content that is predicted to be effective and engaging for patients. This enables patients to engage in their health and support education, self-management, and positive behavior change as well as to enhance clinical outcomes.

(J Med Internet Res 2021;23(10):e25497) doi:10.2196/25497

KEYWORDS

internet; online health information; personalized content; patient education; machine learning

Introduction

The internet is a key medium in the consumption of health care–related content. Two-thirds of internet users in the United Kingdom and the United States access health-related information on the internet [1-3]. Furthermore, patients are increasingly motivated and able to participate in developing this growing repository of information by sharing their lived experiences [4]. Health care professionals also consume as well as create and share web-based health care information [5].

A vast array of web-based content types and delivery media and channels are available, including videos, webpages, podcasts, images, online discussion groups and communities, and social media [6-10]. A systematic review reporting on the assessment of web-based content quality identified key domains, including accuracy, completeness, accessibility, presentation, and design, which were important overall in determining how useful and engaging content was for patients [11]. However, there is considerable variability in the quality of such content [12]. Berland et al [13] demonstrated that using search engines for common health conditions retrieved relevant content in only one out of five searches, suggesting that patients are likely to come across irrelevant content when seeking information about health. Moreover, only half of the topics that physicians thought important to convey were accurately and appropriately covered [13].

The concept of content personalization is a powerful approach that addresses the previously described features of quality by presenting the user with relevant information that is both appropriate and engaging. A more engaged patient is more likely to understand information about their health, partake in healthy behaviors, and adhere to treatment, leading to better health outcomes [14].

Machine learning (ML) is a subset of artificial intelligence that uses algorithms to study patterns in data and develop models that improve predictions about the data over time through supervised learning, unsupervised learning, or reinforcement learning [15]. Many industries use ML techniques to analyze accrued *big data* to personalize content for users [16]. The health care sector may be well served by considering these advances in other industries to personalize experiences for people seeking health care content. ML-assisted personalization can be considered for both large groups and populations or for individuals.

In this review, we aim to first outline why the health care sector should recognize the importance of personalizing content (*Why Personalizing Web-Based Health Care Content Is Important*). We then explore the current landscape of content personalization (including ML and non-ML) both within and outside health care (*Content Personalization*). Finally, we discuss practical applications of personalization in health care, outline a model that demonstrates how ML can personalize web-based content, and consider the anticipated benefits and drawbacks (*Potential Lessons to Learn for Health Care*).

Why Personalizing Web-Based Health Care Content Is Important

There has been an increased focus on empowering patients to engage with their own health. The delivery of information to patients has been recognized as a tenet of health care policy, resulting in almost universally positive outcomes for patients, health care staff, and communities [17]. The UK National Health Service Five Year Forward View outlined the need to facilitate patient activation by improving access to information, supporting self-management, and increasing patient control over the care they receive, with particular emphasis placed on harnessing digital technology [18]. This aligns with the *patient-centered* model [19], which improves patient satisfaction, quality of life, and quality of care provided [20]. Personalization facilitates the patient-centered model by delivering health care content that accounts for the preferences and needs of individual patients. The proliferation of easily accessible web-based content provides an opportunity to enable *patient-centered* information delivery at scale.

A randomized controlled trial of the provision of computer-based information to cancer patients reported that patients preferred to receive personalized information (based on their medical records) as opposed to generalized information [21]. They were more likely to share these resources with family members, and additionally, this approach was associated with a reduction in anxiety levels. A similar effect was demonstrated with personalization of booklets [22] and tailored information packs [23].

It is well established that the health care content needs to vary between different patients and also change over time. Uncertainty, the inability to determine the meaning of illness-related events, has been shown to have a deleterious effect on patient experience and outcomes [24,25]; therefore, timely and accurate delivery of information is important to address information needs. However, patients' information needs vary according to stage of disease, stage of patient journey, age, previous experiences, and coping styles [26]. A blanket *one-size-fits-all* strategy for designing and delivering health care content is unlikely to be effective.

Another advantage of personalizing health care content is its potential to improve health-related choices. One of the principles of the *patient-centered* model is sharing responsibility for clinical decisions with patients (shared decision-making) [27]. Patient decision aids are evidence-based tools designed to assist in shared decision-making. They facilitate information exchange by helping patients understand the clinical conditions and the available options for treatment. They have been demonstrated to improve patient knowledge and facilitate decision-making that is more aligned with patient values and preferences [28,29].

A study on improving patient decision-making related to prostate cancer screening found that personalizing a patient decision aid based on a number of factors that patients considered important (eg, survival, unnecessary biopsy, overdiagnosis, quality of life, burden of treatment, and burden on caregivers) improved patient opinion on screening and the quality of their decision [30]. Decision quality was assessed using an instrument that allows patients to self-rate and weigh separate elements of decision quality, including the perceived clarity of options provided, relative importance and likelihood of possible outcomes, trust toward the information delivered, support received throughout the decision-making process, sense of control over the decision, and commitment toward acting on the decision [31].

It is increasingly recognized that delivering health information without consideration for personalization and the relevance of content experienced limits the potential to change health behavior [32,33]. A meta-analysis on behavior choices from 40 web-based interventions, which used personalized strategies including interactive multimedia content, tailored feedback, discussion groups, and personalized management plans, showed a positive impact on behavior outcomes related to smoking cessation, alcoholism, physical activity, diet, and chronic disease management [34]. These findings are corroborated by other meta-analyses evaluating tailored content for similar health-related behavior outcomes [35-38]. However, given the significant heterogeneity in the intervention modality, design, and features, it is challenging to identify the specific factors that are most associated with behavior change.

With a greater understanding of these factors, there is significant scope to integrate personalized content into both large-scale public health initiatives as well as individual treatment plans to encourage self-management, adherence to treatment, and positive lifestyle changes.

Content Personalization

Content Personalization in Health Care—Current State

The paths patients take to encounter web-based content can be described by a number of discrete patient journeys. First, patients can independently find web-based information using internet search engines. Although this offers patients a plethora of information, quality (as previously discussed in the *Introduction* section) is variable [12]. Without strict content moderation and regulation, patients may struggle to parse out factual and relevant content, instead relying on content that is superficially engaging (*clickbait*) or appears credible. Furthermore, subtle differences in search terms can significantly alter the quality of the retrieved information [39].

Health care organizations and services hold repositories of quality-controlled content and can serve as gateway sites for other similar websites [40]. These provide credible and accurate information but hold limited quantities of content and may not be directly relevant to every patient. Health care professionals can assess individual information-seeking needs during consultations and refer patients to high-quality and engaging content [5]. However, this solution lacks scalability because most web-based health care searching encounters are *unsupervised* by health care providers. Limitations on how patients access health information can be addressed with content personalization, which mandates an understanding of what factors may be important in personalizing content.

Patients' information needs are affected by several factors that may influence how patients respond to web-based content, as discussed in *Why Personalizing Web-Based Health Care Content Is Important* section. For example, in the context of age, older patients often report difficulty in accessing useful web-based content because of complex website layouts, lack of navigational aids or instructional tools, and too much information being presented [41]. Younger patients may be more prone to uncertainty and worry about their health, resulting in information-seeking behavior [42]. A study that allowed cancer patients to self-tailor web-based educational content based on text, visual, and audio-visual modes demonstrated increased satisfaction among younger patients in comparison with nontailored content [43].

With regard to factors that affect the decision to select or reject web-based content, a study found several content and design features that influenced whether patients trust web-based information related to hormone replacement therapy [44]. An initial poor impression of design factors—including inappropriate website name, complex layout, poor navigation aids, dull design, small print, and excessive text—constituted 94% of cited reasons for rejection. Content features were then comparatively more important in selecting trustworthy websites. This consisted of informative content, accessible explanations, illustrations, breadth of topics covered, unbiased information, age-related information, clear language, discussion groups, and a frequently answered questions section. Source factors were also key, such as explicit author or organization credibility and authors with similar social identities.

Other studies have evaluated the design and content factors that influence patients' engagement in web-based videos, particularly on the video streaming website YouTube, which is one of the most popular websites with over 2 billion daily views [45]. These include educational resources on a range of medical topics for both patients and health care professionals [46-51]. These studies also assessed the quality of content uploaded on YouTube, which is not strictly regulated and is liable to misinformation [7]. However, the correlation between engagement and quality of content is conflicting [7], suggesting that other factors are important for gaining user attention in educational resources.

An analysis of 390 scientific communication videos on YouTube found that user-generated content, videos with regular presenters, and rapidly paced videos were more engaging than their counterparts [52].

Similarly, another study concluded that patient experience videos were more popular than videos created by health care professionals, as assessed by the video power index [53]. The video power index is an innovative tool that measures video performance by assessing its effectiveness on all platforms, comparing it with industry leaders, and aiding strategies to engage target audiences [54]. In terms of webpage content, Finnegan et al [55] found that engaging content categories were first-person narrative articles, articles that answer questions posed by readers, and articles with videos embedded in the webpage. These are all potential factors that can be considered when personalizing video content toward patients.

Sorice et al [56] examined patients' preferred social media content related to plastic surgery on six social media platforms (Facebook, Instagram, Pinterest, Snapchat, Twitter, and YouTube) [56]. Patients used Facebook and YouTube as the most favored posts relating to before and after photographs and the surgery practice information. Second, the content that engaged plastic surgeons and patients differed. The authors concluded that this information should guide the web-based activity of plastic surgeons to effectively target the desired patients.

A systematic review evaluating factors associated with engaging web-based content revealed the following key categories: textual information, discussion boards and web-based groups, video content, visual or pictographs, device accessibility, stage of patient journey, credibility, and completeness of information [57]. A framework was developed for each category describing the factors that should be considered when designing an effective content. Evidently, the manner in which users engage with health care content is influenced by both design and content factors, many of which are likely not yet identified.

Content Personalization Outside Health Care—Current State

With increasing volumes of web-based data available for extraction, storage, and processing, ML is useful in improving the efficiency and accuracy of data processing models without human input. Its application spans a wide range of disciplines, including marketing, engineering, computer science, finance, bioinformatics, and health care. In the context of personalizing

health care content, ML applications may fall into the following categories: facilitating market segmentation, content analysis, and recommender systems.

In marketing, maximizing user—or customer—engagement is obviously a key driver. Customer segmentation and personalization of content in these segments in a competitive environment is easy to appreciate. Furthermore, 59% of customers believe that personalization influences their purchasing habits [58]. A study reporting over 30,000 campaigns by one company revealed that targeted campaigns resulted in greater customer retention, engagement, and conversion into active users compared with generic campaigns [59]. Audience segmentation for web-based marketing aims to split the customer population based on characteristic features (eg, demographic, psychographic, geographic, behavior, and product preference) [60]. Individual customer segments can be targeted with specific content and products predicted to elicit the most attention, resulting in sales and profits [61]. However, customer segmentation performed by human marketers is limited by the amount of data that can be amassed, analytical methods that can be used, and the number of conclusions drawn. ML using clustering techniques can process larger volumes of data and uncover complex patterns to draw more practical conclusions and create better-defined segments for targeting. Infamously, this approach can also be used to target groups with messages that may affect behavior, such as political elections [62], but is less likely to be a useful method to personalize health care content for individuals, as there will still be differences in the needs and preferences of individuals within segments.

Recommender systems are used by the entertainment, e-commerce, and marketing industry to personalize content discovery and information retrieval in the context of massive item repositories [63-66]. Established methods include collaborative filtering, which applies the behavior of similar users to suggest new items of interest; content-based methods, which analyze content similarities with previous user preferences to produce recommendations; and hybrid methods, which combine both. Although the research landscape has predominantly focused on collaborative filtering [67], increased interest has gathered around content-based filtering with techniques emerging to identify content features [68], including user-generated tags and reviews [69], and advances in video [70] and image [71] analysis capabilities.

As one of the largest platforms for creating and sharing content, the YouTube recommender system uses deep learning to generate and rank candidate videos by incorporating a rich set of user and video features, such as the user's history, context, and interaction with similar videos [72]. This facilitates access to a small set of engaging personalized content from an ever-increasing repository of videos. Other studies have demonstrated several content factors that can also influence personalization. For example, a study incorporated textual content features including video metadata and nontextual features consisting of audio, scenes, and motion to enhance personalized recommendations for videos; this was more accurate in effective personalized video recommendation from large video data sets (Netflix and MovieLens) over existing models that use single specific content features [73].

Social media recommender systems provide insights into how companies personalize other media content discovery for users. Instagram analyzes content that users have previously interacted with and uses natural language processing to identify similar accounts to recommend content that the user is likely to interact with on their *Explore* page [74]. In addition, content analysis of social media pages reveals several factors that also influence user engagement and may further refine content personalization. In a study on over 13,000 Instagram posts, using an image application programming interface (API) to extract visual features from posts, several creator-related, context, and content factors predicted user engagement [75]. In particular, images containing people, scenery, and emoticons associated with positive emotions engaged users more strongly. Other content features on Instagram that correlate with user attention are photos with faces [76] and filters enhancing warmth, exposure, and contrast [77]. An analysis using a natural language processing API on over 100,000 messages on Facebook found that emotional and philanthropic content enhances engagement, whereas informative content reduces engagement in isolation, but further invokes attention when combined with persuasive features [78].

Advances in recommender systems have further improved the personalized recommendations. For instance, movie recommender systems traditionally use higher semantic features (eg, tags, plot, genre, and actors) suggested by users or experts to personalize recommendations [79]. A recent work using a deep learning neural network found that extracting low-level stylistic features (eg, colors, texture, and lighting) outperformed traditional semantic-based methods in recommending content [70]. With developments in algorithmic approaches and deep learning [68], high- and low-level content features can be integrated to generate more personalized content recommendations.

Recently, open-source services that leverage ML have become available on commercial platforms with the Google Cloud Artificial Intelligence as a foremost example [80,81]. These services require minimal ML expertise and consist of custom models using AutoML and pretrained models, which include video intelligence API (analyze video metadata), natural language API (analyze text), vision API (image segmentation and classification), and speech API (transcribing audio). Similar platforms exist with Amazon Rekognition image and video analysis [82], Microsoft Azure video indexer, text analytics and personalizer [83], and IBM Watson video content analysis and natural language understanding [84]. Amazon's predictive user engagement service offers to improve user engagement by analyzing real-time activity to personalize recommendations and notifications for users [85,86]. The prospect of designing custom ML may have been prohibitive for many industries previously, but these open-source platforms provide an opportunity to adopt it into the mainstream of a variety of disciplines for large-scale data processing.

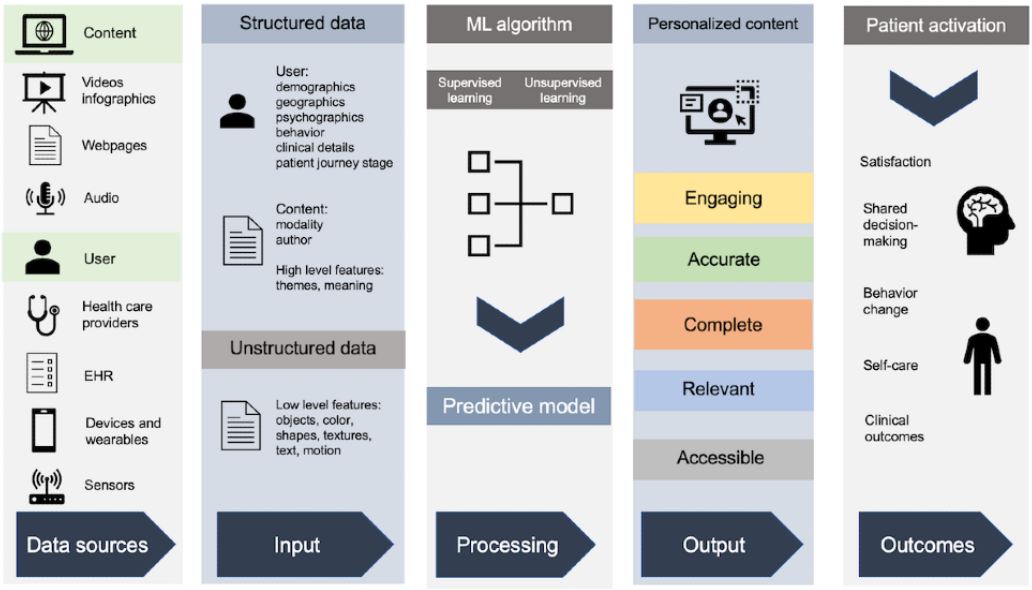
Potential Lessons to Learn for Health Care

The previous sections described user segmentation, targeted advertisements, and personalization based on recommender systems using ML techniques. With the vast amount of web-based health care content readily accessible to patients, cross-disciplinary collaboration and the use of open-source platforms indicate that these techniques may be feasible. If this is achieved, the aim of personalizing web-based content and enhancing outcomes is possible. However, clinical studies and clinical applications related to this are sparse.

Big data in health care can transform the field of health marketing (an established concept in public health medicine), drawing principles from traditional marketing to create, communicate, and deliver information in a patient-centered manner [87]. This aims to identify population segments and market health care messages to them in terms of the segments that are likely to respond [88]. A systematic review of health marketing research identified a number of studies that used hierarchical and nonhierarchical clustering techniques to segment health consumers in unique ways [88]. However, the studies did not explore whether these segments were meaningful (predictive segmentation) or whether personalized interventions affected outcomes. Furthermore, there was a reliance on rudimentary data such as survey, service, and basic clinical data, which limits the clustering process as opposed to truly *big* data. Although these strategies may have beneficial effects for groups of people, it is difficult to imagine their utility to individuals.

We propose a model that leverages ML algorithms to personalize content for an individual person (Figure 1). Health care big data consists of diverse data types, including clinical data, electronic patient records, biometrics, sensor-generated data, population data, social media posts, and webpages [16]. Electronic health records are accumulating data at an exponential rate. With the increasing use of medical devices, sensors, wearable technology, and social media, more personal data can be recorded [89]. These consist of potential sources of structured and unstructured data that may be fed into ML algorithms. Structured data include labeled user features such as demographics, geographics, psychographics, behavior, and clinical details, as well as content features consisting of modality, themes, and author information. Unstructured data, comprising 80% of all health care data [90], can be processed by video, image, and natural language processing APIs into structured formats [91]. ML algorithms using supervised and unsupervised learning can process these data to produce a predictive model for content personalization.

Figure 1. Suggested model using machine learning to personalize web-based health care content. EHR: electronic health record; ML: machine learning.



User features can be matched to content features (whether video, text, infographics, or audio) to create a model predicting which content is likely to be engaging to which people. Content features need not be limited to the content or design features identified in the *Content Personalization* section. Meta-level information encompassing object identification (colors, shapes, and texture), person or face identification, motion features, patterns, textual analysis, medical tags, higher semantic meaning, and significantly more may be extracted and analyzed. The content for patients can be created with these specific features in mind. Recommender systems could automatically predict other content that is useful and engaging to patients, conveying education that is likely to affect them.

Metrics related to view count, likes, shares, and positive comments have traditionally been used as an indicator of popularity, but they may only provide a superficial measure of engagement and fail to capture key outcomes for patients. Similarly, no single outcome metric is likely to be sufficient. Possible surrogate measures to consider include shared decision-making [27], patient satisfaction [92], objective clinical outcomes and symptoms [93], changes in attitude and behavior [94,95], and physiological signals [96]. These factors can aid in content personalization.

Harnessing data from personal digital devices such as wearables, phones, and computers has led to research into digital phenotyping and personal sensing, which refer to the analysis

of data streams from personal devices to build a human phenotype by identifying behaviors, traits, thoughts, and feelings [97,98]. This field has been adopted predominantly in psychiatry, where the objective identification of behavior patterns can aid in the diagnosis and stratification of mental health conditions, as well as their treatment (digital health interventions) [97]. In a recent study of internet-based cognitive behavioral therapy, ML was used to identify different behavior patterns among segments of patients, consisting of low engagers, late engagers, high engagers with rapid disengagement, and the highest engagers [99]. Each patient subtype was more likely to engage with different intervention tools (eg, core modules, goal-based activities, mood trackers, and mindfulness tools), leading to varying improvements in depression and anxiety symptoms. The authors concluded that this information could be used to tailor specific intervention types to different patient subtypes to improve engagement and adherence to treatment.

There are clear similarities between these digital health concepts and the proposed model to personalize web-based health care content. In particular, ML can be used to analyze data streams that include sensor measurements, user activity on personal devices, and user-generated content to identify individual behavior patterns. This can then be used to personalize interventions, of which personalized content could form a part of the intervention, or, at the very least, to inform patients about their health and engage them in making healthy behavior choices.

The successful implementation of big data and ML in personalizing web-based content requires the input and collaboration of several multidisciplinary stakeholders [100]. Health care professionals must produce accurate and engaging user-centered content, which is consumed by patients who can use recommender systems to discover related content and are also able to create content on their own. ML algorithms based on the model described in Figure 1 were designed by computer scientists and ML engineers and further optimized by several data streams provided by patients and health care organizations. There should be ongoing collaborative research between clinicians and computer scientists to take advantage of developments in ML, such as the use of deep learning.

However, current inadequacies in the digital infrastructure of health care systems can pose a significant challenge to this process. For example, as outlined in the UK government policy paper on their future digital strategy plan [101], patient data are often stored in disparate systems between different hospitals and health care settings that are unable to communicate with each other. One of the priorities should, therefore, be to create data standards that facilitate the interoperability of patient health records, which would enable seamless access, storage, and processing at scale. It is promising that government agencies have already taken steps to outline frameworks to achieve secure access, interoperability, and sharing of health-related patient data [101,102].

Other drawbacks of big data and personalized health care must also be considered in addition to the benefits. Maintaining the privacy and security of sensitive patient data is paramount and poses significant challenges with the volume of data recorded from an increasing number of sources. No single legal or ethical framework covers all aspects of health information privacy [103]. Furthermore, many laws are outdated and insufficient for the current era of big data, which includes user-generated data (eg, wearables and sensors) and nonhealth information that can lead to health inferences (eg, social media habits) [104]. Therefore, governments and health care bodies must also act as key stakeholders to ensure that laws are updated to allow ML to be harnessed for the benefit of patients while maintaining privacy and security. This may necessitate the development of oversight agencies to strictly regulate the use of ML, as well as collaboration with cybersecurity experts [100]. The principles of consent in digital data research and use need to be established and will require input from governments, national data regulators, medical ethicists, legal experts, and, most importantly, patients [105].

There are several principles for maintaining private and secure data, including collecting data from trusted sources, encrypting and anonymizing stored data, maintaining strict authorization and access control, and securing processing environments [106]. However, a cybersecurity report in 2016 revealed a 320% year-on-year increase in breaches of protected health information in US hospitals, with 81% of breached records resulting from hacking attacks [107]. This compromised over 16 million individual patient health records, indicating a pressing need to continue monitoring and developing security systems in the face of both malicious and unintentional data breaches.

Conclusions

The proliferation of web-based content and increased participation of patients in interacting with said content provides an opportunity to understand what features of content are engaging to people. Harnessing ML technologies to process *big data* in health care will allow health care providers and other users to create and contribute to personalized content. These insights may be leveraged to facilitate patient activation and enable patients to make healthy choices, ultimately improving outcomes.

Authors' Contributions

UJ constructed themes for the manuscript. AG and PN wrote the paper. PN, AHD, and UJ critically revised the manuscript. All authors reviewed and approved the manuscript.

Conflicts of Interest

None declared.

References

1. O'Neill B, Ziebland S, Valderas J, Lupiáñez-Villanueva F. User-generated online health content: a survey of internet users in the United Kingdom. J Med Internet Res 2014;16(4):e118 [FREE Full text] [doi: 10.2196/jmir.3187] [Medline: 24784798]
2. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, et al. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. Arch Intern Med 2005;165(22):2618-2624. [doi: 10.1001/archinte.165.22.2618] [Medline: 16344419]
3. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]
4. Kamel BMN, Wheeler S. The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. Health Info Libr J 2007 Mar;24(1):2-23. [doi: 10.1111/j.1471-1842.2007.00701.x] [Medline: 17331140]
5. Podichetty VK, Booher J, Whitfield M, Biscup RS. Assessment of internet use and effects among healthcare professionals: a cross sectional survey. Postgrad Med J 2006 Apr;82(966):274-279 [FREE Full text] [doi: 10.1136/pgmj.2005.040675] [Medline: 16597816]
6. Pershad Y, Hangge PT, Albadawi H, Oklu R. Social Medicine: Twitter in Healthcare. J Clin Med 2018 May 28;7(6) [FREE Full text] [doi: 10.3390/jcm7060121] [Medline: 29843360]
7. Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK. Healthcare information on YouTube: a systematic review. Health Informatics J 2015 Sep;21(3):173-194. [doi: 10.1177/1460458213512220] [Medline: 24670899]
8. Boulos MN, Maramba I, Wheeler S. Wikis, blogs and podcasts: a new generation of web-based tools for virtual collaborative clinical practice and education. BMC Med Educ 2006;6:41 [FREE Full text] [doi: 10.1186/1472-6920-6-41] [Medline: 16911779]
9. Jin J, Yan X, Li Y, Li Y. Int J Med Inform 2016 Feb;86:91-103. [doi: 10.1016/j.ijmedinf.2015.11.002] [Medline: 26616406]
10. Rupert DJ, Moultrie RR, Read JG, Amoozegar JB, Bornkessel AS, O'Donoghue AC, et al. Perceived healthcare provider reactions to patient and caregiver use of online health communities. Patient Educ Couns 2014 Sep;96(3):320-326. [doi: 10.1016/j.pec.2014.05.015] [Medline: 24923652]
11. Eysenbach G, Powell J, Kuss O, Sa E. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. J Am Med Assoc 2002;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
12. Benigeri M, Pluye P. Shortcomings of health information on the internet. Health Promot Int 2003 Dec;18(4):381-386. [doi: 10.1093/heapro/dag409] [Medline: 14695369]
13. Berland GK, Elliott MN, Morales LS, Algazy JI, Kravitz RL, Broder MS, et al. Health information on the Internet: accessibility, quality, and readability in English and Spanish. J Am Med Assoc 2001;285(20):2612-2621 [FREE Full text] [doi: 10.1001/jama.285.20.2612] [Medline: 11368735]
14. Hibbard JH, Greene J. What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs. Health Aff (Millwood) 2013 Feb;32(2):207-214. [doi: 10.1377/hlthaff.2012.1061] [Medline: 23381511]
15. Murphy K. Machine Learning: A Probabilistic Perspective. Cambridge, Massachusetts: MIT press; 2012.
16. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2:3 [FREE Full text] [doi: 10.1186/2047-2501-2-3] [Medline: 25825667]
17. Farrell C. Patient and public involvement in health : the evidence for policy implementation. Department of Health. 2004. URL: https://www.semanticscholar.org/paper/Patient-and-Public-Involvement-in-Health-%3A-The-for-Farrell/19d9d7a92bf348b04179b2b83a70cbc77ef6773d [accessed 2021-07-24]
18. Five year forward view. National Health Sevice. 2014. URL: https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf [accessed 2020-10-22]
19. Levenstein JH, McCracken EC, McWhinney IR, Stewart MA, Brown JB. The patient-centred clinical method. 1. A model for the doctor-patient interaction in family medicine. Fam Pract 1986 Mar;3(1):24-30. [doi: 10.1093/fampra/3.1.24] [Medline: 3956899]
20. Rathert C, Wyrwich MD, Boren SA. Patient-centered care and outcomes: a systematic review of the literature. Med Care Res Rev 2013 Aug;70(4):351-379. [doi: 10.1177/1077558712465774] [Medline: 23169897]
21. Jones R, Pearson J, McGregor S, Cawsey AJ, Barrett A, Craig N, et al. Randomised trial of personalised computer based information for cancer patients. Br Med J 1999 Nov 06;319(7219):1241-1247 [FREE Full text] [doi: 10.1136/bmj.319.7219.1241] [Medline: 10550090]
22. Jones RB, Pearson J, Cawsey AJ, Bental D, Barrett A, White J, et al. Effect of different forms of information produced for cancer patients on their use of the information, social support, and anxiety: randomised trial. Br Med J 2006 Apr 22;332(7547):942-948 [FREE Full text] [doi: 10.1136/bmj.38807.571042.68] [Medline: 16597660]
23. O'Connor G, Coates V, O'Neill S. Randomised controlled trial of a tailored information pack for patients undergoing surgery and treatment for rectal cancer. Eur J Oncol Nurs 2014 Apr;18(2):183-191. [doi: 10.1016/j.ejon.2013.10.011] [Medline: 24275208]
24. Mishel MH. Reconceptualization of the uncertainty in illness theory. Image J Nurs Sch 1990;22(4):256-262. [doi: 10.1111/j.1547-5069.1990.tb00225.x] [Medline: 2292449]

25. Arora NK. Interacting with cancer patients: the significance of physicians' communication behavior. Soc Sci Med 2003 Sep;57(5):791-806. [doi: 10.1016/s0277-9536(02)00449-5] [Medline: 12850107]
26. Nanton V, Docherty A, Meystre C, Dale J. Finding a pathway: information and uncertainty along the prostate cancer patient journey. Br J Health Psychol 2009 Sep;14(Pt 3):437-458. [doi: 10.1348/135910708X342890] [Medline: 18718111]
27. Barry MJ, Edgman-Levitan S. Shared decision making--pinnacle of patient-centered care. N Engl J Med 2012 Mar 01;366(9):780-781. [doi: 10.1056/NEJMp1109283] [Medline: 22375967]
28. Normahani P, Sounderajah V, Harrop-Griffiths W, Chukwuemeka A, Peters NS, Standfield NJ, et al. Achieving good-quality consent: review of literature, case law and guidance. BJS Open 2020 May 31:757-763 [FREE Full text] [doi: 10.1002/bjs5.50306] [Medline: 32475083]
29. Stacey D, Légaré F, Col NF, Bennett CL, Barry MJ, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2014;1:CD001431. [doi: 10.1002/14651858.CD001431.pub4] [Medline: 24470076]
30. Salkeld G, Cunich M, Dowie J, Howard K, Patel MI, Mann G, et al. The role of personalised choice in decision support: a randomized controlled trial of an online decision aid for prostate cancer screening. PLoS One 2016;11(4):e0152999 [FREE Full text] [doi: 10.1371/journal.pone.0152999] [Medline: 27050101]
31. Kaltoft M, Cunich M, Salkeld G, Dowie J. Assessing decision quality in patient-centred care requires a preference-sensitive measure. J Health Serv Res Policy 2014 Apr;19(2):110-117 [FREE Full text] [doi: 10.1177/1355819613511076] [Medline: 24335587]
32. Lustria ML, Cortese J, Noar SM, Glueckauf RL. Computer-tailored health interventions delivered over the Web: review and analysis of key components. Patient Educ Couns 2009 Feb;74(2):156-173. [doi: 10.1016/j.pec.2008.08.023] [Medline: 18947966]
33. Zufferey MC, Schulz PJ. Self-management of chronic low back pain: an exploration of the impact of a patient-centered website. Patient Educ Couns 2009 Oct;77(1):27-32. [doi: 10.1016/j.pec.2009.01.016] [Medline: 19321286]
34. Lustria ML, Noar SM, Cortese J, Van Stee SK, Glueckauf RL, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
35. Kroeze W, Werkman A, Brug J. A systematic review of randomized trials on the effectiveness of computer-tailored education on physical activity and dietary behaviors. Ann Behav Med 2006 Jun;31(3):205-223. [doi: 10.1207/s15324796abm3103_2] [Medline: 16700634]
36. Neville LM, O'Hara B, Milat AJ. Computer-tailored dietary behaviour change interventions: a systematic review. Health Educ Res 2009 Aug;24(4):699-720 [FREE Full text] [doi: 10.1093/her/cyp006] [Medline: 19286893]
37. Neville LM, O'Hara B, Milat A. Computer-tailored physical activity behavior change interventions targeting adults: a systematic review. Int J Behav Nutr Phys Act 2009;6:30 [FREE Full text] [doi: 10.1186/1479-5868-6-30] [Medline: 19490649]
38. Strecher VJ. Computer-tailored smoking cessation materials: a review and discussion. Patient Educ Couns 1999 Feb;36(2):107-117. [doi: 10.1016/s0738-3991(98)00128-1] [Medline: 10223016]
39. Fabricant P, Dy C, Patel R, Blanco J, Doyle S. Internet search term affects the quality and accuracy of online information about developmental hip dysplasia. J Pediatr Orthop 2013 Jun;33(4):361-365. [doi: 10.1097/BPO.0b013e31827d0dd2] [Medline: 23653022]
40. Shepperd S, Charnock D, Gann B. Helping patients access high quality health information. Br Med J 1999 Sep 18;319(7212):764-766 [FREE Full text] [doi: 10.1136/bmj.319.7212.764] [Medline: 10488009]
41. Bolle S, Romijn G, Smets EM, Loos EF, Kunneman M, van Weert JC. Older cancer patients' user experiences with web-based health information tools: a think-aloud study. J Med Internet Res 2016 Jul 25;18(7):e208 [FREE Full text] [doi: 10.2196/jmir.5618] [Medline: 27457709]
42. Basevitz P, Pushkar D, Chaikelson J, Conway M, Dalton C. Age-related differences in worry and related processes. Int J Aging Hum Dev 2008;66(4):283-305. [doi: 10.2190/AG.66.4.b] [Medline: 18507331]
43. Nguyen MH, Smets EM, Bol N, Loos EF, van Laarhoven HW, Geijsen D, et al. Tailored web-based information for younger and older patients with cancer: randomized controlled trial of a preparatory educational intervention on patient outcomes. J Med Internet Res 2019 Oct 01;21(10):e14407 [FREE Full text] [doi: 10.2196/14407] [Medline: 31573911]
44. Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? Soc Sci Med 2007 May;64(9):1853-1862. [doi: 10.1016/j.socscimed.2007.01.012] [Medline: 17328998]
45. The top 500 sites on the web. Alexa. 2020. URL: https://www.alexa.com/topsites [accessed 2020-10-07]
46. Szmuda T, Rosvall P, Hetzger TV, Ali S, Słoniewski P. Youtube as a source of patient information for hydrocephalus: a content-quality and optimization analysis. World Neurosurg 2020 Jun;138:469-477. [doi: 10.1016/j.wneu.2020.02.149] [Medline: 32147561]
47. Szmuda T, Syed MT, Singh A, Ali S, Özdemir C, Słoniewski P. YouTube as a source of patient information for Coronavirus Disease (COVID-19): a content-quality and audience engagement analysis. Rev Med Virol 2020 Sep;30(5):e2132 [FREE Full text] [doi: 10.1002/rmv.2132] [Medline: 32537771]

48. Fode M, Nolsøe AB, Jacobsen FM, Russo GI, Østergren PB, Jensen CF, EAU YAU Men's Health Working Group. Quality of information in YouTube videos on erectile dysfunction. Sex Med 2020 Sep;8(3):408-413 [FREE Full text] [doi: 10.1016/j.esxm.2020.05.007] [Medline: 32593674]
49. Ajumobi AB, Malakouti M, Bullen A, Ahaneku H, Lunsford TN. YouTube™ as a source of instructional videos on bowel preparation: a content analysis. J Cancer Educ 2016 Dec;31(4):755-759. [doi: 10.1007/s13187-015-0888-y] [Medline: 26224241]
50. Abdulghani HM, Haque S, Ahmad T, Irshad M, Sattar K, Al-Harbi MM, et al. A critical review of obstetric and gynecological physical examination videos available on YouTube: content analysis and user engagement evaluation. Medicine (Baltimore) 2019 Jul;98(30):e16459 [FREE Full text] [doi: 10.1097/MD.0000000000016459] [Medline: 31348247]
51. Keelan J, Pavri-Garcia V, Tomlinson G, Wilson K. YouTube as a source of information on immunization: a content analysis. J Am Med Assoc 2007 Dec 5;298(21):2482-2484. [doi: 10.1001/jama.298.21.2482] [Medline: 18056901]
52. Welbourne DJ, Grant WJ. Science communication on YouTube: factors that affect channel and video popularity. Public Underst Sci 2016 Aug;25(6):706-718. [doi: 10.1177/0963662515572068] [Medline: 25698225]
53. Ferhatoglu MF, Kartal A, Ekici U, Gurkan A. Evaluation of the reliability, utility, and quality of the information in sleeve gastrectomy videos shared on open access video sharing platform YouTube. Obes Surg 2019 May;29(5):1477-1484. [doi: 10.1007/s11695-019-03738-2] [Medline: 30706318]
54. Improve video performance with shareablee's new Video Power Index (VPI). Shareablee. 2020. URL: https://www.shareablee.com/blog/2017/06/27/improve-video-performance-with-shareablee-s-new-video-power-index-vpi [accessed 2020-10-07]
55. Finnegan G, Holt D, English PM, Glismann S, Thomson A, Salisbury DM, et al. Lessons from an online vaccine communication project. Vaccine 2018 Oct 22;36(44):6509-6511. [doi: 10.1016/j.vaccine.2018.05.007] [Medline: 29921491]
56. Sorice SC, Li AY, Gilstrap J, Canales FL, Furnas HJ. Social media and the plastic surgery patient. Plast Reconstr Surg 2017;140(5):1047-1056. [doi: 10.1097/prs.0000000000003769]
57. Oktay LA, Abdelwahed A, Houbby N, Lampridou S, Normahani P, Peters N, et al. Factors affecting engagement in online healthcare patient information: a systematic review of the literature. J Med Internet Res 2020:In Press. [doi: 10.2196/preprints.19896]
58. STUDY: Rethinking retail, insights from consumers and retailers into an omni-channel shopping experience. Infosys. 2013. URL: https://www.infosys.com/newsroom/press-releases/Documents/genome-research-report.pdf [accessed 2020-10-15]
59. Appboy data analysis of over 30,000 campaigns. Braze. 2016. URL: https://www.braze.com/blog/in-app-conversion-5-tips-infographic/ [accessed 2020-10-15]
60. Cooil B, Aksoy L, Keiningham TL. Approaches to customer segmentation. J Relation Mark 2008 Jan 14;6(3-4):9-39. [doi: 10.1300/j366v06n03_02]
61. An J, Kwak H, Jung S, Salminen J, Jansen BJ. Customer segmentation using online platforms: isolating behavioral and demographic segments for persona creation via aggregated user data. Soc Netw Anal Min 2018 Aug 23;8(1):54. [doi: 10.1007/s13278-018-0531-0]
62. Murray G, Scime A. Microtargeting and electorate segmentation: data mining the American National Election Studies. J Polit Mark 2010 Jul;9(3):143-166. [doi: 10.1080/15377857.2010.497732]
63. Gomez-Uribe CA, Hunt N. The Netflix recommender system: algorithms, business value, and innovation. ACM Trans Manage Inf Syst 2016 Jan 14;6(4):1-19. [doi: 10.1145/2843948]
64. Davidson J, Liebald B, Liu J, Nandy P, Van Vleet T, Gargi U, et al. The YouTube video recommendation system. In: Proceedings of the fourth ACM conference on Recommender Systems. 2010 Presented at: RecSys '10: Fourth ACM Conference on Recommender Systems; September 26 - 30, 2010; Barcelona Spain p. 293-296. [doi: 10.1145/1864708.1864770]
65. Wei K, Huang J, Fu S. A survey of e-commerce recommender systems. In: Proceedings of the International Conference on Service Systems and Service Management. 2007 Presented at: International Conference on Service Systems and Service Management; June 9-11, 2007; Chengdu, China. [doi: 10.1109/icsssm.2007.4280214]
66. Linden G, Smith B, York J. Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Comput 2003 Jan;7(1):76-80. [doi: 10.1109/mic.2003.1167344]
67. Jannach D, Zanker M, Ge M, Gröning M. Recommender systems in computer science and information systems – a landscape of research. In: E-Commerce and Web Technologies. Berlin: Springer; 2012:76-87.
68. Lops P, Jannach D, Musto C, Bogers T, Koolen M. Trends in content-based recommendation. User Model User Adap Interact 2019 Mar 7;29(2):239-249. [doi: 10.1007/s11257-019-09231-w]
69. Chen L, Chen G, Wang F. Recommender systems based on user reviews: the state of the art. User Model User Adap Interact 2015 Jan 22;25(2):99-154. [doi: 10.1007/s11257-015-9155-5]
70. Deldjoo Y, Elahi M, Quadrana M, Cremonesi P. Using visual features based on MPEG-7 and deep learning for movie recommendation. Int J Multimed Info Retr 2018 Jun 14;7(4):207-219. [doi: 10.1007/s13735-018-0155-1]
71. McAuley J, Targett C, Shi Q, Van Den HA. Image-based recommendations on styles and substitutes. In: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2015 Presented

at: SIGIR '15: The 38th International ACM SIGIR conference on research and development in Information Retrieval; August 9 - 13, 2015; Santiago Chile p. 43-52. [doi: 10.1145/2766462.2767755]
72. Deep neural networks for YouTube recommendations. Google. 2016. URL: https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/45530.pdf [accessed 2020-10-30]
73. Du X, Yin H, Chen L, Wang Y, Yang Y, Zhou X. Personalized video recommendation using rich contents from videos. IEEE Trans Knowl Data Eng 2020 Mar 1;32(3):492-505. [doi: 10.1109/tkde.2018.2885520]
74. Powered by AI: Instagram's explore recommender system. Facebook AI. 2019. URL: https://ai.facebook.com/blog/powered-by-ai-instagrams-explore-recommender-system/ [accessed 2020-10-29]
75. Jaakonmäki R, Müller O, vom Brocke BJ. The impact of content, context, and creator on user engagement in social media marketing. ScholarSpace 2017:1-9. [doi: 10.24251/HICSS.2017.136]
76. Bakhshi S, Shamma D, Gilbert E. Faces engage us: photos with faces attract more likes and comments on Instagram. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: CHI '14: CHI Conference on Human Factors in Computing Systems; April 26 - May 1, 2014; Toronto Ontario Canada p. 965-974. [doi: 10.1145/2556288.2557403]
77. Bakhshi S, Shamma D, Kennedy L, Gilbert E. Why we filter our photos and how it impacts engagement. ICWSM. 2015. URL: http://comp.social.gatech.edu/papers/icwsm15.why.bakhshi.pdf [accessed 2021-07-24]
78. Lee D, Hosanagar K, Nair H. The effect of social media marketing content on consumer engagement: evidence from Facebook. Stanford University, Graduate School of Business, CA. 2014. URL: https://econpapers.repec.org/paper/eclstabus/3087.htm [accessed 2021-07-24]
79. Elahi M, Deldjoo Y, Bakhshandegan MF, Cella L, Cereda S, Cremonesi P. Exploring the semantic gap for movie recommendations. In: Proceedings of the Eleventh ACM Conference on Recommender Systems. 2017 Presented at: RecSys '17: Eleventh ACM Conference on Recommender Systems; August 27 - 31, 2017; Como Italy p. 326-330. [doi: 10.1145/3109859.3109908]
80. Bisong E. An overview of Google Cloud platform services. In: Building Machine Learning and Deep Learning Models on Google Cloud Platform. Berkeley, CA: Apress; 2019:7-10.
81. AI and machine learning products. Google Cloud. URL: https://cloud.google.com/products/ai [accessed 2020-10-07]
82. Amazon Rekognition. Amazon Web Services. URL: https://aws.amazon.com/rekognition [accessed 2020-10-07]
83. Azure products. Microsoft. URL: https://azure.microsoft.com/en-us/services/ [accessed 2020-10-07]
84. IBM Watson products and solutions. IBM. URL: https://www.ibm.com/watson/products-services [accessed 2020-10-07]
85. Predictive user engagement. Amazon Web Services. 2019. URL: https://docs.aws.amazon.com/solutions/latest/predictive-user-engagement [accessed 2020-10-07]
86. Predictive user engagement solution. GitHub. URL: https://github.com/awslabs/predictive-user-engagement [accessed 2020-10-07]
87. What is health marketing? Centers for Disease Control and Prevention. 2011. URL: https://www.cdc.gov/healthcommunication/toolstemplates/WhatIsHM.html [accessed 2020-10-15]
88. Swenson ER, Bastian ND, Nembhard HB. Healthcare market segmentation and data mining: a systematic review. Health Mark Q 2018;35(3):186-208. [doi: 10.1080/07359683.2018.1514734] [Medline: 30470165]
89. Maddox TM, Matheny MA. Natural language processing and the promise of big data: small step forward, but many miles to go. Circ Cardiovasc Qual Outcomes 2015 Sep;8(5):463-465. [doi: 10.1161/CIRCOUTCOMES.115.002125] [Medline: 26286870]
90. O'Dowd E. Unstructured healthcare data needs advanced machine learning tools. HIT Infrastructure. 2018. URL: https://hitinfrastructure.com/news/unstructured-healthcare-data-needs-advanced-machine-learning-tools [accessed 2020-10-26]
91. Ohno-Machado L. Realizing the full potential of electronic health records: the role of natural language processing. J Am Med Inform Assoc 2011;18(5):539 [FREE Full text] [doi: 10.1136/amiajnl-2011-000501] [Medline: 21846784]
92. Gupta D, Rodeghier M, Lis CG. Patient satisfaction with service quality as a predictor of survival outcomes in breast cancer. Support Care Cancer 2014 Jan;22(1):129-134. [doi: 10.1007/s00520-013-1956-7] [Medline: 24013568]
93. Armstrong AW, Kim RH, Idriss NZ, Larsen LN, Lio PA. Online video improves clinical outcomes in adults with atopic dermatitis: a randomized controlled trial. J Am Acad Dermatol 2011 Mar;64(3):502-507. [doi: 10.1016/j.jaad.2010.01.051] [Medline: 21236514]
94. Hoffman J, Salzman C, Garbaccio C, Burns SP, Crane D, Bombardier C. Use of on-demand video to provide patient education on spinal cord injury. J Spinal Cord Med 2011;34(4):404-409 [FREE Full text] [doi: 10.1179/2045772311Y.0000000015] [Medline: 21903014]
95. Zhao J, Han H, Zhong B, Xie W, Chen Y, Zhi M. Health information on social media helps mitigate Crohn's disease symptoms and improves patients' clinical course. Comput Hum Behav 2021 Feb;115:106588. [doi: 10.1016/j.chb.2020.106588]
96. Belle A, Hobson R, Najarian K. A physiological signal processing system for optimal engagement and attention detection. In: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW). 2011 Presented at: IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW); Nov. 12-15, 2011; Atlanta, GA, USA. [doi: 10.1109/bibmw.2011.6112429]

97. Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. NPJ Digit Med 2019;2:88 [FREE Full text] [doi: 10.1038/s41746-019-0166-1] [Medline: 31508498]
98. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017 May 08;13:23-47. [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
99. Chien I, Enrique A, Palacios J, Regan T, Keegan D, Carter D, et al. A machine learning approach to understanding patterns of engagement with internet-delivered mental health interventions. JAMA Netw Open 2020 Jul 01;3(7):e2010791 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.10791] [Medline: 32678450]
100. Char DS, Abràmoff MD, Feudtner C. Identifying ethical considerations for machine learning healthcare applications. Am J Bioeth 2020 Nov;20(11):7-17. [doi: 10.1080/15265161.2020.1819469] [Medline: 33103967]
101. The future of healthcare: our vision for digital, data and technology in health and care. Department of Health & Social Care. 2018. URL: https://www.gov.uk/government/publications/the-future-of-healthcare-our-vision-for-digital-data-and-technology-in-health-and-care [accessed 2021-01-17]
102. Exchange of electronic health records across the EU. European Commission. URL: https://ec.europa.eu/digital-single-market/en/exchange-electronic-health-records-across-eu [accessed 2021-01-11]
103. Gray EA, Thorpe JH. Comparative effectiveness research and big data: balancing potential with legal and ethical considerations. J Comp Eff Res 2015 Jan;4(1):61-74 [FREE Full text] [doi: 10.2217/cer.14.51] [Medline: 25565069]
104. Price WN, Cohen IG. Privacy in the age of medical big data. Nat Med 2019 Jan;25(1):37-43 [FREE Full text] [doi: 10.1038/s41591-018-0272-7] [Medline: 30617331]
105. No authors listed. Time to discuss consent in digital-data studies. Nature 2019 Aug;572(7767):5. [doi: 10.1038/d41586-019-02322-z] [Medline: 31367033]
106. Abouelmehdi K, Beni-Hessane A, Khaloufi H. Big healthcare data: preserving security and privacy. J Big Data 2018 Jan 9;5(1):1. [doi: 10.1186/s40537-017-0110-7]
107. 2016 Breach Report. Cynergistik. 2016. URL: https://insights.cynergistek.com/reports/2016-breach-report [accessed 2020-10-15]

Abbreviations

API: application programming interface
ML: machine learning

Edited by R Kukafka; submitted 04.11.20; peer-reviewed by P Dattathreya, A Teles; comments to author 03.01.21; revised version received 19.01.21; accepted 16.03.21; published 19.10.21.

Please cite as:
Guni A, Normahani P, Davies A, Jaffer U
Harnessing Machine Learning to Personalize Web-Based Health Care Content
J Med Internet Res 2021;23(10):e25497
URL: https://www.jmir.org/2021/10/e25497
doi:10.2196/25497
PMID:34665146

Viewpoint

Algorithm Change Protocols in the Regulation of Adaptive Machine Learning–Based Medical Devices

Stephen Gilbert[1,2], BSc, BVMS, MRes, PhD; Matthew Fenech[1,3], MD, PhD; Martin Hirsch[1,4], PhD; Shubhanan Upadhyay[1], MD; Andrea Biasiucci[5,6], PhD; Johannes Starlinger[7], MD, PhD

[1]Ada Health GmbH, Berlin, Germany

[2]Else Kröner-Fresenius Center for Digital Health, Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

[3]Una Health GmbH, Berlin, Germany

[4]Institute for AI in Medicine, University Hospital of Giessen and Marburg, Marburg, Germany

[5]Laboratory for Investigative Neurophysiology, University Hospital Center and University of Lausanne, Lausanne, Switzerland

[6]confinis ag, Sursee, Switzerland

[7]Howto Health GmbH, Berlin, Germany

Corresponding Author:
Stephen Gilbert, BSc, BVMS, MRes, PhD
Else Kröner-Fresenius Center for Digital Health
Faculty of Medicine Carl Gustav Carus
Technische Universität Dresden
Postfach 151
Fetscherstraße 74
Dresden, 01307
Germany
Phone: 49 17680396015
Email: ra.ekfz@tu-dresden.de

Abstract

One of the greatest strengths of artificial intelligence (AI) and machine learning (ML) approaches in health care is that their performance can be continually improved based on updates from automated learning from data. However, health care ML models are currently essentially regulated under provisions that were developed for an earlier age of slowly updated medical devices—requiring major documentation reshape and revalidation with every major update of the model generated by the ML algorithm. This creates minor problems for models that will be retrained and updated only occasionally, but major problems for models that will learn from data in real time or near real time. Regulators have announced action plans for fundamental changes in regulatory approaches. In this Viewpoint, we examine the current regulatory frameworks and developments in this domain. The status quo and recent developments are reviewed, and we argue that these innovative approaches to health care need matching innovative approaches to regulation and that these approaches will bring benefits for patients. International perspectives from the World Health Organization, and the Food and Drug Administration's proposed approach, based around oversight of tool developers' quality management systems and defined algorithm change protocols, offer a much-needed paradigm shift, and strive for a balanced approach to enabling rapid improvements in health care through AI innovation while simultaneously ensuring patient safety. The draft European Union (EU) regulatory framework indicates similar approaches, but no detail has yet been provided on how algorithm change protocols will be implemented in the EU. We argue that detail must be provided, and we describe how this could be done in a manner that would allow the full benefits of AI/ML-based innovation for EU patients and health care systems to be realized.

(J Med Internet Res 2021;23(10):e30545) doi:10.2196/30545

KEYWORDS

artificial intelligence; machine learning; regulation; algorithm change protocol; healthcare; regulatory framework; health care

Introduction

Automated image analysis and segmentation [1], autonomous soft tissue suturing [2], and brain-machine interfaces [3]: these are technologies that until recently were only science fiction imaginings. They now all represent the state of the art in ML-based health care tools and all share one characteristic: all these systems are trained on patient data and can be quickly and automatically improved through retraining and on the basis of new patient data. ML is a subset of AI approaches, and this Viewpoint deals with the subset of ML applications that are classified as medical devices. The concept that an ML model will remain static over time is anathema to the concept of learning. Technologies are rapidly advancing to allow true real-time machine learning, and when the regulatory regimes allow it, ML-based health care tools have the potential to "learn" from new observations and continuous use and to retrain their models fully "on the job" [4,5].

Many ML-based health care tools are classified in the European Union (EU) and United States (and most other jurisdictions) as Software as a Medical Device (referred to in this Viewpoint as ML-based SaMD, with this term used to refer to technologies that learned from patient data sets and that will be further trained after being placed on the market). Learning and updating models pose a regulatory challenge, more so for ML-based SaMD that will learn from data in real time or near real time. Changes of this type pose a new regulatory challenge: the changes will often affect the fundamental clinical safety, clinical performance, and clinical benefit of the algorithm. Should they require full regulatory reassessment, a process that generally takes many months? Alternatively, can novel, faster, robust methods of quality oversight and approval be established? This Viewpoint compares the differing proposals put forward by the US and EU regulatory bodies for adapting the existing medical device frameworks to include consideration of learning ML-based SaMD. The crux of our argument is that highly proactive responses from regulators are required. The US Food and Drug Administration (FDA) and the EU have proposed strategies. The FDA approach is structured and comprehensive, while the EU approach overlaps the US approach but lacks detail on requirements and has not involved detailed stakeholder consultation.

There is evidence that more nuanced regulation of ML-based SaMD is being developed. A 2021 position paper of the American Medical Informatics Association recommended proactive regulatory approaches to bring improvement in clinical decision support (CDS) regulation including transparency standards, real-world performance (RWP) monitoring requirements, and improved postmarket surveillance (PMS) strategies [6]. A recent external validation of a widely implemented proprietary sepsis prediction model, the Epic Sepsis Model (ESM), used in hundreds of hospitals throughout the United States, found that it has poorly identified the onset of sepsis [7]. The authors concluded that the widespread adoption of this CDS, despite its poor performance, raised fundamental concerns about sepsis management. In our view, it also raises regulatory oversight and PMS concerns. The ESM is a penalized logistic regression model, included as part of a widely used electronic health record system, and it was developed and validated based on data from 405,000 patient encounters in three health systems between 2013 and 2015 [7]. Although only limited information is publicly available about the ESM, we recognize that it is not an example of an adaptive CDS (as defined in [6]); however, the ML approach used has the potential to be used in future adaptive CDS systems. Even as an example of a static CDS system, periodic updates based on PMS/RWP monitoring would form part of the lifecycle of this CDS, greatly increasing patient safety. Many of the considerations in this Viewpoint are applicable to this example.

The regulation of ML-based SaMD has been identified as one of the more substantial barriers to their clinical adoption [8]. Explorations of ML-based medical software in the United States found that many tools are not regulated by the FDA, that there is no FDA-maintained public record/database of approved ML-based SaMD, that many devices are approved through the 510(k)-clearance route (claim of substantial equivalence to an already-approved device), and where specific clinical evidence was provided for approval, this was exclusively from retrospective rather than prospective data [7,8]. Some of the medical applications of ML discussed in [7,8] were classified by the FDA as low risk or not classified as ML-based SaMD. For low-risk applications, existing regulatory frameworks may be sufficient. However, for the higher risk class devices discussed by [8,9], and also for adaptive ML-based SaMD and autonomous applications, there is a requirement for smarter regulation in the EU, United States, and worldwide, both to ensure patient safety and to remove a hurdle to adoption and advancement of the technologies [10-13].

The Current Regulatory Framework for Learning ML-Based SaMD

ML-based health care tools with a role in individual patient diagnosis or therapy are currently regulated in the United States and EU as Software as a Medical Device. The regulation of medical devices has been included in US legislation since the 1938 Federal Food, Drug, and Cosmetic Act (FD&C Act), and more comprehensively in the 1976 Medical Device Amendments to the FD&C Act and subsequent updates [14]. In the EU, a legislative framework has been in place since the Medical Devices Directive 93/42/EEC and the Active Implantable Medical Devices Directive 90/385/EEC and was recently updated to the Medical Device Regulation (MDR) [15,16]. Both US and EU frameworks heavily rely on guidelines and harmonized norms, which define a set of standardized best practices for the development and deployment of medical devices. Neither of the current regulatory frameworks in their present form adequately considers the special properties of ML-based systems.

Historically, as hardware medical devices preceded software medical devices, the principles of a relatively static product (ie, following linear steps from initial concept to early and later stages of development, verification, validation, clinical testing, approval, and market release) were logical. As software in/as medical devices became established, the fundamental principles of medical device regulation were adopted largely unchanged

to software. Allowance was provided for special properties of software; however, the level of detail in the legislation itself was low [15] and was instead provided through a set of international standards, including standards for the software life cycle [17], software usability testing, validation, and release and update. Initially, these processes were required to proceed in a linear "waterfall" cascade that had to be followed for software updates, release, and approval [17]. Later international guidance (see [18]) provided approaches for the development of SaMD within agile frameworks, the generally recognized optimal approach to software design [19], which is an iterative and incremental model of software development. Whether using waterfall or agile approaches, the international guidelines provide methodologies to adequately verify and validate SaMD software, a prerequisite for its safety and effectiveness.

Limitations of Current Approaches

This raises the question, "does the current legislation provide a regulatory approval framework for ML-based SaMD?" Although burdensome, the current EU approach can be (and has been) adopted for ML-based SaMD (see [20]). SaMD manufacturers can optimize their software development processes to maximize their efficiency in this linear process, particularly for documenting the effects of model change on SaMD performance between updates (these are aspects of "change control," a fundamental principle of medical device quality management systems). When applied to ML-based SaMD, software verification and validation are not in themselves sufficient, as they do not ensure that the ML model is safe; it could have safety problems related either to low-quality input data or to a poorly designed ML algorithm.

Manufacturers tend to relegate information about ML model updates to software development life cycle (SDLC) activities and postmarket clinical follow-up (PMCF). Although conventional PMCF and SDLC are highly valuable activities, provided they are executed in the correct environment and phase, they are generally inadequate to address ML problems. SDLC activities focus on software design controls that do not ensure clinically acceptable ML performance (ie, a software can be perfectly written and documented and yet the ML that it hosts can fail because of ML model problems). One reason for the inadequacy of current PMCF practices for ML model updates is that they typically generate data on a sufficient number of patients only months to years after changes are made to the SaMD. In addition, historically, the data quantity typically explored in PMCF approaches has been insufficient for the requirements of modern data-driven learning algorithms. As discussed later in this Viewpoint, PMCF can be adapted to allow for the rapid gathering of detailed data related to ML model updates. The adaptation of PMCF to this purpose and the definition of a systematic "protocol" for the implications of this data stream on device regulatory status are the foundation of proposed novel regulatory approaches.

Proposed Solutions: The US FDA Action Plan

The issues of the appropriateness of the ML algorithm and input data and the safety of the derived ML model are tackled to a degree in recent standards, some of which are still under development (see for example ISO [International Organization for Standardization]/IEC [International Electrotechnical Commission] TR 24028 on trustworthiness in AI [21] and ISO/IEC DTR 24027 on bias in AI systems and AI-aided decision-making [22]) but have not yet been addressed in a joined-up fashion in legislation. However, these issues are addressed by novel and comprehensive proposals in the US FDA's 2021 action plan [23], which effectively provides a roadmap for ML model validation. The FDA has conducted a structured consultation and has published a comprehensive action plan on regulatory approval strategies for adaptive ML-based SaMD [23,24], which has been accompanied by a high degree of engagement with the themes in the literature [25-27]. The action plan does not yet fully resolve the problems described in this Viewpoint, as the action plan is not complete or implemented. Nevertheless, the proactive and open approach of the FDA is commendable.

The action plan clearly recognizes that adaptive ML-based SaMD presents a challenge to traditional approaches: "The FDA's traditional paradigm of medical device regulation was not designed for adaptive artificial intelligence and machine learning technologies" [23]. The FDA started the formal online consultation process in April 2019 [28]—with contributions guided by a well-conceived detailed consultation document—and conducted a public workshop in February 2020, followed by the publication of the action plan in January 2021 [23]. The consultation document and action plan are based on the FDA's premarket programs, and took into consideration the International Medical Device Regulators Forum's (IMDRF) medical device risk categorization principles [29], the benefit-risk framework, software modifications guidance, and the organization-based total product life cycle approach. Good machine learning practice (GMLP) principles will be used to ensure rigorous ML-based SaMD development. Algorithm changes will be transparently labeled for users, while methodologies for ensuring robustness and identification and elimination of bias will be incorporated. For each device, a two-component predetermined change control plan (PCCP) is envisioned. This will include a SaMD prespecification (SPS)—a predetermined change control plan setting out the scope of the permissible modifications—and an algorithm change protocol (ACP; note that it is the prediction model that changes, see Textbox 1), which sets out the methodology used in the ML-based SaMD to implement the defined changes within the scope of the SPS. The ACP is a step-by-step delineation of procedures to be followed so that the modification achieves its goals and the ML-based SaMD remains safe and effective. The action plan is notable for its strengths in harnessing the iterative improvement power of ML-based SaMD, while at the same time ensuring patient safety through continuous RWP monitoring. As a next step, the FDA will publish a complete draft guidance on the PCCP in 2021 [23].

Textbox 1. Definitions of machine learning algorithm software and models. These definitions are set out explicitly here as some of the regulatory discussion documents use machine learning terminology imprecisely.

Artificial intelligence/machine learning algorithm

Machine learning algorithms are mathematical procedures that are implemented in code and are run on data to create an output machine learning model. Machine learning algorithms perform pattern recognition tasks to learn from data. For example, a machine learning algorithm could be trained on physician-labeled radiographs used to develop a machine learning model for tumor detection.

Artificial intelligence/machine learning software

Machine learning software is the code (ie, programming language) implementation of the machine learning algorithm. It is possible to implement a machine learning algorithm in many alternative ways or different programming languages.

Artificial intelligence/machine learning model

The machine learning model is created by running a machine learning algorithm on data. The machine learning model represents what was learned by a machine learning algorithm. The machine learning model consists of model data and a prediction algorithm, which can be regarded as an automatically created computer program. Once created, the machine learning model can be used for a specific task (eg, the machine learning model can be applied to unlabeled radiographs to locate possible tumors).

Do the Innovative ACP-Based Approaches Adequately Ensure Safety?

The status quo that the FDA action plan will alter has the foundational principle that a medical device should be clearly defined, definitively tested, and meticulously documented before approval and then should effectively have unchanged clinical safety, performance, and benefit on the market, and this should be ensured through tight change control processes and postmarket surveillance. Any substantial change in clinical behavior would require reapproval. This framework has advantages in the simplicity of traceability and maintenance of safety oversight. A disadvantage is that this framework prohibits the rapid change of ML-based SaMD. The FDA action plan effectively proposes the same system, except that a boundary of change of the clinical behavior of the adaptive ML-based SaMD can be predefined, along with methods to oversee the degree of change and the resulting effects while on the market. If this is a paradigm shift, it is a small one—it effectively shifts the approval at the stage of readiness of a new product revision and the postmarket evaluation of change to a premarket comprehensive consideration of the changes that would be acceptable for the device.

By definition, changes that do not fall within the predefined risk-assessed thresholds are not allowed and require the normal processes of examination by the regulator before approval for the market. The FDA noted in the action plan that "stakeholders provided specific feedback about the elements that might be included in the SPS/ACP to support safety and effectiveness as the SaMD and its associated algorithm(s) change over time" and has reacted to this by promising detailed guidance on what should be included in an SPS and ACP to support the safety and effectiveness of ML-based SaMD algorithms [23]. It is the view of the authors that the overall principles set out in the FDA action plan (ie, predefining acceptable clinical safety, performance, and benefit on the market, and conducting RWP monitoring of these) represent an approach that is both rational and proportionate, and one that would ensure patient safety, provided the regulator is sufficiently involved in the oversight of RWP monitoring data and evaluation of this data in the context of the PCCP.

The EU Artificial Intelligence Act

Following a European Commission white paper on AI in February 2020 and a subsequent public consultation [30,31], the European Commission published a draft Artificial Intelligence Act in April 2021 [32]. This draft legislation lays down harmonized rules for AI applications and it extends classical EU product conformity and CE marking concepts to all "high-risk" AI applications. The draft legislation is very similar to MDR in its core approaches, which are based around product intended use and postmarket monitoring systems. All use of AI in medical devices is defined as "high-risk," and the draft legislation is designed to be compatible with MDR [16], to be overseen by the same Notified Bodies as MDR (although the detail on how oversight will operate remains to be established), and devices are to be covered by a single CE-mark representing conformity to both MDR and the Artificial Intelligence Act [16,32]. It is striking that there is no single mention of ML in MDR, its annexes, or its associated guidance (MDR [16] and Medical Device Coordination Group guidelines [33,34]), despite the fact these documents were released in 2017 or later. Essentially, the new draft Artificial Intelligence Act extends MDR [16], bringing it into the AI era.

The FDA approach had a clear and focused published proposal on ML-based SaMD regulation to frame the discussion for the public consultation. In contrast, the European Commission consultation on the 2020 white paper that preceded the draft legislation did not have an associated published proposal and was broad, bringing in all high-risk AI applications, not just health care. We studied the contributions to the consultation, and although there are some well-considered submissions, overall, there was little focused discussion on precisely how ML-based SaMD should be overseen in the EU. This lack of detail in proposals is also reflected in the draft legislation. For the first time in EU medical device legislation, the draft describes the concept of ACPs, but—unlike in the FDA action plan [23]—these are implied, rather than being specifically named or their requirements being set out in detail. Likewise, the draft legislation does not set out an analogue to the FDA's PCCP approach, although again, the need for this is implied.

The critical clause in the draft legislation is as follows:

> *In line with the commonly established notion of substantial modification for products regulated by Union harmonisation legislation, it is appropriate that an AI system undergoes a new conformity assessment whenever a change occurs which may affect the compliance of the system with this Regulation or when the intended purpose of the system changes. In addition, as regards AI systems which continue to 'learn' after being placed on the market or put into service (i.e., they automatically adapt how functions are carried out), it is necessary to provide rules establishing that changes to the algorithm and its performance that have been pre-determined by the provider and assessed at the moment of the conformity assessment should not constitute a substantial modification.*

The importance of postmarket performance monitoring is described, but no details are provided on special considerations for this in the context of ML-based SaMD: "all providers should have a post-market monitoring system in place. This system is also key to ensure that the possible risks emerging from AI systems which continue to 'learn' after being placed on the market or put into service can be more efficiently and timely addressed."

EU Regulatory Oversight: What Is Still Needed?

In September 2020, a thorough analysis of the EU legal requirements for ML-based SaMD [35] was carried out by the European medical devices trade association (European Coordination Committee of the Radiological, Electromedical and Healthcare IT Industry [COCIR]), who concluded that deployment is possible in a way that is consistent with MDR, but recommended that practical guidance should be made available, supported by the development of international standards. More specifically, the group recommends that the international standard describing software life cycle processes (IEC 62304) [17] should be updated, requiring manufacturers to define an ACP for adaptive ML-based SaMD. COCIR's recommendation has been included in the text of the new "under review" edition of the standard [36]. We agree that the updating of standards is an important stepping-stone toward a clear framework, but changes can be long in the incubation. Simply updating standards documents may not bring the clarity required by EU Notified Bodies to allow them to make the judgment calls required to approve learning ML-based SaMD. The modifications to the IEC standard are also discrete, and do not guarantee cohesiveness of the EU regulatory framework. Moreover, updates in standards are generally conducted by a narrow group of domain experts; in the example above, the expertise group will largely consist of medical device software life cycle experts. As fully acknowledged by the FDA, consultation on the design of new regulatory frameworks for adaptive ML-based SaMD should also bring together experts on postmarket surveillance, RWP measurement, clinical evaluation, and labelling, as well as patient representatives. Although not specifically stated by the FDA, we argue that experts in real-time adaptive ML approaches, which are likely to be increasingly proposed for ML-based SaMD, should also be a key part of these discussions. The awaited EU guidance was not published with the draft Artificial Intelligence Act, and as such there will be a clear legal requirement for ACPs for adaptive ML-based SaMD, but without detail on how ACPs should be implemented.

There are several major implications for the EU if standardized procedures are not provided for premarket review of adaptive ML-based SaMD, for ACP, or for manufacturer oversight of systems on the market. Unclear or unspecified regulatory approaches required to fulfil requirements could lead to frameworks that are too burdensome to be worthwhile for manufacturers to deploy their technologies in a particular region. This may put patients there at considerable disadvantage, as they may not be able to access new diagnostic, therapeutic, or preventive modalities, or may only be able to access them after a significant delay. Unclear regulatory strategies could also significantly disadvantage the growth and prosperity of EU AI businesses. Lastly, as discussed in the general context of EU and US medical device harmonization and regulation in [6,37], unclear regulatory requirements are unlikely to function to ensure safety as they will likely lead to highly uneven regulatory oversight and enforcement.

It is unclear the degree to which the detailed EU approaches to adaptive ML-based SaMD will piggyback on the results of the already well-progressed consultative process undertaken by the US FDA. Other international approaches, such as those of the IMDRF and the joint WHO/International Telecommunication Union strategy for an independent standard evaluation framework for ML model benchmarking [38], could also provide input to an EU approach; however, concerted and prompt action is required on the part of the European Commission, to consult on and define EU-specific guidelines for providers to enable them to "establish[...] that changes to the algorithm and its performance that have been pre-determined." The benefits of the US FDA approach have been discussed at length in this Viewpoint but as described in detail in a July 2020 viewpoint by Cohen et al [19], there are aspects of the US approach that cannot easily be translated to the EU. EU-specific solutions are required in three domains: (1) EU data protection considerations relating to the update problem, (2) the relatively less established system in the EU for RWP monitoring, and (3) EU differences in public perceptions and stated community values regarding the role of AI. Point 2 may be partly addressed through complaint and incident registration in the proposed EU database for stand-alone high-risk AI systems but this is not yet sufficiently defined in the draft legislation to determine this with certainty.

It is our view that the EU needs to provide specific guidelines for adaptive ML-based SaMD ACPs and RWP monitoring. Waiting for coordination to be achieved through alignment of international standards is an approach without a proven track record of success, and it is unclear whether international approaches alone are sufficient for the EU's special circumstances. This should not be the basis for the development of the EU's health care ML-based SaMD ecosystem, on which we depend to bring the benefits of health care AI to European society and its economy. What is needed is a clear standardized

approach, similar to the now 2-year-old approach of the FDA, which sets clear procedures required for ML-based SaMD approval and postmarket provider and regulatory oversight. This could be achieved, with or without a focused public consultation, through published guidance from the EU Medical Devices Coordination Group, which should bring together the aspects addressed in the FDA action plan, the COCIR report, and the developing harmonized standards [23,35,36].

The EU approach to regulation of medical devices has faced criticism for lacking both harmonization and approaches for ensuring patient safety [37]. Although both of these aspects have been improved due to the MDR [16] through the introduction of greater transparency for patients and a central vigilance and PMS database (EUDAMED), the key underlying problems of market fragmentation and lack of clarity and harmonization still exist [37]. The main issue that this Viewpoint addresses is the potential for continued lack of standardized adoption of proactive oversight of health AI, and therefore the potential for the EU Artificial Intelligence Act [32] to fail in its objectives of harmonizing AI regulation and ensuring future-ready oversight and the safety of EU ML-based SaMD. We call for a detailed action plan and public consultation, echoing the FDA's 2020 approach, that will allow manufacturers and other stakeholders in the medical community to inform the discussion on requirements and to definitively understand requirements as they relate to oversight, market surveillance, and checkpoints of safety and performance, reactive to ML-based SaMD adoption (ie, the details of ACPs and associated RWP monitoring and ACP mechanisms).

Authors' Contributions

SG wrote the initial version of the manuscript and MF, MH, SU, AB, and JS made extensive revisions and contributed to the editing process.

Conflicts of Interest

SG, MF, MH and SU are or were employees, contractors, or equity holders in Ada Health GmbH. AB is a shareholder of companies active in neurotechnology and regulatory consulting and is a paid employee of confinis ag. All authors are considered to have an interest in Ada Health GmbH.

References

1. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. BMJ 2020 Mar 25;368:m689 [FREE Full text] [doi: 10.1136/bmj.m689] [Medline: 32213531]
2. Shademan A, Decker RS, Opfermann JD, Leonard S, Krieger A, Kim PCW. Supervised autonomous robotic soft tissue surgery. Sci Transl Med 2016 May 04;8(337):337ra64-337ra64. [doi: 10.1126/scitranslmed.aad9398] [Medline: 27147588]
3. Collinger JL, Wodlinger B, Downey JE, Wang W, Tyler-Kabara EC, Weber DJ, et al. High-performance neuroprosthetic control by an individual with tetraplegia. The Lancet 2013 Feb;381(9866):557-564. [doi: 10.1016/s0140-6736(12)61816-9]
4. Nishihara R, Moritz P, Wang S, Tumanov A, Paul W, Schleier-Smith J, et al. Real-Time Machine Learning: The Missing Pieces. In: Proceedings of the 16th Workshop on Hot Topics in Operating Systems. 2017 May 10 Presented at: HotOS '17; May 7-10, 2017; Whistler, BC. [doi: 10.1145/3102980.3102998]
5. Vokinger KN, Feuerriegel S, Kesselheim AS. Continual learning in medical devices: FDA's action plan and beyond. The Lancet Digital Health 2021 Jun;3(6):e337-e338 [FREE Full text] [doi: 10.1016/s2589-7500(21)00076-5]
6. Petersen C, Smith J, Freimuth RR, Goodman KW, Jackson GP, Kannry J, et al. Recommendations for the safe, effective use of adaptive CDS in the US healthcare system: an AMIA position paper. J Am Med Inform Assoc 2021 Mar 18;28(4):677-684. [doi: 10.1093/jamia/ocaa319] [Medline: 33447854]
7. Wong A, Otles E, Donnelly JP, Krumm A, McCullough J, DeTroyer-Cooley O, et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. JAMA Intern Med 2021 Aug 01;181(8):1065-1070. [doi: 10.1001/jamainternmed.2021.2626] [Medline: 34152373]
8. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ Digit Med 2020 Sep 11;3(1):118 [FREE Full text] [doi: 10.1038/s41746-020-00324-0] [Medline: 34508167]
9. Wu E, Wu K, Daneshjou R, Ouyang D, Ho DE, Zou J. How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. Nat Med 2021 Apr;27(4):582-584. [doi: 10.1038/s41591-021-01312-x] [Medline: 33820998]
10. Higgins D, Madai VI. From Bit to Bedside: A Practical Framework for Artificial Intelligence Product Development in Healthcare. Advanced Intelligent Systems 2020 Jul 02;2(10):2000052 [FREE Full text] [doi: 10.1002/aisy.202000052]
11. Madai V, Higgins DC. Artificial Intelligence in Healthcare: Lost In Translation? arXiv Preprint posted online on July 28, 2021 [FREE Full text]
12. Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. The Lancet Digital Health 2021 Mar;3(3):e195-e203. [doi: 10.1016/s2589-7500(20)30292-2]

13. Lyell D, Coiera E, Chen J, Shah P, Magrabi F. How machine learning is embedded to support clinician decision making: an analysis of FDA-approved medical devices. BMJ Health Care Inform 2021 Apr 14;28(1):e100301 [FREE Full text] [doi: 10.1136/bmjhci-2020-100301] [Medline: 33853863]
14. US PUBLIC LAW 94-295—MAY 28, 1976, 94th Congress. An Act to amend the Federal Food, Drug, and Cosmetic Act to provide for the safety and effectiveness of medical devices intended for human use, and for other purposes. 1976. URL: https://www.govinfo.gov/content/pkg/STATUTE-90/pdf/STATUTE-90-Pg539.pdf [accessed 2021-02-24]
15. Council Directive 90/385/EEC of 20 June 1990 on the approximation of the laws of the Member States relating to active implantable medical devices. Official Journal of the European Communities. 1990 Jun 20. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31990L0385&from=EN [accessed 2021-02-24]
16. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices. Official Journal of the European Communities. 2017 Apr 05. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0745 [accessed 2021-02-24]
17. IEC 62304:2006/AMD 1:2015 Medical device software — Software life cycle processes — Amendment 1. ISO. 2006. URL: https://www.iso.org/standard/64686.html [accessed 2021-02-24]
18. Guidance on the use of AGILE practices in the development of medical device software AAMI TIR45:2012 (R2018). AAMI. 2012. URL: https://webstore.ansi.org/standards/aami/aamitir452012r2018 [accessed 2021-02-24]
19. Dikert K, Paasivaara M, Lassenius C. Challenges and success factors for large-scale agile transformations: A systematic literature review. Journal of Systems and Software 2016 Sep;119:87-108. [doi: 10.1016/j.jss.2016.06.013]
20. Gilbert S, Fenech M, Idris A, Türk E. Periodic Manual Algorithm Updates and Generalizability: A Developer's Response. Comment on "Evaluation of Four Artificial Intelligence-Assisted Self-Diagnosis Apps on Three Diagnoses: Two-Year Follow-Up Study". J Med Internet Res 2021 Jun 16;23(6):e26514 [FREE Full text] [doi: 10.2196/26514] [Medline: 34132641]
21. ISO/IEC TR 24028:2020 Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence. ISO. 2020 May. URL: https://www.iso.org/standard/77608.html [accessed 2021-10-11]
22. ISO/IEC TR 24027 Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making. ISO. 2020 Nov. URL: https://www.iso.org/standard/77607.html [accessed 2021-10-11]
23. Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan. Food and Drug Administration. URL: https://www.fda.gov/media/145022/download [accessed 2021-02-24]
24. Artificial Intelligence and Machine Learning in Software as a Medical Device. Food and Drug Administration. URL: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device [accessed 2021-02-24]
25. Minssen T, Gerke S, Aboy M, Price N, Cohen G. Regulatory responses to medical machine learning. J Law Biosci 2020 May 01;7(1):lsaa002 [FREE Full text] [doi: 10.1093/jlb/lsaa002] [Medline: 34221415]
26. Babic B, Gerke S, Evgeniou T, Cohen IG. Algorithms on regulatory lockdown in medicine. Science 2019 Dec 06;366(6470):1202-1204. [doi: 10.1126/science.aay9547] [Medline: 31806804]
27. Cohen IG, Evgeniou T, Gerke S, Minssen T. The European artificial intelligence strategy: implications and challenges for digital health. The Lancet Digital Health 2020 Jul;2(7):e376-e379. [doi: 10.1016/s2589-7500(20)30112-6]
28. Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) - Discussion Paper and Request for Feedback (2019). Food and Drug Administration. 2019. URL: https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf [accessed 2021-02-24]
29. "Software as a Medical Device": Possible Framework for Risk Categorization and Corresponding Considerations. IMDRF Software as a Medical Device (SaMD) Working Group. 2014 Dec 18. URL: http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf [accessed 2021-02-24]
30. Legal and regulatory implications of Artificial Intelligence (AI): the case of autonomous vehicles, e-health and data mining, Brussels (2018). European Commission. 2019. URL: https://publications.jrc.ec.europa.eu/repository/bitstream/JRC116235/jrc116235_report_on_ai_%281%29.pdf [accessed 2021-02-24]
31. White Paper On Artificial Intelligence - A European approach to excellence and trust. European Commission. 2020 Feb. URL: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf [accessed 2021-02-24]
32. Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts. EUR-Lex. 2021 Apr 21. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206 [accessed 2021-10-11]
33. MDCG 2019-11 - Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 MDR and Regulation (EU) 2017/746 IVDR (2019). Medical Device Coordination Group. 2019 Oct. URL: https://ec.europa.eu/health/sites/health/files/md_sector/docs/md_mdcg_2019_11_guidance_qualification_classification_software_en.pdf [accessed 2021-02-24]
34. MDCG 2020-1 - Guidance on Clinical Evaluation (MDR) / Performance Evaluation (IVDR) of Medical Device Software. Medical Device Coordination Group. 2020 Mar. URL: https://ec.europa.eu/health/sites/health/files/md_sector/docs/md_mdcg_2020_1_guidance_clinic_eva_md_software_en.pdf [accessed 2021-02-24]

35. Artificial Intelligence in Healthcare. COCIR, the European Coordination Committee of the Radiological, Electromedical and Healthcare IT Industry. 2019 Apr. URL: https://www.cocir.org/uploads/media/COCIR_White_Paper_on_AI_in_Healthcare.pdf [accessed 2021-02-24]
36. IEC/DIS 62304 Health software — Software life cycle processes. ISO. 2021. URL: https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/07/16/71604.html [accessed 2021-02-24]
37. Jarman H, Rozenblum S, Huang TJ. Neither protective nor harmonized: the crossborder regulation of medical devices in the EU. HEPL 2020 Jul 07;16(1):51-63. [doi: 10.1017/s1744133120000158]
38. Wiegand T, Krishnamurthy R, Kuglitsch M, Lee N, Pujari S, Salathé M, et al. WHO and ITU establish benchmarking process for artificial intelligence in health. The Lancet 2019 Jul;394(10192):9-11. [doi: 10.1016/s0140-6736(19)30762-7]

Abbreviations

ACP: algorithm change protocol
AI: artificial intelligence
CDS: clinical decision support
COCIR: European Coordination Committee of the Radiological, Electromedical and Healthcare IT Industry
FDA: Food and Drug Administration (US)
GMLP: good machine learning practices
IEC: International Electrotechnical Commission
IMDRF: International Medical Device Regulators Forum
ISO: International Organization for Standardization
MDR: Medical Device Regulation
ML: machine learning
PCCP: predetermined change control plan
PMCF: postmarket clinical follow-up
PMS: postmarket surveillance
RWP: real-world performance
SaMD: Software as a Medical Device
SDLC: software development life cycle
SPS: SaMD prespecification

Edited by R Kukafka; submitted 19.05.21; peer-reviewed by X Liu, F Magrabi, A Teles, L Guo; comments to author 05.08.21; revised version received 23.08.21; accepted 22.09.21; published 26.10.21.

Please cite as:
Gilbert S, Fenech M, Hirsch M, Upadhyay S, Biasiucci A, Starlinger J
Algorithm Change Protocols in the Regulation of Adaptive Machine Learning–Based Medical Devices
J Med Internet Res 2021;23(10):e30545
URL: https://www.jmir.org/2021/10/e30545
doi:10.2196/30545
PMID:34697010

Original Paper

Improved Environment-Aware–Based Noise Reduction System for Cochlear Implant Users Based on a Knowledge Transfer Approach: Development and Usability Study

Lieber Po-Hung Li[1,2,3,4], MD, PhD; Ji-Yan Han[5], BSc; Wei-Zhong Zheng[5], MS; Ren-Jie Huang[5], MS; Ying-Hui Lai[5], PhD

[1]Department of Otolaryngology, Cheng Hsin General Hospital, Taipei, Taiwan

[2]Faculty of Medicine, Institute of Brain Science, National Yang Ming Chiao Tung University, Taipei, Taiwan

[3]Department of Medical Research, China Medical University Hospital, China Medical University, Taichung, Taiwan

[4]Department of Speech Language Pathology and Audiology, College of Health Technology, National Taipei University of Nursing and Health Sciences, Taipei, Taiwan

[5]Department of Biomedical Engineering, National Yang Ming Chiao Tung University, Taipei, Taiwan

Corresponding Author:
Ying-Hui Lai, PhD
Department of Biomedical Engineering
National Yang Ming Chiao Tung University
No 155, Sec 2, Linong Street
Taipei, 112
Taiwan
Phone: 886 228267021
Fax: 886 228210847
Email: yh.lai@nycu.edu.tw

Abstract

Background: Cochlear implant technology is a well-known approach to help deaf individuals hear speech again and can improve speech intelligibility in quiet conditions; however, it still has room for improvement in noisy conditions. More recently, it has been proven that deep learning–based noise reduction, such as noise classification and deep denoising autoencoder (NC+DDAE), can benefit the intelligibility performance of patients with cochlear implants compared to classical noise reduction algorithms.

Objective: Following the successful implementation of the NC+DDAE model in our previous study, this study aimed to propose an advanced noise reduction system using knowledge transfer technology, called NC+DDAE_T; examine the proposed NC+DDAE_T noise reduction system using objective evaluations and subjective listening tests; and investigate which layer substitution of the knowledge transfer technology in the NC+DDAE_T noise reduction system provides the best outcome.

Methods: The knowledge transfer technology was adopted to reduce the number of parameters of the NC+DDAE_T compared with the NC+DDAE. We investigated which layer should be substituted using short-time objective intelligibility and perceptual evaluation of speech quality scores as well as *t*-distributed stochastic neighbor embedding to visualize the features in each model layer. Moreover, we enrolled 10 cochlear implant users for listening tests to evaluate the benefits of the newly developed NC+DDAE_T.

Results: The experimental results showed that substituting the middle layer (ie, the second layer in this study) of the noise-independent DDAE (NI-DDAE) model achieved the best performance gain regarding short-time objective intelligibility and perceptual evaluation of speech quality scores. Therefore, the parameters of layer 3 in the NI-DDAE were chosen to be replaced, thereby establishing the NC+DDAE_T. Both objective and listening test results showed that the proposed NC+DDAE_T noise reduction system achieved similar performances compared with the previous NC+DDAE in several noisy test conditions. However, the proposed NC+DDAE_T only required a quarter of the number of parameters compared to the NC+DDAE.

Conclusions: This study demonstrated that knowledge transfer technology can help reduce the number of parameters in an NC+DDAE while keeping similar performance rates. This suggests that the proposed NC+DDAE_T model may reduce the implementation costs of this noise reduction system and provide more benefits for cochlear implant users.

(J Med Internet Res 2021;23(10):e25460) doi:10.2196/25460

KEYWORDS

cochlear implants; noise reduction; deep learning; noise classification; hearing; deaf; sound; audio; cochlear

Introduction

Cochlear implants (CIs) are implanted electronic medical devices that can enable patients with profound-to-severe hearing loss to obtain a sense of sound. In their study, Gifford et al [1] showed that 28% of individuals equipped with CI achieved 100% speech intelligibility. Sladen et al [2] also reported similar results in their study: after undergoing CI implantation, the word accuracy of CI users was 80% in a quiet environment. Although CI users have few obstacles in a quiet environment, there is still scope for improvement in a noisy environment [2].

Noise reduction (NR) is one of classical methods to alleviate the effect of background noise for CI users. Over the past few decades, many statistical signal processing NR methods have been proposed, such as log minimum mean squared error [3], Karhunen-Loéve transform [4], Wiener filter-based on a priori signal-to-noise ratio (SNR) estimation [5], generalized maximum a posteriori spectral amplitude [6], and SNR-based [7] approaches. Loizou et al [8] proposed a single-channel algorithm to conduct NR, and the results showed that the sentence recognition scores in 14 participants with CI improved significantly over their daily performances. Dawson et al [7] evaluated a real-time NR algorithm which used the noise estimation to pick up 1 NR approach out of 2 different levels of NR approaches according to the SNR. The study results showed that the proposed NR algorithm could benefit CI users in speech a reception threshold under 3 kinds of noise. Mauger et al [9] optimized the gain function to achieve a better SNR-based NR, and the results showed that with the optimized gain function, a 27% improvement was achieved for CI users in speech-weighted noise. Although classical NR function can improve speech intelligibility for CI users in stationary noise conditions [7-9], improvements are still needed in nonstationary noise conditions [10].

Deep learning (DL)–based NR methods have recently shown better performance than classical statistical-based NR methods [11-17]. Lai et al [18] used a deep denoising autoencoder (DDAE)–based NR using vocoder simulation to perform NR function for CI users; the listening test showed that the speech intelligibility was better with DDAE-based NR than with convectional single-microphone NR approaches, whether in stationary or nonstationary noise conditions. Goehring et al [19,20] used neural and recurrent neural networks to perform the NR function for CI users, and the results showed that the proposed NR function could significantly improve speech intelligibility in babbling noise conditions. In DL methods, the nonstationary noise can be processed well, but this needs a huge amount of training data in different noise types and SNR levels. However, when a mismatch exists, such as when there is a difference in data between the training and testing phase, the performance of the DL method is usually degraded [10,18].

An environment-aware–based NR system called noise classifier (NC) +DDAE (NC+DDAE) was proposed to alleviate the above issue [21]. The NC+DDAE NR system combines *n*-specific noise-dependent (ND)-DDAE NR models and a noise-independent (NI)-DDAE NR model. The NC function (ie, deep neural network model) was used to distinguish *n* different typical noises and select a suitable DDAE model to perform the NR function for CI users. Hence, the NC function made the NC+DDAE an environment-aware–based NR system. The objective measures and listening test showed that the NC+DDAE model had a much higher performance than did the other NR methods. Although the NC+DDAE model has proven to benefit the CI user and have the flexibility of customization, the NC+DDAE model requires several parameters, which increase the requirements for device implementation. Therefore, the NC+DDAE model needs to be modified to have fewer requirements while maintaining the performance at the same level.

Recently, the knowledge transfer (so called transfer learning) approach [22] has been used in many speech signal processing tasks (eg, speech emotion detection [23], text-to-speech system [24,25], and speech enhancement [26]) and has proven to provide benefits for the DL-based model. Knowledge transfer is a machine learning method developed for a specific task that reuses the initial parameters for a new model for the target task. In other words, the knowledge transfer technology transfers the domain knowledge based on the source domain to the target domain to help the DL-based model achieve better performance; furthermore, it can speed up the time needed to develop and train a model by reusing these pieces or modules that have already been developed [22]. Following the concept of knowledge transfer technology, we proposed an improved NC+DDAE NR model, called NC+DDAE_transfer (NC+DDAE_T). We first analyzed the differences between features in each layer of DDAE to choose the most suitable layer for NR adaptation. Next, we compared the performance between NC+DDAE and NC+DDAE_T with 2 well-known objective metrics: perceptual evaluation of speech quality (PESQ) [27] and short-time objective intelligibility (STOI) [28]. The PESQ shows the result of comparing the clean and processed speech by mean opinion score. In the mean opinion score, 5 is the highest score while 1 is the lowest. According to a previous study [27], a score over 4 is high enough for most people to listen comfortably and a score of 3.6 is an acceptable boundary for those with normal hearing. The STOI represents the speech intelligibility by a correlation coefficient derived from comparing the energy of clean and processed speech in each frame. STOI ranges from 0 to 1, with a higher score representing more clear and understandable speech. Finally, the clinical effectiveness of NC+DDAE_T with the NC+DDAE and DDAE NR systems for patients with CI was evaluated in noisy listening conditions.

Methods

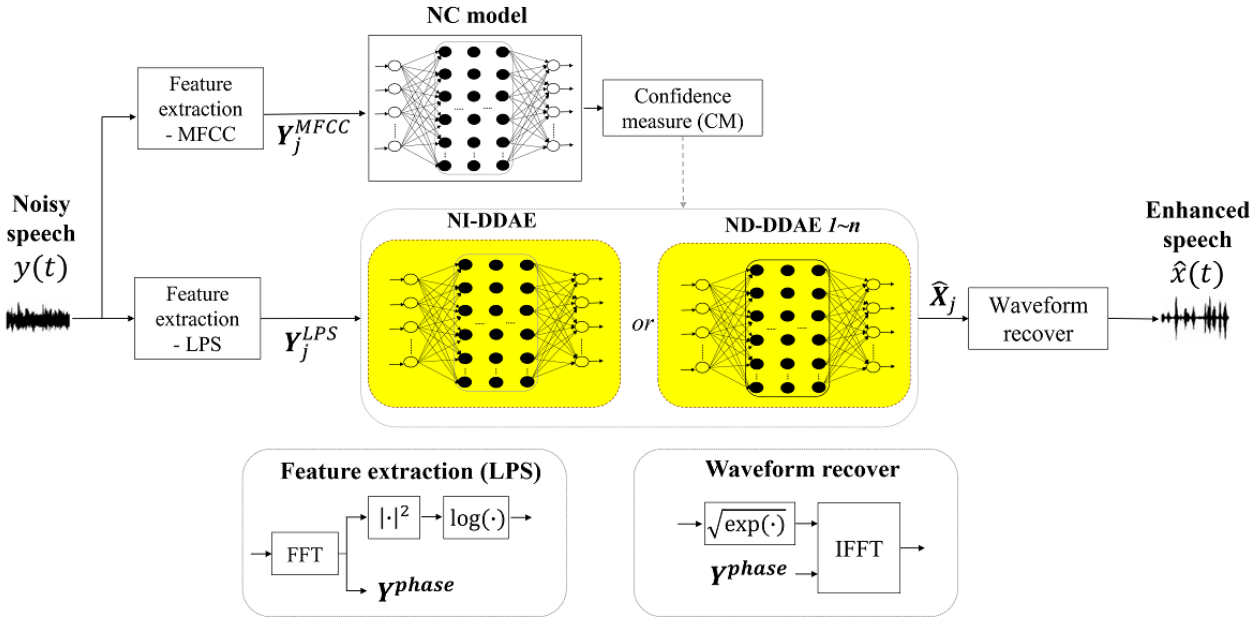
In this section, we describe first the NC+DDAE approach. We then introduce the NC+DDAE_T method, the transfer learning–based NC+DDAE NR modified in this study. Finally, we describe the experimental setting and material to prove the

benefits of the proposed NC+DDAE_T compared to 2 well-known DL-based NR systems (ie, DDAE and NC+DDAE).

NR Based on the NC+DDAE Approach

Figure 1 shows the proposed NC+DDAE model in our previous study [21], where 2 critical units, NC and DDAE, were included. In this approach, first, the noisy speech signals ***y(t)*** are processed by feature extraction units to obtain $\boldsymbol{Y_j^{MFCC}}$ and $\boldsymbol{Y_j^{LPS}}$, which denote log power spectra (LPS) [29] and Mel-frequency cepstral coefficients [30], respectively, with *j* denoting the frame in the short-time Fourier transform. $\boldsymbol{Y_j^{MFCC}}$ is the input of the NC model to determine the current type of background noise and to select a suitable DDAE model for NR, which includes multiple ND-DDAE models each trained by a model-specific noise type and a single NI-DDAE model trained by 120 noise types [15]. When the noisy input signal is similar to one of the specific noise types, the specific ND-DDAE model is chosen for NR; otherwise, the NI-DDAE is used. Afterward, the selected DDAE model processes $\boldsymbol{Y_j^{LPS}}$ to obtain the enhanced features.  is combined with the noisy phase $\boldsymbol{Y^{phase}}$ to finally reconstruct the enhanced speech  . The NC+DDAE NR system has been defined in detail previously [21].

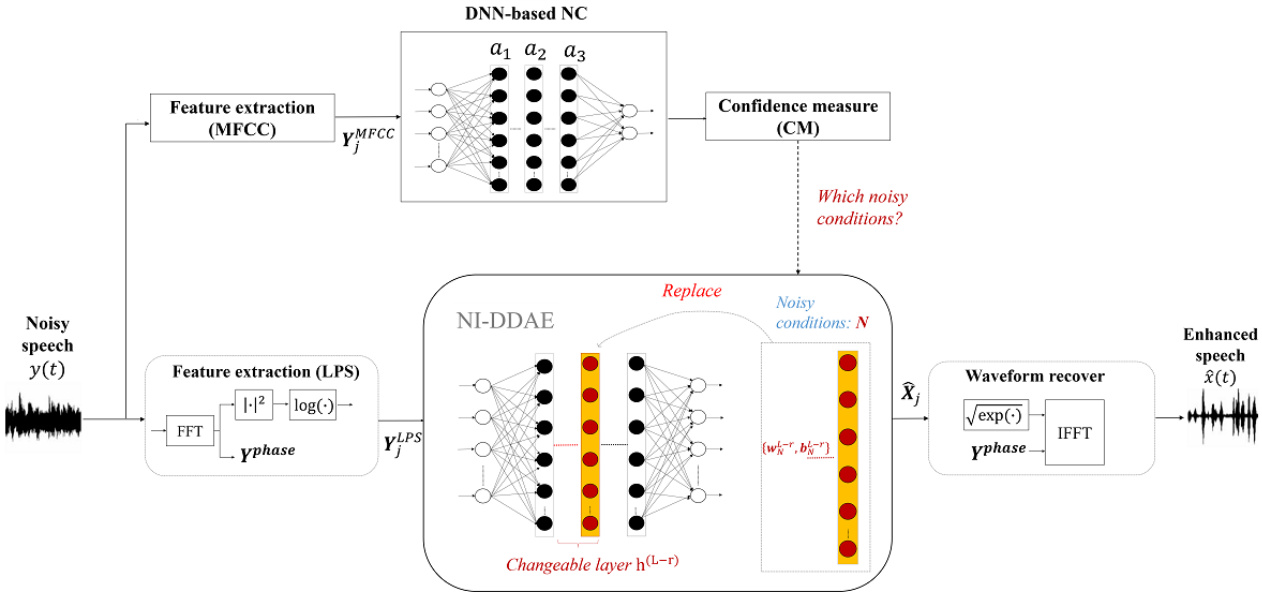
Figure 1. Structure of the noise classifier with a deep denoising autoencoder (NC+DDAE) system. DDAE: deep denoising autoencoder; FFT: fast Fourier transform; IFFT: inverse fast Fourier transform; LPS: log power spectra; NC: noise classifier; ND: noise-dependent; NI: noise-independent; MFCC: Mel-frequency cepstral coefficient.



NR With the Proposed NC+DDAE_T Approach

Figure 2 shows the pipeline of the NC+DDAE_T NR approach proposed in this study. The signal processing procedure of the NC+DDAE_T is similar to that of the above-mentioned NC+DDAE. The major difference lies in the NR model as described in the following sections.

Figure 2. Structure of the proposed noise classifier system with DDAE and knowledge transfer. DDAE: deep denoising autoencoder; DNN: deep neural network; FFT: fast Fourier transform; IFFT: inverse fast Fourier transform; LPS: log power spectra; NC: noise classifier; NI: noise-independent; MFCC: Mel-frequency cepstral coefficient.



NC Model

The NC model of the proposed NC+DDAE_T is the same as that in our previously described system. Initially, the system receives a noisy speech ***y(t)*** and computes the $\boldsymbol{Y_j^{MFCC}}$ and $\boldsymbol{Y_j^{LPS}}$ features separately. $\boldsymbol{Y_j^{MFCC}}$ is then sent to the NC model. The NC model is a deep neural network (DNN) composed of 3 hidden layers. Each layer consists of 100 neurons and an output layer adapting the softmax function [30]. The output at the *j*-th node of the *l*-th layer in a DNN $h_j^{(l)}$ is produced according to equation 1:

(1)

where the term $h_j^{(l-1)}$ denotes the output from the *i*-th node in the (*l*−1)-th layer, $b_j^{(l)}$ is the bias of index *j*, and $W_{ij}^{\ l}$ is the weight between hidden unit *j* and *i*. σ(·) is the activation function [30], which is the logistic function described in equation 2:

(2)

Next, the trained DNN model is used in the NC function. The output of the last layer is converted into the probability by the softmax function [31] to obtain the normalized probability-based output. The back propagation algorithm [32,33] is then applied to parameter set θ in equation 3, where $L(\cdot)$ is the loss function, N_i denotes the correct noise class, and  is the output class of the DNN-based NC.

(3)

To avoid substantial variance in the DNN output, we use the confidence measurement [34] to analyze the output of the DNN-based NC. Based on the confidence measurement score, a threshold is used to determine the classification results. In other words, when the confidence measurement score is higher than the threshold, the result predicted by the NC model is considered trustworthy. Nevertheless, if the confidence measurement score is not concrete to one noise type, then the NI-DDAE is chosen for NR; on the other hand, if the confidence measurement is solid, the ND-DDAE is selected.

DDAE-based NR Model

In the training phase, the noisy LPS feature $\boldsymbol{Y_j^{LPS}}$ and clean LPS feature $\boldsymbol{X_j^{LPS}}$ are the input and output, respectively, of the DDAE–based NR model. The details for training the DDAE NR model with *L* hidden layers mapping $\boldsymbol{Y_j^{LPS}}$ to $\boldsymbol{X_j^{LPS}}$ are available elsewhere [21]. The difference between NC+DDAE and NC+DDAE_T is that only the parameters of a specific layer (ie, $w^{L\text{-}r}$ and $b^{L\text{-}r}$) are trainable as shown in equation 4, whereas the other parameters remain untrainable in the fine-tuning process. The constant *L* denotes the number of layers, and we used 5 layers (ie, *L*=5) in this study.

(4)

where $\{W^1 \ldots W^{(L\text{-}r)} \ldots W^L\}$ and $\{b^1 \ldots b^{(L\text{-}r)} \ldots b^L\}$ are the matrices of weights and bias vectors of the DDAE NR model, respectively, whereas *Relu* represents the activation function rectified linear unit [35]. The constant *r* is the index to identified the specific trainable layer. In this study, the second layer (ie, *r*=3) was chosen because, on average, substituting the second layer achieved the best performance in our pilot study. The detailed experimental results are shown in Multimedia Appendix 1.

Based on the above idea, the original NI-DDAE, trained with a huge database of noise samples, can be transformed into many ND-DDAE models according to the type of background noise. In this study, 12 common types of background noise were used; hence, 12 ND-DDAE models were derived from the NI-DDAE model. More specifically, each ND-DDAE model was determined by optimizing the following objective function:

(5)

(6)

where M is the total number of training samples and F( ) is the loss function derived from and $\boldsymbol{X}_j^{LPS}$. is the vector that contains the logarithmic amplitudes of the enhanced speech corresponding to the paired noisy LPS feature $\boldsymbol{Y}_j^{LPS}$. Subsequently, the trained NI-DDAE provides the initial parameters for the ND-DDAE model, and the noise data of the specific environment are used to fine-tune this ND-DDAE model. Finally, the transformed LPS feature is sent to the waveform recovery unit to reconstruct the waveform. More specifically, is first processed using square root and exponential operations. The waveform recovery function then reconstructs the enhanced speech with the noisy phase $\boldsymbol{Y}^{phase}$.

Training and Evaluation Procedure

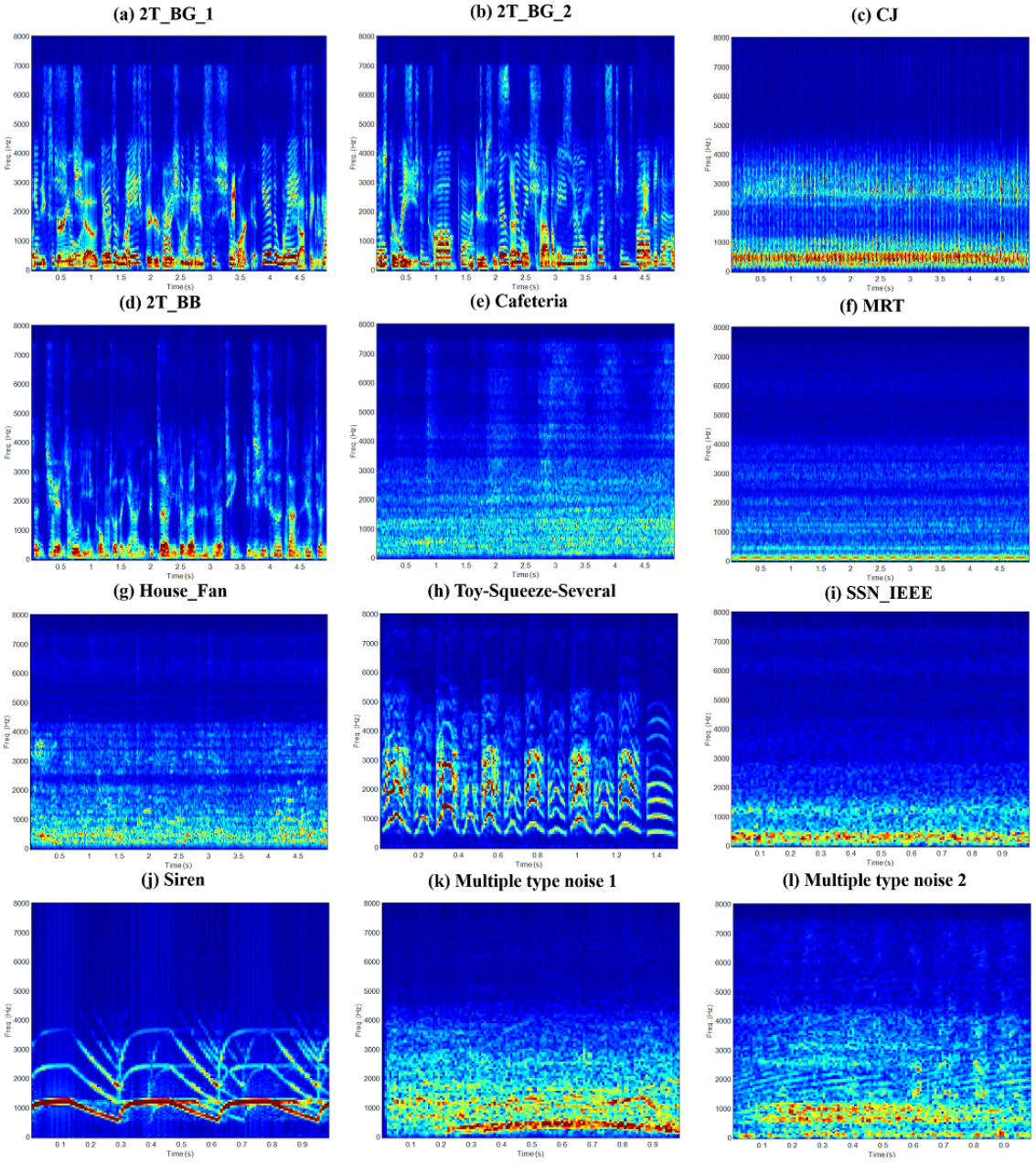
In this section, we show how the NC, DDAE, and NC+DDAE_T models were trained. First, we trained a new NC model according to the 12 common background noises, 2talker_unseen1, 2talker_unseen2, Construction Jackhammer (CJ), 2 Talker, Cafeteria, MRT (Mass Rapid Transit), cafeteria, Toy-Squeeze-Several, speech shape noise from the Institute of Electrical and Electronics Engineers (SSN_IEEE), Siren, Multiple type noise 1, and Multiple type noise 2, which are shown in Figure 3. Note that the training approach is described in the previous section “NC Model”. After the training, the prediction accuracy of the 12 noises was 100%. The detailed results of the confusion matrix are shown in Multimedia Appendix 2.

To train the DDAE NR model, the Taiwan Mandarin version of the hearing in noise test (TMHINT) corpus [36] was selected to conduct all experiments, including the training and evaluation parts. All 320 sentences, each consisting of 10 characters, were recorded at a 16 kHz sampling rate, after which 120 utterances among the TMHINT corpus were selected and corrupted by 120 noise types [15] at 7 SNR levels (−10, −7, −4, −1, 1, 4, 7, and 10 dB) as the training set for the DDAE model. The other 200 utterances were also corrupted with the 12 common background noises—as mentioned in the description of NC training—at 6 SNR levels (-6, -3, 0, 3, and 6 dB) as the outside testing set. In our previous study, this trained model was defined as the NI-DDAE.

Next, we combined the NC with NI-DDAE and fine-tuned the model with each noise type in the NC, and the NI-DDAE was transformed into NC+DDAE_T. In the fine-tuning step, we could freeze or adopt each layer in the NI-DDAE. Previously, we had studied which layer of the NI-DDAE model had to be replaced to achieve the best performance. We substituted each layer by modifying r in the range from 1 to 5; meanwhile, we conducted 2 well-known objective speech evaluations, PESQ [27] and STOI [28], to identify the most appropriate layer. On average, replacing the middle layer of the NI-DDAE model (ie, the second layer this study) achieved a better performance than did substituting other layers. The detailed results can be found in Multimedia Appendix 1. Hence, we uniformly replaced the parameters of the third layer in all subsequent tests. As the 2 DL-based NR systems, DDAE and NC+DDAE, achieved better performances in our previous studies [18,21] than did the well-known unsupervised NR algorithms, the log minimum mean squared error [3] and Karhunen-Loéve transform [37], we used the DDAE and NC+DDAE algorithms for comparisons to evaluate the NC+DDAE_T in this study.

Subsequently, we enrolled 10 CI users to conduct speech intelligibility tests, and details of these subjects are shown in the Multimedia Appendix 3. This study protocol was approved by the Research Ethics Review Committee of Cheng Hsin Hospital under the following approval number: CHGH-IRB (645) 107A-17-2. The first author, LPHL, explained the study to the patients and collected the signed institutional review board informed consent before the experiment. All participants used their own clinical speech processors and temporarily disabled the built-in NR functions during the test. The test signals of noisy and enhanced speech were played at 65 dB sound pressure level by a speaker and were then processed through a CI processor to simulate the performance of each NR approach for CI users. To ensure that fatigue did not affect the study participants, each individual only heard a total of 16 test conditions (2 background noise [2 talker and CJ] × 2 SNR levels [0 and 3 dB] × 4 signal processing systems [noisy, DDAE, NC+DDAE, and NC+DDAE_T]) with 10 sentences of 10 words in each test condition. The participants were instructed to repeat verbally what they had heard. We evaluated the speech intelligibility under each test condition using the word correct rate (WCR) [38-42] calculated as the ratio between the number of correctly identified words and the total number of words. To further prevent participant fatigue, tests were paused for 5 minutes every 30 minutes. Moreover, we calculated the statistical power to see whether the sample size (10 patients in this study) was large enough to obtain a significant difference in the result. The statistical power of this study is 1. According to Cohen et al [43] a statistical power over 0.8 is sufficiently high to conclude that there is a significant difference in the hypothesis.

Figure 3. Spectrograms of the 12 noise signals: (a) 2T_BG_1, (b) 2T_BG_2, (c) CJ, (d) 2T_BB, (e) Cafeteria, (f) MRT, (g) House Fan, (h) Toy-Squeeze-Several, (i) SSN_IEEE, (j) Siren, (k) Multiple type noise 1, and (l) Multiple type noise 2. 2T_BG_1 is a noise that mixes the speech of a girl and a boy both speaking repeatedly in English. 2T_BG_1 is a noise that mixes the speech of a girl and a boy both speaking repeatedly in English. The speakers in 2T_BG_2 are the same as those in 2T_BG_1 but with different sentences. 2T_BB is a noise that overlays 2 sentences in Chinese spoken by the same male speaker. Multiple type noise 1 is a mix of the sound of sirens and cheering crowd, whereas Multiple type noise 2 is a sound combining scratching and booing. The other samples are common background noises from daily life. 2T_BB: 2 Talker; 2T_BG_1: 2 talker_unseen1; 2T_BG_2: 2 talker_unseen2; CJ: Construction Jackhammer; MRT: Mass Rapid Transit; SSN_IEEE: speech shape noise from the Institute of Electrical and Electronics Engineers.



Results

Objective Evaluation Using PESQ and STOI Scores

We compared the newly proposed NC+DDAE_T with the previously established NR systems, DDAE and NC+DDAE. The PESQ and STOI scores of these tests are shown in Figures 4 and 5, respectively. As demonstrated in Figure 4, the PESQ scores of the proposed NC+DDAE_T are generally similar to those of the NC+DDAE. The details regarding the average scores of each approach (ie, noisy, DDAE, NC+DDAE, and NC+DDAE_T) for the 12 background noises at 6 different SNR levels can be found in Table A1 of Multimedia Appendix 4. In the STOI scores, the NC+DDAE_T model also achieved the same level as did the NC+DDAE (Figure 5). The detailed STOI scores are listed in Table A2 of Multimedia Appendix 4. These objective evaluation results proved that the NC+DDAE_T could provide almost the same speech intelligibility performance as the NC+DDAE.

Figure 4. Mean perceptual evaluation of speech quality (PESQ) scores of the 4 noise reduction approaches. 2T_BB: 2 Talker; 2T_BG_1: 2 talker_unseen1; 2T_BG_2: 2 talker_unseen2; CJ: Construction Jackhammer; dB: decibel; DDAE: deep denoising autoencoder; NC: noise classifier; NC+DDAE_T: noise classifier + deep denoising autoencoder with knowledge transfer; MRT: Mass Rapid Transit; PESQ: perceptual evaluation of speech quality; SNR: signal-to-noise ratio; SSN_IEEE: speech shape noise from the Institute of Electrical and Electronics Engineers.

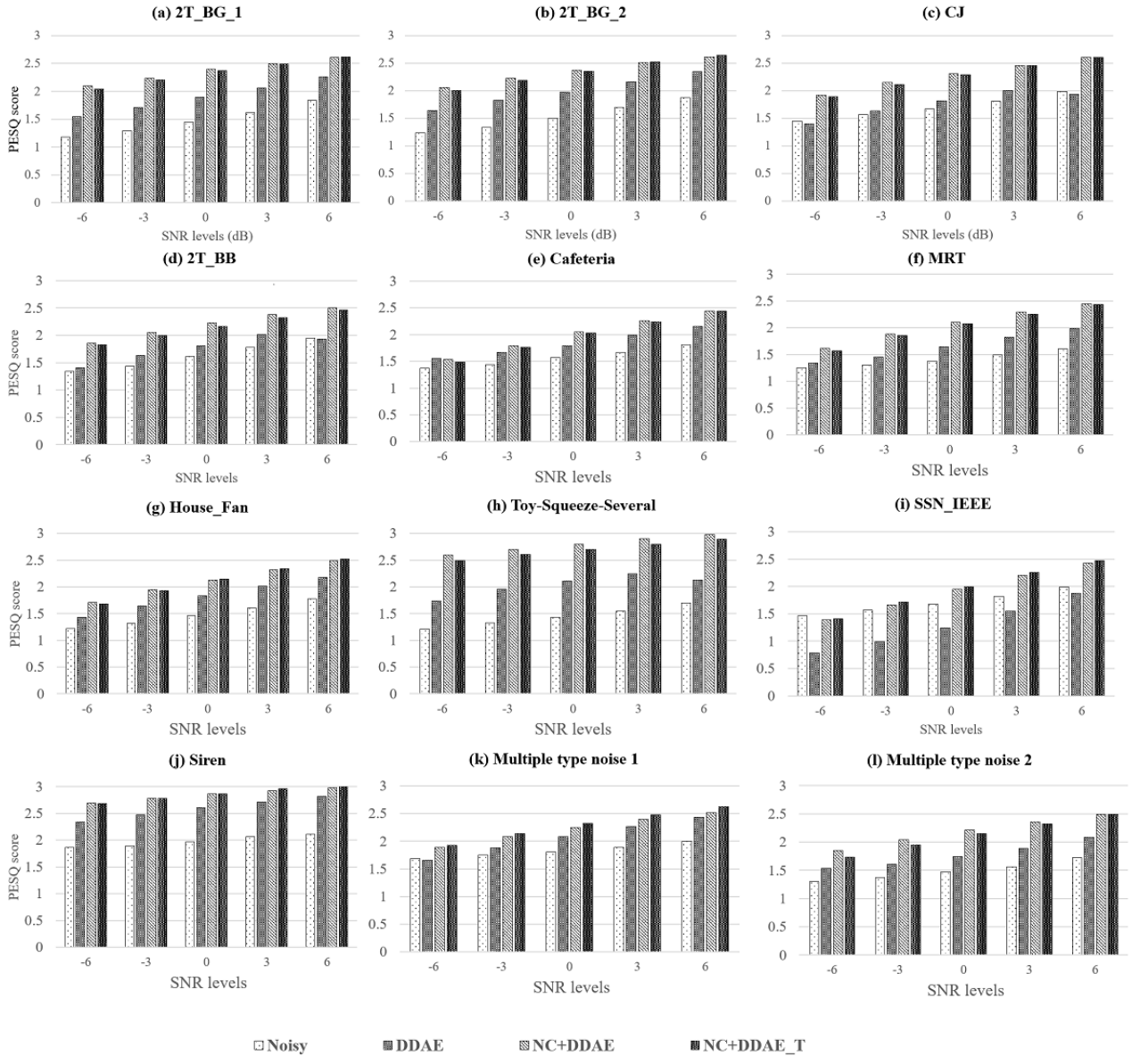
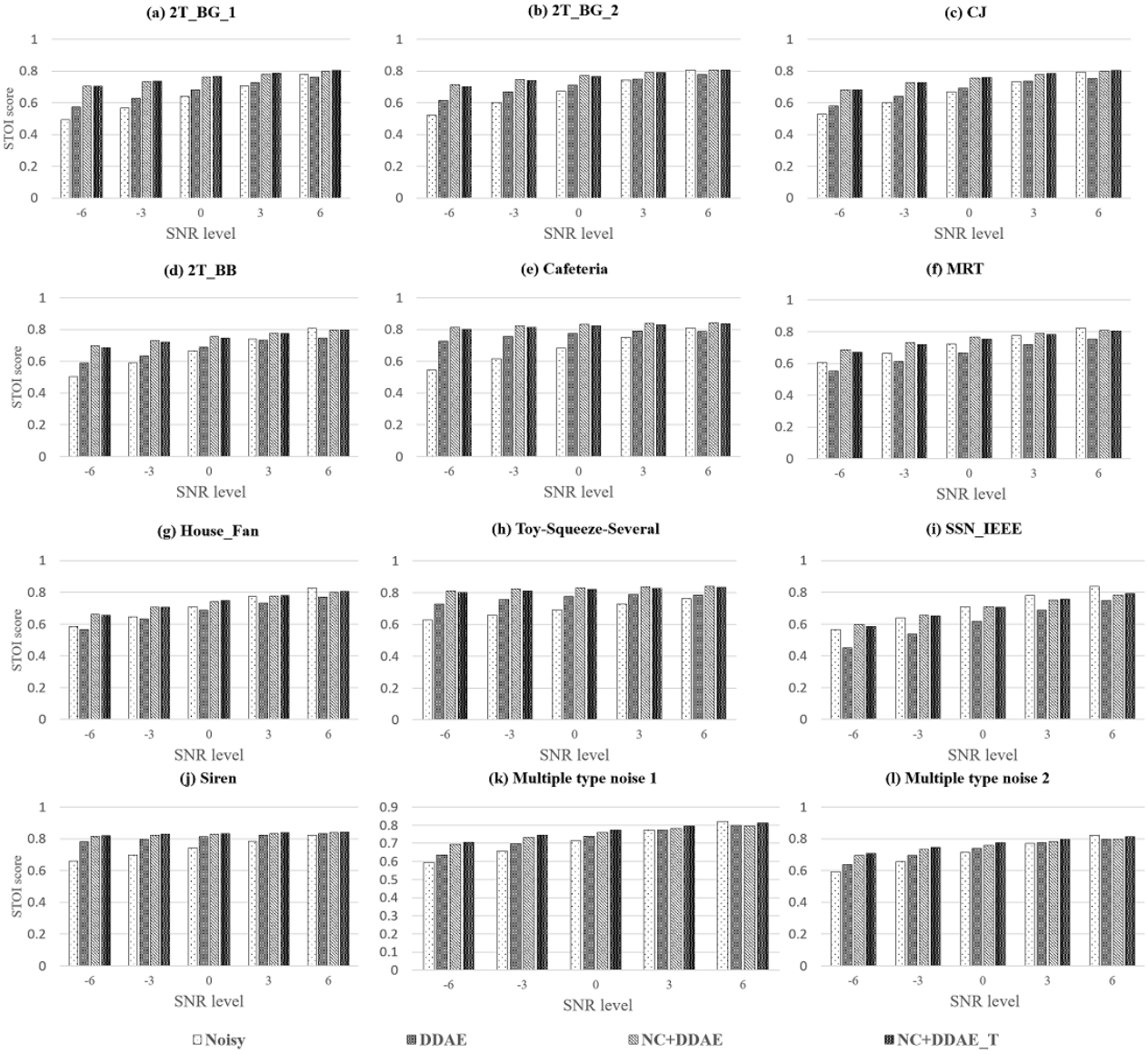


Figure 5. Mean short-time objective intelligibility (STOI) scores of the different noise reduction approaches. 2T_BB: 2 Talker; 2T_BG_1: 2 talker_unseen1; 2T_BG_2: 2 talker_unseen2; CJ: Construction Jackhammer; DDAE: deep denoising autoencoder; NC: noise classifier; NC+DDAE_T: noise classifier + deep denoising autoencoder with knowledge transfer; MRT: Mass Rapid Transit; SNR: signal-to-noise ratio; SSN_IEEE: speech shape noise from the Institute of Electrical and Electronics Engineers; STOI: short-time objective intelligibility.



Recognition in Listening Tests

Figure 6 shows the average WCR scores of 10 individuals with CI in the 2 Talker and CJ noise conditions each at 0- and 3-dB SNR levels. The detailed results are as follows: The respective average WCR scores and standard error of the mean (SEM) for noisy, DDAE, NC+DDAE, and NC+DDAE_T with 2 Talker background noise were 4.1 (SEM 1.87), 27.8 (SEM 5.42), 38.9 (SEM 8.83), and 43.2 (SEM 9.33) at the 0-dB SNR level; and 10.3 (SEM 3.84), 27.7 (SEM 5.24), 48.2 (SEM 9.69), and 50.3 (SEM 8.98) at the 3-dB SNR level. In the CJ background noise, the respective average scores and SEMs were 19.3 (SEM 5.76), 27.7 (SEM 5.24), 42.2 (SEM 9.64), and 50.6 (SEM 10.0) at the 0-dB SNR level; and 37.1 (SEM 9.84), 38.8 (SEM 8.41), 49.3 (SEM9.31), and 50.9 (SEM 10.13) at the 3-dB SNR level. These results demonstrated that the NC+DDAE_T provided better speech intelligibility scores than did noisy speech. Moreover, the newly developed NC+DDAE_T model achieved slightly higher intelligibility performances than did the NC+DDAE approach under most test conditions. The 1-way analysis of variance (ANOVA) [44] with least significant difference post hoc comparison [45] was used to analyze the results of the 4 NR systems (noisy, DDAE, NC+DDAE, and NC+DDAE_T) in the 4 test conditions. The 1-way ANOVA result confirmed that the WCR scores differed significantly among the 4 systems (F=13.256; P<.001). The least significant difference post hoc comparisons (Table 1) further revealed that the noisy condition was significantly different from the other 3 systems (DDAE: P=.16; NC+DDAE: P<.001; NC+DDAE_T: P<.001). Meanwhile, the differences between the NC+DDAE and NC+DDAE_T models were not significant (P=.50).

Figure 6. Mean intelligibility scores of 10 participants with cochlear implants in 4 types of simulated test conditions. 2T_BB: 2 Talker; CJ: Construction Jackhammer; dB: decibel; DDAE: deep denoising autoencoder; NC: noise classifier; NC+DDAE_T: noise classifier + deep denoising autoencoder with knowledge transfer.

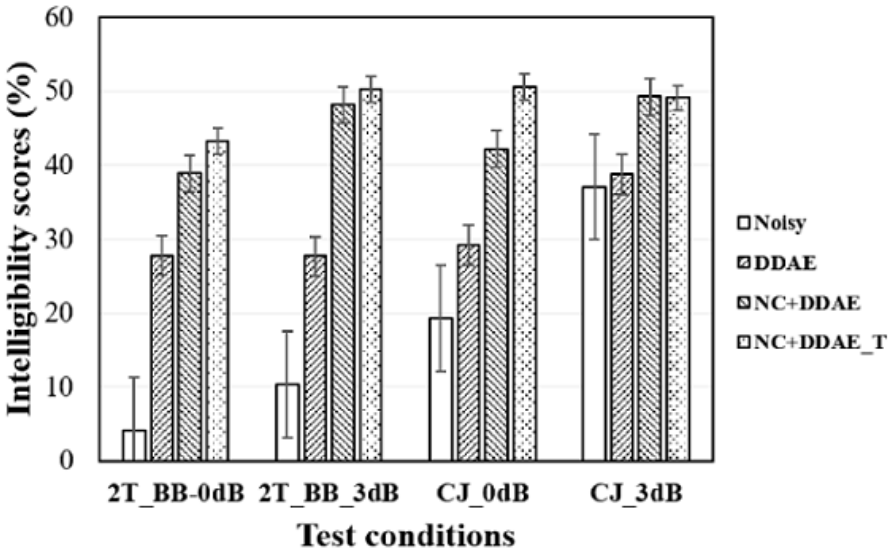


Table 1. The mean difference, standard error, and significance of the listening test in each noise reduction system.

Method (I) by test (J)	Mean difference (I–J) (standard error)	*P* value[a]
Noisy (I)		
DDAE[b] (J)	–13.18 (5.428)	*.016* [c]
NC[d]+DDAE (J)	–26.95 (5.428)	*< .001*
NC+DDAE_T[e] (J)	–30.60 (5.428)	*< .001*
DDAE (I)		
Noisy (J)	13.18 (5.428)	*. 02*
NC+DDAE (J)	–13.78 (5.428)	*.01*
NC+DDAE_T (J)	–17.43 (5.428)	*.002*
NC+DDAE (I)		
Noisy (J)	26.95 (5.428)	*< .001*
DDAE (J)	13.78 (5.428)	*.01*
NC+DDAE_T (J)	–3.65 (5.428)	.50
NC+DDAE_T (I)		
Noisy (J)	30.60 (5.428)	*< .001*
DDAE (J)	17.43 (5.428)	*. 002*
NC+DDAE (J)	3.65 (5.428)	.50

[a]*P* values are significant at $\alpha = .05$. Least significant difference was selected to conduct post hoc testing.

[b]DDAE: deep denoising autoencoder.

[c]Values in italics represent significant values.

[d]NC: noise classifier.

[e]NC+DDAE_T: noise classifier + deep denoising autoencoder with knowledge transfer.

Comparison of the Numbers of Parameters

The original structure of the NC+DDAE system used 12 ND+DDAEs and 1 NI+DDAE for the NR. In this study, the newly developed NC+DDAE_T system only needed 1 NI+DDAE and 12 different layer parameters to achieve the same performance as the previous NC+DDAE system. We further compared the numbers of parameters between the NC+DDAE and NC+DDAE_T approaches. The NC+DDAE_T approach required only 0.1 million parameters while the

previous NC+DDAE system needed 4.4 million parameters. The number of parameters was thus reduced by 76.5% compared to the previous approach.

Discussion

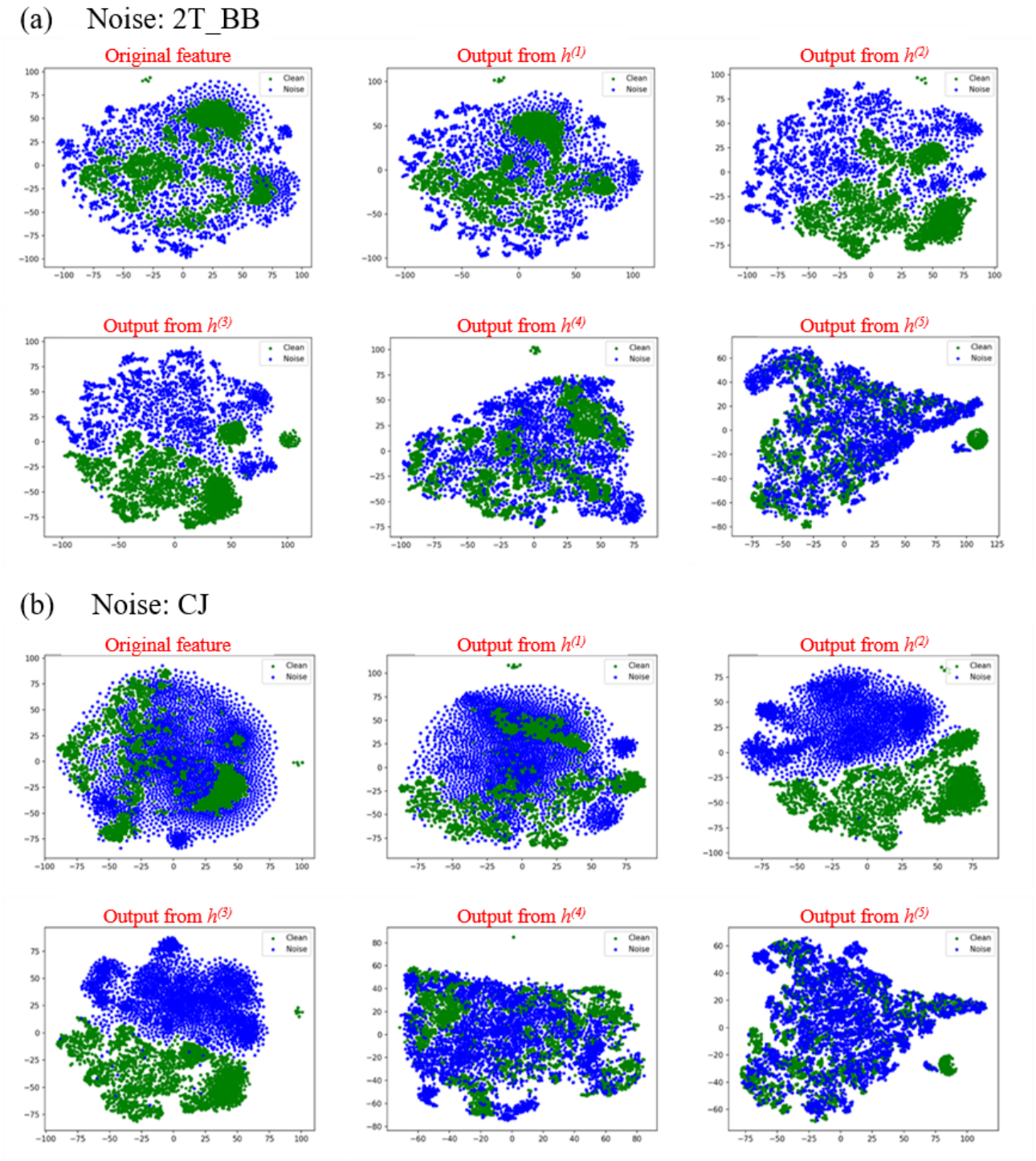
Layers for Substitution

This study proposed a new NC+DDAE_T NR model that helps CI users to improve speech intelligibility in noisy listening conditions. Knowledge transfer technology was used to reduce the parameter requirements in comparison to the previous NC+DDAE approach. The experimental results of the objective evaluation and the subjective listening tests demonstrated that the NC+DDAE_T achieved performances comparable to those of the NC+DDAE approach, while the number of parameters used by the NC+DDAE_T was reduced by 76.5% compared to the NC+DDAE. Therefore, knowledge transfer technology could be a useful approach to further improve the benefits of NC+DDAE in reducing the cost of implementation in the future.

The architecture of the NC+DDAE_T, (ie, which layer is substituted) is the basis for achieving higher performance with this novel system compared to the NC-DDAE. According to the objective evaluation by PESQ and STOI scores (Multimedia Appendix 1), the substitution of the middle layer can achieve better performances. To further analyze why the middle layer was so important, *t*-distributed stochastic neighbor embedding (*t*-SNE) [46] was used to visualize the features that output by each layer. The acoustic features of noisy and clean speech (ie, LPS) were the inputs for the trained NI-DDAE NR model. The output features of each NI-DDAE layer were analyzed using *t*-SNE, which can project the distribution of each layer onto a 2D plane. Figure 7 shows the results of this feature visualization. Green dots represent the output features of clean speech, whereas blue dots indicate features of noisy speech. The less overlap is apparent between the green and blue areas, the better the layer can separate the features. These results indicate that clean and noisy data were primarily separated in the output from $h^{(2)}$ and $h^{(3)}$, implying that the front layers help to distinguish noisy speech from clean features and thus could be the most important layers. This interpretation is also consistent with the objective evaluation results in Multimedia Appendix 1.

To explain the phenomenon illustrated in Figure 7, we suggest that the NC+DDAE_T model may work similarly to the human brain. The first layers of the model may try to separate the noise from the speech features. Therefore, these features would diverge completely in the middle layers of this NR model. The model would then try to reconstruct the enhanced speech and lower the volume of the noise in the final layers of the model; hence, the features would converge again in the *t*-SNE analysis. Based on these hypotheses, the second layer may be the key to feature separation because the features are well separated after the second layer. Therefore, to adapt the NR model to a specific type of noise, substituting the second layer would be the best choice, which corresponds to the results of the objective evaluation. The other parts of the NC+DDAE_T model may work as preprocessing and vocoder units. These parts are common units of all NR models; thus, different ND-DDAEs can share the same weight and bias values. Therefore, the concept of knowledge transfer can be used in this part to decrease the size of each model.

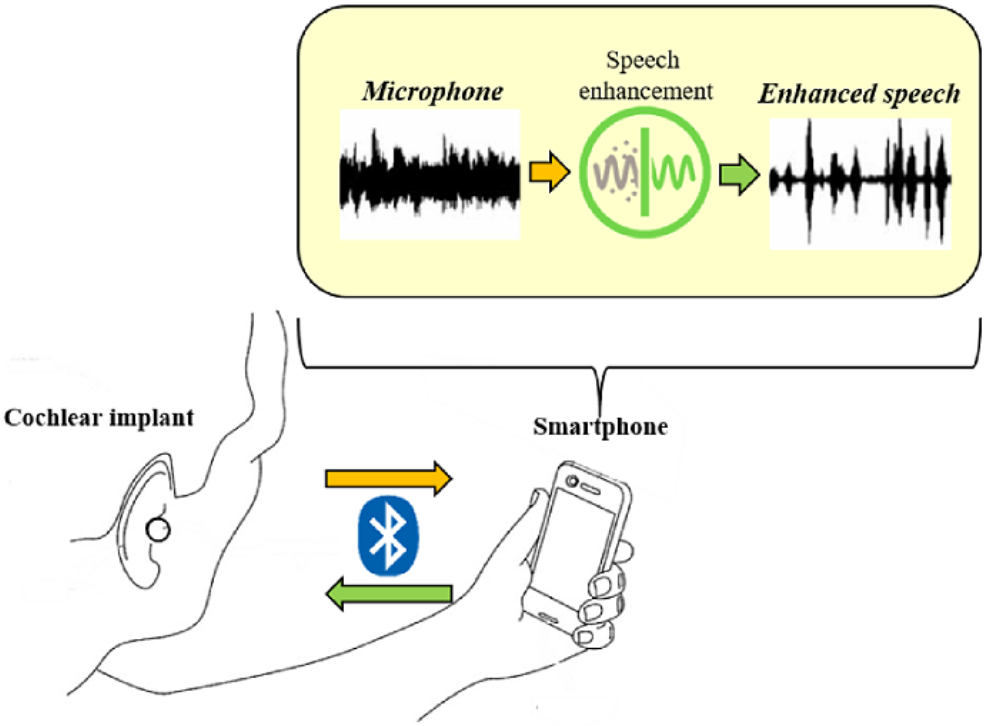
Figure 7. *t*-distributed stochastic neighbor embedding (*t*-SNE) feature analysis of each layer in the noise-independent deep denoising autoencoder (NI-DDAE) model with noisy and clean speech data. The green dots represent the output features of clean speech and the blue dots indicate features of noisy speech. 2T_BB: 2 Talker; CJ: Construction Jackhammer.



Future Perspectives

Based on previous and current results of objective evaluation and listening tests, we can conclude that the proposed NC+DDAE_T performs comparably to the NC+DDAE. In addition, the NC+DDAE_T needs only a quarter of the number of parameters compared to the 12 ND-DDAE models. These characteristics suggest a great potential for future implementation of the NC+DDAE_T model. With the decreased number of parameters, an implemented device would require less memory. To prove this concept, we have implemented the NC+DDAE_T architecture in an app on an iPhone XR mobile phone (Apple Inc) as shown in Figure 8. The processing time could satisfy the maximum group delay requirement of assistive listening devices. With this advantage of edge computing, the proposed NC+DDAE_T may become a new kind of hearing assistive technology in the near future.

Figure 8. Schematic of the noise classifier deep denoising autoencoder with knowledge transfer (NC+DDAE_T) implementation.



Limitations

The proposed NC+DDAE_T is an adaptable NR system, which means that the system benefits may be affected by the training data (eg, background noise types, speakers). Therefore, if the proposed system faces noisy conditions that are very different from the training data (ie, mismatch conditions), the proposed system would require major improvements, and new recordings of noise data may be needed. Overcoming this issue requires future study. Additionally, although the proposed system was implemented in an app, the full implementation of the proposed system in the hardware of currently used CI devices is still a way off. However, as studies increasingly focus on the acceleration of DL-based models in microprocessors [47,48], there is a greater chance that DL technologies may be implemented into CI devices in the near future.

Conclusions

This study proposed a novel NC+DDAE_T system for NR in CI devices. The knowledge transfer approach was used to lower the number of parameters of the DDAE model. The experimental results of the objective evaluations, along with the listening tests, showed that the proposed NC+DDAE_T model provided comparable performance to the previously established NC+DDAE NR model. These results suggest that the proposed NC+DDAE_T model may be a new NR system that can enable CI users to hear well in noisy conditions.

Acknowledgments

This study was supported by the Ministry of Science and Technology of Taiwan (project #110-2218-E-A49A-501, #110-2314-B-350-003, #109-2218-E-010-004, and #108-2314-B-350 -002-MY2) and Cheng Hsin General Hospital (#CY10933).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Results Following Replacement of Each Layer of weight and bias of the Deep Denoising Autoencoder Model.
[DOCX File , 126 KB - jmir_v23i10e25460_app1.docx]

Multimedia Appendix 2
Confusion matrix of the 12 noise classifications.
[DOCX File , 46 KB - jmir_v23i10e25460_app2.docx]

Multimedia Appendix 3
Individual biographical data of the attended cochlear implant subjects.
[DOCX File , 17 KB - jmir_v23i10e25460_app3.docx]

Multimedia Appendix 4
Perceptual evaluation of speech quality and short-time objective intelligibility scores of different noise reduction systems.
[DOCX File , 29 KB - jmir_v23i10e25460_app4.docx]

References

1. Gifford RH, Shallop JK, Peterson AM. Speech recognition materials and ceiling effects: considerations for cochlear implant programs. Audiol Neurootol 2008;13(3):193-205. [doi: 10.1159/000113510] [Medline: 18212519]
2. Sladen DP, Ricketts TA. Frequency importance functions in quiet and noise for adults with cochlear implants. Am J Audiol 2015 Dec;24(4):477-486. [doi: 10.1044/2015_AJA-15-0023] [Medline: 26650018]
3. Ephraim Y, Malah D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE Trans. Acoust., Speech, Signal Process 1985 Apr;33(2):443-445. [doi: 10.1109/tassp.1985.1164550]
4. Rezayee A, Gazor S. An adaptive KLT approach for speech enhancement. IEEE Trans. Speech Audio Process 2001;9(2):87-95. [doi: 10.1109/89.902276]
5. Scalart P. Speech enhancement based on a priori signal to noise estimation. 1996 Presented at: Speech enhancement based on a priori signal to noise estimation. In IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings(Vol. 2, pp.). IEEE; May 5, 1996; Atlanta, GA, USA p. 629-632. [doi: 10.1109/icassp.1996.543199]
6. Lai YH, Su YC, Tsao Y, Young ST. Evaluation of generalized maximum a posteriori spectral amplitude (GMAPA) speech enhancement algorithm in hearing aids. 2013 Presented at: 2013 IEEE International Symposium on Consumer Electronics (ISCE); June 2013; Hsinchu, Taiwan p. 245-246. [doi: 10.1109/isce.2013.6570208]
7. Dawson PW, Mauger SJ, Hersbach AA. Clinical evaluation of signal-to-noise ratio–based noise reduction in Nucleus cochlear implant recipients. Ear and hearing 2011;32(3):382. [doi: 10.1097/aud.0b013e318201c200]
8. Loizou PC, Lobo A, Hu Y. Subspace algorithms for noise reduction in cochlear implants. The Journal of the Acoustical Society of America 2005 Nov;118(5):2791-2793. [doi: 10.1121/1.2065847] [Medline: 16334894]
9. Mauger SJ, Dawson PW, Hersbach AA. Perceptually optimized gain function for cochlear implant signal-to-noise ratio based noise reduction. The Journal of the Acoustical Society of America 2012 Jan;131(1):327-336. [doi: 10.1121/1.3665990] [Medline: 22280595]
10. Tu Y, Du J, Lee C. Speech enhancement based on teacher? Student deep learning using improved speech presence probability for noise-robust speech recognition. IEEE/ACM Trans. Audio Speech Lang. Process 2019 Dec;27(12):2080-2091. [doi: 10.1109/taslp.2019.2940662]
11. Wang D, Chen J. Supervised speech separation based on deep learning: an overview. IEEE/ACM Trans Audio Speech Lang Process 2018 Oct;26(10):1702-1726 [FREE Full text] [doi: 10.1109/TASLP.2018.2842159] [Medline: 31223631]
12. Wu M, Wang D. A one-microphone algorithm for reverberant speech enhancement. 2003 Presented at: IEEE International Conference on Acoustics, Speech, and Signal Processing; May 2003; Hong Kong, China. [doi: 10.1109/icassp.2003.1198925]
13. Healy EW, Delfarah M, Vasko JL, Carter BL, Wang D. An algorithm to increase intelligibility for hearing-impaired listeners in the presence of a competing talker. The Journal of the Acoustical Society of America 2017 Jun;141(6):4230-4239. [doi: 10.1121/1.4984271] [Medline: 28618817]
14. Kumar A, Florencio D. Speech enhancement in multiple-noise conditions using deep neural networks. arXiv:1605.02427. 2016. URL: https://arxiv.org/abs/1605.02427 [accessed 2016-05-09]
15. Xu Y, Du J, Dai L, Lee C. A regression approach to speech enhancement based on deep neural networks. IEEE/ACM Trans. Audio Speech Lang. Process 2015 Jan;23(1):7-19. [doi: 10.1109/taslp.2014.2364452]
16. Lu X, Tsao Y, Matsuda S, Hori C. Speech enhancement based on deep denoising autoencoder. In: Interspeech. 2013 Aug Presented at: Interspeech 2013; 25-29 August 2013; Lyon, France.
17. Xu Y, Du J, Dai LR, Lee CH. Dynamic noise aware training for speech enhancement based on deep neural networks. 2014 Presented at: Fifteenth Annual Conference of the International Speech Communication Association; Sept 14, 2014; Singapore.
18. Lai Y, Chen F, Wang S, Lu X, Tsao Y, Lee C. A deep denoising autoencoder approach to improving the intelligibility of vocoded speech in cochlear implant simulation. IEEE Trans. Biomed. Eng 2017 Jul;64(7):1568-1578. [doi: 10.1109/tbme.2016.2613960]
19. Goehring T, Bolner F, Monaghan JJ, van Dijk B, Zarowski A, Bleeck S. Speech enhancement based on neural networks improves speech intelligibility in noise for cochlear implant users. Hearing Research 2017 Feb;344:183-194. [doi: 10.1016/j.heares.2016.11.012] [Medline: 27913315]
20. Goehring T, Keshavarzi M, Carlyon RP, Moore BCJ. Using recurrent neural networks to improve the perception of speech in non-stationary noise by people with cochlear implants. The Journal of the Acoustical Society of America 2019 Jul;146(1):705-718. [doi: 10.1121/1.5119226] [Medline: 31370586]

21. Lai YH, Tsao Y, Lu X, Chen F, Su YT, Chen KC, et al. Deep learning–based noise reduction approach to improve speech intelligibility for cochlear implant recipients. Ear and hearing 2018;39(4):795-809. [doi: 10.1097/aud.0000000000000537]
22. Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans. Knowl. Data Eng 2010 Oct;22(10):1345-1359. [doi: 10.1109/tkde.2009.191]
23. Latif S, Rana R, Younis S, Qadir J, Epps J. Transfer learning for improving speech emotion classification accuracy. arXiv. 2018. URL: https://arxiv.org/abs/1801.06353 [accessed 2018-01-19]
24. Fan Y, Qian Y, Soong FK, He L. Multi-speaker modeling and speaker adaptation for DNN-based TTS synthesis. 2015 Presented at: International Conference on Acoustics, Speech and Signal Processing; April 2015; Toronto, Ontario, Canada p. 4475-4479. [doi: 10.1109/icassp.2015.7178817]
25. Jia Y, Zhang Y, Weiss R, Wang Q, Shen J, Ren F, et al. A transfer learning and progressive stacking approach to reducing deep model sizes with an application to speech enhancement. Advances in neural information processing systems 2018:4480-4490 [FREE Full text] \
26. Wang S, Li K, Huang Z, Siniscalchi SM, Lee CH. Transfer learning and progressive stacking approach to reducing deep model sizes with an application to speech enhancement. 2017 Presented at: International Conference on Acoustics, Speech and Signal Processing; March 2017; New Orleans, LA, USA p. 5575-5579. [doi: 10.1109/icassp.2017.7953223]
27. Rix AW, Beerends JG, Hollier MP, Hekstra AP. Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs. 2001 May Presented at: International Conference on Acoustics, Speech, and Signal Processing; May 7-11, 2001; Salt Lake City, UT, USA p. 749-752. [doi: 10.1109/icassp.2001.941023]
28. Taal CH, Hendriks RC, Heusdens R, Jensen J. A short-time objective intelligibility measure for time-frequency weighted noisy speech. 2010 Presented at: International Conference on Acoustics, Speech and Signal Processing; April 2010; Dallas, TX, USA p. 4214-4217. [doi: 10.1109/icassp.2010.5495701]
29. Du J, Huo Q. A speech enhancement approach using piecewise linear approximation of an explicit model of environmental distortions. 2008 Presented at: Ninth Annual Conference of Tthe International Speech Communication Association; Sept 2008; Brisbane, Australia.
30. Davis S, Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Trans. Acoust., Speech, Signal Process 1980 Aug;28(4):357-366. [doi: 10.1109/tassp.1980.1163420]
31. Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. arXiv. 2015. URL: https://arxiv.org/abs/1503.02531 [accessed 2015-03-09]
32. Bengio Y. Learning deep architectures for AI. In: FNT in Machine Learning. Norwell, MA: Now Publishers Inc; 2009:1-127.
33. Mohamed A, Dahl GE, Hinton G. Acoustic modeling using deep belief networks. IEEE Trans. Audio Speech Lang. Process 2012 Jan;20(1):14-22. [doi: 10.1109/tasl.2011.2109382]
34. Mengusoglu E, Ris C. Use of acoustic prior information for confidence measure in ASR applications. 2001 Presented at: Seventh European Conference on Speech Communication and Technology; Sept 13, 2021; Virtual.
35. Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. ICML. 2010 Jan. URL: https://openreview.net/forum?id=rkb15iZdZB [accessed 2019-07-17]
36. Wong LLN, Soli SD, Liu S, Han N, Huang MW. Development of the Mandarin Hearing in Noise Test (MHINT). Ear Hear 2007 Apr;28(2 Suppl):70S-74S. [doi: 10.1097/AUD.0b013e31803154d0] [Medline: 17496652]
37. Mittal U, Phamdo N. Signal/noise KLT based approach for enhancing speech degraded by colored noise. IEEE Trans. Speech Audio Process 2000;8(2):159-167. [doi: 10.1109/89.824700]
38. Chen F, Loizou PC. Predicting the intelligibility of vocoded and wideband Mandarin Chinese. The Journal of the Acoustical Society of America 2011 May;129(5):3281-3290. [doi: 10.1121/1.3570957] [Medline: 21568429]
39. Chen F, Wong LLN, Qiu J, Liu Y, Azimi B, Hu Y. The contribution of matched envelope dynamic range to the binaural benefits in simulated bilateral electric hearing. J Speech Lang Hear Res 2013 Aug;56(4):1166-1174. [doi: 10.1044/1092-4388(2012/12-0255)] [Medline: 23926330]
40. Chen F, Hu Y, Yuan M. Evaluation of noise reduction methods for sentence recognition by Mandarin-speaking cochlear implant listeners. Ear Hear 2015 Jan;36(1):61-71. [doi: 10.1097/AUD.0000000000000074] [Medline: 25127321]
41. Lai Y, Tsao Y, Chen F. Effects of adaptation rate and noise suppression on the intelligibility of compressed-envelope based speech. PLoS ONE 2015 Jul 21;10(7):e0133519. [doi: 10.1371/journal.pone.0133519]
42. Wang SS, Tsao Y, Wang HLS, Lai YH, Li PH. A deep learning based noise reduction approach to improve speech intelligibility for cochlear implant recipients in the presence of competing speech noise. 2017 Presented at: Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC); Dec 2017; Kuala Lumpur, Malaysia p. 808-812. [doi: 10.1109/APSIPA.2017.8282144]
43. Cohen J. Statistical power analysis for the behavioural sciences. In: Hillsdale, NJ: Laurence Erlbaum Associates. Cambridge, Massachusetts: Academic press; 1988:273-406.
44. Dien J. Issues in the application of the average reference: Review, critiques, and recommendations. Behavior Research Methods, Instruments, & Computers 1998 Mar;30(1):34-43. [doi: 10.3758/bf03209414]
45. Williams LJ, Abdi H. Fisher's least significant difference (LSD) test. Encyclopedia of research design 2010:840-853. [doi: 10.1007/978-1-4020-6754-9_9279]

46. Williams LVD, Hinton G. Visualizing data using t-SNE. Journal of machine learning research. Journal of Machine Learning Research 2008:2579-2605.
47. Georgiev P, Lane ND, Mascolo C, Chu D. Accelerating mobile audio sensing algorithms through on-chip gpu offloading. 2017 Presented at: Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services; June 2017; New York, NY, USA p. 306-318. [doi: 10.1145/3081333.3081358]
48. Chen J, Ran X. Deep learning with edge computing: a review. Proc. IEEE 2019 Aug;107(8):1655-1674. [doi: 10.1109/jproc.2019.2921977]

Abbreviations

ANOVA: analysis of variance
CI: cochlear implant
CJ: Construction Jackhammer
DDAE: deep denoising autoencoder
DL: deep learning
DNN: deep neural network
LPS: log power spectra
MRT: Mass Rapid Transit
NC: noise classifier
NC+DDAE_T: noise classifier + deep denoising autoencoder with knowledge transfer
ND: noise-dependent
NI: noise-independent
NR: noise reduction
PESQ: perceptual evaluation of speech quality
SEM: standard error of the mean
SNR: signal-to-noise ratio
SSN_IEEE: speech shape noise from the Institute of Electrical and Electronics Engineers
STOI: short-time objective intelligibility
TMHINT: Taiwan Mandarin version of the hearing in noise test
***t*-SNE:** *t*-distributed stochastic neighbor embedding
WCR: word correct rate

Edited by R Kukafka; submitted 09.11.20; peer-reviewed by YC Chu, ST Tang; comments to author 30.11.20; revised version received 11.02.21; accepted 27.04.21; published 28.10.21.

Please cite as:
Li LPH, Han JY, Zheng WZ, Huang RJ, Lai YH
Improved Environment-Aware–Based Noise Reduction System for Cochlear Implant Users Based on a Knowledge Transfer Approach: Development and Usability Study
J Med Internet Res 2021;23(10):e25460
URL: https://www.jmir.org/2021/10/e25460
doi:10.2196/25460
PMID:34709193

Review

Applications of Blockchain in the Medical Field: Narrative Review

Yi Xie[1,2], MD; Jiayao Zhang[1,2], MD; Honglin Wang[1,2], MD; Pengran Liu[1,2], MD; Songxiang Liu[1,2], MD; Tongtong Huo[2,3], MD; Yu-Yu Duan[2,4], MD; Zhe Dong[5], MD; Lin Lu[1,2*], MD; Zhewei Ye[1*], MD, Prof Dr

[1]Department of Orthopedics Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[2]Laboratory of Intelligent Medicine, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[3]School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China
[4]Hubei University of Chinese Medicine, Wuhan, China
[5]Wuhan Academy of Intelligent Medicine, Wuhan, China
[*]these authors contributed equally

Corresponding Author:
Zhewei Ye, MD, Prof Dr
Department of Orthopedics Surgery, Union Hospital
Tongji Medical College
Huazhong University of Science and Technology
1277 Jiefang Dadao
Wuhan, 430022
China
Phone: 86 17771413685
Email: yezhewei@hust.edu.cn

Abstract

Background: As a distributed technology, blockchain has attracted increasing attention from stakeholders in the medical industry. Although previous studies have analyzed blockchain applications from the perspectives of technology, business, or patient care, few studies have focused on actual use-case scenarios of blockchain in health care. In particular, the outbreak of COVID-19 has led to some new ideas for the application of blockchain in medical practice.

Objective: This paper aims to provide a systematic review of the current and projected uses of blockchain technology in health care, as well as directions for future research. In addition to the framework structure of blockchain and application scenarios, its integration with other emerging technologies in health care is discussed.

Methods: We searched databases such as PubMed, EMBASE, Scopus, IEEE, and Springer using a combination of terms related to blockchain and health care. Potentially relevant papers were then compared to determine their relevance and reviewed independently for inclusion. Through a literature review, we summarize the key medical scenarios using blockchain technology.

Results: We found a total of 1647 relevant studies, 60 of which were unique studies that were included in this review. These studies report a variety of uses for blockchain and their emphasis differs. According to the different technical characteristics and application scenarios of blockchain, we summarize some medical scenarios closely related to blockchain from the perspective of technical classification. Moreover, potential challenges are mentioned, including the confidentiality of privacy, the efficiency of the system, security issues, and regulatory policy.

Conclusions: Blockchain technology can improve health care services in a decentralized, tamper-proof, transparent, and secure manner. With the development of this technology and its integration with other emerging technologies, blockchain has the potential to offer long-term benefits. Not only can it be a mechanism to secure electronic health records, but blockchain also provides a powerful tool that can empower users to control their own health data, enabling a foolproof health data history and establishing medical responsibility.

(J Med Internet Res 2021;23(10):e28613) doi:10.2196/28613

KEYWORDS

blockchain; smart health care; health care; health data; review; COVID-19; electronic health records

Introduction

Background

With the development of medical informatization, the amount of available health care data is increasing at an extremely fast rate. The sharing and use of medical information have played an important role in the optimization of medical resource allocation, clinical decision-making assistance, medical quality monitoring, precision medicine, and disease risk assessment and prediction [1-3]. However, such data sharing comes with the risks of data security and privacy concerns, data dictatorship, insufficient autonomy of the subject, increased social unfairness, and others. Moreover, the sudden development of the COVID-19 pandemic has also posed new challenges for personal health data sharing and mining. As another world-changing technology based on cloud computing, the Internet of Things (IoT), and artificial intelligence (AI), blockchain may provide a solution to the above-mentioned problems owing to its unique characteristics such as decentralization, autonomy, credibility, and transparency [4,5]. Therefore, it is necessary to explore the impact of blockchain on the medical industry to further clarify the potential value of the medical application of blockchain technology in the context of medical informatization.

Objectives

Although previous studies have analyzed blockchain applications from the perspective of technology, business, or patient care, few studies have focused on its actual use-case scenarios in health care. In particular, the outbreak of COVID-19 brought about some ideas for the application of blockchain in medical practice. Therefore, this paper aims to provide a systematic review of the use of blockchain technology in health care. In addition to summarizing the basic principles and framework, this review highlights the different characteristics based on blockchain and the application of blockchain in clinical practice. Furthermore, integration with other technologies is discussed, which provides a reference for future research. Toward this end, we first describe the framework and perform specific technical analyses from a theoretical standpoint. We then summarize the application of blockchain in medical scenarios and sort them into three parts according to the characteristics of blockchain from a practical point of view. Furthermore, we summarize the use cases of blockchain in the fight against the COVID-19 pandemic, including the prevention of infectious diseases, location sharing, and contact tracing, and the supply chain of injectable medicines. Finally, we explore the integration of blockchain with new technology and highlight some of the associated challenges. By depicting a blueprint of interconnected ecosystems in health care, we aim to provide some reflections for engineers and decision-makers in the medical industry.

Methods

Design

A systematic review design with narrative methods was used to analyze the existing evidence. More precisely, a review methodology was followed to form a conception of the application of blockchain technology in health care.

Search Strategy

We performed a comprehensive literature search on May 10, 2021. The following electronic databases were searched with the assistance of an information specialist at a medical library: PubMed, EMBASE, Scopus, IEEE, and Springer. The review was limited to articles published in English between 2016 and 2021 for which abstracts were available. This time frame was chosen based on the dramatic improvement in information technology that occurred during this period. The review was also limited to studies of blockchain technology in the health care domain. The initial search terms used were as follows: (blockchain) OR (distributed ledger technology) OR (smart contract) AND (health care). After reviewing the literature identified through these search terms, we added the search terms “health data,” “clinical,” “biomedical research,” “supply chain,” “drug safety,” and “health monitoring” to capture the relevant studies found in the references of the articles retrieved during the initial search (Table 1).

Table 1. Databases and search terms used, and number of references found for each (N=1647).

Database	Search terms	Number of references retrieved
PubMed (MEDLINE)	(((blockchain) OR (distributed ledger technology)) OR (smart contract)) AND (health care)	232
Embase	((blockchain or distributed ledger technology) and health care).af.	185
Scopus	(TITLE-ABS-KEY (block AND chain) AND TITLE-ABS-KEY (health AND care))	259
IEEE	((“All Metadata”: blockchain or distributed ledger technology) AND “All Metadata”: health care)	295
Springer	blockchain AND in AND health AND care AND “distributed ledger technology” AND (blockchain)	676

Inclusion and Exclusion Criteria

A total of 1647 search results were screened for relevance using their titles or abstracts, leaving 60 articles that were fully reviewed and summarized. Our inclusion criteria were as follows: (1) application studies that demonstrated the effectiveness of blockchain technology, including data preservation and sharing, medical insurance and supply chain, clinical and biomedical research, drug safety, medical education, electronic prescription anticounterfeiting, wearable devices, and epidemic prevention; (2) English language studies published in scientific journals; (3) studies for which the full text was

available; and (4) completed studies. The review was not restricted according to the study location, and any international study written in English was eligible. Our exclusion criteria were papers describing the process of blockchain design, books or book chapters, letters, statistical reviews, dissertations, editorials, and study protocols.

Study Selection

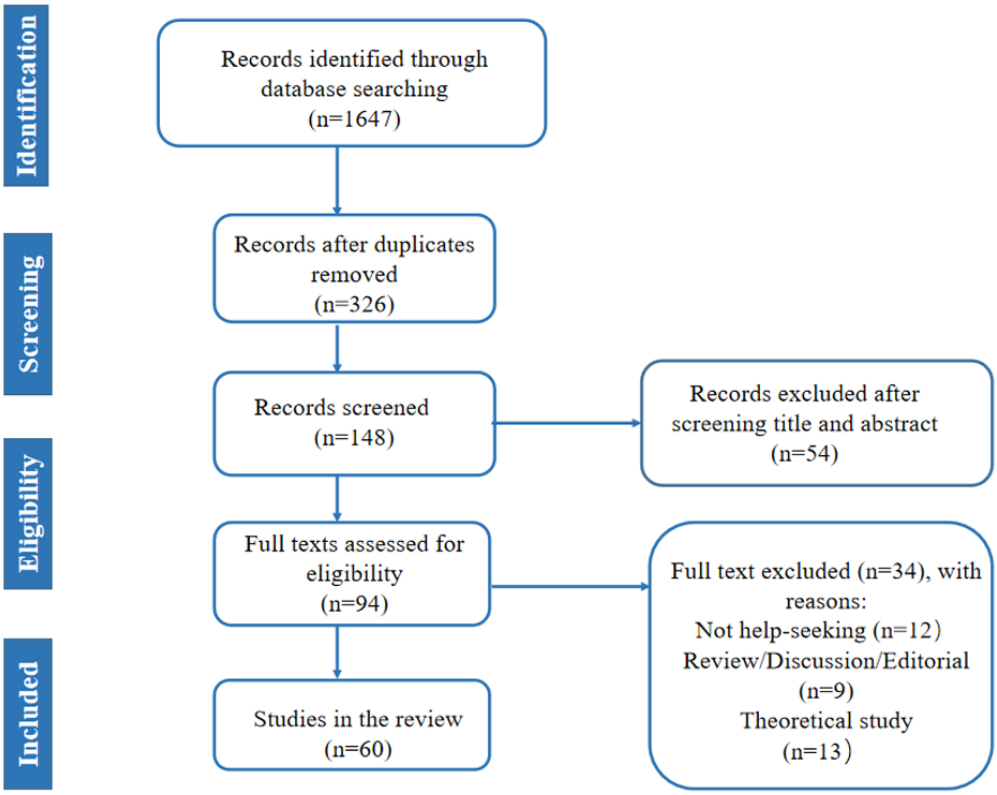
The research selection included four steps. First, three authors (XY, JZ, and HW) independently screened all titles and abstracts related to the systematic review (N=1647). Second, the abstracts of all related articles were qualified by five authors (PL, SL, TH, YD, and ZD). Third, the full texts of eligible publications were obtained and screened by two authors (XY and LL) according to the inclusion and exclusion criteria. If there were any different opinions in the decisions made, the controversial documents were discussed until a consensus was reached, with the support of ZY. Fourth, the reference list of all included papers and the system overview identified in the original search were checked to identify other publications that met our inclusion criteria.

Results

Search Findings

The search identified 1647 potentially relevant documents after a review of titles and abstracts, 60 of which satisfied the inclusion criteria after a full-text review (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) flow diagram outlining the review process.



Application Principle of Blockchain Technology in Health Care

Distributed Ledger and the Characteristics of Blockchain

Although traditional databases can have built-in redundancy, they do not have the advantage of replication on every node [6-8]. In the nodes of blockchain, only cryptography and programs are used to realize point-to-point transactions, complete cooperation, and establish tasks. In this mode, the problems of low efficiency, high cost, and data security caused by centralized systems can be solved [7-9]. Some of the characteristics of blockchain and closely related applications are summarized in Table 2. By maintaining an immutable, tamper-proof, consecutive list of transactional data in a distributed network, blockchain has created several disruptions in incumbent business processes, and provides a promising new distributed framework for amplifying the integration of health care information across a range of stakeholders [1-3,10,11].

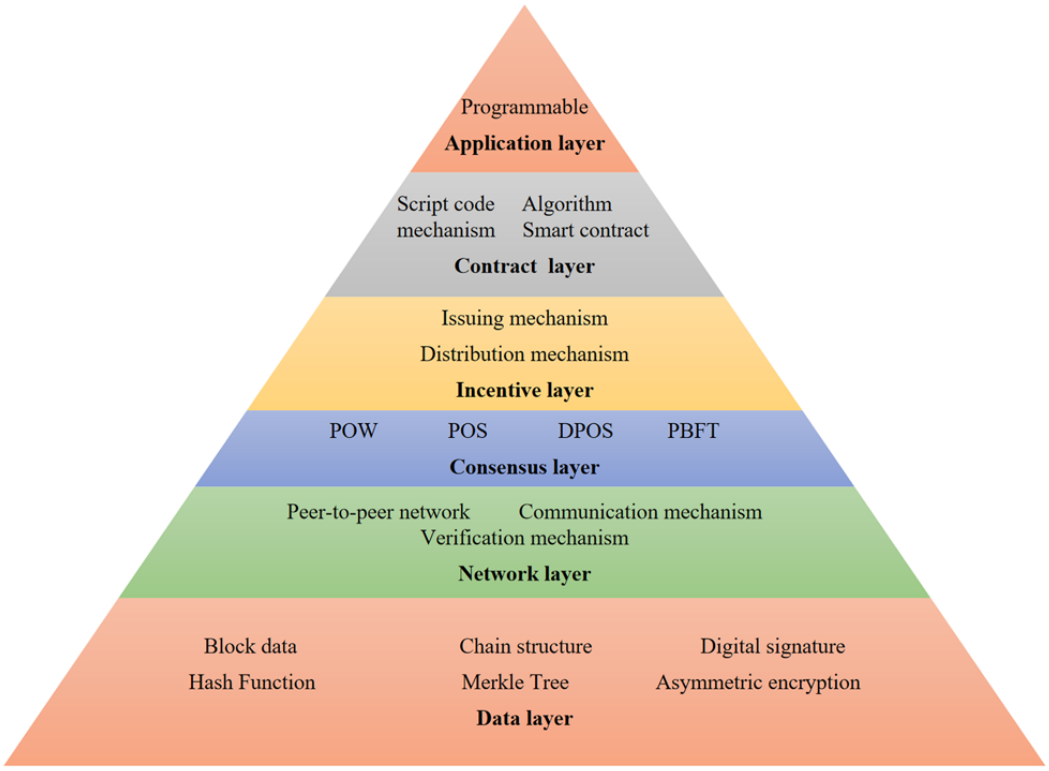
Table 2. Characteristics of blockchain and related applications.

Characteristics	Description	Related applications
Decentralization	There is no centralized management organization in the whole network, but rather a distributed end-to-end network structure [4-7].	Health data preservation and authorization; the preservation and authorization of health data
Autonomy	Using consensus-based specifications and protocols to enable all nodes to exchange data freely and safely in a detrusted environment [4,5,7,9,10].	Medical insurance; health status monitoring and tracking (wearable devices)
Credibility	Asymmetric cryptography is used to encrypt transaction data, and with the help of workload proof mechanism to ensure that the data are difficult to tamper with in theory [5-7,10].	Controlling of fake drugs; anticounterfeiting of electronic prescriptions; digital ledger of students' scores
Transparency	All transaction records are open and transparent in the whole network, breaking the information asymmetry [5-10,12].	Supply chain; clinical trials; biomedical research

Framework Structure of Blockchain

The framework and structure of blockchain can be divided into six layers (Figure 2). The application layer carries out the accounting, transferring, and verifying functions on the client side. Then, the contract layer behind the application layer includes the script code, algorithm mechanism, and smart contract, which performs transaction identification. The incentive layer involves the distribution mechanisms. The consensus layer ensures the consistency of the distributed systems and that consensus can be reached even if there are malicious nodes in the network. This layer also avoids the problem of a "double-spend attack" and ensures the generation time of blocks. In addition, the network layer is the mechanism for which we use the P2P network to complete communication and confirmation, and the data layer includes a series of encryption and storage technologies [6,7,9,12,13].

Figure 2. Framework structure and classification of blockchain. DPOS: delegated proof of stake; PBFT: practical Byzantine fault tolerance; POS: proof of stake; POW: proof of work.



Classification of Blockchain

Generally speaking, blockchain can be divided into public chains, consortium chains, and private chains, as shown in Table 3 [14]. Hasselgren et al [15] counted the current medical use of various types of blockchains and frameworks, and found that the consortium chain is the most widely used. Currently, ethernet and hyper ledger fabric are the most popular frameworks. Hasavari et al [16] showed that hyper ledger fabric is the most effective frame structure combined with medical treatment at present. One solution to the problem regarding the combination of blockchain and medical treatment is to replace patient care reports with electronic health records (EHRs) so that they can interoperate with other EHRs. Another solution suggests

integration of the hyper ledger platform with an interplanetary file system (IPFS), which is a point-to-point method for storing and sharing media in distributed file systems that uses BitTorrent technology. The data themselves are stored on the IPFS, and the hash pointer is stored on the blockchain. MedRec [17] is a decentralized record management system utilizing blockchain for authentication, confidentiality, responsibility, and data sharing. At present, most medical data-sharing and distribution solutions use the allowed blockchain technology and rely on business process integration; that is, customers run code on each node and go through a specific process until they store the data in the ledger [18].

Table 3. Classification of blockchain [14].

Property	Public blockchain	Consortium blockchain	Private blockchain
Consensus determination	All miners	Selected set of nodes	One organization
Read permission	Public	Could be public or restricted	Could be public or restricted
Immutability	Nearly impossible to tamper	Could be tampered	Could be tampered
Efficiency	Low	High	High
Consensus process	Permissionless	Permissioned	Permissioned

Application Scenarios in Health Care

Principal Scenarios

Blockchain technology can use cryptography to program and operate smart contracts composed of data encryption and automated script codes, and also provides distributed infrastructure and economic incentives [4,5,10,11,13,14,17-19]. Based on previous studies, we summarize three main blockchain-based medical scenarios (Table 4) and discuss the integration of blockchain and emerging technologies in the future.

Table 4. Medical application scenarios of blockchain and the details of related studies (N=60).

Applications according to technical characteristics	Number of studies	References
Preservation and authorization of personalized health data		
Preservation of health data	10	Kim et al [6], Shi et al [20], Zhou et al [21], Dubovitskaya et al [22], Jones et al [23], Chen et al [24], Hylock et al [25], Lo et al [26], Xiao et al [27], Yue et al [28]
Medical data sharing	8	Xia et al [29], Patel et al [30], Fan et al [31], Yazdinejad et al [32], Zhu et al [33], Dubovitskaya et al [34], Khurshid et al [35], Cheng et al [36]
Medical insurance	1	Zhou et al [37]
Health status monitoring and tracking	3	Brogan et al [38], Griggs et al [39], Ichikawa et al [40]
Promoting the management of social and public health		
Administration of medicine	8	Mao et al [41], Fernández et al [42], Sylim et al [43], Lohmer et al [44], Hoy et al [45], Vruddhula et al [46], Tseng et al [47], Mackey et al [48]
Anticounterfeiting of electronic prescriptions	2	Aldughayfiq et al [49], Li et al [50]
Prevention and control of pandemic	10	Raghavendra et al [51], Nandi et al [52], Marbouh et al [53], Mashamba et al [54], Bansal et al [55], Khurshid et al [56], Abdel-Basset et al [57], Resiere et al [58], Chang et al [59], Garg et al [60]
Empowering the credibility of medical education and research		
Medical education	3	Verde et al [61], Funk et al [62], Durant et al [63]
Clinical trials	7	Benchoufi et al [64], Wong et al [65], Omar et al [66], Hirano et al [67], Zhuang et al [68], Nugent et al [69], Wan et al [70]
Biomedical research	8	Jin et al [71], Kuo et al [72-74], Ozercan et al [75], Johnson et al [76], Chen et al [77], Mamoshina et al [78]

Personalized Health Data Preservation and Authorization

Overview

With the accumulation of a large amount of individual health information, a reliable storing and sharing approach is needed to ensure the safety of patients' private information. The existing medical data management systems are generally based on a set of centralized servers, which build a large site system or centralized relational database system. Blockchain is an open distributed ledger based on peer-to-peer networks and consensus algorithms with natural advantages in solving these problems.

Preservation of Electronic Medical Records

With the increased specialization of health care services and high levels of patient mobility, accessing health care services across multiple hospitals or clinics has become very common for diagnosis and treatment, particularly for patients with chronic diseases. Based on blockchain, Dubovitskaya et al [22] developed ACTION-EHR, an EHR data management system for the radiation treatment of cancer. The synchronous nodes in the blockchain network can immediately find data changes and prevent malicious tampering with the data. Similarly, HealthChain, a novel patient-centered blockchain framework designed by Hylock and Zeng [25], offers patients and providers access to consistent and comprehensive medical records. To integrate patient referral data from the National Health Insurance Administration national medical referral system, Lo et al [26] developed a blockchain-enabled framework for acquiring electronic medical record (EMR) and EHR data of patients in hospitals and community-based clinics. The framework assists in the establishment of an alliance-based medical referral service to promote trusting relationships and transaction security among patients, family doctors, and specialists. In addition, Yue et al [28] proposed the Health Care Data Gateway, a medical data network that not only enables a patient to control their own data easily but also allows untrusted third parties to process health data securely. By expounding the working principle and process of HealthChain in detail, Xiao et al [27] verified the feasibility of using blockchain in EMRs. As a major direction of telemedicine, the blockchain-based storing system of health data will play a vital role in protecting the privacy of patients and ensuring credibility.

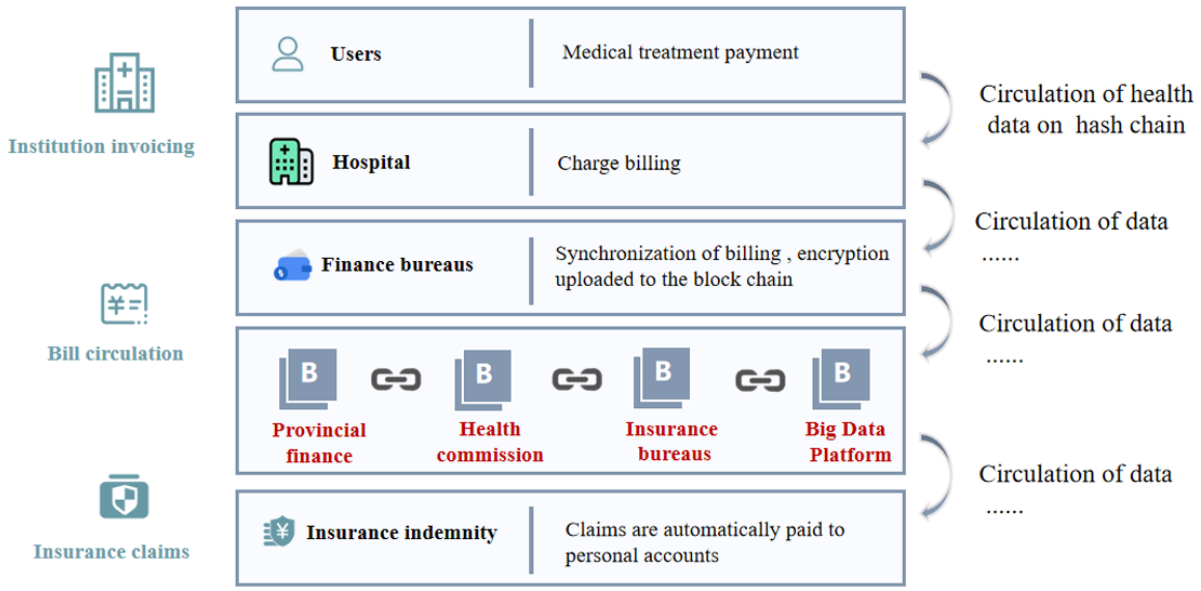
Exchange of Patients' Medical Data

Traditional medical records are stored in the central database of various hospitals, which leads to the phenomenon of an "isolated island of information" in the medical field [1-3,29,31,32,35,36,79-83]. Blockchain allows patients to have access to their medical information while authorizing them to grant access of their EHRs to third parties as they see fit. Health care providers have easy access to all patients' medical data, regardless of when and where health care services are provided. Guardtime [79], a Netherlands-based data security firm, partnered with the government of Estonia to create a blockchain-based framework to validate patient identities when sharing their health records. Xia et al [29] introduced MeDShare, a system that addresses the trust issue of medical data sharing among medical big data custodians in an untrusted network environment. For example, it may be urgently necessary to know the radiation dose received during treatment to avoid possible harmful consequences for the patient. In a clinical imaging department, a smart framework for cross-domain radiological image sharing and patient-defined access permissions was developed. Zhu et al [33] proposed a cloud resource–sharing model and explored the cloud service of breast tumor diagnosis based on a consensus-oriented blockchain to protect the privacy of cancer patients. In clinical work, timely and accurate sharing of private data related to personal health is necessary for patient treatment, and the significance of blockchain in this respect should be realized by more doctors and decision-makers.

Simplifying the Process of Medical Insurance

In the field of medical insurance, the insurance process involves patients, medical institutions, and insurance service providers; however, the speed of information exchange among them is slow. Based on blockchain, the process can be simplified. Zhou et al [37] proposed a blockchain-based medical insurance storage system named MIStore, which was deployed on the Ethereum blockchain and provided a platform between insurance companies and hospitals. The system improved the efficiency of the information storage process so that insurance companies can quickly settle claims and preauthorize payments to patients. Based on blockchain, the real-time circulation and sharing of bills can be realized, which can ensure the payment of claims and reduce user advances. The technical framework of medical insurance based on blockchain is depicted in Figure 3.

Figure 3. Medical insurance on blockchain.



Monitoring Health Status Based on Personal Wearable Devices

Chronic disease is the dominant cause of morbidity and mortality worldwide. The cost of treating traditional chronic diseases such as cardiovascular disease was US $555 billion in the United States in 2014 and is estimated to rise to US $1.1 trillion by 2035. With the extensive development of wearable devices, smartphones, clouds, and wireless systems, the integration of these devices can be applied to tracking the health status of patients with chronic diseases. By using blockchain-based transmission and storage mechanisms, health information can be uploaded to the cloud more accurately and in a more timely manner [84-86]. Brogan et al [38] demonstrated that it was feasible to use a distributed ledger to receive authenticated activity data from a wearable device. Griggs et al [39] proposed the utilization of blockchain-based smart contracts to evaluate information collected by health care devices and log transactions in a wireless body area network. A quick-response system based on the blockchain can detect emergencies such as asthma attacks and alert the closest emergency services in real time to provide immediate relief. Continuous monitoring can not only detect adverse health events early and reduce the risk of such events but can also improve the monitoring of medication compliance and reduce unnecessary treatments [40,87]. Based on blockchain, wearable devices can potentially reduce patients' needs for more complex interventions, which in turn reduce the number of emergency department visits and hospitalizations, caregiver burden, and health care costs.

Promoting the Management of Social and Public Health

Overview

The management of social and public health is involved in many aspects, including disease prevention and control, management of drugs, authentication of health records, and medical insurance. Due to the lack of information exchange, the process often becomes inefficient. Based on the characteristics of blockchain such as traceability and immutability, we can apply it to simplify the process of insurance, administration of medicine, anticounterfeiting electronic prescriptions, and controlling pandemic issuance.

Administration of Medicine

The Center for the Public Interest in the United States estimates that global sales of counterfeit drugs are likely to exceed US $75 billion this year, representing an increase of 90% within the past 5 years. According to the World Health Organization, 10% of the world's medicines are counterfeit drugs, 30% of which are found in developing countries. Owing to their substandard dose and purity, these counterfeit drugs pose a great threat to people's physical and mental health [41,88,89]. In the supply chain, supply chain management systems have played important roles and have great importance for an enormous number of industries and organizations. With the development of mobile technologies, the supply chain of drugs is considered relatively mature and easy to adopt; the existing solution is to track the logistics information and check the authenticity of drugs through quick response codes, radio frequency identification, or SMS text messaging; however, these methods do not prevent counterfeit drugs from entering the logistics chain. The combination of supply chain and blockchain technologies can revolutionize the medical field, bringing about the benefits of vast objects' connectivity, and features to process and record a large amount of medical information with more efficiency, privacy, and security [89].

Due to the particularity of drug production, the legitimacy and authenticity of drugs can be effectively guaranteed by using blockchain anticounterfeiting technology [45,46]. Using blockchain, a supply chain can maintain the privacy and security of information and provide great immutability for transactions. For example, Mao et al [41] provided a blockchain-based credit evaluation system to strengthen the effectiveness of supervision and management in the food and drug supply chain. Based on

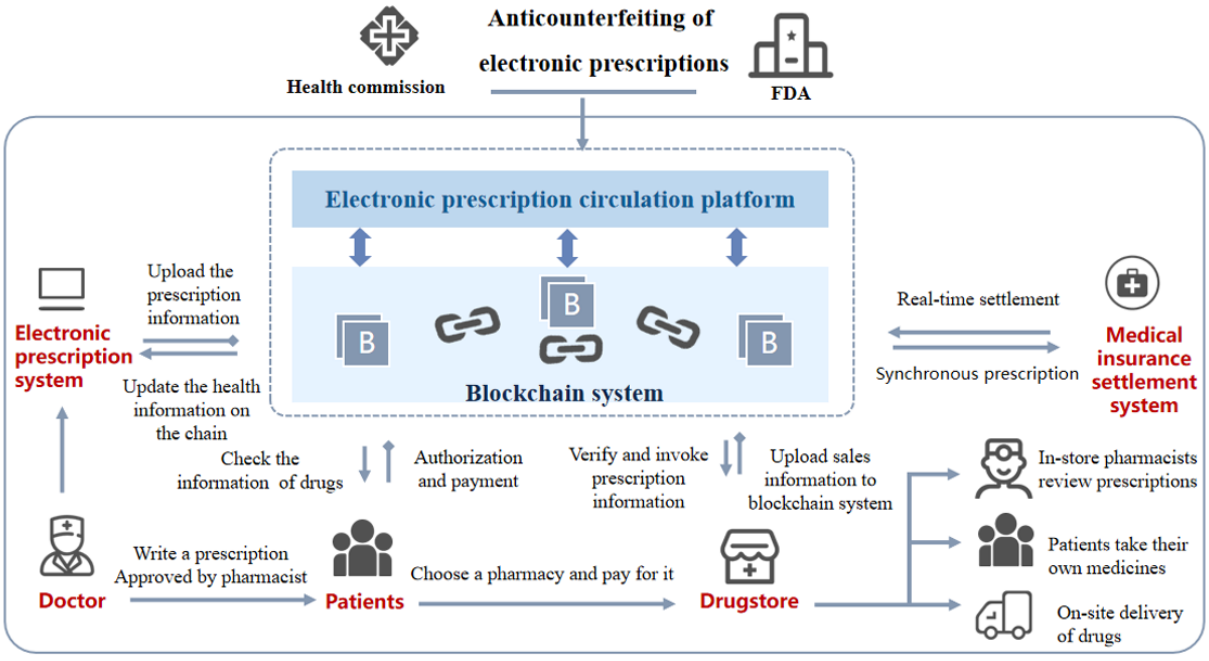
the concept of Industry 4.0, Fernández-Caramés et al [42] presented the design of an unmanned aerial vehicle and blockchain-based supply chain system for traceability applications. To develop a supply regulatory system for drugs, Sylim et al [43] developed a distributed application based on smart contracts. Blockchain has interesting features of validation and smart contracts, which are very useful for medical supply chains to manage millions of transactions in more authentic ways [44]. In the pharmaceutical industry, Chronicled Inc launched a prototype technology combining near field communication–embedded adhesive seals that were registered and verified on the blockchain. Vruddhula [46] launched a project called Oggic to prevent the entry of counterfeit drugs into the supply chain. Tseng et al [47] chose the Gcoin blockchain to address the anticounterfeiting problem in the pharmaceutical world. Each transaction would be recorded on the blockchain, making it tamper-proof, decentralized, time-stamped, and highly secure. Mackey et al [48] proposed a framework for improving the claims process by using blockchain, which makes the adjudication process more patient-centric and prevents drug fraud and abuse.

The primary function of blockchain in the circulation of drugs can be summarized as follows: (1) track and trace pharmaceutical raw materials and finished products in an immutable digital ledger, (2) provide greater transparency of fake drugs by allowing participants to verify their authenticity, (3) integrate anticounterfeit devices into the IoT and provide better authentication, and (4) serve as an underlying technology to enhance information exchange across different actors in the drug supply chain [47,48].

Anticounterfeiting of Electronic Prescriptions

With the continuous development of medical treatment on the internet, the establishment of online consultation platforms has facilitated consultation needs. Currently, individuals can conduct medical consultations through online platforms, purchase drugs from certified online pharmacies, and provide services such as drug delivery in the same city [49]. However, fake symptoms and irregular electronic prescriptions continue to emerge, which lead to the behavior of relying on fake prescriptions to buy drugs. Based on the decentralized and traceable blockchain system, a blockchain online consultation platform can be established to ensure the storage of large sample data and the sharing of health information [2,3,5,11]. The framework is shown in Figure 4. Such a project mainly realizes the functions of hospital management, physician management, user management, online consultation, and prescriptions through a doctor-patient online consultation platform, and achieves information credibility with the help of blockchain digital authentication, integral management, and other technologies. In this framework, users can be divided into doctors, ordinary users (patients), and administrators. Doctors need to pass the qualification examination to achieve the functions of prescribing medicine and diagnosing diseases, while patients need to register their personal information and submit a description of their illness to obtain a prescription. The consultation information is uploaded and recorded in personal medical records to prevent patients from fabricating false health information [50]. According to the results of physical examination records, medical diagnosis, and prescription opinions uploaded, the above-mentioned process can ultimately be realized in this blockchain-based tamper-proof system.

Figure 4. Structure for anticounterfeiting of electronic prescriptions. FDA: US Food and Drug Administration.



Pandemic Prevention and Control (Case of COVID-19)

COVID-19 spread rapidly around the world during 2020, which has attracted substantial attention on methods to prevent and control epidemics. Scientists are struggling to find a model to fight COVID-19. Blockchain technology has several potential use cases that can help tackle the current pandemic crisis. Blockchain can be used to simplify the clinical trial processes for vaccines and drugs, raise public awareness, transparently track donations and fundraising activities, and act as a reliable data tracker. Compared with traditional supply information management, blockchain smart contracts can guarantee the rights and obligations between nodes and the judgment of contract execution [51-53,56,58,59]. A joint Walmart-IBM project demonstrated how tracking the sources of contamination in green vegetables, a task that previously took months, could be achieved within seconds using blockchain [53]. Thus, many experts have made meaningful attempts in the control of epidemics and pandemics. Mashamba et al [54] proposed a blockchain-based and AI-coupled self-testing and tracking system for infectious diseases. Based on blockchain, International Classification of Diseases codes (current and previous versions) can be used to map the causes of death and observe disease trends and pattern changes across geographical locations over time. Bansal et al [55] suggested the use of a blockchain-based system to mitigate the falsification of test reports and encourage people to contact individuals with immunity-based licenses. Blockchain also enables information to be collected from individuals without identifying them by using a system of public and private keys. The DeepTrace system based on the blockchain can provide anonymized personal identification while allowing regulators and health care providers to contact people who are at risk of being potentially infected [56]. In addition, Abdel-Basset et al [57] proposed a framework integrating different disruptive technologies and blockchain to provide governance with an integrated vision toward managing the COVID-19 pandemic. Although epidemics and pandemics have seriously affected our lives, they also promoted the advancement of scientific technologies in the process of combating viruses and enhanced our capabilities in facing such emergencies [58-60].

Empowering the Credibility of Medical Education and Research

Overview

Medical education and research are promising areas where the introduction of blockchain may bring about benefits. Based on a blockchain system, reports and procedures at different stages of the study, as well as test results, and congress and course attendance in medical education could be easily archived and serve as digital proof of acquired competencies performed [62,63]. Since real clinical and experimental data are inseparable from high-quality medical research, the source and reliability of data are essential for researchers [64-66,68,90]. Based on the immutability and transparency of blockchain, the uploading of experimental data records can accelerate clinical data sharing and prevent academic misconduct [69].

Medical Education

Medical education is constantly changing and must adapt to address advances in biomedical sciences, improvements in learning theory, new regulatory policy, technological innovation, and efforts to have health care professionals perform at the highest level of competency. Blockchain technology in medical education has the potential to help solve many of the challenges currently faced by academic administrators, faculty, learners, and institutions. Since medical education is a lifelong learning process, a blockchain framework and measurable public exchanges between learners and teachers allow for the transmission of content, feedback about instructional designs, evaluation of learners, competency assessment, and certification.

A blockchain-based structure for the recording, crediting, and appraisal of educational deliverables could be a robust way for educators to track the value that their academic and system achievements create. In medical schools, blockchain can be used to store and track students' scores and abilities acquired through a range of different clinical environments. Doctors can then decide whether they are willing to share such information so that verified certificates and diplomas can be issued more easily and the process can be more cost-effective and tamper-proof [61,62]. Verde et al [61] proposed that the introduction of blockchain in academic radiology settings can be valuable for monitoring resident progress over the years. Massachusetts Institute of Technology launched a pilot program in 2017 involving the issuance of digital diplomas to students' smartphones via an app called Blockcerts Wallet, which is based on blockchain [63]. Blockchain could optimize the use of administrative resources by reducing the bureaucratic workload, with the added benefit of increased transparency, as records stored via blockchain can be automatically verified. Objectively speaking, blockchain implementation in the educational system could be in the assessment of faculty member competencies and academic performance in a secure and unalterable fashion.

Clinical Trials

The integrity of data in a clinical trial is essential; however, the current data management process is too complex and highly labor-intensive. By applying blockchain technology to medical research, the data can be time-stamped and transparent [64,65]. Wong et al [65] illustrated that a blockchain-based file and data structure could be used to reliably safeguard data in a clinical trial network. Omar et al [66] proposed a blockchain-based framework for computed tomography (CT) data management in clinical trials. Hirano et al [67] performed a project to demonstrate data management under a regulatory sandbox and tested the system through a clinical trial for breast cancer. Cichosz et al [91] explored the operational concept of the use of blockchain to improve data management and analyze diabetes in clinical observations. Nugent et al [69] showed that smart contracts can act as trusted administrators, which can improve the transparency of data reporting in clinical trials. In addition, Engel et al [92] proposed that blockchain can play an essential role in improving surgical outcome research and trial design. Even before clinical trials begin, all plans, agreements, scenarios, and possible results can be stored on blockchain [70]. This approach can transform our thinking about trial design and

produce truly verifiable and immutable data, which in turn can lead to better data reproducibility.

Biomedical Research

Biomedical data sharing has always been a cornerstone of scientific development. In the open world of science, it is inevitable to share, access, analyze, and learn from different sources of data for a meaningful result [73,74,77,93]. Blockchain can not only help in clinical trials but can also accelerate biomedical research and reduce reporting selectivity and fabrication, which are widespread problems in today's science fields. This can be achieved by integrating the consensus model of blockchain into current solutions to decentralized data storage and analysis. Jin et al [71] introduced LifeCODE.ai, a blockchain-based genomics big data platform, which aims to provide relatively safe and trustworthy data storage for genomic stakeholders. This is a decentralized approach in which each owner has complete control over their data, including where it is stored, who can access it, and when it is updated. This approach may be the best way of sharing scientific data.

The majority of the recent approaches to personalized medicine in oncology and other diseases have relied on various data types, including multiple types of genomic, transcriptomic, microRNA, proteomic, antigen, imaging, physiological, and other data. Research institutions can use DNA data stored in the blockchain to perform advanced searches to find topics of interest for potential genomic research. However, biomedical data are often personal, private, and sensitive, and should thus be treated carefully. There are currently a few similar proposals to help protect the data for academia. The first is the Cancer Gene Trust being developed by the Global Alliance for Genomics and Health Consortium, and the second is the CrypDist project. Both projects have similar properties, where summary data such as somatic cancer variation data are kept and distributed in a blockchain system [75]. In addition, Johnson et al [76] described a decentralized app to build a secure biomedical data-sharing system in biomedical and health care communities. This unprecedented progress has brought us into an era of genomic data–driven medicine and drug development, and blockchain technology will bring us into an era of genomics in an all-around way [77,78,93].

Integration with Emerging Technology in Health Care

Currently, the combination of the unsustainable cost of care, an aging population, the need for improved access to care, and the growth of precision medicine has ignited the ideal platform for disruptive innovation through blockchain and digital health [1-3,77,78,94-97]. Based on the distributed ledger, blockchain can be integrated with AI, cloud computing, big data, and the IoT, gaining more application scenarios and aiding the development of the health industry [78,94]. We summarize the practical usage of blockchain in Table 5.

Table 5. The integration of blockchain and emerging technology.

Emerging technology	Integrating with blockchain	The role of blockchain	Reference
AI[a]	Disease diagnosis and prediction, medical image–assisted reading, intelligent devices, new drug research, health management, and gene sequencing	Blockchain forms a natural foundation to standardize health data structures for AI training, clinical trials, and regulatory purposes. Blockchain-based AI could thereby accelerate the definition of phenotype-specific outcomes of orphan diseases, improve the representation of racial minorities, and reduce sex-specific inequalities	[94,95,98,99]
Cloud computing	Supporting communication and sharing data among stakeholders in health care	Based on the blockchain and cloud computing, the prediction model of disease evolution will be constructed, which plays a role in the prevention and control of major infectious diseases and tumors involving personal information	[82,83,100]
Big data	Offering a huge amount of data in real time to reduce health risks and optimize the outcome	Big data are stored and shared through a blockchain, thus avoiding the defect that centralized storage is easy to be lost and attacked	[77]
Wearable devices	Collecting personal data in an intimate and timely manner, the source of health data for telemedicine	Blockchain allows continuous individual monitoring, and requires robust, rapid, real-time analysis of physiological signals to avoid the storage of large amounts of data in a centralized system	[96]
Internet of Things	Supply chain and tracing of drugs, apparatuses, and wearable devices	Tracking the source and use of drugs and medical devices in the Internet of Things through blockchain technology to ensure authenticity and reliability	[88,100]
5G	Improving the speed of medical information sharing	The distributed processing of medical data in the 5G environment is carried out by blockchain technology, and access and use can only be achieved by obtaining multiparty permissions	[97]

[a]AI: artificial intelligence.

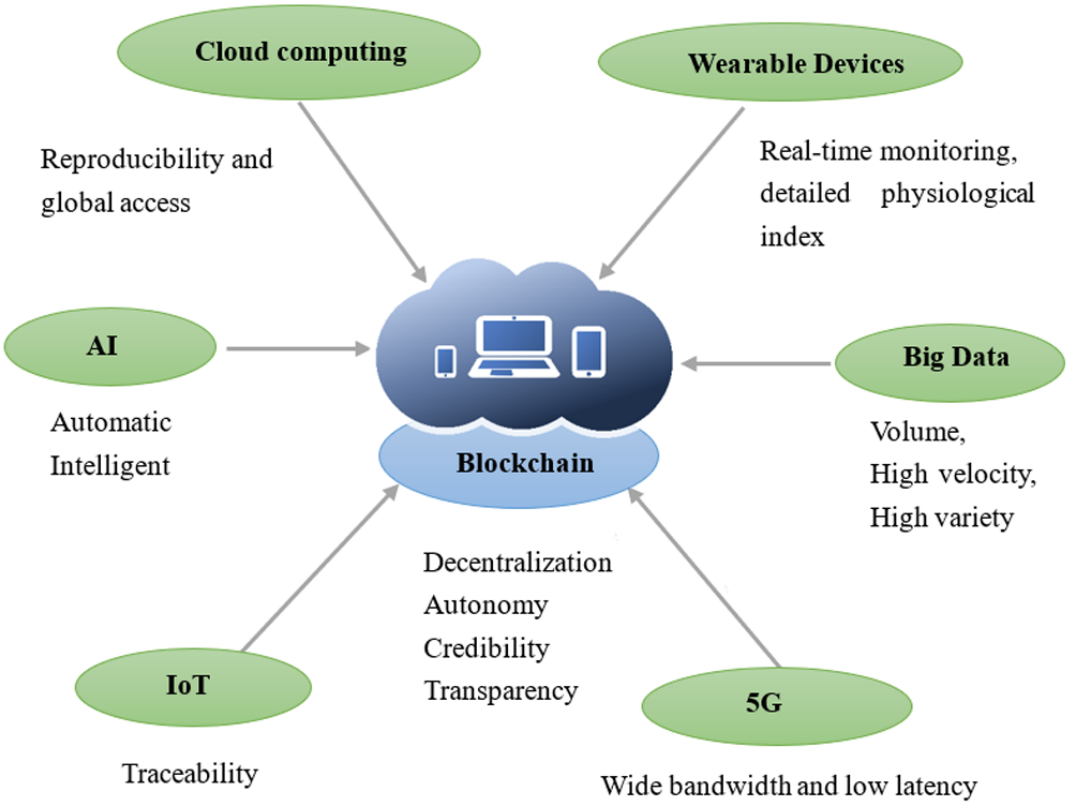
As shown in Table 5, AI can play a more sophisticated role in patient care through diagnostics, treatment, prevention, and predictive modeling [95,98,99]. For instance, in the cardiovascular arena, convolutional neural networks were recently trained using electrocardiograms (ECGs) from one specific wearable monitor in patients and provided cardiologist-level diagnostic accuracy for arrhythmias. An example is the smartphone app KardiaBand from AliveCor [94] based on machine learning for the identification of atrial fibrillation episodes from ECG data. Pilozzi et al [95] used natural language processing and blockchain-based storage systems to alleviate Alzheimer disease stigma and fears among

patients. To diagnose early cancer from CT images, Kumar et al [98] proposed a novel multimodel method combining deep learning and blockchain technology, which showed a perfect effect in practice.

Additionally, cloud computing is a new technique that provides different services by minimizing cost and infrastructure, and can be used with the blockchain system to support communication and sharing data among stakeholders in health care [82]. Liang et al [83] presented ProvChain, a blockchain-based data provenance architecture, to provide assurance of data operations in a cloud computing and storage application, while simultaneously enhancing privacy and availability.

Furthermore, big data can be generated from different sources such as wearable devices, EHRs, magnetic resonance imaging, and CT imaging. By integrating information about multiple features of diseases, big data offer an enormous amount of data in real time to reduce health risks and optimize health outcomes [77,82,83,96]. As we discussed in a previous study [85], wearable devices can be used to perceive, record, analyze, regulate, and intervene to maintain health, and can even be used to treat diseases with the support of various technologies for identification, sensing, connection, cloud services, and storage [96,100]. IoT can then play an essential role in the supply chain and tracing of drugs, apparatuses, and data from wearable devices. Based on blockchain and IoT, Fernández-Caramés et al [100] described the design and implementation of a system that enhances continuous glucose monitoring by adding IoT capabilities to allow for monitoring patients remotely and warning them about potentially dangerous situations. To motivate users to add new data to the system, an incentive system based on a digital cryptocurrency can be devised to reward the users that contribute to the system by providing their own data. This is a meaningful attempt to control chronic diseases. Furthermore, in the 5G environment, the blockchain-based cloud system is expected to take full charge of data transfer, storage, and processing. By combining all of these technologies, we can build a healthy database for developing a complete medical ecosystem (Figure 5).

Figure 5. Advantage of integration based on blockchain and emerging technologies. AI: artificial intelligence; IoT: Internet of Things.



Discussion

Principal Findings

The results of this narrative literature review suggest that blockchain is an innovative technology with great potential in health care. Prior reviews included conceptual papers, industry reports, and empirical research that primarily focused on technology, business, or patient care. In this paper, we demonstrate the use of blockchain in managing medical data, confirming the traceability of the supply chain, and in anticounterfeiting electronic prescriptions and clinical or biomedical research. We also demonstrate the important role that blockchain has played during the COVID-19 outbreak, which provides a reference for the prevention of major infectious diseases in the future. The integration and application of new technologies and blockchain were explored, attempting to

describe a blueprint of the interconnected ecosystem in health care.

From a practical point of view, the application of EMR exchange appears to be relatively mature. For example, in 10 studies, there were some related systems such as ACTION-EHR [22], HealthChain [25], Healthcare Data Gateway [28], Guardtime [79], and MeDShare [29], which are based on blockchain and used for preservation and exchange of health data. In five studies [42-45,47], the supply chain of medicine appeared to be another important area where blockchain may function. It would make sense to certify medical devices and monitor health status in combination with wearable devices. Moreover, the usage of tracing medicine will prevent the spread of fake drugs. In addition, based on a highly transparent scientific system in clinical and biomedical research, we can enhance the efficiency and confirm the integrity of the study. The situation for the prevention and control of COVID-19 remains grim, and many scholars (eg, [51-60]) have proposed using blockchain technology for preventing and controlling the epidemic, tracking asymptomatic infected individuals, and distinguishing former infected individuals. This decentralized technology ensures the patients' privacy and protects their rights and interests. The progress of electronic information technology has brought about the development of interdisciplinary tools with a benefit to the medical industry [1,3,64,95,97-99,101-103]. Undeniably, blockchain is booming under the influence of disruptive technologies. These current new technologies are constantly merging with blockchain to innovate medical models and systems. We listed the six typical integrations (AI, cloud computing, big data, wearable devices, IoT, and 5G) with blockchain and wish to provide readers with some inspiration.

However, some concerns were also identified when applying blockchain to health care. The first was the interoperability issue. When encountering a problem requiring cooperation, it is necessary for blockchain-based service providers and users to connect seamlessly; however, the standards among different institutions are not unified. There still exist great differences in the supervision modes of blockchain across countries. In the European Union, individual countries may be willing to use blockchain technology for public plans, but it is not clear how blockchain projects meet the EU General Data Protection Regulation privacy standards [1-4,96,97,104]. The general standard of blockchain will accelerate the industry to reach an agreement on blockchain and contribute to the formation of a large-scale ecosystem of social blockchain [101-103,105-107]. The second is efficiency issues. With the amount of data growing exponentially, the blockchain database has higher requirements for network speed, and the efficiency of data dissemination and real-time acquisition of data will be affected. More reasonable frameworks need to be designed to avoid blockchain efficiency problems across sectors. Moreover, security issues must be taken into consideration. It is not clear whether blockchain is truly the solution for all issues regarding highly sensitive data. The decentralized networks may be subject to 51% of attacks when there are fewer honest nodes than malicious nodes in the network, and the entire network will be taken over by malicious attackers. The last but not least concern is related to regulatory policies, which are important factors for the stable development of blockchain. Decentralization is one of the advantages of blockchain; however, after diluting government regulations, blockchain may have an impact on the existing electronic health system of a country [1-3,97,101,102,104-106]. Therefore, related policies need to be introduced systematically as soon as possible.

Limitations

The results of this review must be interpreted with caution owing to multiple limitations. First, the findings of this scoping review are mainly intended for health care entities and are not as applicable to other domains such as business and marketing. Second, for practical reasons, the search strategy was restricted to studies reported in the English language, which could have overlooked other benefits and threats reported in other studies in languages other than English. Third, owing to the broad variety of application scenarios of blockchain in the medical field, we cannot demonstrate every aspect of principle and framework in detail.

Conclusions

With the continuous improvement and development of new technologies, blockchain may become increasingly closely integrated with the contemporary development of the financial sector and health industry. Since each country and region has different attitudes toward this technology, we need to conduct in-depth exploration and research on the blockchain according to individual situations. In medical applications, many startup companies are actively exploring and promoting the development of blockchain in the fields of posttransaction settlement, smart contracts, supply chains, and identity authentication. From the theoretical perspective, blockchain-based theoretical foundations have been established for enhancing trust in an intelligent medicine environment. In the future, the issuance of digital currency will change the traditional economic transaction mode, and the introduction of blockchain will reshape the value exchange system, increase trust and privacy, and efficiently complete economic transactions and medical records [108]. Although countries and regions worldwide have diverse attitudes toward the blockchain, along with skeptical attitudes, these will not affect the research and further development of this technology. Throughout this work, we have also highlighted the principles and major challenges concerning distributed ledger technology. The great value of blockchain-based health care systems will gradually emerge in the coming years. In future work, technical personnel and researchers need to cooperate and incorporate the blockchain into the design of the medical framework. In this review, we demonstrated all of the potential scenarios of blockchain technology for patients and health care providers, offering large samples for further research. This only represents the beginning of the blockchain, and its development is expected to be more of a marathon than a sprint.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (grant 81974355), Major Technical Innovation Project of Hubei Province (grant 2016-176), and Establishment of National Intelligent Medical Clinical Research Centre (Establish a national-level innovation platform cultivation plan, 02.07.20030019). We greatly appreciate the anonymous reviewers for the insightful comments that improved this manuscript greatly.

Authors' Contributions

YX, JZ, and HW contributed equally as co-first authors of this paper. LL and ZY contributed equally as corresponding authors of this paper.

Conflicts of Interest

None declared.

References

1. Mackey TK, Kuo T, Gummadi B, Clauson KA, Church G, Grishin D, et al. 'Fit-for-purpose?' - challenges and opportunities for applications of blockchain technology in the future of healthcare. BMC Med 2019 Mar 27;17(1):68 [FREE Full text] [doi: 10.1186/s12916-019-1296-7] [Medline: 30914045]
2. Kamel Boulos MN, Wilson JT, Clauson KA. Geospatial blockchain: promises, challenges, and scenarios in health and healthcare. Int J Health Geogr 2018 Jul 05;17(1):25 [FREE Full text] [doi: 10.1186/s12942-018-0144-x] [Medline: 29973196]
3. Justinia T. Blockchain technologies: opportunities for solving real-world problems in healthcare and biomedical sciences. Acta Inform Med 2019 Dec;27(4):284-291 [FREE Full text] [doi: 10.5455/aim.2019.27.284-291] [Medline: 32055097]
4. Porsdam Mann S, Savulescu J, Ravaud P, Benchoufi M. Blockchain, consent and prosent for medical research. J Med Ethics 2020 May 04:244 [FREE Full text] [doi: 10.1136/medethics-2019-105963] [Medline: 32366703]
5. Ahmad SS, Khan S, Kamal MA. What is blockchain technology and its significance in the current healthcare system? A brief insight. Curr Pharm Des 2019;25(12):1402-1408. [doi: 10.2174/1381612825666190620150302] [Medline: 31258067]
6. Kim M, Yu S, Lee J, Park Y, Park Y. Design of secure protocol for cloud-assisted electronic health record system using blockchain. Sensors (Basel) 2020 May 21;20(10):2913 [FREE Full text] [doi: 10.3390/s20102913] [Medline: 32455635]
7. Rathore H, Mohamed A, Guizani M. A survey of blockchain enabled cyber-physical systems. Sensors (Basel) 2020 Jan 03;20(1):282 [FREE Full text] [doi: 10.3390/s20010282] [Medline: 31947860]
8. Kuo T, Zavaleta Rojas H, Ohno-Machado L. Comparison of blockchain platforms: a systematic review and healthcare examples. J Am Med Inform Assoc 2019 May 01;26(5):462-478 [FREE Full text] [doi: 10.1093/jamia/ocy185] [Medline: 30907419]
9. Alladi T, Chamola V, Rodrigues JJPC, Kozlov SA. Blockchain in smart grids: a review on different use cases. Sensors (Basel) 2019 Nov 08;19(22):4862 [FREE Full text] [doi: 10.3390/s19224862] [Medline: 31717262]
10. Abu-Elezz I, Hassan A, Nazeemudeen A, Househ M, Abd-Alrazaq A. The benefits and threats of blockchain technology in healthcare: A scoping review. Int J Med Inform 2020 Oct;142:104246 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104246] [Medline: 32828033]
11. Yli-Huumo J, Ko D, Choi S, Park S, Smolander K. Where is current research on blockchain technology? A systematic review. PLoS One 2016;11(10):e0163477 [FREE Full text] [doi: 10.1371/journal.pone.0163477] [Medline: 27695049]
12. Shu H, Qi P, Huang Y, Chen F, Xie D, Sun L. An efficient certificateless aggregate signature scheme for blockchain-based medical cyber physical systems. Sensors (Basel) 2020 Mar 10;20(5):1521 [FREE Full text] [doi: 10.3390/s20051521] [Medline: 32164220]
13. McBee MP, Wilcox C. Blockchain technology: principles and applications in medical imaging. J Digit Imaging 2020 Jun;33(3):726-734 [FREE Full text] [doi: 10.1007/s10278-019-00310-3] [Medline: 31898037]
14. Zheng Z, Xie S, Dai H. An overview of blockchain technology: architecture, consensus, and future trends. 2017 Jun 25 Presented at: 6th IEEE International Congress on Big Data; June 25, 2017; Honolulu, HI.
15. Hasselgren A, Kralevska K, Gligoroski D, Pedersen SA, Faxvaag A. Blockchain in healthcare and health sciences-A scoping review. Int J Med Inform 2020 Feb;134:104040 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.104040] [Medline: 31865055]
16. Hasavari S, Song YT. A secure and scalable data source for emergency medical care using blockchain technology. 2019 Jun 20 Presented at: 2019 IEEE 17th International Conference on Software Engineering Research, Management and Applications (SERA); 2019; Nigeria.
17. Ekblaw A, Azaria A, Halamka JD, Lippman A. MedRec: A case study for blockchain in healthcare. Digital Currency Initiative MIT Media Lab. URL: https://dci.mit.edu/research/blockchain-medical-records [accessed 2016-08-22]
18. Motohashi T, Hirano T, Okumura K, Kashiyama M, Ichikawa D, Ueno T. Secure and scalable mHealth data management using blockchain combined with client hashchain: system design and validation. J Med Internet Res 2019 May 16;21(5):e13385 [FREE Full text] [doi: 10.2196/13385] [Medline: 31099337]
19. Gordon WJ, Catalini C. Blockchain technology for healthcare: facilitating the transition to patient-driven interoperability. Comput Struct Biotechnol J 2018;16:224-230 [FREE Full text] [doi: 10.1016/j.csbj.2018.06.003] [Medline: 30069284]

20. Shi S, He D, Li L, Kumar N, Khan MK, Choo KR. Applications of blockchain in ensuring the security and privacy of electronic health record systems: A survey. Comput Secur 2020 Oct;97:101966 [FREE Full text] [doi: 10.1016/j.cose.2020.101966] [Medline: 32834254]
21. Zhou T, Li X, Zhao H. Med-PPPHIS: blockchain-based personal healthcare information system for national physique monitoring and scientific exercise guiding. J Med Syst 2019 Aug 14;43(9):305. [doi: 10.1007/s10916-019-1430-2] [Medline: 31410583]
22. Dubovitskaya A, Baig F, Xu Z, Shukla R, Zambani PS, Swaminathan A, et al. ACTION-EHR: patient-centric blockchain-based electronic health record data management for cancer care. J Med Internet Res 2020 Aug 21;22(8):e13598 [FREE Full text] [doi: 10.2196/13598] [Medline: 32821064]
23. Jones M, Johnson M, Shervey M, Dudley JT, Zimmerman N. Privacy-preserving methods for feature engineering using blockchain: review, evaluation, and proof of concept. J Med Internet Res 2019 Aug 14;21(8):e13600 [FREE Full text] [doi: 10.2196/13600] [Medline: 31414666]
24. Chen Y, Ding S, Xu Z, Zheng H, Yang S. Blockchain-based medical records secure storage and medical service framework. J Med Syst 2018 Nov 22;43(1):5. [doi: 10.1007/s10916-018-1121-4] [Medline: 30467604]
25. Hylock RH, Zeng X. A blockchain framework for patient-centered health records and exchange (HealthChain): evaluation and proof-of-concept study. J Med Internet Res 2019 Aug 31;21(8):e13592 [FREE Full text] [doi: 10.2196/13592] [Medline: 31471959]
26. Lo Y, Yang C, Chien H, Chang S, Lu C, Chen R. Blockchain-enabled iWellChain Framework integration with the national medical referral system: development and usability study. J Med Internet Res 2019 Dec 04;21(12):e13563 [FREE Full text] [doi: 10.2196/13563] [Medline: 31799935]
27. Xiao Y, Xu B, Jiang W, Wu Y. The HealthChain blockchain for electronic health records: development study. J Med Internet Res 2021 Jan 22;23(1):e13556 [FREE Full text] [doi: 10.2196/13556] [Medline: 33480851]
28. Yue X, Wang H, Jin D, Li M, Jiang W. Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control. J Med Syst 2016 Oct;40(10):218. [doi: 10.1007/s10916-016-0574-6] [Medline: 27565509]
29. Xia Q, Sifah EB, Asamoah KO, Gao J, Du X, Guizani M. MeDShare: trust-less medical data sharing among cloud service providers via blockchain. IEEE Access 2017;5:14757-14767. [doi: 10.1109/ACCESS.2017.2730843]
30. Patel V. A framework for secure and decentralized sharing of medical imaging data via blockchain consensus. Health Informatics J 2019 Dec;25(4):1398-1411 [FREE Full text] [doi: 10.1177/1460458218769699] [Medline: 29692204]
31. Fan K, Wang S, Ren Y, Li H, Yang Y. MedBlock: efficient and secure medical data sharing via blockchain. J Med Syst 2018 Jun 21;42(8):136. [doi: 10.1007/s10916-018-0993-7] [Medline: 29931655]
32. Yazdinejad A, Srivastava G, Parizi RM, Dehghantanha A, Choo KR, Aledhari M. Decentralized authentication of distributed patients in hospital networks using blockchain. IEEE J Biomed Health Inform 2020 Aug;24(8):2146-2156. [doi: 10.1109/JBHI.2020.2969648] [Medline: 31995507]
33. Zhu X, Shi J, Lu C. Cloud health resource sharing based on consensus-oriented blockchain technology: case study on a breast tumor diagnosis service. J Med Internet Res 2019 Jul 23;21(7):e13767 [FREE Full text] [doi: 10.2196/13767] [Medline: 31339106]
34. Dubovitskaya A, Novotny P, Xu Z, Wang F. Applications of blockchain technology for data-sharing in oncology: results from a systematic literature review. Oncology 2020;98(6):403-411 [FREE Full text] [doi: 10.1159/000504325] [Medline: 31794967]
35. Khurshid A, Rajeswaren V, Andrews S. Using blockchain technology to mitigate challenges in service access for the homeless and data exchange between providers: qualitative study. J Med Internet Res 2020 Jun 04;22(6):e16887 [FREE Full text] [doi: 10.2196/16887] [Medline: 32348278]
36. Cheng X, Chen F, Xie D, Sun H, Huang C. Design of a secure medical data sharing scheme based on blockchain. J Med Syst 2020 Jan 08;44(2):52. [doi: 10.1007/s10916-019-1468-1] [Medline: 31915982]
37. Zhou L, Wang L, Sun Y. MIStore: a blockchain-based medical insurance storage system. J Med Syst 2018 Jul 02;42(8):149 [FREE Full text] [doi: 10.1007/s10916-018-0996-4] [Medline: 29968202]
38. Brogan J, Baskaran I, Ramachandran N. Authenticating health activity data using distributed ledger technologies. Comput Struct Biotechnol J 2018;16:257-266 [FREE Full text] [doi: 10.1016/j.csbj.2018.06.004] [Medline: 30101004]
39. Griggs KN, Ossipova O, Kohlios CP, Baccarini AN, Howson EA, Hayajneh T. Healthcare blockchain system using smart contracts for secure automated remote patient monitoring. J Med Syst 2018 Jun 06;42(7):130. [doi: 10.1007/s10916-018-0982-x] [Medline: 29876661]
40. Ichikawa D, Kashiyama M, Ueno T. Tamper-resistant mobile health using blockchain technology. JMIR Mhealth Uhealth 2017 Jul 26;5(7):e111 [FREE Full text] [doi: 10.2196/mhealth.7938] [Medline: 28747296]
41. Mao D, Wang F, Hao Z, Li H. Credit evaluation system based on blockchain for multiple stakeholders in the food supply chain. Int J Environ Res Public Health 2018 Aug 01;15(8):1627 [FREE Full text] [doi: 10.3390/ijerph15081627] [Medline: 30071695]
42. Fernández-Caramés TM, Blanco-Novoa O, Froiz-Míguez I, Fraga-Lamas P. Towards an autonomous Industry 4.0 Warehouse: A UAV and blockchain-based system for inventory and traceability applications in big data-driven supply chain management. Sensors (Basel) 2019 May 25;19(10):2394 [FREE Full text] [doi: 10.3390/s19102394] [Medline: 31130644]

43. Sylim P, Liu F, Marcelo A, Fontelo P. Blockchain technology for detecting falsified and substandard drugs in distribution: pharmaceutical supply chain intervention. JMIR Res Protoc 2018 Sep 13;7(9):e10163 [FREE Full text] [doi: 10.2196/10163] [Medline: 30213780]
44. Lohmer J, Bugert N, Lasch R. Analysis of resilience strategies and ripple effect in blockchain-coordinated supply chains: An agent-based simulation study. Int J Prod Econ 2020 Oct;228:107882 [FREE Full text] [doi: 10.1016/j.ijpe.2020.107882] [Medline: 32834505]
45. Hoy MB. An introduction to the blockchain and its implications for libraries and medicine. Med Ref Serv Q 2017;36(3):273-279. [doi: 10.1080/02763869.2017.1332261] [Medline: 28714815]
46. Vruddhula S. Application of on-dose identification and blockchain to prevent drug counterfeiting. Pathog Glob Health 2018 Jun;112(4):161 [FREE Full text] [doi: 10.1080/20477724.2018.1503268] [Medline: 30213253]
47. Tseng J, Liao Y, Chong B, Liao S. Governance on the drug supply chain via Gcoin blockchain. Int J Environ Res Public Health 2018 May 23;15(6):1055 [FREE Full text] [doi: 10.3390/ijerph15061055] [Medline: 29882861]
48. Mackey TK, Miyachi K, Fung D, Qian S, Short J. Combating health care fraud and abuse: conceptualization and prototyping study of a blockchain antifraud framework. J Med Internet Res 2020 Sep 10;22(9):e18623 [FREE Full text] [doi: 10.2196/18623] [Medline: 32909952]
49. Aldughayfiq B, Sampalli S. Digital health in physicians' and pharmacists' office: a comparative study of e-prescription systems' architecture and digital security in eight countries. OMICS 2021 Feb;25(2):102-122 [FREE Full text] [doi: 10.1089/omi.2020.0085] [Medline: 32931378]
50. Li P, Nelson SD, Malin BA, Chen Y. DMMS: A decentralized blockchain ledger for the management of medication histories. Blockchain Healthc Today 2019;2:38 [FREE Full text] [doi: 10.30953/bhty.v2.38] [Medline: 32524086]
51. Raghavendra M. Can blockchain technologies help tackle the opioid epidemic: a narrative review. Pain Med 2019 Oct 01;20(10):1884-1889. [doi: 10.1093/pm/pny315] [Medline: 30848821]
52. Nandi S, Sarkis J, Hervani AA, Helms MM. Redesigning supply chains using blockchain-enabled circular economy and COVID-19 experiences. Sustain Prod Consum 2021 Jul;27:10-22 [FREE Full text] [doi: 10.1016/j.spc.2020.10.019] [Medline: 33102671]
53. Marbouh D, Abbasi T, Maasmi F, Omar IA, Debe MS, Salah K, et al. Blockchain for COVID-19: review, opportunities, and a trusted tracking system. Arab J Sci Eng 2020 Oct 12:1-17 [FREE Full text] [doi: 10.1007/s13369-020-04950-4] [Medline: 33072472]
54. Mashamba-Thompson TP, Crayton ED. Blockchain and artificial intelligence technology for novel coronavirus disease-19 self-testing. Diagnostics (Basel) 2020 Apr 01;10(4):198 [FREE Full text] [doi: 10.3390/diagnostics10040198] [Medline: 32244841]
55. Bansal A, Garg C, Padappayil RP. Optimizing the implementation of COVID-19 "Immunity Certificates" using blockchain. J Med Syst 2020 Jul 19;44(9):140 [FREE Full text] [doi: 10.1007/s10916-020-01616-4] [Medline: 32683501]
56. Khurshid A. Applying blockchain technology to address the crisis of trust during the COVID-19 pandemic. JMIR Med Inform 2020 Sep 22;8(9):e20477 [FREE Full text] [doi: 10.2196/20477] [Medline: 32903197]
57. Abdel-Basset M, Chang V, Nabeeh NA. An intelligent framework using disruptive technologies for COVID-19 analysis. Technol Forecast Soc Change 2021 Mar;163:120431 [FREE Full text] [doi: 10.1016/j.techfore.2020.120431] [Medline: 33162617]
58. Resiere D, Resiere D, Kallel H. Implementation of medical and scientific cooperation in the Caribbean using blockchain technology in coronavirus (Covid-19) pandemics. J Med Syst 2020 May 26;44(7):123 [FREE Full text] [doi: 10.1007/s10916-020-01589-4] [Medline: 32458071]
59. Chang MC, Park D. How can blockchain help people in the event of pandemics such as the COVID-19? J Med Syst 2020 Apr 16;44(5):102 [FREE Full text] [doi: 10.1007/s10916-020-01577-8] [Medline: 32300906]
60. Garg C, Bansal A, Padappayil RP. COVID-19: prolonged social distancing implementation strategy using blockchain-based movement passes. J Med Syst 2020 Aug 11;44(9):165 [FREE Full text] [doi: 10.1007/s10916-020-01628-0] [Medline: 32780276]
61. Verde F, Stanzione A, Romeo V, Cuocolo R, Maurea S, Brunetti A. Could blockchain technology empower patients, improve education, and boost research in radiology departments? An open question for future applications. J Digit Imaging 2019 Dec;32(6):1112-1115 [FREE Full text] [doi: 10.1007/s10278-019-00246-8] [Medline: 31197561]
62. Funk E, Riddell J, Ankel F, Cabrera D. Blockchain technology: a data framework to improve validity, trust, and accountability of information exchange in health professions education. Acad Med 2018 Dec;93(12):1791-1794. [doi: 10.1097/ACM.0000000000002326] [Medline: 29901658]
63. Durant E, Trachy A. Digital diploma debuts at MIT. MIT News. URL: https://news.mit.edu/2017/mit-debuts-secure-digital-diploma-using-bitcoin-blockchain-technology-1017 [accessed 2021-08-22]
64. Benchoufi M, Ravaud P. Blockchain technology for improving clinical research quality. Trials 2017 Jul 19;18(1):335 [FREE Full text] [doi: 10.1186/s13063-017-2035-z] [Medline: 28724395]
65. Wong DR, Bhattacharya S, Butte AJ. Prototype of running clinical trials in an untrustworthy environment using blockchain. Nat Commun 2019 Feb 22;10(1):917. [doi: 10.1038/s41467-019-08874-y] [Medline: 30796226]

66. Omar IA, Jayaraman R, Salah K, Simsekler MCE, Yaqoob I, Ellahham S. Ensuring protocol compliance and data transparency in clinical trials using Blockchain smart contracts. BMC Med Res Methodol 2020 Sep 07;20(1):224 [FREE Full text] [doi: 10.1186/s12874-020-01109-5] [Medline: 32894068]
67. Hirano T, Motohashi T, Okumura K, Takajo K, Kuroki T, Ichikawa D, et al. Data validation and verification using blockchain in a clinical trial for breast cancer: regulatory sandbox. J Med Internet Res 2020 Jun 02;22(6):e18938 [FREE Full text] [doi: 10.2196/18938] [Medline: 32340974]
68. Zhuang Y, Sheets L, Shae Z, Tsai JJP, Shyu C. Applying blockchain technology for health information exchange and persistent monitoring for clinical trials. AMIA Annu Symp Proc 2018;2018:1167-1175 [FREE Full text] [Medline: 30815159]
69. Nugent T, Upton D, Cimpoesu M. Improving data transparency in clinical trials using blockchain smart contracts. F1000Res 2016;5:2541 [FREE Full text] [doi: 10.12688/f1000research.9756.1] [Medline: 28357041]
70. Wan PK, Satybaldy A, Huang L, Holtskog H, Nowostawski M. Reducing alert fatigue by sharing low-level alerts with patients and enhancing collaborative decision making using blockchain technology: scoping review and proposed framework (MedAlert). J Med Internet Res 2020 Oct 28;22(10):e22013 [FREE Full text] [doi: 10.2196/22013] [Medline: 33112253]
71. Jin X, Zhang M, Zhou Z, Yu X. Application of a blockchain platform to manage and secure personal genomic data: a case study of LifeCODE.ai in China. J Med Internet Res 2019 Sep 10;21(9):e13587 [FREE Full text] [doi: 10.2196/13587] [Medline: 31507268]
72. Kuo T, Kim H, Ohno-Machado L. Blockchain distributed ledger technologies for biomedical and health care applications. J Am Med Inform Assoc 2017 Nov 01;24(6):1211-1220 [FREE Full text] [doi: 10.1093/jamia/ocx068] [Medline: 29016974]
73. Kuo T, Gabriel RA, Cidambi KR, Ohno-Machado L. EXpectation Propagation LOgistic REgRession on permissioned blockCHAIN (ExplorerChain): decentralized online healthcare/genomics predictive model learning. J Am Med Inform Assoc 2020 May 01;27(5):747-756 [FREE Full text] [doi: 10.1093/jamia/ocaa023] [Medline: 32364235]
74. Kuo T, Jiang X, Tang H, Wang X, Bath T, Bu D, et al. iDASH secure genome analysis competition 2018: blockchain genomic data access logging, homomorphic encryption on GWAS, and DNA segment searching. BMC Med Genomics 2020 Jul 21;13(Suppl 7):98 [FREE Full text] [doi: 10.1186/s12920-020-0715-0] [Medline: 32693816]
75. Ozercan HI, Ileri AM, Ayday E, Alkan C. Realizing the potential of blockchain technologies in genomics. Genome Res 2018 Sep;28(9):1255-1263 [FREE Full text] [doi: 10.1101/gr.207464.116] [Medline: 30076130]
76. Johnson M, Jones M, Shervey M, Dudley JT, Zimmerman N. Building a secure biomedical data sharing decentralized app (DApp): tutorial. J Med Internet Res 2019 Oct 23;21(10):e13601 [FREE Full text] [doi: 10.2196/13601] [Medline: 31647475]
77. Chen X, Zhu H, Geng D, Liu W, Yang R, Li S. Merging RFID and blockchain technologies to accelerate big data medical research based on physiological signals. J Healthc Eng 2020;2020:2452683. [doi: 10.1155/2020/2452683] [Medline: 32351676]
78. Mamoshina P, Ojomoko L, Yanovich Y, Ostrovski A, Botezatu A, Prikhodko P, et al. Converging blockchain and next-generation artificial intelligence technologies to decentralize and accelerate biomedical research and healthcare. Oncotarget 2018 Jan 19;9(5):5665-5690 [FREE Full text] [doi: 10.18632/oncotarget.22345] [Medline: 29464026]
79. Mettler M. Blockchain technology in healthcare: the revolution starts here. 2016 Presented at: IEEE 18th International Conference on e-Health Networking; September 14-16, 2016; Munich, Germany URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7749510 [doi: 10.1109/healthcom.2016.7749510]
80. Aitken M, de St Jorre J, Pagliari C, Jepson R, Cunningham-Burley S. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. BMC Med Ethics 2016 Nov 10;17(1):73 [FREE Full text] [doi: 10.1186/s12910-016-0153-x] [Medline: 27832780]
81. Esmaeilzadeh P, Mirzaei T. The potential of blockchain technology for health information exchange: experimental study from patients' perspectives. J Med Internet Res 2019 Jun 20;21(6):e14184 [FREE Full text] [doi: 10.2196/14184] [Medline: 31223119]
82. Kaur H, Alam MA, Jameel R, Mourya AK, Chang V. A proposed solution and future direction for blockchain-based heterogeneous Medicare data in cloud environment. J Med Syst 2018 Jul 10;42(8):156. [doi: 10.1007/s10916-018-1007-5] [Medline: 29987560]
83. Liang X, Shetty S, Tosh DK, Kamhoua CA, Kwiat KA, Njilla LY. ProvChain: A blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability. 2017 May 16 Presented at: IEEE/ACM CCGRID; May 14-17, 2017; Madrid, Spain p. A. [doi: 10.1109/ccgrid.2017.8]
84. Greenberg H, Pi-Sunyer FX. Preventing preventable chronic disease: An essential goal. Prog Cardiovasc Dis 2019;62(4):303-305. [doi: 10.1016/j.pcad.2019.08.002] [Medline: 31421079]
85. Lu L, Zhang J, Xie Y, Gao F, Xu S, Wu X, et al. Wearable health devices in health care: narrative systematic review. JMIR Mhealth Uhealth 2020 Nov 09;8(11):e18907 [FREE Full text] [doi: 10.2196/18907] [Medline: 33164904]
86. Guk K, Han G, Lim J, Jeong K, Kang T, Lim E, et al. Evolution of wearable devices with real-time disease monitoring for personalized healthcare. Nanomaterials (Basel) 2019 May 29;9(6):813 [FREE Full text] [doi: 10.3390/nano9060813] [Medline: 31146479]
87. Haghi M, Thurow K, Stoll R. Wearable devices in medical Internet of Things: scientific research and commercially available devices. Healthc Inform Res 2017 Jan;23(1):4-15 [FREE Full text] [doi: 10.4258/hir.2017.23.1.4] [Medline: 28261526]

88. Shahzad A, Zhang K, Gherbi A. Intuitive development to examine collaborative IoT supply chain system underlying privacy and security levels and perspective powering through proactive blockchain. Sensors (Basel) 2020 Jul 05;20(13):3760 [FREE Full text] [doi: 10.3390/s20133760] [Medline: 32635589]
89. O'Donoghue O, Vazirani AA, Brindley D, Meinert E. Design choices and trade-offs in health care blockchain implementations: systematic review. J Med Internet Res 2019 May 10;21(5):e12426 [FREE Full text] [doi: 10.2196/12426] [Medline: 31094344]
90. Welch MJ, Lally R, Miller JE, Pittman S, Brodsky L, Caplan AL, et al. The ethics and regulatory landscape of including vulnerable populations in pragmatic clinical trials. Clin Trials 2015 Oct;12(5):503-510 [FREE Full text] [doi: 10.1177/1740774515597701] [Medline: 26374681]
91. Cichosz SL, Stausholm MN, Kronborg T, Vestergaard P, Hejlesen O. How to use blockchain for diabetes health care data and access management: an operational concept. J Diabetes Sci Technol 2019 Mar;13(2):248-253 [FREE Full text] [doi: 10.1177/1932296818790281] [Medline: 30047789]
92. Engel A. Could blockchain technology add value to surgical outcomes research? Colorectal Dis 2018 May;20(5):369-370. [doi: 10.1111/codi.14066] [Medline: 29717560]
93. Ismail L, Materwala H, Karduck AP, Adem A. Requirements of health data management systems for biomedical care and research: scoping review. J Med Internet Res 2020 Jul 07;22(7):e17508 [FREE Full text] [doi: 10.2196/17508] [Medline: 32348265]
94. Skaria R, Satam P, Khalpey Z. Opportunities and challenges of disruptive innovation in medicine using artificial intelligence. Am J Med 2020 Jun;133(6):e215-e217. [doi: 10.1016/j.amjmed.2019.12.016] [Medline: 31981487]
95. Pilozzi A, Huang X. Overcoming Alzheimer's disease stigma by leveraging artificial intelligence and blockchain technologies. Brain Sci 2020 Mar 23;10(3):183 [FREE Full text] [doi: 10.3390/brainsci10030183] [Medline: 32210011]
96. Shahbazi Z, Byun Y. Towards a secure thermal-energy aware routing protocol in wireless body area network based on blockchain technology. Sensors (Basel) 2020 Jun 26;20(12):3604 [FREE Full text] [doi: 10.3390/s20123604] [Medline: 32604851]
97. Zhang C, Xu C, Sharif K, Zhu L. Privacy-preserving contact tracing in 5G-integrated and blockchain-based medical applications. Comput Stand Interfaces 2021 Aug;77:103520 [FREE Full text] [doi: 10.1016/j.csi.2021.103520] [Medline: 33584007]
98. Kumar R, Wang W, Kumar J, Yang T, Khan A, Ali W, et al. An Integration of blockchain and AI for secure data sharing and detection of CT images for the hospitals. Comput Med Imaging Graph 2021 Jan;87:101812. [doi: 10.1016/j.compmedimag.2020.101812] [Medline: 33279761]
99. Krittanawong C, Rogers AJ, Aydar M, Choi E, Johnson KW, Wang Z, et al. Integrating blockchain technology with artificial intelligence for cardiovascular medicine. Nat Rev Cardiol 2020 Jan;17(1):1-3 [FREE Full text] [doi: 10.1038/s41569-019-0294-y] [Medline: 31605093]
100. Fernández-Caramés TM, Froiz-Míguez I, Blanco-Novoa O, Fraga-Lamas P. Enabling the internet of mobile crowdsourcing health things: a mobile fog computing, blockchain and IoT based continuous glucose monitoring system for diabetes mellitus research and care. Sensors (Basel) 2019 Jul 28;19(15):3319 [FREE Full text] [doi: 10.3390/s19153319] [Medline: 31357725]
101. Mackey T, Bekki H, Matsuzaki T, Mizushima H. Examining the potential of blockchain technology to meet the needs of 21st-century Japanese health care: viewpoint on use cases and policy. J Med Internet Res 2020 Jan 09;22(1):e13649 [FREE Full text] [doi: 10.2196/13649] [Medline: 31917371]
102. El-Gazzar R, Stendal K. Blockchain in health care: hope or hype? J Med Internet Res 2020 Jul 10;22(7):e17199 [FREE Full text] [doi: 10.2196/17199] [Medline: 32673219]
103. Hasselgren A, Hanssen Rensaa J, Kralevska K, Gligoroski D, Faxvaag A. Blockchain for increased trust in virtual health care: proof-of-concept study. J Med Internet Res 2021 Jul 30;23(7):e28496 [FREE Full text] [doi: 10.2196/28496] [Medline: 34328437]
104. Durneva P, Cousins K, Chen M. The current state of research, challenges, and future research directions of blockchain technology in patient care: systematic review. J Med Internet Res 2020 Jul 20;22(7):e18619 [FREE Full text] [doi: 10.2196/18619] [Medline: 32706668]
105. Leeming G, Ainsworth J, Clifton DA. Blockchain in health care: hype, trust, and digital health. Lancet 2019 Jun 22;393(10190):2476-2477. [doi: 10.1016/S0140-6736(19)30948-1] [Medline: 31232356]
106. Shuaib K, Saleous H, Shuaib K, Zaki N. Blockchains for secure digitized medicine. J Pers Med 2019 Jul 13;9(3):35 [FREE Full text] [doi: 10.3390/jpm9030035] [Medline: 31337080]
107. Gaynor M, Tuttle-Newhall J, Parker J, Patel A, Tang C. Adoption of blockchain in health care. J Med Internet Res 2020 Sep 17;22(9):e17423 [FREE Full text] [doi: 10.2196/17423] [Medline: 32940618]
108. Bublitz FM, Oetomo A, Sahu KS, Kuang A, Fadrique LX, Velmovitsky PE, et al. Disruptive technologies for environment and health research: an overview of artificial intelligence, blockchain, and Internet of Things. Int J Environ Res Public Health 2019 Oct 11;16(20):3847 [FREE Full text] [doi: 10.3390/ijerph16203847] [Medline: 31614632]

Abbreviations

AI: artificial intelligence
CT: computed tomography
ECG: electrocardiogram
EHR: electronic health record
EMR: electronic medical record
IoT: Internet of Things
IPFS: interplanetary file system
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses

Edited by C Basch; submitted 08.03.21; peer-reviewed by YC Chen, Y Chen, Y Ying; comments to author 23.04.21; revised version received 12.08.21; accepted 10.09.21; published 28.10.21.

Please cite as:
Xie Y, Zhang J, Wang H, Liu P, Liu S, Huo T, Duan YY, Dong Z, Lu L, Ye Z
Applications of Blockchain in the Medical Field: Narrative Review
J Med Internet Res 2021;23(10):e28613
URL: https://www.jmir.org/2021/10/e28613
doi:10.2196/28613
PMID:34533470

Original Paper

Feasibility of Indirect Secondary Distribution of HIV Self-test Kits via WeChat Among Men Who Have Sex With Men: National Cross-sectional Study in China

Shangcao Li[1,2,3*], MPH; Jing Zhang[1,2,3*], PhD; Xiang Mao[1,2,3], PhD; Tianyi Lu[1,2,3], MPH; Yangyang Gao[1,2,3], MD; Wenran Zhang[1,2,3], MPH; Hongyi Wang[1,2,3], PhD; Zhenxing Chu[1,2,3], PhD; Qinghai Hu[1,2,3], PhD; Yongjun Jiang[1,2,3], PhD; Wenqing Geng[1,2,3], PhD; Hong Shang[1,2,3*], PhD; Junjie Xu[1,2,3*], PhD

[1]NHC Key Laboratory of AIDS Immunology, National Clinical Research Center for Laboratory Medicine, The First Affiliated Hospital of China Medical University, Shenyang, Liaoning Province, China

[2]Key Laboratory of AIDS Immunology, Chinese Academy of Medical Sciences, Shenyang, China

[3]Key Laboratory of AIDS Immunology of Liaoning Province, Shenyang, China

[*]these authors contributed equally

Corresponding Author:
Junjie Xu, PhD
NHC Key Laboratory of AIDS Immunology
National Clinical Research Center for Laboratory Medicine
The First Affiliated Hospital of China Medical University
No 155, Nanjing North Street, Heping District
Shenyang, Liaoning Province, 110001
China
Phone: 86 8328 2634
Email: xjjcmu@163.com

Abstract

Background: HIV self-testing (HIVST) kits are common in key sexually active populations. Direct secondary distribution of HIVST kits (DSDHK) is effective in improving the uptake of HIVST. However, there are concerns about the various limitations of DSDHK, including limited geographic reach, payment problems, and need for face-to-face interactions.

Objective: In this study, we aim to evaluate the feasibility and characteristics of indirect secondary distribution of HIVST kits (ISDHK) via WeChat (distributing HIVST application links and follow-up HIVST kits to partners) among men who have sex with men (MSM).

Methods: From October 2017 to September 2019, an HIVST recruitment advertisement was disseminated on the WeChat social media platform to invite MSM to apply for the HIVST kits (referred to as *index participants* [IPs]). All MSM participants were encouraged to distribute the HIVST application link to their friends and sexual partners (referred to as *alters*) through their social networks. All the alters were further encouraged to continue distributing the HIVST application link. All participants paid a deposit (US $7), which was refundable upon completion of the questionnaire, and uploaded the test results via a web-based survey system.

Results: A total of 2263 MSM met the criteria and successfully applied for HIVST. Of these, 1816 participants returned their HIVST results, including 1422 (88.3%) IPs and 394 (21.7%) alters. More alters had condomless anal intercourse, a higher proportion of them had never previously tested for HIV, and they showed a greater willingness to distribute HIVST kits to their sexual partners (*P*=.002) than the IPs. After controlling for age, education, and income, the alters had a greater proportion of MSM who had never tested for HIV before (adjusted odds ratio [aOR] 1.29, 95% CI 1.00-1.68), were more willing to distribute the HIVST application link (aOR 1.71, 95% CI 1.21-2.40), had a lower number of sexual partners (aOR 0.71, 95% CI 0.57-0.90), and were less likely to search for sexual partners on the web (aOR 0.78, 95% CI 0.60-1.02) than IPs. In comparison, the rates of reactive HIVST results, conducting HIV confirmatory tests, HIV seropositivity, and initiation of HIV antiretroviral therapy were similar for IPs and alters.

Conclusions: The ISDHK model of distributing HIVST application links among the MSM population via social media is feasible. The ISDHK model should be used to supplement the DSDHK model to enable a greater proportion of the MSM population to know their HIV infection status.

(J Med Internet Res 2021;23(10):e28508) doi:10.2196/28508

KEYWORDS

secondary distribution; HIV; men who have sex with men; WeChat; HIV self-testing

Introduction

Background

HIV testing is the first step in HIV treatment and care and is a crucial part of HIV prevention [1]. However, as of 2019, approximately 20% of people living with HIV globally still do not know their serological status [2], and this figure is approximately 25% in China. Men who have sex with men (MSM) continue to be disproportionately impacted by the HIV pandemic [3]. The proportion of MSM among the annual newly reported HIV infections in China increased from 16% in 2011 to 23.4% in 2018 [4,5]. A national survey in 2016 showed that only 47% of Chinese MSM had ever tested for HIV, and only 38% had tested for HIV in the last 12 months [6]. There are barriers to traditional facility-based HIV testing for MSM, including concerns about discrimination, privacy, time, and transportation [7-9]. Hence, increasing people's awareness of HIV status through innovative methods has become the focus of HIV prevention campaigns globally.

HIV self-testing (HIVST) is a new strategy recommended by the World Health Organization to improve the uptake of HIV testing among key populations [10]. Evidence shows that HIVST has the potential to overcome facility-based barriers to HIV testing, improve testing conditions in areas where HIV testing opportunities are insufficient, and increase accessibility to testing [11].

Direct secondary distribution of HIVST kits (DSDHK) is a social network–based approach, giving *index participants* (IPs) multiple HIVST kits for direct distribution to their sexual partners or others in their social network (referred to as *alters*) [12]. Several studies have demonstrated greater acceptability and effectiveness of DSDHK, in which IPs share HIVST kits with alters. In South Africa and Uganda, 2 cohort studies showed that DSDHK is acceptable and feasible for MSM, increasing the frequency of HIV testing among MSM and improving early detection of HIV [13,14]. In Kenya, Uganda, and South Africa, 4 randomized controlled trials showed that DSDHK promoted HIV testing by helping MSM and female sex workers to adopt safer sex practices with their sexual partners through secondary distribution of self-test kits by HIV-negative female sex workers and women receiving antenatal and postpartum care [12,15-17]. A recent study conducted in China, which used social media to recruit IPs, showed that DSDHK increased the level of diagnosis of HIV-positive cases and the coverage of HIV testing among MSM [18]. However, the abovementioned DSDHK study recruited only participants living around the study sites, and only a limited number of alters were able to obtain HIVST kits from IPs. The sharing of HIVST kits requires the IPs to live in the same city as alters or for it to be shared by mail, which will arouse various concerns, including concerns regarding geographic location for distributing HIVST kits, payment for mailing HIVST kits, or face-to-face interaction [19]. This is an instance of a lack of innovative social determinants of health theory, which highlights the limitations of DSDHK practice. Moreover, no studies that we are aware of have explored the feasibility and effectiveness of an indirect secondary distribution of HIVST kits (ISDHK) model of HIVST recruitment and self-test kit distribution.

WeChat is the most popular mobile social media platform in use in China today, with more than 1 billion registered user accounts [20]. We hypothesize that an ISDHK model could boost the distribution of HIVST among MSM who were willing to promote HIV testing to their sexual partners or friends by sharing HIVST application links (websites or QR codes for web-based applications, mailing, and uploading of test results).

Objectives

This study aims to evaluate the feasibility of social media–based ISDHK among MSM and explore differences in sexual behaviors, HIV testing behaviors, willingness to distribute HIV testing further, and linkage-to-HIV care characteristics between IPs and alters.

Methods

Study Setting and Participants

From October 2017 to September 2019, we performed a WeChat-based cross-sectional study of ISDHK among MSM in mainland China (Hong Kong, Macao, and Taiwan were not included). Participants were eligible if they were male, aged ≥16 years, had reported having engaged in anal sex with another man on at least one occasion, willing to provide their sexual behavior information, and willing to apply for an HIVST kit for HIVST, provide the results, and sign the digital informed consent form.

The ISDHK research was carried out in Shenyang, Liaoning Province, China (First Affiliated Hospital of China Medical University).

IP Recruitment

The recruitment advertisements for this HIVST study were disseminated by 6 key opinion leaders of MSM through 6 WeChat public accounts, and detailed information about the participants' recruitment was published [21]. The primary subscribers to the 6 WeChat public accounts were from the MSM population. The advertisements included an introduction to and information about informed consent and links to apply for HIVST. Those MSM interested in the HIVST project who

successfully completed the application questionnaire were provided with HIVST kits (the IPs) and encouraged to share the HIVST kit application links with their peer MSM friends and sexual partners.

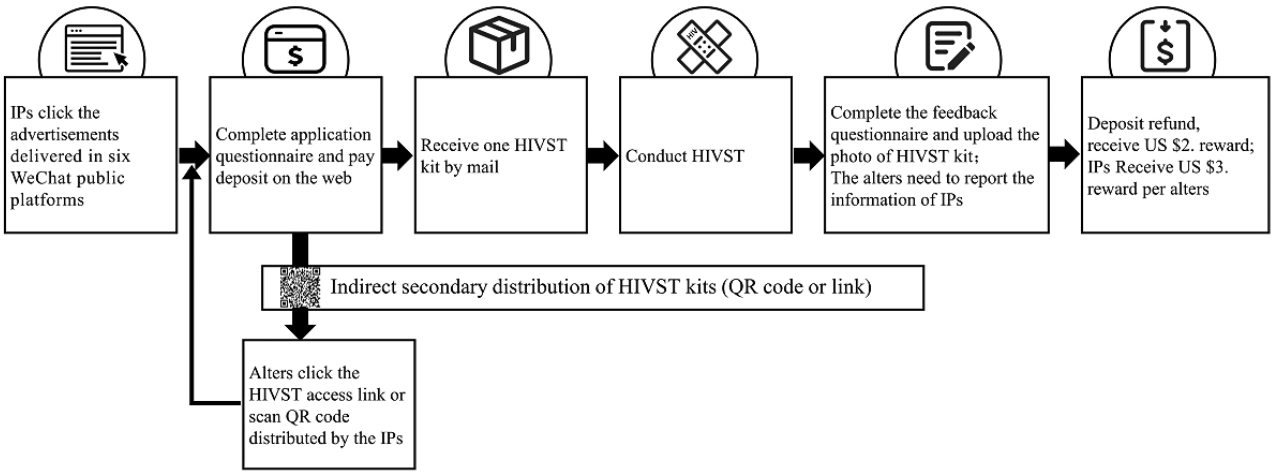
Indirect Secondary Distribution of HIVST Kits

IPs were recruited through the 6 WeChat public accounts and encouraged to share the HIVST service links on social media (such as Blued, WeChat Moments, etc) or with their sexual partners and friends (the *alters*). The alters obtained HIVST via the HIVST application link on various social media platforms shared by IPs. The alters were also encouraged to further share the HIVST application link on the social networks that they used.

Provision of HIVST Kits

Following the completion of their HIVST applications, information about participants' sociodemographic and sexual behaviors was collected by means of *Golden Data*, a web-based survey software (Xi'an Data Rujin Information Technology Co, Ltd). The study investigators reviewed the eligibility of the HIVST applications from all applicants. Then, to ensure the uploading of a proportion of HIVST test results, the enrolled eligible MSM participants were asked to pay a deposit of US $7, which would be refunded following the uploading of the test results and conducting of the HIV posttesting counseling. Fingerstick whole blood HIVST kits, and supporting tools (a disposable retractable blood needle, capillary tube, Wondlfo solution, alcohol tables, disposable fingertip blood collection needle, woundplast, and condoms) were sent to eligible participants by express mail. A reward of US $2 was paid to participants who uploaded their test results to compensate them for their time spent participating in the survey and to increase the feedback rate of HIVST results. We also asked the alters to fill in the IP information (phone number, WeChat name, or name) to understand the distribution network and pay fees as an incentive. IPs received an additional reward of US $3 if they helped recruit an alter to attend the HIVST project and upload their HIVST results (Figure 1).

Figure 1. WeChat-based model of indirect secondary distribution of HIV self-testing kits. HIVST: HIV self-testing; IP: index participant.



Postcounseling, Linkage-to-Care, and Follow-up

If a participant had a *positive* or *indeterminate* HIVST test result, they were immediately contacted by trained researchers for an HIV posttest consultation via WeChat or over the phone to help with the interpretation of the test results and referral to services for clinical confirmatory testing and antiviral treatment. If the participant was located in Shenyang or surrounding cities, they were invited to the Voluntary Counseling and Test of the First Hospital of China Medical University to provide 10 mL of venous blood for HIV confirmation, diagnosis, and treatment. The local Chinese Center for Disease Control information about further testing was also provided to them.

Measurements

After conducting the HIVST, participants were required to provide feedback on their HIVST results by uploading the HIVST result pictures and completing a computer-assisted self-administered questionnaire. The web-based questionnaire system is equipped with logic verification, which automatically checks the survey object before it is submitted, avoiding missing numbers or logic errors. The questionnaire included questions about the following: (1) sociodemographic information, including age, occupation, educational level, and monthly income; (2) sexual behavior, including ways of finding sexual partners, sexual behavioral roles with other males, numbers of sexual partners, questions about condomless anal intercourse (CAI), and questions about chemsex (a term used by MSM to describe sex that occurs under the influence of drugs such as rush [poppers or alkyl nitrites], MDMA [3,4-methylenedioxymethamphetamine; ecstasy], ice, amphetamines, tramadol, or ketamine) in the last 6 months; (3) HIV testing, including history of HIV testing, types of HIV testing, and frequency of HIV testing; (4) photo of the HIVST result uploaded to the system; and (5) the source of their test paper for this test. If the participants reported, "Apply through the link shared by your friends" or "Your friends help you apply," they were identified as alters and were asked to report the IPs' names and phone numbers. IPs and alters were asked to complete the same questionnaire.

Sample Size

We calculated the sample size of MSM participants using a simple random sampling method to test 2 independent proportions. According to the preliminary study, the ratio of

alters to IPs is 0.3, the estimated share of IPs who had never tested for HIV was 20% and the estimated share of alters who had never tested for HIV was 30%. We used the parameters of 80% power at a 2-sided significance level of .05 and calculated that the smallest sample size was 600 IPs and 200 alters. We used PASS (Power Analysis and Sample Size) software, version 11 (NCSS) to calculate the sample size.

Statistical Analysis

Quantitative variables are classified; for example, age is divided into <24 years and ≥24 years, and the number of sexual partners is divided into <2 and >2. The distribution of demography, sexual behavior, and HIV testing–related characteristics of IPs and alters was compared using chi-square analysis or Fisher exact test. A multivariable logistic regression model was used to analyze the correlation between alters and IPs. Age, educational level, and monthly income were adjusted in the multivariable logistic model. Data analysis was performed using SPSS 25.0 software (IBM Corporation). Two-way *P* values that were <.05 were considered to be statistically significant. See the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist for this cross-sectional study in Multimedia Appendix 1.

Ethical Permission

This study was approved by the Institutional Review Board of the First Affiliated Hospital of China Medical University (Project No. 2018-174-2). A signed digital informed consent form was obtained from all the participants before collecting any study information or specimens. Participants joined the study voluntarily and were free to withdraw from the study at any time. When collecting data through the *Golden Data* web-based survey platform, only the researcher was provided with an account and password to ensure the participants' data security and privacy. The data were downloaded directly and saved on a local computer. Identifiable variables (such as nicknames, phone numbers, and addresses) were encrypted to protect the privacy of the participants. HIV test results were disclosed only to the person concerned and not to anyone else.

Results

Sociodemographic Characteristics of the Participants

In total, 2263 MSM participants attended the HIVST project and applied for the HIVST kits. Of these, 80.25% (1816/2263) were unique participants from 29 provinces and 229 cities in mainland China (Figure 2) who completed the feedback questionnaire and uploaded their HIVST results. A total of 1816 eligible MSM were included in this cross-sectional study, of which 78.3% (1422/1816) were categorized as IPs and 21.69% (394/1816), as alters. Alters were recruited by IPs through ISDHK (Figure 3).

Figure 2. Distribution of participants in 29 provinces of China who fed back their HIV self-testing.

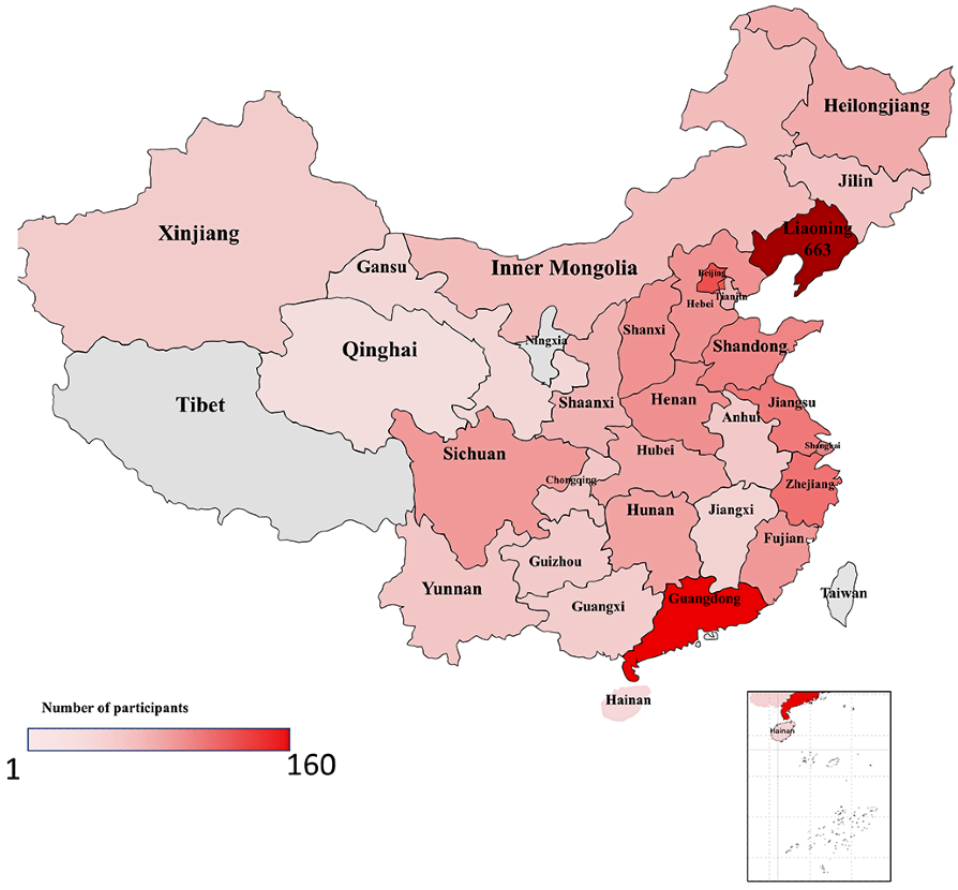


Figure 3. Flowchart of results and linkage to care of HIV self-testing secondary distribution indirectly among men who have sex with men in 29 provinces in China, 2017-2019. "*" indicates participants were eligible and paid the deposit; ART: antiretroviral therapy; HIVST: HIV self-testing; IP: index participant.

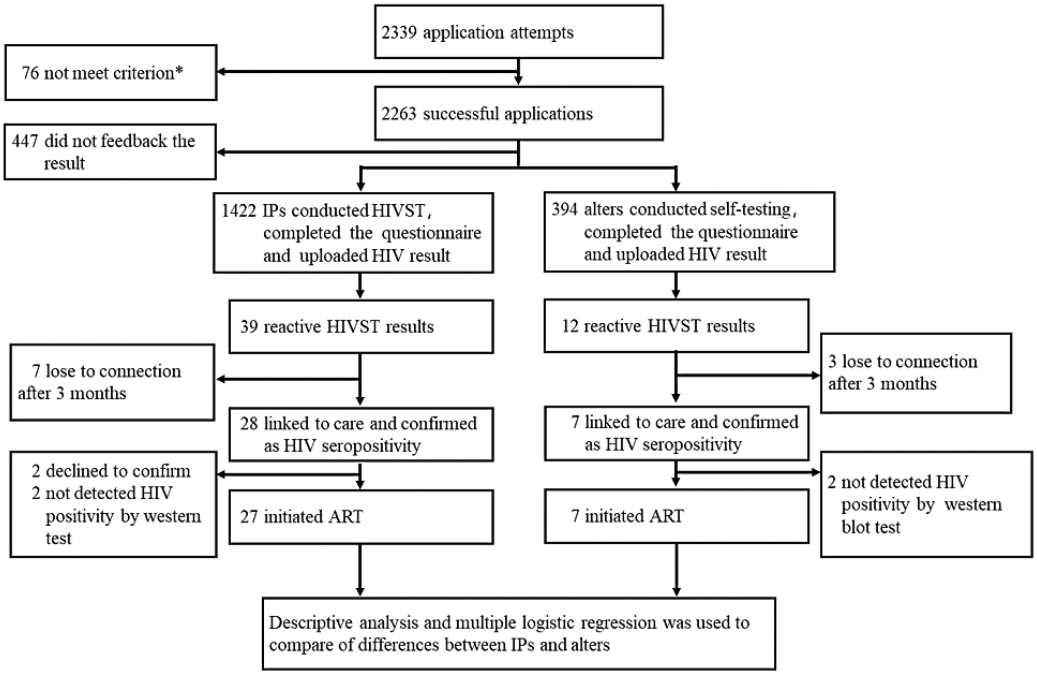


Table 1 summarizes the sociodemographic characteristics, sexual behaviors, and HIV testing behaviors of alters and IPs. Alters were aged ≥24 years (251/394, 64.3%), had college-level or higher-level education (320/394, 81.2%), sought sexual partners through offline means (303/394, 76.9%), had more than 2 sexual partners (220/394, 55.8%), had been tested for HIV before (283/394, 71.8%), were willing to share HIVST results with sexual partners (349/394, 88.6%), and applied regularly for HIVST (353/394, 89.6%; Table 1).

Table 1. Demographic characteristics, sexual behaviors, and HIV prevalence of index participants and alters (N=1816).

Characteristic	Alters (n=394), n (%)	Index participants (n=1422), n (%)
Age above 24 years	251 (63.7)	918 (64.6)
Monthly income (US $)		
<300	125 (32.6)	453 (31.9)
300-900	168 (43.8)	550 (38.7)
900	91 (23.7)	361 (25.4)
Having a university degree or above	320 (81.2)	1239 (87.1)
Student	129 (32.7)	478 (33.6)
Homosexual	265 (67.3)	992 (69.8)
Sexual role		
Insertive	125 (31.7)	423 (29.7)
No preference	111 (28.2)	389 (27.4)
Acceptive	115 (29.2)	443 (31.2)
Oral	43 (10.9)	167 (11.7)
Sexual behavior in the last six months		
Sought sexual partners on the web	303 (76.9)	1151 (80.9)
Had >2 sexual partners	220 (55.8)	889 (62.5)
Engaged in commercial anal intercourse	8 (2)	37 (2.6)
Engaged in chemsex[a]	145 (36.8)	50 (35.5)
Engaged in group sex with males	32 (8.1)	123 (8.6)
Engaged in CAI[b] with males	184 (46.7)	580 (40.8)
Experienced anal bleeding	134 (34)	498 (35)
Experienced symptoms of an STI[c]	33 (8.4)	170 (12)
HIV testing frequency		
Never tested for HIV before	111 (28.2)	329 (23.1)
Every 3 months or more often	53 (18.7)	216 (20)
Every 6 months	77 (27.2)	338 (31)
Every 12 months	67 (23.7)	219 (20)
Every ≥12 months	18 (6.4)	71 (6.5)
No regular frequency	68 (24)	249 (22.8)
Willingness to share HIVST[d] with sexual partners	349 (88.6)	1167 (82.1)
Apply for HIVST regularly	353 (89.6)	1306 (91.8)
Positive HIVST result	12 (3)	39 (2.7)
HIV confirmatory test		
Positive	7 (58.3)	28 (71.8)
Negative or unknown	5 (41.7)	6 (28.2)

[a]Used rush (poppers or alkyl nitrites), MDMA (3,4-methylenedioxymethamphetamine; ecstasy), ice, amphetamines, tramadol, or ketamine in the past six months.

[b]CAI: condomless anal intercourse.

[c]STI: sexually transmitted infection.

[d]HIVST: HIV self-testing.

Differences Between IPs and Alters

Chi-square analysis showed that alters had a higher likelihood of having engaged in CAI (184/394, 46.6% vs 580/1422, 40.78%; *P*=.04), a higher proportion of never having previously tested for HIV (111/394, 28.2% vs 329/1422, 23.14%; *P*=.04), and greater willingness to recommend or promote HIVST to sexual partners (349/394, 88.6% vs 1167/1422, 82.07%; *P*=.002) than IPs (Multimedia Appendix 2).

Multivariable logistic regression analysis showed that alters had fewer sexual partners (adjusted odds ratio [aOR] 0.71, 95% CI 0.57-0.90), a lower proportion with symptoms of sexually transmitted infections (aOR 0.67, 95% CI 0.45-1.00), a greater willingness to distribute the HIVST application link to partners and friends (aOR 1.71, 95% CI 1.21-2.40), a marginally higher proportion of never having previously tested for HIV (aOR 1.29, 95% CI 1.00-1.68), and a marginally lower proportion of MSM seeking sexual partners through offline means (aOR 0.78, 95% CI 0.60-1.02; Multimedia Appendix 2).

HIVST Results and Linkage-to-Care

Overall, out of the 1816, 51 (2.81%) study participants had a reactive HIVST result. Of these, 39 were IPs and 12 were alters. The reactive HIV rates among the 2 groups showed no significant differences (39/1422, 2.74% vs 12/394, 3%; *P*=.75). Among those who recorded a reactive HIVST result, the percentage of participants who sought HIV care and were followed up was similar between IPs and alters (32/39, 82% vs 9/12, 75%; *P*=.70). The percentages of participants who had an HIV-positive confirmatory test (28/32, 93% vs 7/9, 77%; *P*=.99), and who initiated HIV ART (antiretroviral therapy; 27/28, 96% vs 7/7, 100%; *P*=.99) were also similar between IPs and alters (Table 2).

Table 2. The proportions of index participants and alters who accessed different HIV testing services.

HIV care cascade	Index participants, n (%)	Alters, n (%)	*P* value[a]
Reactive HIVST[b]	39 (2.7)	12 (3)	.75
Follow-up	32 (82)	9 (75)	.70
Confirmatory testing	28 (88)	7 (78)	.99
Initiate ART[c]	27 (96)	7 (100)	.99

[a]Fisher exact test.

[b]HIVST: HIV self-testing.

[c]ART: antiretroviral therapy.

Discussion

Principal Findings

To the best of our knowledge, this study is the first to clarify that the indirect secondary distribution of HIVST application links through the WeChat social media app is feasible and effective for expanding the coverage of HIV testing among the MSM population in China. This new ISDHK model can reach out to those MSM who have a limited HIV testing history and a higher likelihood of having engaged in CAI. The ISDHK model can expand the coverage of HIV testing among the social networks of MSM. The rates of HIVST-positive reactive results, HIV seropositivity, and the linkage-to-care ratio of MSM recruited by ISDHK were similar to those directly recruited from the MSM population. This finding indicates that ISDHK could fill in the gaps in our knowledge of DSDHK theory and practice.

Our study found that the ISDHK model increased the coverage of HIVST distribution to MSM partners and friends on the internet among MSM participants. However, DSDHK can continue to expand the continued distribution of HIVST among key populations through the social networks of participants and increase the coverage of HIV testing. However, the direct method faces the limitation of limited distribution. Most HIVST kits can only be distributed to friends and sexual partners in person. The influence of the DSDHK is relatively limited, and it is difficult to influence some key groups. Instead of providing the HIVST kit directly to MSM, we provided a social media–based HIVST service link, which could be easily forwarded via the social networks of MSM. We encouraged all recruited MSM to share the HIVST service link to help their partners receive HIVST kits. Through voluntary application and mailing, HIVST kits can reach beyond the recruited participants and expand the testing coverage to wider social networks [22]. In addition, the indirect model has the advantage of helping the participants avoid face-to-face interactions and associated challenging encounters, such as partner violence faced by the participants of DSDHK [23]. During the global COVID-19 pandemic, both MSM and medical workers need to maintain physical distancing, and facility-based HIV testing services are restricted, making it more difficult for MSM to obtain good facility-based HIV testing services [24]. In addition, the ISDHK model can access a greater number of MSM from 29 different provinces and municipalities across China than facility-based HIV tests; whereas facility-based HIV tests provide more accurate testing and better consulting services, 1 facility-based testing center is usually only suitable for the region around the city in which it is based [17]. Combining direct and indirect secondary distribution models by giving MSM multiple HIVST kits and application links may influence more MSM to get tested for HIV. Hence, it is necessary to integrate the ISDHK model into the HIV testing service to improve the coverage of HIV testing.

To the best of our knowledge, this study is the first to find that a higher proportion of MSM who received ISDHK had never previously been tested for HIV and preferred to find sexual

partners via offline means. MSM who received ISDHK were more likely to have never previously been tested for HIV than those who received DSDHK (111/394, 28.2% vs 329/1422, 23.14%). The results of this study are consistent with those of Wu et al [18] obtained in Guangdong, China, through social networks for DSDHK (IPs: 21% vs alters: 40%). The ISDHK model may help MSM who have never been tested for HIV, thus reducing HIV transmission. A meta-analysis and previous studies have shown that MSM who find sexual partners via offline means are generally older and have lower HIV test rates, higher CAI incidence, and higher HIV prevalence [25-27]. Hence, providing HIV testing services and interventions to MSM who seek male sexual partners offline is very important for the prevention and control of HIV transmission in the overall service of public health. The results of this study indicate that ISDHK plays an essential role in expanding the coverage of HIV testing and in interventions for high-risk sexual practices (for HIV) among MSM. It is crucial to increase the coverage and frequency of HIV testing for MSM seeking male sexual partners offline and to further reduce high-risk sexual behavior.

This study also describes the HIV linkage-to-care characteristics of the MSM received ISDHK. There is no statistically significant difference in the HIV-linkage-to-care rates between MSM who received secondary distribution and those directly recruited via WeChat (27/28, 96% vs 7/7, 100%). These rates are higher than the HIV-positive referral rates of a social network–based DSDHK in Guangdong (12/15, 80%) [18] and a randomized controlled trial study in Uganda, where a women's facility-based study delivered HIVST kits to their male partners (6/26, 23%) [28]. A vast majority of MSM who were confirmed to be positive began ART within 3 months. This may have been for the following reasons: following a reactive HIVST result, using the working platform of this project, the staff were able to provide immediate web-based HIV testing consultation and linkage-to-care services through WeChat and participants suspected of being HIV-infected were recommended to facility-based tests in order to provide them with free HIV re-examination, confirmation, and ART. This study showed that the ISDHK model operated via WeChat could increase the rate of HIV testing and promote the completion of the first third of the 90% global target and the second and third AIDS prevention and control targets. Although this study and previous similar studies in China have shown that ISDHK has obvious advantages in improving the coverage of HIV testing and can compensate for the limitations of traditional facilities-based HIV testing, the current Chinese HIVST guidelines have no relevant technical specifications or guidance on ISDHK [29]. The results of this study provide the data that support the updating of the Chinese HIVST guidelines, with the aim of increasing the geographic and societal distribution of HIVST.

Strengths and Limitations

There are 3 strengths of this study. This is the first national-wide study carried out on the WeChat social media platform using ISDHK among MSM in China. Second, it is the first study to report the behavioral characteristics of alters among MSM. Third, this study is the first to encourage participants to distribute HIVST among their social media.

There are 3 limitations of this study. First, approximately one fifth of the survey respondents did not return their HIV results after receiving the self-test kit. Therefore, there may be reporting bias, and the results of this study did not reflect the characteristics of those who did not feedback their HIVST results and this may lead to a lower HIV-positive rate. Further research is needed to improve the feedback rate of the results from those participating in the HIVST. Second, only 76.6% (302/394) of participants reported 246 IPs' information; therefore, information about the complete secondary distribution network is unknown. Third, each participant could obtain only one HIVST kit for each application and pay a Chinese ¥ 50 deposit (approximately US $7), thus incentivizing the participant to feed back their test results. However, poorer participants may be discouraged or even prevented from applying for HIVST. Therefore, there may be selection bias, and the requirement for paying this deposit may have affected the representativeness of the results.

Conclusions

The indirect secondary distribution model of distributing HIVST application links via the WeChat social media platform is feasible for the MSM population. It can increase the coverage of HIV testing among MSM, especially among those who have a history of not being tested for HIV. It is necessary to integrate DSDHK with ISDHK to help increase the coverage of HIVST among MSM and other key populations.

Acknowledgments

The authors would like to thank all the participants in this study, the Rainbow Harbor staff (HIV voluntary Counseling of The First Affiliated Hospital of China Medical University, Shenyang, China), and Qiang Kang (Sunshine Voluntary Services Center, Shenyang, China) for their valuable efforts. This work was financially supported by the National Natural Science Foundation of China (81872674), the National Science and Technology Major Project (2018ZX10101-001-001-003), and Mega-Projects of National Science Research (13th Five-Year Plan [2017ZX10201101-002-007]). The funders of the study played no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

SL and JZ contributed equally to this work as first authors. HS, JX, YJ, and WG conceived and designed the study. MX, YG, TL, WZ, HW, ZC, JZ, and QH collected the data. JX and SL cleaned and analyzed the data. JX, SL, and JZ interpreted the study

results and wrote and revised the manuscript. HS and JX contributed equally as corresponding authors. All authors have read and approved the final manuscript as submitted.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist for this cross-sectional study.
[PDF File (Adobe PDF File), 92 KB - jmir_v23i10e28508_app1.pdf]

Multimedia Appendix 2
Comparison between index participants and alters (N=1816).
[DOCX File , 25 KB - jmir_v23i10e28508_app2.docx]

References

1. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, HPTN 052 Study Team. Antiretroviral therapy for the prevention of HIV-1 transmission. N Engl J Med 2016 Sep 01;375(9):830-839 [FREE Full text] [doi: 10.1056/NEJMoa1600693] [Medline: 27424812]
2. UNAIDS. URL: https://www.unaids.org/en/resources/fact-sheet [accessed 2021-09-10]
3. Beyrer C, Baral SD, van Griensven F, Goodreau SM, Chariyalertsak S, Wirtz AL, et al. Global epidemiology of HIV infection in men who have sex with men. The Lancet 2012 Jul 28;380(9839):367-377 [FREE Full text] [doi: 10.1016/S0140-6736(12)60821-6] [Medline: 22819660]
4. NCAIDS, NCSTD, China CDC. Update on the AIDS/STD epidemic in China in December 2017. Chin J AIDS STD 2014;20:75. [doi: 10.13419/j.cnki.aids.2018.02.01]
5. Chow E, Wilson D, Zhang L. The rate of HIV testing is increasing among men who have sex with men in China. HIV Med 2012 May 18;13(5):255-263 [FREE Full text] [doi: 10.1111/j.1468-1293.2011.00974.x] [Medline: 22252151]
6. Cao B, Liu C, Durvasula M, Tang W, Pan S, Saffer AJ, et al. Social media engagement and HIV testing among men who have sex with men in china: a nationwide cross-sectional survey. J Med Internet Res 2017 Jul 19;19(7):e251 [FREE Full text] [doi: 10.2196/jmir.7251] [Medline: 28724510]
7. Ma W, Wu G, Zheng H, Zhang W, Peng Z, Yu R, et al. Prevalence and risk factors of HIV and syphilis, and knowledge and risk behaviors related to HIV/AIDS among men who have sex with men in Chongqing, China. J Biomed Res 2016 Mar;30(2):101-111 [FREE Full text] [doi: 10.7555/JBR.30.20150057] [Medline: 28276665]
8. Zhang C, Li X, Heilemann M, Chen X, Wang H, Koniak-Griffin D. Facilitators and barriers of HIV self-testing among Chinese men who have sex with men: a qualitative study. J Assoc Nurses AIDS Care 2020 Nov 06;32(5):599-609 (forthcoming) [FREE Full text] [doi: 10.1097/JNC.0000000000000218] [Medline: 33165181]
9. Ye R, Liu C, Tan S, Li J, Simoni JM, Turner D, et al. Factors associated with past HIV testing among men who have sex with men attending university in China: a cross-sectional study. Sex Health 2021 Mar;18(1):58-63. [doi: 10.1071/SH20088] [Medline: 33639685]
10. Crepaz N, Marks G, Liau A, Mullins MM, Aupont LW, Marshall KJ, HIV/AIDS Prevention Research Synthesis (PRS) Team. Prevalence of unprotected anal intercourse among HIV-diagnosed MSM in the United States: a meta-analysis. AIDS 2009 Aug 24;23(13):1617-1629. [doi: 10.1097/QAD.0b013e32832effae] [Medline: 19584704]
11. Krishnaratne S, Hensen B, Cordes J, Enstone J, Hargreaves JR. Interventions to strengthen the HIV prevention cascade: a systematic review of reviews. Lancet HIV 2016 Jul;3(7):e307-e317 [FREE Full text] [doi: 10.1016/S2352-3018(16)30038-8] [Medline: 27365205]
12. Thirumurthy H, Masters SH, Mavedzenge SN, Maman S, Omanga E, Agot K. Promoting male partner HIV testing and safer sexual decision making through secondary distribution of self-tests by HIV-negative female sex workers and women receiving antenatal and post-partum care in Kenya: a cohort study. Lancet HIV 2016 Jun;3(6):e266-e274 [FREE Full text] [doi: 10.1016/S2352-3018(16)00041-2] [Medline: 27240789]
13. Okoboi S, Lazarus O, Castelnuovo B, Nanfuka M, Kambugu A, Mujugira A, et al. Peer distribution of HIV self-test kits to men who have sex with men to identify undiagnosed HIV infection in Uganda: A pilot study. PLoS One 2020;15(1):e0227741 [FREE Full text] [doi: 10.1371/journal.pone.0227741] [Medline: 31971991]
14. Lippman S, Lane T, Rabede O, Gilmore H, Chen YH, Mlotshwa N, et al. High acceptability and increased HIV-testing frequency after introduction of HIV self-testing and network distribution among South African MSM. J Acquir Immune Defic Syndr 2018 Mar 01;77(3):279-287 [FREE Full text] [doi: 10.1097/QAI.0000000000001601] [Medline: 29210826]
15. Masters SH, Agot K, Obonyo B, Napierala Mavedzenge S, Maman S, Thirumurthy H. Promoting partner testing and couples testing through secondary distribution of HIV self-tests: a randomized clinical trial. PLoS Med 2016 Nov;13(11):e1002166 [FREE Full text] [doi: 10.1371/journal.pmed.1002166] [Medline: 27824882]

16. Gichangi A, Wambua J, Mutwiwa S, Njogu R, Bazant E, Wamicwe J, et al. Impact of HIV self-test distribution to male partners of ANC clients: results of a randomized controlled trial in Kenya. J Acquir Immune Defic Syndr 2018 Dec 01;79(4):467-473 [FREE Full text] [doi: 10.1097/QAI.0000000000001838] [Medline: 30148731]
17. Pettifor A, Lippman SA, Kimaru L, Haber N, Mayakayaka Z, Selin A, et al. HIV self-testing among young women in rural South Africa: a randomized controlled trial comparing clinic-based HIV testing to the choice of either clinic testing or HIV self-testing with secondary distribution to peers and partners. EClinicalMedicine 2020 Apr;21:100327 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100327] [Medline: 32322811]
18. Wu D, Zhou Y, Yang N, Huang S, He X, Tucker J, et al. Social media-based secondary distribution of HIV/syphilis self-testing among Chinese men who have sex with men. Clin Infect Dis 2020 Jun 26:ciaa825 (forthcoming)(forthcoming). [doi: 10.1093/cid/ciaa825] [Medline: 32588883]
19. Bwalya C, Simwinga M, Hensen B, Gwanu L, Hang'andu A, Mulubwa C, HPTN 071 (PopART) study team. Social response to the delivery of HIV self-testing in households: experiences from four Zambian HPTN 071 (PopART) urban communities. AIDS Res Ther 2020 Jun 11;17(1):32 [FREE Full text] [doi: 10.1186/s12981-020-00287-y] [Medline: 32527261]
20. WeChat. Wikipedia. URL: https://en.wikipedia.org/wiki/WeChat [accessed 2021-02-09]
21. Zhang W, Hu Q, Tang W, Jin X, Mao X, Lu T, et al. HIV self-testing programs to men who have sex with men delivered by social media key opinion leaders and community-based organizations are both effective and complementary: a national pragmatic study in China. J Acquir Immune Defic Syndr 2020 Aug 15;84(5):453-462. [doi: 10.1097/QAI.0000000000002375] [Medline: 32692103]
22. Freeman AE, Sullivan P, Higa D, Sharma A, MacGowan R, Hirshfield S, eSTAMP Study Group. Perceptions of HIV self-testing among men who have sex with men in the United States: a qualitative analysis. AIDS Educ Prev 2018 Feb;30(1):47-62. [doi: 10.1521/aeap.2018.30.1.47] [Medline: 29481298]
23. Matovu JK, Kisa R, Buregyeya E, Chemusto H, Mugerwa S, Musoke W, et al. 'If I had not taken it [HIVST kit] home, my husband would not have come to the facility to test for HIV': HIV self-testing perceptions, delivery strategies, and post-test experiences among pregnant women and their male partners in Central Uganda. Glob Health Action 2018;11(1):1503784 [FREE Full text] [doi: 10.1080/16549716.2018.1503784] [Medline: 30092155]
24. Lagat H, Sharma M, Kariithi E, Otieno G, Katz D, Masyuko S, et al. Impact of the COVID-19 pandemic on HIV testing and assisted partner notification services, Western Kenya. AIDS Behav 2020 Nov;24(11):3010-3013 [FREE Full text] [doi: 10.1007/s10461-020-02938-7] [Medline: 32488552]
25. Wei L, Chen L, Zhang H, Yang Z, Liu S, Tan W, et al. Relationship between gay app use and HIV testing among men who have sex with men in Shenzhen, China: a serial cross-sectional study. BMJ Open 2019 Aug 24;9(8):e028933 [FREE Full text] [doi: 10.1136/bmjopen-2019-028933] [Medline: 31446409]
26. Zou H, Fan S. Characteristics of men who have sex with men who use smartphone geosocial networking applications and implications for HIV interventions: a systematic review and meta-analysis. Arch Sex Behav 2017 May;46(4):885-894. [doi: 10.1007/s10508-016-0709-3] [Medline: 27043837]
27. Rice E. Sex risk among young men who have sex with men who use Grindr, a smartphone geosocial networking application. J AIDS Clinic Res 2012;01(S4):05. [doi: 10.4172/2155-6113.s4-005]
28. Korte J, Kisa R, Vrana-Diaz C, Malek AM, Buregyeya E, Matovu JK, et al. HIV oral self-testing for male partners of women attending antenatal care in central Uganda: uptake of testing and linkage to care in a randomized trial. J Acquir Immune Defic Syndr 2020 Jul 01;84(3):271-279. [doi: 10.1097/QAI.0000000000002341] [Medline: 32168168]
29. Xiaoxia H, Yan J, Weng X. Brief introduction of AIDS self-testing guide manual (1st edition). Chin J AIDS STD 2020 Nov 06;26(06):653-655. [doi: 10.13419/j.cnki.aids.2020.06.25]

Abbreviations

aOR: adjusted odds ratio
ART: antiretroviral therapy
CAI: condomless anal intercourse
DSDHK: direct secondary distribution of HIV self-testing kits
HIVST: HIV self-testing
IP: index participant
ISDHK: indirect secondary distribution of HIV self-testing kits
MSM: men who have sex with men
PASS: Power Analysis and Sample Size
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by R Kukafka; submitted 05.03.21; peer-reviewed by W Tang, W Li, TH Kwan; comments to author 27.04.21; revised version received 30.05.21; accepted 05.07.21; published 26.10.21.

Please cite as:
Li S, Zhang J, Mao X, Lu T, Gao Y, Zhang W, Wang H, Chu Z, Hu Q, Jiang Y, Geng W, Shang H, Xu J
Feasibility of Indirect Secondary Distribution of HIV Self-test Kits via WeChat Among Men Who Have Sex With Men: National Cross-sectional Study in China
J Med Internet Res 2021;23(10):e28508
URL: https://www.jmir.org/2021/10/e28508
doi:10.2196/28508
PMID:34698651

Original Paper

Identifying Opportunities, and Motivation to Enhance Capabilities, Influencing the Development of a Personalized Digital Health Hub Model of Care for Hip Fractures: Mixed Methods Exploratory Study

Lalit Yadav[1,2,3*], BDS, MPH; Tiffany K Gill[2*], BAppSc, MAppSc, MBA, PhD; Anita Taylor[3], MSc; Jennifer De Young[3], MSc; Mellick J Chehade[1,3*], MBBS, PhD, FRACS

[1]NHMRC Centre for Research Excellence in Frailty and Healthy Ageing, Adelaide Medical School, University of Adelaide, Adelaide, Australia

[2]Adelaide Medical School, Faculty of Health and Medical Sciences, University of Adelaide, Adelaide, Australia

[3]Department of Orthopaedics & Trauma, Royal Adelaide Hospital, Adelaide, Australia

[*]these authors contributed equally

Corresponding Author:
Lalit Yadav, BDS, MPH
NHMRC Centre for Research Excellence in Frailty and Healthy Ageing
Adelaide Medical School
University of Adelaide
Level 5G 581, Royal Adelaide Hospital
Adelaide, 5000
Australia
Phone: 61 870742006
Email: Lalit.yadav@adelaide.edu.au

Abstract

Background: Most older people after a hip fracture injury never return to their prefracture status, and some are admitted to residential aged care facilities. Advancement of digital technology has helped in optimizing health care including self-management and telerehabilitation.

Objective: This study aims to understand the perspectives of older patients with hip fracture and their family members and residential aged caregivers on the feasibility of developing a model of care using a personalized digital health hub.

Methods: We conducted a mixed methods study in South Australia involving patients aged 50 years and older, their family members, and residential aged caregivers. Quantitative data analysis included basic demographic characteristics, and access to digital devices was analyzed using descriptive statistics. Spearman rank-order correlation was used to examine correlations between the perceived role of a personalized digital health hub in improving health and the likelihood of subsequent use. Findings from qualitative analysis were interpreted using constructs of capability, opportunity, and motivation to help understand the factors influencing the likelihood of potential personalized digital health hub use.

Results: This study recruited 100 participants—55 patients, 13 family members, and 32 residential aged caregivers. The mean age of the patients was 76.4 (SD 8.4, range 54-88) years, and 60% (33/55) of the patients were female. Approximately 50% (34/68) of the patients and their family members had access to digital devices, despite less than one-third using computers as part of their occupation. Approximately 72% (72/100) of the respondents thought that personalized digital health hub could improve health outcomes in patients. However, a moderate negative correlation existed with increasing age and likelihood of personalized digital health hub use (Spearman ρ=–0.50; P<.001), and the perceived role of the personalized digital health hub in improving health had a strong positive correlation with the likelihood of personalized digital health hub use by self (Spearman ρ=0.71; P<.001) and by society, including friends and family members (Spearman ρ=0.75; P<.001). Most patients (54/55, 98%) believed they had a family member, friend, or caregiver who would be able to help them use a personalized digital health hub. Qualitative analysis explored capability by understanding aspects of existing knowledge, including willingness to advance digital navigation skills. Access could be improved through supporting opportunities, and factors influencing intrinsic motivation were considered crucial for designing a personalized digital health hub–enabled model of care.

Conclusions: This study emphasized the complex relationship between capabilities, motivation, and opportunities for patients, their family members, and formal caregivers as a *patient networked unit*. The next stage of research will continue to involve a

cocreation approach followed by iterative processes and understand the factors influencing the development and successful integration of complex digital health care interventions in real-world scenarios.

(J Med Internet Res 2021;23(10):e26886) doi:10.2196/26886

KEYWORDS

digital health; mixed-methods; hip fractures; behavior change; patient education; model of care; mobile phone; patient networked units

Introduction

Background

The population of South Australia is older than that of all the mainland states and territories in Australia, except Tasmania. According to the Australian Bureau of Statistics, the current population of South Australia is approximately 1.7 million [1]. It is expected to increase to 1.85 million by 2026 and to 2 million by 2038 using current population projections. This increase was reflected by a significant increase in the older population. The number of retirees in the 65-79 years age group is projected to increase by 40% by 2041, using 2016 as the baseline. Moreover, the population aged 80 years and older is projected to increase by 117% over the same period [1], which will require an increased need for appropriate health and social care [2]. Although there has been greater realization that the skills, knowledge, and experience of older people could be better used with regard to their health care, there remains a view that they are a drain on society given their health problems and service needs [3,4]. The United Nations Economic Commission for Europe has suggested that altering this view is a key strategy to improve the integration and participation of older people in society [5].

Hip Fractures and Multimorbidity

Fragility fractures mostly occur in older people owing to low-trauma falls, which often result from multimorbidity [6,7]. Multimorbidity is the presence of more than one chronic disease in an individual and is influenced not only by health-related characteristics but also by socioeconomic, cultural, and environmental factors, as well as patient behavior [8]. Hip fractures are among the most devastating fragility fractures, and their management becomes challenging because of the required involvement of several disciplines within health and social care. This cohort not only represents healthy older people at one end of the spectrum but also comprises people with frailty, sarcopenia, osteoporosis, and dementia at the other end of the spectrum. This makes management of an acute event such as a hip fracture complex, with wide-ranging outcomes within the health care systems involving multiple disciplines and service providers [9-11]. It is made even more complex with the crossover between different levels of care, ranging from acute tertiary to primary and residential aged care [12,13]. Most patients who are admitted to acute hospital care are unable to return to their prefracture level of independence [14-16]. Although some patients return to independent living in their own homes, a significant number are either newly admitted or return to residential aged care [17]. Thus, we believe that individual patient outcomes can only be improved by envisaging a model of care that ensures a holistic and integrated approach to health service delivery while empowering patients and their caregivers.

Digital Health–Enabled Models of Care

Models of care (MoCs) are frameworks mutually agreed by key stakeholders accountable for delivering evidence-informed quality health care. Such frameworks must be functional, outlining the optimal manner in which condition-specific care should be made available and delivered to consumers while addressing issues related to specific aspects of service provision [18]. They go beyond clinical practice guidelines to incorporate practical delivery issues of who, when, where, and how care is best delivered and evaluated [18]. Thus, MoCs become complex due to their multidisciplinary workforce links to secondary and tertiary care services, the biopsychosocial needs of the patients, and frequently changing organizational structures. Although the mandate of primary care is to offer a generalist approach for dealing with older adults with multimorbidity, the coordination of community services is difficult. It is often left to the patients and their caregivers to coordinate and navigate through a range of services into which their individual social circumstances and priorities also need to be factored [18]. Provision of accurate, timely, and adequate information by educating patients plays a vital role in improving engagement and participation in the recovery and rehabilitation processes within the MoC. Health professionals often overlook patients' health literacy during routine practice, incorrectly assuming that the health information and instructions provided to patients and their family members have been understood [19,20]. Patient education, which also involves family members and residential aged care staff, is crucial for empowerment and improving health literacy [7,21-23].

Technological advancements have led to the evolution of clinical decision support systems and a myriad of consumer mobile apps to target different stakeholders, with the intention of optimizing health care and self-management of chronic disease conditions and maintaining a healthy lifestyle [12,13,24]. Nevertheless, there remains a need to build on the knowledge exchange process between health care providers and patients, along with their family members and caregivers, acting as facilitators [13]. By targeting different multimorbidities, which correspond to the internal capacity of individual older people, care can be personalized. This aligns with the World Health Organization (WHO) guidelines on community-level interventions to manage declines in intrinsic capacity through an integrated care approach for older people (WHO-Integrated Care for Older People) [13,23]. The WHO describes digital health as a broad umbrella term encompassing eHealth, mobile health, and emerging areas, such as the use of advanced computing sciences in big data, genomics, artificial intelligence, and machine learning [25].

Commitment and strategic engagement of stakeholders, including patients and the community, is required to improve health care services across all stages, from inception to operation or implementation [25]. Further advancement of these technological solutions can bypass some of the care disparities imposed by sociodemographic and geographic barriers and support the move toward universal health coverage [26,27].

Objective

The aim of this study is to understand the perspectives of older patients with hip fracture, their family members, and formal caregivers in residential aged care facilities to inform the development of a personalized digital health hub by understanding their current access to digital devices and factors affecting the likelihood of future use [8].

Methods

Setting and Study Design

We examined patients with hip fracture aged 50 years and older. This mixed-methods study [28] was conducted at the Royal Adelaide Hospital, a tertiary trauma care center in Adelaide, South Australia. This is one of the busiest hospitals in Australia for acute hip fractures, with local estimates suggesting approximately 500 to 600 patients treated annually [29].

Digital Health Hub Initial Concept

In this study, a digital health hub scenario, which is currently under development, was described to generate appropriate responses from the study participants. This proposed web-based health information portal, or a website, is intended for patients who can access all relevant information about their hip fractures. It includes details in multimedia formats of diagnosis and treatment options, medications, wound management and rehabilitation exercises, potential problems encountered during the hospital admission and post discharge, information on how to deal with difficulties, as well as how and when to attend follow-up appointments or seek more help from the health care team. It is interactive, enabling patients and their caregivers to provide both targeted and patient-initiated information to their health care provider, which is captured digitally. It also allows users to selectively make information available to family members or other people involved in their care (either formally or informally).

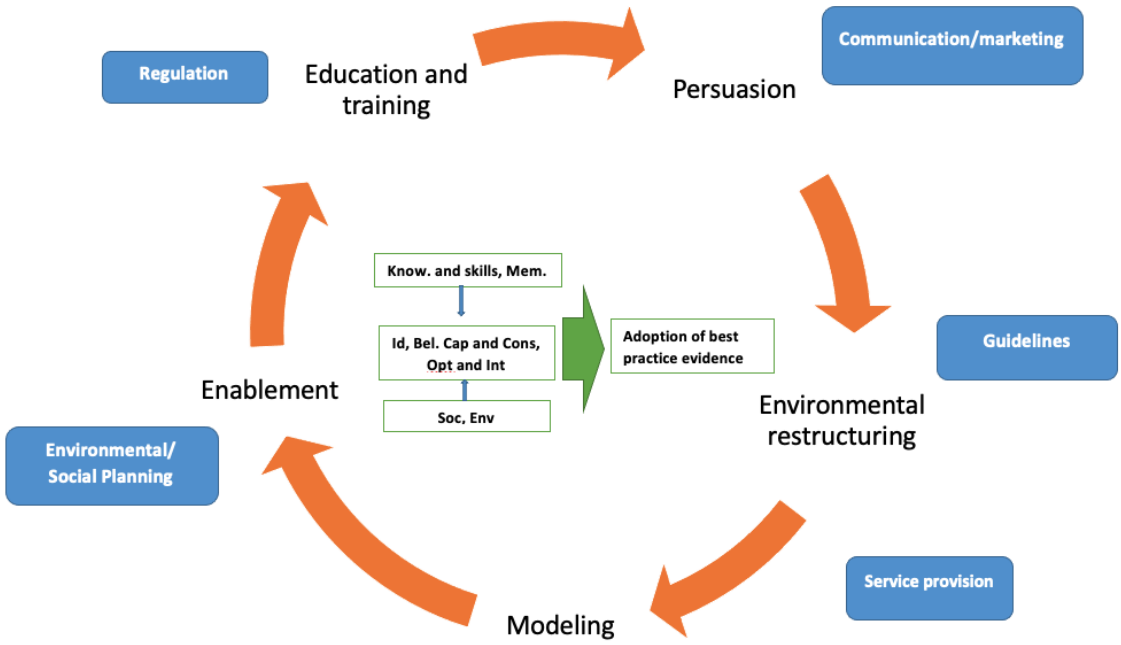
Data Collection and Analysis

Participants in this study were recruited from a previous prospective cohort study that focused on the delivery of fracture liaison service, undertaken between January and December 2016. Patients were contacted consecutively, and those who consented were invited to participate in the study. If participants in the original study had caregivers participating on their behalf, they were approached in a similar manner. Family members were represented as informal caregivers, whereas formal caregivers were caregivers of older people in residential aged care facilities. The data for this study were collected over a period of 6 months, from January 2017 to June 2017, using face-to-face interactions or telephone calls on the basis of individual preferences. A semistructured questionnaire consisting of closed and open-ended questions was developed and administered (Multimedia Appendix 1). Participants' responses to each question were entered into a hard copy Word (Microsoft Corporation) document by the research staff while administering the survey questionnaire. These responses were then compiled on an Excel (Microsoft Corporation) spreadsheet and stored on a password-protected folder on the secured server of SA Health.

Quantitative data analysis included basic demographic characteristics, and access to computers and the internet (digital devices) at home and in the workplace were analyzed using descriptive statistics. Spearman rank-order correlation was used to examine correlations between the perceived role of a personalized digital health hub in improving health and the likelihood of subsequent use of a personalized digital health hub. Fisher exact test and odds ratios were calculated for comparisons across respondent groups with respect to previous access to computers at the workplace, gender differences, and likelihood of potential personalized digital health hub use.

Qualitative data analysis included a series of open-ended questions to identify potential barriers and facilitators for accessing personalized digital health hub. Responses to the open-ended questions were analyzed deductively and aligned with the tenets of capability, opportunity, and motivation [30,31]. These findings interpreted using constructs of capability, opportunity, and motivation embedded within a theoretical Behavior Change Wheel (BCW) framework and helped to understand factors influencing the likelihood of potential personalized digital health hub use. The use of this framework to explore multidisciplinary stakeholder engagement within hip fracture management has been described elsewhere (Figure 1) [31,32].

Figure 1. Constructs of capability, opportunity, and motivation embedded within a Behaviour Change Wheel Framework [27]. Sources of Behaviour box and arrow in green; Intervention functions orange; Policy categories blue. Know: knowledge; Mem.: memory, attention and decision processes (capability); Id: social or professional role and identity; Bel. Cap. and Cons: beliefs about capabilities and consequences; Opt. and Int.: Optimism and intentions (motivation); Soc.: social influences; Env.: environmental context and resources (opportunity).



Ethics

This study was approved by the Human Research Ethics Committee of the Central Adelaide Local Health Network (RAH protocol number R20080704, HREC reference: 080704, ethics approval amendment on 12/12/2016, CALHN reference number: 8977, SSA approval 23/1/2017).

Results

Quantitative Findings

Overall, 100 participants were recruited in the study (Table 1). These included 55 patients, 13 family members as informal caregivers, and 32 residential aged care workers. The age (in years) of patients (mean 76.4, SD 8.4; range 54-88) was similar to that of their family members (mean 77.2, SD 10.0), whereas the residential aged caregivers were younger (mean 45.2, SD 11.6). Females represented 60% (33/55), 54% (7/13), and 86% (28/32) of patients, family members or informal caregivers, and residential aged formal caregivers, respectively. Around a quarter to a third of the participants were from *professional* occupations across all 3 groups. The residential care group included 5 registered nurses and 2 enrolled nurses. Within the patients' group, the common occupations were laborers, clerical and administrative workers, homemakers, machinery operators, and drivers.

With respect to digital access, approximately half of the patients (28/55, 51%) and their families (6/13, 46%) had access to their own computer devices through the internet. A quarter (13/55, 24%) of the patients used a computer as part of their occupation in comparison to 31% (4/13) of family members, whereas more than 91% (29/32) of the residential aged caregivers had computer access.

Patients who reported using a computer as part of their work were 8 times more likely to have access to a computer with internet access at home compared with those who did not use a computer at work (odds ratio [OR] 8.08, 95% CI 1.58-41.18; Fisher exact test=0.0095; *P*=.05). The mean age of those with access to a computer was 4.6 years less than that of those without access (74.2 vs 78.8, *P*=.04).

Approximately 40% (25/68) of the patients and their family members reported using basic operational tools such as email and Google (or other search engines) in comparison to 100% (32/32) of the residential caregivers who used these functions (Table 1). Of these, more than 85% (46/53) of the patients found it *reasonably easy* to *very easy* to operate these basic functionalities through the internet. Skype or other video calling programs were used by only a quarter of patients and their family members, whereas 44% (14/32) of the residential caregivers reported that they used these programs. Among the patient group, men were 3 times more likely to report having used email than women (OR 3.75, 95% CI 1.17-11.9; *P*=.02). However, given the opportunity, 42% (23/55), 38% (5/13), and 56% (18/32) of the patients, their family members, and residential aged caregivers, respectively, expressed their willingness to learn or advance their skills in these areas. While exploring this aspect further, all patients except 1 (54/55, 98%)

also said they had a family member, friend, or caregiver who would be able to help them use a digital health platform.

Approximately 72% (72/100) of the respondents thought that personalized digital health hub could improve the health of patients. Although a moderate negative correlation existed with increasing age and likelihood of personalized digital health hub use (Spearman ρ=–0.50; *P*<.001), the perceived role of the digital health hub in improving health had a strong positive correlation with the likelihood of personalized digital health hub use by self (Spearman ρ=0.71; *P*<.001) and by society, including friends and family members (Spearman ρ=0.75; *P*<.001). Furthermore, those participants who thought that the support content and services provided through personalized digital health hub would improve their health were more likely to use such a platform by themselves (OR 33.80, 95% CI 7.33-155.76; *P*<.001), and their friends and family members (OR 27.23, 95% CI 8.06-91.95; *P*<.001).

In terms of intention to buy a computer, 65% (36/55) of the patients said they would not be willing to purchase a computer or other device to enable them to access a web-based portal. Of the 35% (19/55) who would be willing to purchase a computer or device, 13% (7/55) said they would be willing to spend up to Aus $200 (US $144), 18% (10/55) said they would spend up to Aus $500 (US $360), and 4% (2/55) said they would spend up to Aus $1000 (US $720).

Table 1. Basic demographics and computer access characteristics.

Demographics and access to digital technology	Patients (n=55)	Family members (n=13)	Residential aged caregivers (n=32)
Age (years), mean (SD)	76.4 (8.4)	77.2 (10.0)	45.2 (11.6)
Female, n (%)	33 (60)	7 (54)	28 (86)
Occupation[a], n (%)			
Managers	2 (4)	1 (8)	N/A[b]
Professionals	13 (24)	4 (30)	7 (22)
Technicians and trade workers	6 (11)	2 (14)	N/A
Community and personal service workers	1 (2)	N/A	25 (78)
Clerical and administrative workers	7 (13)	1 (8)	N/A
Sales workers	3 (5)	1 (8)	N/A
Machinery operators and drivers	5 (9)	1 (8)	N/A
Laborers	8 (14)	1 (8)	N/A
Homemaker	5 (9)	1 (8)	N/A
Unemployed	1 (2)	N/A	N/A
Did not respond	4 (7)	1 (8)	N/A
Access to digital technology			
Use of computer	13 (24)	4 (31)	29 (91)
Own computer with internet access	28 (51)	6 (46)	32 (100)
Device access but no internet	4 (7)	0 (0)	0 (0)
Use email	20 (36)[c]	5 (38)	32 (100)
Use Google or other search engines	20 (36)	5 (38)	32 (100)
Use Skype or other video calling programs	13 (24)	3 (23)	14 (44)
Willingness to learn[d]	23 (42)	5 (38)	18 (56)

[a]Occupation groups as defined by the Australian and New Zealand Standard Classification of Occupations.

[b]N/A: not applicable.

[c]Men versus women, odds ratio 3.75 (95% CI 1.17-11.9; *P*=.02).

[d]Willingness to learn how to use email or internet search engines such as Google or a video calling program such as Skype, if the respondents have not used any of them before.

Qualitative Findings

The respondents answered the two open-ended questions within the survey instrument to explore barriers (Q15) and facilitators (Q16) influencing the likelihood of using a personalized digital health hub to educate and empower patients, their family members, and caregivers within residential aged care (Multimedia Appendix 1). These factors were interpreted using constructs of capability, opportunity, and motivation within a BCW framework, and relevant quotes from the study respondents (R) were also provided. Capability in this study was defined as an individual's psychological and physical

capacity to engage with the potential personalized digital health hub, which included having the necessary knowledge and skills. Opportunity considered all the factors lying outside the individual that make the behavior possible or prompt it, such as the likelihood of engaging with the potential personalized digital health hub. Motivation included processes that energize and direct behavior, not just goals but also habitual processes, emotional responses, and analytical decision-making. These constructs influence each other, as they work dynamically, such as access to opportunity can drive motivation, whereas enacting behavior can alter capability, motivation, and opportunity [32]. There were 59 participants who responded to questions corresponding to barriers with further breakdown of 38, 9, and 12 as patients, their family members, and residential aged caregivers, respectively. In contrast, 40 participants responded to questions corresponding to facilitators, representing 24, 3, and 13 patients, their family members, and residential aged caregivers, respectively.

Capability

Some patients recognized that possessing the necessary knowledge and skills while accessing digital devices can help explore relevant web-based health information, which could enable a better understanding of their health condition. Conversely, there were some family members and caregivers who lacked confidence in using digital devices. There was no one available to teach them and also felt inadequate about understanding the patient's medical condition. The patients' comments reflect that the information gathered through such a digital health platform would actually help improve their decision-making during the recovery process:

> *Can't use the computer, no one to teach me to use it at the moment.* [R53]
>
> *Would have been very useful (internet), always looking things up anyway.* [R49]
>
> *Keeping up with computers and technology keeps me sharp [informing decisions].* [R38]

Opportunity

Patients and their family members considered their personal environment and the affordability of resources, such as digital devices, as a major limiting factor. Residential aged caregivers saw digital health platforms as an opportunity to provide general health information, including healthy lifestyle, diet, and exercise. Furthermore, it was suggested that a platform such as a personalized digital health hub (or similar) would be more efficient or easier than existing options and would provide a potential solution with resources consisting of videos of exercises such as yoga and tai chi. In addition, several patients were of the opinion that a list of available services (eg, allied health professionals, exercise classes, alternative therapies) along with health management information could be well received by the community:

> *Can't afford computer on aged pension.* [R15]
>
> *Lots of people interested in general information about their health as well ie healthy lifestyle, diet, exercise.* [R72]
>
> *Videos [exercises] would be very helpful, would like information about how long hip replacements last and how to take care of them, any information is good information? Could include tai chi or yoga.* [R24]
>
> *Knowing it's there [digital platform] to fall back on, list of things/services that are available.* [R47]
>
> *Rehab exercises, suggestions and options, a list of services.* [R67]

Motivation

Being older was identified by both patients and their family members as one of the main hurdles to using the potential personalized digital health hub platform. Residential caregivers, however, identified practical issues such as lack of time in their existing role to use such a solution, which is currently not a part of their job. Participants across all 3 categories identified their existing capabilities as a limiting factor. However, they were also positive about the potential capabilities of a digital solution, such as the availability of information that would reduce the need to visit a physician and access to trustworthy interventions. These interventions include videos and information about health and instructions from reputable sources such as physiotherapists that they can follow in their own time as well as the potential to more easily track their appointments with different health care providers:

> *Hard for other elderly people.* [R8]
>
> *Just not practical, responsible for more than one person at a time so time using this would take away from actually caring for people.* [R83]
>
> *Access to information on demand, not have to visit doctor, not missing phone calls and use on own time.* [R3]
>
> *Reputable sources would be good, videos of exercises helpful, a realistic timeline for recovery would be useful.* [R43]

The residential aged care staff thought that a digital health solution could potentially improve handover processes through a better exchange of information between specialists and caregivers. Most participants were optimistic about the range of functions that a digital health platform could provide; however, some had reservations such as preferring phone conversations or maintaining conventional face-to-face interactions with the physician. Emotionally, some consumers were unhappy with the services provided through technology-based solutions in comparison with face-to-face interactions. One of the patients identified a potential lack of reinforcement in terms of someone who could teach or handhold, which could be a barrier to using a digital solution. Conversely, some patients thought that it could help them achieve more peace of mind and service satisfaction:

> *Would (digital platform) improve handover of information between specialists and carers or the patient, keep everyone on the same page more.* [R88]
>
> *Prefer phone calls, more personal, know who you're talking to.* [R14]

Good to be able to see exercises [over the internet], peace of mind. [R6]

Discussion

Principal Findings

This study was considered as one of its kind due to the involvement of combined perspectives from patients with hip fragility fractures, their family members, and residential aged caregivers. Older people with hip fractures often have low intrinsic capacity, leading to depletion in physical function, mental health impairment, and increased health care costs [33]. Therefore, it is crucial to address hip fractures among older people, particularly those with multimorbidities, as a whole and in an integrated manner, rather than managing individual issues in isolation or silos, including improving health literacy by connecting with family members and formal caregivers as networked units [34]. This study explored the feasibility of a potential personalized digital health hub model of care in educating, empowering, and integrating health services, including self-management, for older patients with hip fractures in South Australia. Quantitative and qualitative methods were used in synergy to maximize the interpretation of findings. The BCW framework was applied through constructs of capability, opportunity, and motivation. These constructs are embedded within the BCW framework, which has been used in many contemporary scenarios for developing complex health interventions, including stroke rehabilitation [35] and multiple lifestyle issues [36,37].

Quantitative findings suggested that patients and their family members were of the same age, and almost half of them had current access to digital devices with the internet, despite only about a quarter of them using computers as part of their occupation (Opportunity). Although significant gaps existed with respect to operating emails, video calling, and exploring search engines (Capability), many of them expressed their willingness (motivation) to advance their skills through the supporting environment. The latter can be strengthened as 98% (54/55) of the patients said they had family members, friends, or caregivers (Opportunity) who would be able to help them use such a resource. Furthermore, the findings from our study suggest that 72% (72/100) of the respondents thought that the personalized digital health hub would be useful for improving their health.

Findings from the qualitative analysis explored deeper meanings of individual capability, opportunity, personal circumstances, and motivational factors varying within each group. Capability mainly focuses on knowledge, skills, and decision-making processes, whereas opportunities could be in the form of availability and access to digital devices and holistic care [10], including healthy lifestyle, diet, and exercise [36]. Some consumers may have better knowledge and skills to understand health information and access web-based resources. These people advocated for the personalized digital health hub being available for communicating high-quality and trustworthy health information resources, tracking appointments, and linking relevant services through a single hub. On the other hand, some preferred traditional face-to-face interactions and considered declining individual capacities due to aging as a possible challenge to cope with learning associated with the new technology. For some, personal circumstances, including affordability and access to computer systems, were important aspects to be considered. Furthermore, the findings suggested that motivation to engage with personalized digital health hub could be adversely affected by increasing patient age and contributing to additional workload for caregivers. However, information available in different formats, such as video or interactive, could improve patient engagement, help navigate different service provisions, and improve workforce handover processes delivered through an agreed model of care [18].

The sharing of information between patients and health care professionals is one of the key pillars of therapeutic relationships [38]. Increasingly, this information is being shared on the web, as digital health platforms through which patients can access education regarding medical conditions, information on self-management, and communication of health information to health professionals [39-41]. Clearly accessing and using some of these platforms will challenge certain groups within populations, such as the older adults. Ulrich and Vaccaro [42] described the benefits to patients receiving health information on the internet. These included improved health outcomes, mainly due to fulfillment of expectations and changes in behavior, which are facilitated by improved availability of information and resources. They note that older people in particular prefer audiovisual or pictorial explanations and information [42]. Furthermore, most patients do not have the capacity to distinguish nonbiased and reputable sources of information from commercially biased promotional materials [42]. The personalized digital health hub research program described here presents an opportunity to standardize and ensure the quality and evidence base of information received by patients and caregivers. A contemporary example of a digital technology used to improve access to first-line care for musculoskeletal conditions is the painHEALTH initiative [43]. This platform was codeveloped with consumers in response to the escalating burden of pain management associated with musculoskeletal conditions. The development of content was aligned to best practice recommendations from musculoskeletal MoC [18,43] and calls to action for improved care highlighted in the Australian National Pain Strategy [44].

Not all people have access to the internet, and this is especially true for people older than 65 years. However, according to the Australian Bureau of Statistics and Household Use of Information Technology survey for 2016-2017, the proportion of users accessing the internet for health-related services or research has more than doubled from 22% in 2014-2015 to 46% in 2016-17. Among older people, 55% of those aged 65 or older accessed the internet in a typical week, a 4% rise from the survey conducted in 2014-2015 [45]. Internet use correlated positively with educational attainment and household income and negatively with rurality. However, the survey is likely to be an overestimate of the proportion of older adults in the population who regularly use the internet as it excluded *people living in nonprivate dwellings such as hotels, university residences, students at boarding schools, patients in hospitals, inmates of prisons*, and *residents of other institutions* (eg, *retirement homes*

and *homes for persons with disabilities*) [45]. The survey also noted that 14% of Australian households did not have internet access [45].

Caregivers, spouses, or family members can and should be engaged to assist with the personalized digital health hub platform in consultation with the patient. However, any provision to replace in-person clinical interactions must include a safety net for patients without access. Our study suggests that 46% (46/100) of the participants were willing to learn and develop their skills. Usually, patients accessing public tertiary care facilities are more likely to belong to the lowest socioeconomic status grouping [46]; despite being economically disadvantaged, more than one-third of these patients within our study were willing to buy a computer or other smartphone devices to access the potential personalized digital health hub platform. The majority of them would spend between Aus $200-$500 (US $144-$360), but very few of them could go up to Aus $1000 (US $720). This emerging likelihood of using potential digital health solutions was also supported by another study [47], in which 63% of the participants expressed their intentions as *definitely or probably* to be using a digital health platform as a web-based interface. Such a platform could carry out tasks such as making appointments, asking questions, receiving treatment, information, and providing support for their health and well-being. This study also emphasized the importance of user-friendliness and quickly resolving issues such as bugs in the initial releases [47].

Although digital health care may offer feasible and efficient options for monitoring and securely interacting with patients, an adequate level of engagement with the technology by all stakeholders is critical. In a recent rollout of an Australian opt-out digital health data management system (known as My Health Record), although there was a 90% subscription rate by patients, less than a quarter of health provider organizations were using the system [48]. This was even though 60% ranked clinical integration and improved patient experiences as their top priorities [49]. This mismatch suggests that educating practitioners to use digital systems is as important as patient engagement and compliance [50,51]. Together, this could have an incremental effect on patient outcomes and service delivery.

Limitations

This study has several limitations. One of them was a convenience sample from a single hospital site. However, this site is a major tertiary referral center that receives hip fracture patients across South Australia. Similarly, because of time and resource constraints, we were only able to recruit 100 participants in this study. We recommend that future studies consider a large sample size and further build on the evidence [52,53]. Another limitation is that patients with impaired cognition and high multimorbidity risk were not included. However, we attempted to engage residential aged caregivers who received many of these patients. These caregivers certainly act as facilitators, helping their patients, and are equally important stakeholders in the care process. Our study highlighted, from the perspective of caregivers, that personalized digital health hub could improve the handover process between specialist care and residential age care. [13]. We also acknowledge that these data were collected in 2017; however, we consider the findings from the study to be unique and still relevant, on the basis of the existing literature and development in the area of broader musculoskeletal care and digital health. Nonetheless, studies demonstrate the need to substitute current inefficiencies of siloed health care models with more person-centered and integrated models in which the patients and their caregivers are empowered as a team that works toward a personalized health solution to illness [8,18,51]. We acknowledge that when this study was conducted, the concept of personalized digital health hub was very theoretical; however, with inputs from other study components, we have been able to advance it to the stage of a prototype to be tested in the practice setting in the next stage of our research activity. Another limitation offered by the design of this study was a weak component of qualitative methodology, as the primary data collection tool consisted of only 2 open-ended questions. However, despite this limitation, we attempted to maximize the relevance of findings by applying the analytical behavior change framework. Similarly, in this process, the increased use of digital technologies to support health care is inevitable, particularly in the context of COVID-19, which has not only accelerated the willingness of health care practitioners to adopt telehealth options but also resulted in patients quickly adapting to and embracing these recent changes [54-56].

Conclusions

Recovery from fragility fractures among older people requires input from multiple specialties within medicine and allied health domains depending on the presence of concurrent medical conditions. Rather than approaching patients as isolated individuals, we need to consider them in the context of a network of caregivers and delivery of service as an integrated holistic model of care. Findings from this study contributed to understanding the capabilities, motivation and opportunities of patients, family members and formal caregivers as a *patient networked unit* rather than as siloed groups and provided a proof of concept around a personalized digital health hub [8]. This will provide greater cohesion and opportunities for success while navigating through a complex recovery pathway with multiple caregivers and is critical to the development of a personalized digital health hub–enabled MoC. Future paths will also incorporate perspectives from other relevant stakeholders as part of the *patient networked unit*, evolving through iterative processes and cocreation, to improve our understanding around the successful development of complex health care interventions and its drivers [13,57].

Acknowledgments

The authors would like to acknowledge all the patients, their family members, and residential aged caregivers for their time and valuable participation in this study. Furthermore, we would also like to thank multiple frontline health care practitioners at the

Royal Adelaide Hospital for their incredible support in helping us collect the required data for this study. This study was financially supported by the National Health and Medical Research Council (NHMRC) Centre for Research Excellence (CRE) in Frailty and Healthy Aging at the Adelaide Medical School, Faculty of Health and Medical Sciences, The University of Adelaide, Australia (NHMRC grant ID-1102208). LY is supported through the Commonwealth Government of Australia Research Training Program Scholarship and Top-up scholarship by NHMRC CRE in Frailty and Healthy Aging. No external funding was received for this study.

Authors' Contributions

MJC and TKG conceptualized the initial prospective cohort study, and LY joined in the iterative discussions around the analysis and interpretation of study findings. AT and JDY facilitated data collection for this study. LY wrote the initial draft of this manuscript. The drafts with multiple versions were subsequently reviewed, and constructive inputs were provided by TKG, MJC, AT, and JDY. This final version of the manuscript incorporates comments and edits from the authors and has been approved by all.

Conflicts of Interest

None declared.

Multimedia Appendix 1
In-depth interview schedule (stakeholders).
[DOCX File , 16 KB - jmir_v23i10e26886_app1.docx]

References

1. Understanding south Australia's growing, changing population 2020. Department of Planning, Transport and Infrastructure, Government of South Australia. URL: https://www.saplanningportal.sa.gov.au/data_and_research/population [accessed 2020-02-20]
2. McDougall K, Barrie H. South Australia retirement village survey 2016. The University of Adelaide. 2017. URL: https://adelaide.figshare.com/articles/journal_contribution/South_Australia_Retirement_Village_Survey_2016/5487340 [accessed 2021-09-07]
3. National Research Council, Institute of Medicine, Committee on an Aging Society. America's Aging: Productive Roles in an Older Society. Washington, DC: The National Academies Press; 1986.
4. Fenton S, Draper H. Older people make a huge contribution to society. Some communities and faith groups draw on this contribution in responding to the needs of all their members. University of Birmingham, Birmingham Policy Commission. URL: https://www.birmingham.ac.uk/Documents/research/policycommission/healthy-ageing/7-What-older-people-and-communities-are-contributing-updated.pdf [accessed 2020-02-20]
5. Integration and participation of older persons in society. United Nations Economic Commission for Europe. 2009. URL: https://www.unece.org/fileadmin/DAM/pau/_docs/age/2009/Policy_briefs/4-Policybrief_Participation_Eng.pdf [accessed 2020-02-20]
6. Osteoporosis: Assessing the Risk of Fragility Fracture. London, UK: National Institute for Health and Care Excellence: Clinical Guidelines; 2017.
7. Khow K, McNally C, Shibu P, YU SC, Visvanathan R, Chehade MJ. Getting patients back on their feet after a hip fracture. Medicine Today. 2016. URL: https://medicinetoday.com.au/sites/default/files/cpd/MT2016-05-030-KHOW.pdf [accessed 2021-09-07]
8. Chehade MJ, Yadav L, Jayatilaka A, Gill TK, Palmer E. Personal digital health hubs for multiple conditions. Bull World Health Organ 2020 Aug 01;98(8):569-575 [FREE Full text] [doi: 10.2471/BLT.19.249136] [Medline: 32773902]
9. Riemen AH, Hutchison JD. The multidisciplinary management of hip fractures in older patients. Orthop Trauma 2016 Apr;30(2):117-122 [FREE Full text] [doi: 10.1016/j.mporth.2016.03.006] [Medline: 27418950]
10. Martin FC, Ranhoff AH. Frailty and Sarcopenia. In: Orthogeriatrics: The Management of Older Patients with Fragility Fractures. Cham: Springer; Aug 21, 2020.
11. Friedman SM, Menzies IB, Bukata SV, Mendelson DA, Kates SL. Dementia and hip fractures: development of a pathogenic framework for understanding and studying risk. Geriatr Orthop Surg Rehabil 2010 Nov;1(2):52-62 [FREE Full text] [doi: 10.1177/2151458510389463] [Medline: 23569663]
12. Yadav L, Haldar A, Jasper U, Taylor A, Visvanathan R, Chehade M, et al. Utilising digital health technology to support patient-healthcare provider communication in fragility fracture recovery: systematic review and meta-analysis. Int J Environ Res Public Health 2019 Oct 22;16(20):4047 [FREE Full text] [doi: 10.3390/ijerph16204047] [Medline: 31652597]
13. Yadav L, Gill TK, Taylor A, Jasper U, De Young J, Visvanathan R, et al. Cocreation of a digital patient health hub to enhance education and person-centred integrated care post hip fracture: a mixed-methods study protocol. BMJ Open 2019 Dec 18;9(12):e033128 [FREE Full text] [doi: 10.1136/bmjopen-2019-033128] [Medline: 31857315]

14. Cooper C, Campion G, Melton LJ. Hip fractures in the elderly: a world-wide projection. Osteoporos Int 1992 Nov;2(6):285-289. [doi: 10.1007/BF01623184] [Medline: 1421796]
15. Gullberg B, Johnell O, Kanis J. World-wide projections for hip fracture. Osteoporos Int 1997;7(5):407-413. [doi: 10.1007/pl00004148] [Medline: 9425497]
16. Johnell O, Kanis JA. An estimate of the worldwide prevalence, mortality and disability associated with hip fracture. Osteoporos Int 2004 Nov;15(11):897-902. [doi: 10.1007/s00198-004-1627-0] [Medline: 15490120]
17. Mitchell R, Draper B, Harvey L, Wadolowski M, Brodaty H, Close J. Comparison of hospitalised trends, treatment cost and health outcomes of fall-related hip fracture for people aged ≥ 65 years living in residential aged care and the community. Osteoporos Int 2019 Feb;30(2):311-321. [doi: 10.1007/s00198-018-4800-6] [Medline: 30569228]
18. Dziedzic K, French S, Davis A, Geelhoed E, Porcheret M. Implementation of musculoskeletal models of care in primary care settings: theory, practice, evaluation and outcomes for musculoskeletal health in high-income economies. Best Pract Res Clin Rheumatol 2016 Jun;30(3):375-397 [FREE Full text] [doi: 10.1016/j.berh.2016.08.004] [Medline: 27886938]
19. Kelly PA, Haidet P. Physician overestimation of patient literacy: a potential source of health care disparities. Patient Educ Couns 2007 Apr;66(1):119-122. [doi: 10.1016/j.pec.2006.10.007] [Medline: 17140758]
20. Rogers ES, Wallace LS, Weiss BD. Misperceptions of medical understanding in low-literacy patients: implications for cancer prevention. Cancer Control 2006 Jul;13(3):225-229 [FREE Full text] [doi: 10.1177/107327480601300311] [Medline: 16885919]
21. Allegrante JP, Peterson MG, Cornell CN, MacKenzie CR, Robbins L, Horton R, et al. Methodological challenges of multiple-component intervention: lessons learned from a randomized controlled trial of functional recovery after hip fracture. HSS J 2007 Feb;3(1):63-70 [FREE Full text] [doi: 10.1007/s11420-006-9036-x] [Medline: 18751772]
22. World Report on Ageing and Health. Geneva, Switzerland: World Health Organization; 2015.
23. Integrated Care for Older People: Guidelines on Community-Level Interventions to Manage Declines in Intrinsic Capacity. Geneva, Switzerland: World Health Organization; 2017.
24. Jin W, Kim DH. Design and implementation of e-Health system based on semantic sensor network using IETF YANG. Sensors (Basel) 2018 Feb 20;18(2):629 [FREE Full text] [doi: 10.3390/s18020629] [Medline: 29461493]
25. WHO guideline: recommendations on digital interventions for health system strengthening. World Health Organization. 2019. URL: https://apps.who.int/iris/bitstream/handle/10665/324998/WHO-RHR-19.7-eng.pdf [accessed 2020-02-10]
26. Atlas of EHealth Country Profiles: The Use of EHealth in Support of Universal Health Coverage: Based on the Findings of the Third Global Survery on EHealth 2015. Geneva, Switzerland: World Health Organization; 2016.
27. WHO Compendium of Innovative Health Technologies for Low Resource Settings, 2011-2014: Assistive Devices, ehealth Solutions, Medical Devices, Other Technologies, Technologies for Outbreaks. Geneva, Switzerland: World Health Organization; 2011.
28. Harwell M. Research design in qualitative/quantitative/mixed methods. In: The SAGE Handbook for Research in Education: Pursuing Ideas as the Keystone of Exemplary Inquiry. Thousand Oaks, CA: Sage; 2011.
29. Taylor A. A day in the life of orthopaedic and trauma practitioners: orthopaedic nurse practitioner - hip fracture. Int J Orthop Trauma Nurs 2015 Feb;19(1):50-51. [doi: 10.1016/j.ijotn.2014.11.003] [Medline: 25787817]
30. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. Int J Qual Methods 2006 Mar 01;5(1):80-92 [FREE Full text] [doi: 10.1177/160940690600500107]
31. Rath S, Yadav L, Tewari A, Chantler T, Woodward M, Kotwal P, et al. Management of older adults with hip fractures in India: a mixed methods study of current practice, barriers and facilitators, with recommendations to improve care pathways. Arch Osteoporos 2017 Dec;12(1):55 [FREE Full text] [doi: 10.1007/s11657-017-0344-1] [Medline: 28577286]
32. Yadav L, Tewari A, Jain A, Essue B, Peiris D, Woodward M, et al. Protocol-based management of older adults with hip fractures in Delhi, India: a feasibility study. Pilot Feasibility Stud 2016 Mar 9;2:15 [FREE Full text] [doi: 10.1186/s40814-016-0056-0] [Medline: 27965835]
33. Duffield SJ, Ellis BM, Goodson N, Walker-Bone K, Conaghan PG, Margham T, et al. The contribution of musculoskeletal disorders in multimorbidity: implications for practice and policy. Best Pract Res Clin Rheumatol 2017 Apr;31(2):129-144 [FREE Full text] [doi: 10.1016/j.berh.2017.09.004] [Medline: 29224692]
34. Araujo de Carvalho I, Epping-Jordan J, Pot AM, Kelley E, Toro N, Thiyagarajan JA, et al. Organizing integrated health-care services to meet older people's needs. Bull World Health Organ 2017 Nov 01;95(11):756-763 [FREE Full text] [doi: 10.2471/BLT.16.187617] [Medline: 29147056]
35. Loft MI, Martinsen B, Esbensen BA, Mathiesen LL, Iversen HK, Poulsen I. Strengthening the role and functions of nursing staff in inpatient stroke rehabilitation: developing a complex intervention using the Behaviour Change Wheel. Int J Qual Stud Health Well-being 2017;12(sup2):1392218 [FREE Full text] [doi: 10.1080/17482631.2017.1392218] [Medline: 29088984]
36. Evangelidis N, Craig J, Bauman A, Manera K, Saglimbene V, Tong A. Lifestyle behaviour change for preventing the progression of chronic kidney disease: a systematic review. BMJ Open 2019 Oct 28;9(10):e031625 [FREE Full text] [doi: 10.1136/bmjopen-2019-031625] [Medline: 31662393]

37. Munir F, Biddle S, Davies M, Dunstan D, Esliger D, Gray LJ, et al. Stand More AT Work (SMArT Work): using the behaviour change wheel to develop an intervention to reduce sitting time in the workplace. BMC Public Health 2018 Mar 06;18(1):319 [FREE Full text] [doi: 10.1186/s12889-018-5187-1] [Medline: 29510715]
38. de Haes H, Bensing J. Endpoints in medical communication research, proposing a framework of functions and outcomes. Patient Educ Couns 2009 Mar;74(3):287-294. [doi: 10.1016/j.pec.2008.12.006] [Medline: 19150197]
39. Bolle S, Romijn G, Smets EM, Loos EF, Kunneman M, van Weert JC. Older cancer patients' user experiences with web-based health information tools: a think-aloud study. J Med Internet Res 2016 Jul 25;18(7):e208 [FREE Full text] [doi: 10.2196/jmir.5618] [Medline: 27457709]
40. Slater H, Campbell JM, Stinson JN, Burley MM, Briggs AM. End user and implementer experiences of mHealth technologies for noncommunicable chronic disease management in young adults: systematic review. J Med Internet Res 2017 Dec 12;19(12):e406 [FREE Full text] [doi: 10.2196/jmir.8888] [Medline: 29233804]
41. Majeed-Ariss R, Baildam E, Campbell M, Chieng A, Fallon D, Hall A, et al. Apps and adolescents: a systematic review of adolescents' use of mobile phone and tablet apps that support personal management of their chronic or long-term physical conditions. J Med Internet Res 2015 Dec 23;17(12):e287 [FREE Full text] [doi: 10.2196/jmir.5043] [Medline: 26701961]
42. Ullrich PF, Vaccaro AR. Patient education on the internet: opportunities and pitfalls. Spine 2002 Apr 01;27(7):E185-E188. [doi: 10.1097/00007632-200204010-00019] [Medline: 11923675]
43. Musculoskeletal health network, models of care and other resources. Government of Western Australia Department of Health. URL: https://ww2.health.wa.gov.au/Articles/J_M/Musculoskeletal-Health-Network [accessed 2020-02-20]
44. National pain strategy. Painaustralia. URL: https://www.painaustralia.org.au/improving-policy/national-pain-strategy [accessed 2020-02-20]
45. Household use of information technology, Australia, 2016-17. The Australian Bureau of Statistics. 2018. URL: https://www.abs.gov.au/ausstats/abs@.nsf/mf/810 [accessed 2020-02-20]
46. Access to hospitals: admitted patient care-socioeconomic status. Australian Institute of Health and Welfare. URL: https://www.aihw.gov.au/reports-data/myhospitals/themes/hospital-access [accessed 2020-12-14]
47. de Veer AJ, Peeters JM, Brabers AE, Schellevis FG, Rademakers JJ, Francke AL. Determinants of the intention to use e-Health by community dwelling older people. BMC Health Serv Res 2015 Mar 15;15(1):103. [doi: 10.1186/s12913-015-0765-8] [Medline: 25889884]
48. My health record failed to manage cybersecurity and privacy risks, audit finds. The Guardian. 2019 Nov 25. URL: https://www.theguardian.com/australia-news/2019/nov/25/my-health-record-failed-to-manage-cybersecurity-and-privacy-risks-audit-finds [accessed 2020-02-20]
49. Taking the pulse of patient experience: why australian healthcare must change. Forrester Opportunity Snapshot: A Custom Study Commissioned by SFDC. 2019. URL: https://c1.sfdcstatic.com/content/dam/web/en_au/www/documents/pdf/patient-experience-australian-healthcare-change.pdf [accessed 2020-02-22]
50. Chehade M, Gill T, Kopansky-Giles D, Schuwirth L, Karnon J, McLiesh P, et al. Building multidisciplinary health workforce capacity to support the implementation of integrated, people-centred Models of Care for musculoskeletal health. Best Pract Res Clin Rheumatol 2016 Jun;30(3):559-584. [doi: 10.1016/j.berh.2016.09.005] [Medline: 27886946]
51. Chehade M, Yadav L, Kopansky-Giles D, Merolli M, Palmer E, Jayatilaka A, et al. Innovations to improve access to musculoskeletal care. Best Pract Res Clin Rheumatol 2020 Oct;34(5):101559. [doi: 10.1016/j.berh.2020.101559] [Medline: 32718885]
52. Rothman K, Greenland S, Lash T. Modern Epidemiology. Philadelphia, PA, USA: Wolters Kluwer Health; 2008.
53. Kirkwood B, Sterne J. Essential Medical Statistics. Oxford, UK: Wiley-Blackwell; 2003.
54. Hollander J, Carr B. Virtually perfect? Telemedicine for covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
55. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
56. Garg SK, Rodbard D, Hirsch IB, Forlenza GP. Managing new-onset type 1 diabetes during the covid-19 pandemic: challenges and opportunities. Diabetes Technol Ther 2020 Jun;22(6):431-439. [doi: 10.1089/dia.2020.0161] [Medline: 32302499]
57. Turner KM, Rousseau N, Croot L, Duncan E, Yardley L, O'Cathain A, et al. Understanding successful development of complex health and healthcare interventions and its drivers from the perspective of developers and wider stakeholders: an international qualitative interview study. BMJ Open 2019 May 30;9(5):e028756 [FREE Full text] [doi: 10.1136/bmjopen-2018-028756] [Medline: 31152042]

Abbreviations

BCW: Behavior Change Wheel
MoC: model of care
OR: odds ratio
WHO: World Health Organization

Edited by R Kukafka; submitted 01.01.21; peer-reviewed by N Milevska-Kostova, C Richardson, I Wilson, LA Gomes; comments to author 20.02.21; revised version received 26.05.21; accepted 14.06.21; published 28.10.21.

Please cite as:
Yadav L, Gill TK, Taylor A, De Young J, Chehade MJ
Identifying Opportunities, and Motivation to Enhance Capabilities, Influencing the Development of a Personalized Digital Health Hub Model of Care for Hip Fractures: Mixed Methods Exploratory Study
J Med Internet Res 2021;23(10):e26886
URL: https://www.jmir.org/2021/10/e26886
doi:10.2196/26886
PMID:34709183

Original Paper

The Association Between Professional Accounts on Social Networks Twitter and ResearchGate and the Number of Scientific Publications and Citations Among Anesthesia Researchers: Observational Study

Thomas Clavier[1], MD, PhD; Emilie Occhiali[1], MD; Zoé Demailly[1], MD; Vincent Compère[1], MD, PhD; Benoit Veber[1], MD, PhD; Jean Selim[1], MD; Emmanuel Besnier[1], MD, PhD

Department of Anesthesiology, Critical Care and Perioperative Medicine, Rouen University Hospital, Rouen, France

Corresponding Author:
Thomas Clavier, MD, PhD
Department of Anesthesiology, Critical Care and Perioperative Medicine
Rouen University Hospital
1 rue de Germont
Rouen, 76000
France
Phone: 33 232881733
Email: thomasclavier76@gmail.com

Abstract

Background: Social networks are now essential tools for promoting research and researchers. However, there is no study investigating the link between presence or not on professional social networks and scientific publication or citation for a given researcher.

Objective: The objective of this study was to study the link between professional presence on social networks and scientific publications/citations among anesthesia researchers.

Methods: We included all the French full professors and associate professors of anesthesia. We analyzed their presence on the social networks Twitter (professional account with ≥1 tweet over the 6 previous months) and ResearchGate. We extracted their bibliometric parameters for the 2016-2020 period via the Web of Science Core Collection (Clarivate Analytics) database in the Science Citation Index-Expanded index.

Results: A total of 162 researchers were analyzed; 42 (25.9%) had an active Twitter account and 110 (67.9%) a ResearchGate account. There was no difference between associate professors and full professors regarding active presence on Twitter (8/23 [35%] vs. 34/139 [24.5%], respectively; *P*=.31) or ResearchGate (15/23 [65%] vs. 95/139 [68.3%], respectively; *P*=.81). Researchers with an active Twitter account (median [IQR]) had more scientific publications (45 [28-61] vs. 26 [12-41]; *P*<.001), a higher h-index (12 [8-16] vs. 8 [5-11]; *P*<.001), a higher number of citations per publication (12.54 [9.65-21.8] vs. 10.63 [5.67-16.10]; *P*=.01), and a higher number of citations (563 [321-896] vs. 263 [105-484]; *P*<.001). Researchers with a ResearchGate account (median [IQR]) had more scientific publications (33 [17-47] vs. 26 [9-43]; *P*=.03) and a higher h-index (9 [6-13] vs. 8 [3-11]; *P*=.03). There was no difference between researchers with a ResearchGate account and those without it concerning the number of citations per publication and overall number of citations. In multivariate analysis including sex, academic status, and presence on social networks, the presence on Twitter was associated with the number of publications (β=20.2; *P*<.001), the number of citations (β=494.5; *P*<.001), and the h-index (β=4.5; *P*<.001).

Conclusions: Among French anesthesia researchers, an active presence on Twitter is associated with higher scientific publication and citations.

(J Med Internet Res 2021;23(10):e29809) doi:10.2196/29809

KEYWORDS

social network; anesthesia; publication; Twitter; ResearchGate; citation; social media; academic; researcher; bibliometrics; research output

Introduction

In a globalized world, social networks have now taken a major place in the scientific field and are essential tools for promoting research and researchers (including the recruitment or the promotion of a researcher) [1-3]. Twitter is a microblogging service that allows its users to blog through short messages (ie, tweets). Each user can "retweet" a tweet from another user and broadcast it to his/her own followers, thus allowing fast dissemination of content. Because of its concise and synthetic nature, and given the possibility to follow specific thematic accounts, it is a social network used for professional purposes by many researchers and physicians [4]. ResearchGate is a social network reserved for researchers, used to promote their works and to connect with people working in the same field of research. It is thus a social network only dedicated to professional use for the international scientific community.

Medical journals use Twitter to increase their visibility within the scientific community. It is by far the most used social network to share publications because more than 20% of published articles receive at least one announcement on Twitter (compared with less than 5% of notifications on other nonprofessional social networks) [5]. It has been recently described that, among a selection of authors publishing in anesthesia journals, 24% have a Twitter account and 72% have a ResearchGate account [6]. Moreover, among French anesthesia, intensive care, and emergency medicine health care workers, 46% use social media to obtain information about medical actuality and 17% consult Twitter at least once a week [7]. This professional social network use should increase with the arrival of younger physicians because it was reported in a single-center study that 35% of medical students used Twitter for teaching purposes [8]. The link between social and usage metrics (altmetrics) and traditional bibliometric indicators is weak and variable, but Twitter's altmetrics indicators seem to perform well in predicting the actual citation rate [9,10]. Twitter users tweet the articles they write and it is known that tweets can predict highly cited articles within the first days of an article publication [10-12]. Finally, recent randomized studies showed that, for a given journal, articles that benefited from exposure on Twitter were more cited than articles that were not tweeted [13,14]. However, there is no study investigating the link between presence or not on professional social networks and scientific publication or citation for a given researcher.

The objective of this study was to study the link between presence on social networks and scientific publication and citation among anesthesia researchers.

Methods

Study Design and Population

We used publicly available data; as a retrospective analysis that did not involve human participants (and in accordance with French laws), this study was exempt from institutional ethics board review [15].

We included all French physicians with an academic function of teaching and research in anesthesia (full professors and associate professors from the 48-01 subsection of the French National Council of Universities, Directory of Members for the year 2019).

Objectives

The main objective of this work was to compare the scientific publication and citation of French anesthesia researchers according to the presence or absence of an active Twitter account. The secondary objectives were:

- to compare, in the same population, the scientific publication and citation according to the presence or absence on ResearchGate;
- to assess if the presence or absence from Twitter and ResearchGate was associated with scientific publication and citation of researchers.

Data Extraction

To limit the impact of profile variations on social networks and publication citations in the bibliometric database, the entire data collection was carried out manually over 10 consecutive days in March 2021.

We analyzed the presence of included researchers on the social networks Twitter (professional account, that is, at least one follow of a profile related to anesthesia or intensive care medicine) and ResearchGate. The screening for finding social network accounts followed a step-by-step procedure:

- The first and last name were entered into the social network search engine. On Twitter, the author was searched on the account search tool and also on the "TOP" (most relevant Tweets for a given search) and "LATEST" (the most recently posted Tweets matching a given search) tabs;
- If no author was found after this first search, only the last name was used in association with the following keywords: "Dr", "Pr", "Anesthésie", "Réanimation" (French keywords), "Anesthesia", and "Intensive care";
- If no author was still found, the first and last name were entered into the Google search engine with the keyword "Twitter" or "ResearchGate";
- If several accounts were found for a given name, all accounts were manually analyzed in search of information on the account, to identify whether or not it was the researcher's account (particularly through her/his hospital and/or academic affiliation).
- When an account was found, the following data were collected:
  - For Twitter: existence of a professional Twitter profile (an account was considered active if it has published at least one professional tweet over the 6 months preceding the data collection date). For each active Twitter account, the recorded data were presence or not of a photograph, information concerning profession (academic or anesthetist), number of tweets, number of followers, and date of creation of the account (to determine the number of tweets and followers by month spent on Twitter);
  - For ResearchGate, existence of an active ResearchGate profile (with a least one research work documented); if there is an existing profile, the following data were

collected: presence or not of a photograph, number of followers, RG score (which is a measure of scientific reputation on ResearchGate), and Total Research Interest score (which is linked to the reading, citation, and recommendation of the researcher's work on ResearchGate).

- Bibliometric parameters were extracted from the Web of Science Core Collection (Clarivate Analytics) database in the Science Citation Index-Expanded index. To limit the risk of errors due to homonyms in other research disciplines, we used a search algorithm focused on medical specialties that correspond to the fields of activity of anesthetists in France. Thus, the analysis focused on the publication of reviews, original articles, and editorials in medical journals over the period 2016-2020 on the advanced search tool of Web of Science with the following search formula:

(SU=CRITICAL CARE MEDICINE OR SU=ANESTHESIOLOGY OR SU=SURGERY OR SU=EMERGENCY MEDICINE OR SU=ALLERGY OR SU=CARDIAC & CARDIOVASCULAR SYSTEMS OR SU=CLINICAL NEUROLOGY OR SU=ENDOCRINOLOGY & METABOLISM OR SU=GASTROENTEROLOGY & HEPATOLOGY OR SU=HEMATOLOGY OR SU=IMMUNOLOGY OR SU=INFECTIOUS DISEASES OR SU=MEDICAL INFORMATICS OR SU=MEDICINE, GENERAL & INTERNAL OR SU=MEDICINE, RESEARCH & EXPERIMENTAL OR SU=MULTIDISCIPLINARY SCIENCES OR SU=NURSING OR SU=NUTRITION & DIETETICS OR SU=OBSTETRICS & GYNECOLOGY OR SU=PERIPHERAL VASCULAR DISEASE OR SU=PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH OR SU=RESPIRATORY SYSTEM OR SU=TRANSPLANTATION OR SU=TOXICOLOGY) AND (PY=2016 OR PY=2017 OR PY=2018 OR PY=2019 OR PY=2020) AND (DT=ARTICLE OR DT=REVIEW OR DT=EDITORIAL MATERIAL) AND AU="NAME OF THE AUTHOR, surname of the author"

For each researcher, the following parameters were recorded: number of publications, h-index, number of citations per publication, and overall number of citations.

Statistical Analysis

The values are presented as n (%) for qualitative variables, and as median (IQR) for quantitative variables. The quantitative variables were compared using a Mann-Whitney *U* test. The qualitative variables were analyzed using a Fisher test. The Pearson correlation test was used to assess the strength of association between 2 quantitative variables. A multivariate analysis using a linear regression model was realized to identify whether the presence or absence from Twitter and ResearchGate was related to the number of citations, the h-index, and the number of publications. The multivariate analysis included the following variables: presence on Twitter, presence on ResearchGate, sex, and academic status (full professor or associate professor). All statistical tests were 2-sided and the .05 probability level was used to establish statistical significance. All statistics and graphs were produced using GraphPad PRISM software (version 9.1.2; GraphPad Software).

Data Availability Statement

The raw data supporting the conclusions of this manuscript can be made available on request by the authors to any qualified researcher.

Results

Population Description

Of the 162 researchers analyzed (147 men and 15 women), 68 (42.0%) had a Twitter account, of which 42 (25.9%) were considered active, and 110 (67.9%) had a ResearchGate account. A total of 36 (22.2%) researchers had both an active Twitter account and a ResearchGate account. The characteristics of the Twitter and ResearchGate accounts identified are presented in Table 1.

There was no difference between associate professor and full professor regarding active presence on Twitter (8/23 (35%) vs. 34/139 (24.5%), respectively; *P*=.31) or on ResearchGate (15/23 (65%) vs. 95/139 (68.3%), respectively; *P*=.81).

Table 1. Characteristics of the Twitter and ResearchGate accounts.

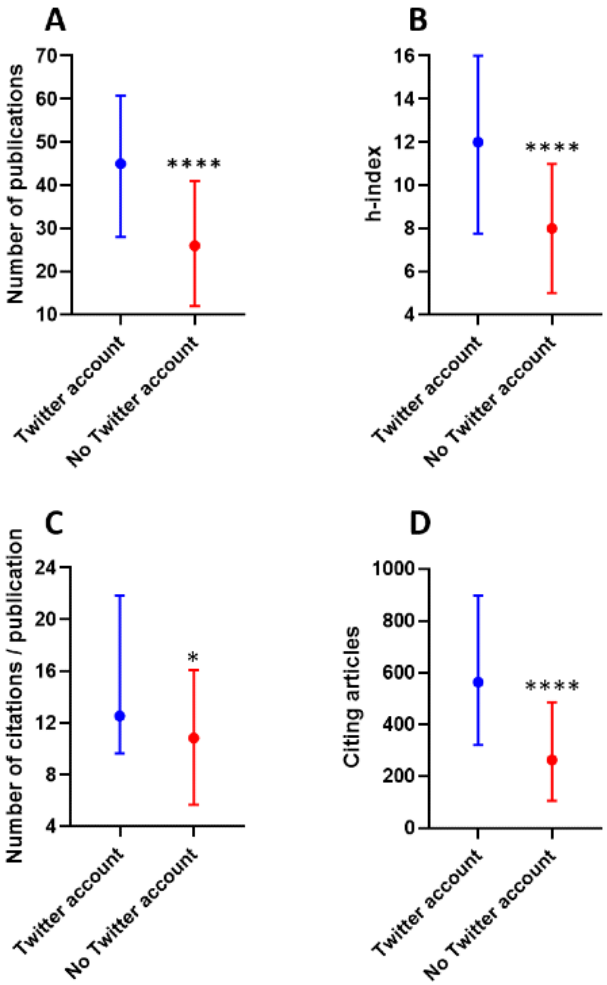
Characteristics	Values[a]
Active Twitter accounts (n=42)	
Specification of an anesthetist or academic function	25 (59.5)
Photograph identifying the account owner	30 (71.4)
Number of tweets/month	4.0 (1.5-11.0)
Number of followers/month	4.4 (1.2-8.6)
ResearchGate accounts (n=110)	
Photograph identifying the account owner	65 (59.1)
RG score	39 (36-42)
Total Research Interest score	1415 (779-2244)
Number of followers	90 (50-168)

[a]Values are presented as n (%) for qualitative variables and as median (IQR) for quantitative variables.

Scientific Publication and Citation According to the Presence or Absence of an Active Twitter Account

Over the period 2016-2020, researchers with an active Twitter account had more (median [IQR]) scientific publications (45 [28-61] vs. 26 [12-41]; *P*<.001; Figure 1A), a higher h-index (12 [8-16] vs. 8 [5-11]; *P*<.001; Figure 1B), a higher number of citations per publication (12.54 [9.65-21.8] vs. 10.63 [5.67-16.10]; *P*=.01; Figure 1C), and a higher number of citations (563 [321-896] vs. 263 [105-484]; *P*<.001; Figure 1D).

Figure 1. Number of scientific publications (A), h-index (B), number of citations per publication (C), and number of citing articles (D) over the period 2016-2020 among researchers with an active Twitter account. Data are presented as as median with interquartile range. *, *P*<.05; ****, *P*<.0001.

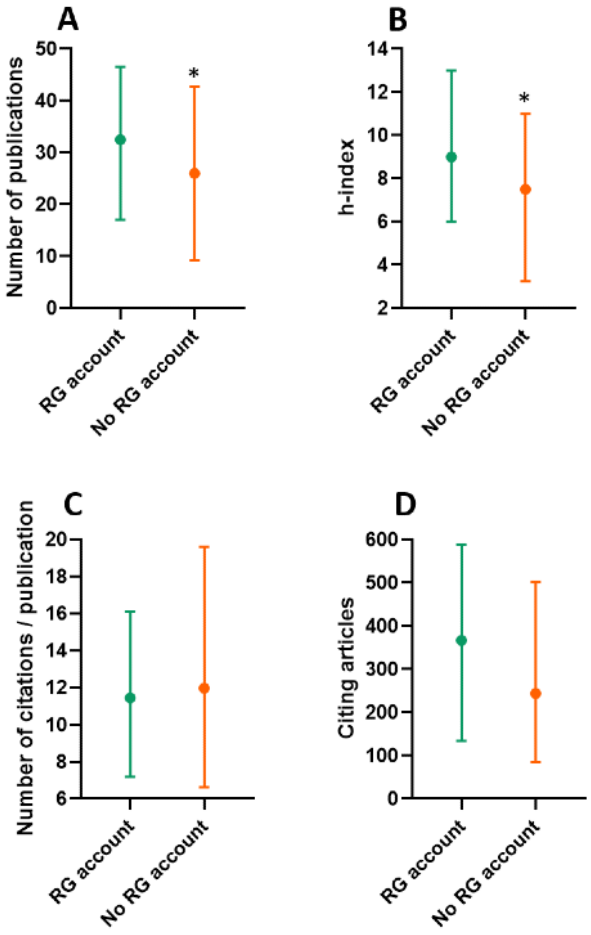


Among researchers with an active Twitter account, there was a correlation between the number of tweets/month and the number of followers/month (r=0.69; 95% CI 0.49-0.82; P<.001), the number of publications (r=0.37; 95% CI 0.08-0.61; P=.02), and the h-index (r=0.37; 95% CI 0.08-0.61; P=.02). There was also a correlation between the number of followers/month and the number of publications (r=0.43; 95% CI 0.14-0.64; P<.01) and the h-index (r=0.47; 95% CI 0.20-0.68; P<.01). There was no correlation between the number of tweets/month or the number of followers/month and the number of citations per publication or the overall number of citations.

Scientific Publication and Citation According to the Presence or Absence of a ResearchGate Account

Over the period 2016-2020, researchers with a ResearchGate account had more (median [IQR]) scientific publications (33 [17-47] vs. 26 [9-43]; P=.03; Figure 2A) and a higher h-index (9 [6-13] vs. 8 [3-11]; P=.03; Figure 2B). There was no difference (median [IQR]) between researchers with a ResearchGate account and those without it concerning the number of citations per publication (11.45 [7.19-21.8] vs. 11.98 [6.62-19.62]; P=.67; Figure 2C) and the overall number of citations (367 [134-589] vs. 244 [85-502]; P=.17; Figure 2D).

Figure 2. Number of scientific publications (A), h-index (B), number of citations per publication (C), and number of citing articles (D) over the period 2016-2020 among researchers with a ResearchGate (RG) account. Data are presented as as median with interquartile range. *, P<.05.



Among researchers with a ResearchGate account, there was a correlation between the number of followers and the number of scientific publications (r=0.78; 95% CI 0.70-0.85; P<.001), the h-index (r=0.72; 95% CI 0.62-0.80; P<.001), the number of citations per publication (r=0.34; 95% CI 0.17-0.50; P<.001), and the overall number of citations (r=0.65; 95% CI 0.53-0.75; P<.001).

Multivariate Analysis

In multivariate analysis, the presence on Twitter (but not on ResearchGate) was associated with the number of publications, the number of citations, and the h-index (Table 2).

Table 2. Multiple linear regression model to predict the number of citations, the h-index, and the number of publications.

Variable	Number of publications			Number of citations			h-index		
	β	95% CI	*P* value	β	95% CI	*P* value	β	95% CI	*P* value
Sex	11.8	–4.3 to 27.9	.15	236.8	–157.7 to 631.2	.24	2.2	–0.8 to 5.3	.15
Active presence on Twitter	20.2	9.2 to 31.2	<.001	494.5	224.8 to 764.1	<.001	4.5	2.4 to 6.6	<.001
Presence on ResearchGate	2.1	–8.2 to 12.3	.69	–158.8	–410.5 to 92.9	.21	0.45	–1.5 to 2.4	.65
Status	12.3	–1.9 to 26.6	.09	184.1	–165.9 to 534.2	.30	1.3	–1.5 to 4.0	.36

Discussion

Preliminary Findings

To our knowledge, we describe for the first time the participation rates and professional use patterns of social networks among all academics from a medical specialty of a given country. We also explore for the first time the link between researchers' medical publication activity and their presence on social networks.

The 2 social networks analyzed in this work have different use in professional life. Twitter, which is not a network designed solely for professional use, allows subscribers to give their opinion, to follow some influencers, and also to exchange personal information. ResearchGate is reserved for researchers, and is used to promote their work and to connect with people working in the same field of research. Twitter is a network that requires active and frequent participation to disseminate information, whereas ResearchGate automatically imports publications from authors (who only has to validate them) and gives visibility to researchers with a high RG score, even if they are not very active on the network, allowing a more passive management of the account once it has been created. This may partially explain why the rate of researchers with a ResearchGate account is higher than that of researchers with an active Twitter account. Another likely explanation is that as ResearchGate is a social network specifically dedicated to research, it makes more sense for a researcher to be present there than to have a professional account on a mainstream network such as Twitter. The rates of use of professional social networks (Twitter: 42% (68/162) and ResearchGate: 67.9% (110/162)) are higher than those recently described among ENT surgeons, neurosurgeons, or pediatric orthopedists (Twitter: 2%-13% and ResearchGate: 23%-36%) [16-18]. The presence on Twitter of French anesthesia researchers also appears to be greater than that of nonacademic professionals from the same country working in the same field (17%) [7]. Nevertheless, the fact that we only included academic scientific authors in our analysis, versus physicians without academic activity in these other works, probably explains this difference. Thus, our rate of active Twitter users among researchers is closer to those recently described among health policy and health services researchers (30%) or among researchers publishing in anesthesia journals (between 22% and 25%) [6,19]. Similarly, the rate of use of ResearchGate appears similar to those described among academic researchers publishing in medical journals (between 45% and 70%) [6,20,21]. We can therefore assume that our data collection was relatively exhaustive in the study population. It is interesting to note that there was no difference in social networks between full professors and associate professors. This result should be interpreted with caution as there are few associate professors in our cohort and therefore probably a great lack of statistical power for this analysis. One explanation could be that, while associate professors belong to a generation more aware of social networks, full professors have had more time in their academic career to discover and use social networks for professional purposes. A significant proportion of Twitter and ResearchGate users do not have a photograph or description (or both in some cases) of their profession on their accounts. This may likely reduce their visibility on these social networks.

Among anesthesia researchers, an active presence on Twitter is associated with better bibliometric parameters, in terms of both number of publications and citations. The same trend is observed on ResearchGate. However, our study methodology does not allow to establish a causal link between scientific activity and presence on social networks. There are several possible explanations for the link we have identified. It is possible that very prolific researchers have the desire to disseminate their numerous publications and are therefore more inclined to use social networks. It is also possible that being on a professional social network allows to widely disseminate publications to the connected scientific community and thus have a greater chance of being read and cited. By analogy with the studies conducted on medical journals and showing the effectiveness of Twitter in increasing citations, it may seem logical that a researcher who posts his/her new publications on Twitter would also have more citations at an individual level [4,13,14]. It is also possible that sharing information on Twitter creates links and networks between some researchers who would be more likely to work together and thus increase their overall research activity and citation of each other. However, all these suggested explanations are only hypothetical and future studies seem necessary to explore a possible causal link between professional presence on social networks and scientific activity.

Limitations

Despite interesting results, our study has several limitations. First, this work was limited to French anesthesia researchers on a 5-year period. Given the academic organization of the medical professions in France (which is centralized by the National Council of Universities), it is easy to have exhaustive access to the list of all academics in a given specialty. In other countries with a more decentralized academic organization (in particular, Anglo-Saxon countries), it is more complex to compile an exhaustive list of all academics in a country. In addition, social networks evolve very quickly (subscription/unsubscription, new followers/loss of followers, etc.), so it was essential to use data

collected over a short period. However, because of the manual standardized procedure needed to detect all the accounts (which is time-consuming), the number of researchers that could be included was limited. Nevertheless, it seems necessary to confirm our observations in the academics of other countries. We chose the 2016-2020 period to analyze bibliometrics parameters because it corresponds to the period where social networks were started to be used in a massive way to promote research in France. It did not seem relevant for us to go further back in time, as the use of social networks was less common and the population studied would have been too disparate (old researchers who had retired or young researchers who had not yet started their academic activity). Second, even a standardized manual account search procedure has its flaws; some authors may use a pseudonym, a diminutive, or misspell their name when they register, etc. Third, some bibliometric parameters of ResearchGate such as RG score and total Research Interest score were not analyzed in our work. However, these scores are based on the number of publications and citations of researchers and have already been shown to be correlated with the h-index, the number of citations, the number of publication, and the academic level of registered researchers [9,18,22,23]. Thus, it appeared futile to search for correlation between these scores and the bibliometric parameters found on Web of Sciences. Moreover, ResearchGate metric parameters are controversial to assess research output of researchers, especially the RG score, which is built from both activity related to asking and answering questions on the website and not just from publication metrics [24]. We therefore did not consider it relevant to include these scores in our analysis. Fourth, we did not analyze other famous social networks (eg, Instagram, Facebook), but because these networks are mainly dedicated to recreational use and are rarely used by physicians in a professional context, it did not seem relevant for us to include them in a study dedicated to professional impact of social networks. We also did not include LinkedIn in our analysis. Although it is a professional social network, it gives little access to profile data: no quantification of the number of posts and no precise quantification of the number of relationships beyond 500 relationships. It therefore seemed difficult to have relevant markers for this network.

Conclusion

Among French anesthesia researchers, an active presence on Twitter is associated with higher scientific publication and citations. Future studies could explore a possible causal link between these parameters.

Acknowledgments

The authors are grateful to Gaetan Kerdelhue (Rouen University Hospital) for his help in the bibliometric analysis methodology. Support for this work was provided solely from departmental sources.

Authors' Contributions

TC was involved in the study conception and design, acquisition of data, analysis and interpretation of data, and manuscript draft preparation. EO, ZD, and JS were involved in acquisition of data, interpretation of data, and manuscript draft preparation. VC, BV, and EB were involved in the study conception and design, interpretation of data, and manuscript revision. All authors read and approved the final manuscript. The corresponding author had full access to all of the data in the study and had final responsibility for the decision to submit for publication.

Conflicts of Interest

None declared.

References

1. Hawkins CM, Hunter M, Kolenic GE, Carlos RC. Social Media and Peer-Reviewed Medical Journal Readership: A Randomized Prospective Controlled Trial. J Am Coll Radiol 2017 May;14(5):596-602. [doi: 10.1016/j.jacr.2016.12.024] [Medline: 28268163]
2. Buckarma EH, Thiels CA, Gas BL, Cabrera D, Bingener-Casey J, Farley DR. Influence of Social Media on the Dissemination of a Traditional Surgical Research Article. J Surg Educ 2017;74(1):79-83. [doi: 10.1016/j.jsurg.2016.06.019] [Medline: 27993626]
3. Boudry C, Bouchard A. [Role of academic social networks in disseminating the scientific production of researchers in biology/medicine: the example of ResearchGate]. Med Sci (Paris) 2017;33(6-7):647-652 [FREE Full text] [doi: 10.1051/medsci/20173306023] [Medline: 28990568]
4. Clavier T, Besnier E, Blet A, Boisson M, Sigaut S, Frasca D, et al. A communication strategy based on Twitter improves article citation rate and impact factor of medical journals. Anaesth Crit Care Pain Med 2020 Dec;39(6):745-746. [doi: 10.1016/j.accpm.2020.11.001] [Medline: 33308590]
5. Haustein S, Costas R, Larivière V. Characterizing social media metrics of scholarly papers: the effect of document properties and collaboration patterns. PLoS One 2015;10(3):e0120495 [FREE Full text] [doi: 10.1371/journal.pone.0120495] [Medline: 25780916]

6. Demailly Z, Brulard G, Selim J, Compère V, Besnier E, Clavier T. Gender differences in professional social media use among anaesthesia researchers. Br J Anaesth 2020 Mar;124(3):e178-e184 [FREE Full text] [doi: 10.1016/j.bja.2019.12.030] [Medline: 31987471]
7. Clavier T, Popoff B, Selim J, Beuzelin M, Roussel M, Compere V, et al. Association of Social Network Use With Increased Anxiety Related to the COVID-19 Pandemic in Anesthesiology, Intensive Care, and Emergency Medicine Teams: Cross-Sectional Web-Based Survey Study. JMIR Mhealth Uhealth 2020 Sep 24;8(9):e23153 [FREE Full text] [doi: 10.2196/23153] [Medline: 32924946]
8. AlFaris E, Irfan F, Ponnamperuma G, Jamal A, Van der Vleuten C, Al Maflehi N, et al. The pattern of social media use and its association with academic performance among medical students. Med Teach 2018 Sep;40(sup1):S77-S82. [doi: 10.1080/0142159X.2018.1465536] [Medline: 29732945]
9. Ortega JL. Relationship between altmetric and bibliometric indicators across academic social sites: The case of CSIC's members. Journal of Informetrics 2015 Jan;9(1):39-49. [doi: 10.1016/j.joi.2014.11.004]
10. Thelwall M, Haustein S, Larivière V, Sugimoto CR. Do altmetrics work? Twitter and ten other social web services. PLoS One 2013;8(5):e64841 [FREE Full text] [doi: 10.1371/journal.pone.0064841] [Medline: 23724101]
11. Eysenbach G. Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. J Med Internet Res 2011 Dec 19;13(4):e123 [FREE Full text] [doi: 10.2196/jmir.2012] [Medline: 22173204]
12. Ortega JL. To be or not to be on Twitter, and its relationship with the tweeting and citation of research papers. Scientometrics 2016 Aug 26;109(2):1353-1364. [doi: 10.1007/s11192-016-2113-0]
13. Ladeiras-Lopes R, Clarke S, Vidal-Perez R, Alexander M, Lüscher TF. Twitter promotion predicts citation rates of cardiovascular articles: a preliminary analysis from the ESC Journals Randomized Study. Eur Heart J 2020 Sep 07;41(34):3222-3225. [doi: 10.1093/eurheartj/ehaa211] [Medline: 32306033]
14. Luc JGY, Archer MA, Arora RC, Bender EM, Blitz A, Cooke DT, et al. Does Tweeting Improve Citations? One-Year Results From the TSSMN Prospective Randomized Trial. Ann Thorac Surg 2021 Jan;111(1):296-300. [doi: 10.1016/j.athoracsur.2020.04.065] [Medline: 32504611]
15. Toulouse E, Lafont B, Granier S, Mcgurk G, Bazin J. French legal approach to patient consent in clinical research. Anaesth Crit Care Pain Med 2020 Dec;39(6):883-885. [doi: 10.1016/j.accpm.2020.10.012] [Medline: 33130015]
16. Baird SM, Marsh PA, Lawrentschuk N, Smart P, Chow Z. Analysis of social media use among Australian and New Zealand otolaryngologists. ANZ J Surg 2019 Jun;89(6):733-737. [doi: 10.1111/ans.14884] [Medline: 30347493]
17. Lander ST, Sanders JO, Cook PC, O'Malley NT. Social Media in Pediatric Orthopaedics. J Pediatr Orthop 2017;37(7):e436-e439 [FREE Full text] [doi: 10.1097/BPO.0000000000001032] [Medline: 28719545]
18. Joshi ND, Lieber B, Wong K, Al-Alam E, Agarwal N, Diaz V. Social Media in Neurosurgery: Using ResearchGate. World Neurosurg 2019 Jul;127:e950-e956. [doi: 10.1016/j.wneu.2019.04.007] [Medline: 30965167]
19. Zhu JM, Pelullo AP, Hassan S, Siderowf L, Merchant RM, Werner RM. Gender Differences in Twitter Use and Influence Among Health Policy and Health Services Researchers. JAMA Intern Med 2019 Dec 01;179(12):1726-1729 [FREE Full text] [doi: 10.1001/jamainternmed.2019.4027] [Medline: 31609393]
20. Nemati-Anaraki L, Razmgir M, Moradzadeh M. Scientific impact of Iran University of Medical Sciences researchers in ResearchGate, Google Scholar, and Scopus: An altmetrics study. Med J Islam Repub Iran 2020;34:142 [FREE Full text] [doi: 10.34171/mjiri.34.142] [Medline: 33437738]
21. Boudry C, Durand-Barthez M. Use of author identifier services (ORCID, ResearcherID) and academic social networks (Academia.edu, ResearchGate) by the researchers of the University of Caen Normandy (France): A case study. PLoS One 2020;15(9):e0238583 [FREE Full text] [doi: 10.1371/journal.pone.0238583] [Medline: 32877458]
22. Yu M, Wu YJ, Alhalabi W, Kao H, Wu W. ResearchGate: An effective altmetric indicator for active researchers? Computers in Human Behavior 2016 Feb;55:1001-1006. [doi: 10.1016/j.chb.2015.11.007]
23. Shrivastava R, Mahajan P. Relationship amongst ResearchGate altmetric indicators and Scopus bibliometric indicators: The case of Panjab University Chandigarh (India). New Library World 2015;116(9/10):564-577. [doi: 10.1108/NLW-03-2015-0017]
24. Orduna-Malea E, Martín-Martín A, Thelwall M, Delgado López-Cózar E. Do ResearchGate Scores create ghost academic reputations? Scientometrics 2017 Apr 28;112(1):443-460. [doi: 10.1007/s11192-017-2396-9]

Edited by R Kukafka; submitted 21.04.21; peer-reviewed by R Vidal-Perez, C Boudry, M Kamiński; comments to author 10.06.21; revised version received 22.07.21; accepted 12.08.21; published 15.10.21.

Please cite as:
Clavier T, Occhiali E, Demailly Z, Compère V, Veber B, Selim J, Besnier E
The Association Between Professional Accounts on Social Networks Twitter and ResearchGate and the Number of Scientific Publications and Citations Among Anesthesia Researchers: Observational Study
J Med Internet Res 2021;23(10):e29809
URL: https://www.jmir.org/2021/10/e29809
doi:10.2196/29809
PMID:34652279

Corrigenda and Addenda

Correction: Artificial Intelligence–Based Prediction of Lung Cancer Risk Using Nonimaging Electronic Medical Records: Deep Learning Approach

Marvin Chia-Han Yeh[1,2], MD, PhD; Yu-Hsiang Wang[3], MD; Hsuan-Chia Yang[4,5], PhD; Kuan-Jen Bai[6,7,8], MD; Hsiao-Han Wang[1,2,4,9*], MD; Yu-Chuan Jack Li[1,2,4,5,9*], MD, PhD

[1]Department of Dermatology, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan

[2]Research Center of Big Data and Meta-analysis, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan

[3]School of Medicine, Taipei Medical University, Taipei, Taiwan

[4]Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan

[5]International Center for Health Information Technology, Taipei Medical University, Taipei, Taiwan

[6]Division of Pulmonary Medicine, Department of Internal Medicine, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan

[7]School of Respiratory Therapy, College of Medicine, Taipei Medical University, Taipei, Taiwan

[8]Pulmonary Research Center, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan

[9]Department of Dermatology, School of Medicine, Taipei Medical University, Taipei, Taiwan

[*]these authors contributed equally

Corresponding Author:
Yu-Chuan Jack Li, MD, PhD
Department of Dermatology
Wan Fang Hospital
Taipei Medical University
No 111, Section 3, Xinglong Road
Wenshan District
Taipei, 116
Taiwan
Phone: 886 29307930 ext 2980
Email: jaak88@gmail.com

Related Article:

Correction of: https://www.jmir.org/2021/8/e26256

(J Med Internet Res 2021;23(10):e33519) doi:10.2196/33519

Correction: Artificial Intelligence–Based Prediction of Lung Cancer Risk Using Nonimaging Electronic Medical Records: Deep Learning Approach

In “Artificial Intelligence–Based Prediction of Lung Cancer Risk Using Nonimaging Electronic Medical Records: Deep Learning Approach” (J Med Internet Res 2021;23(8):e26256), two errors were noted.

Due to a system error, the name of one author, Marvin Chia-Han Yeh, was replaced with the name of another author on the paper, Hsuan-Chia Yang. As well, the formatting of the author name "Yu-Chuan (Jack) Li" has been changed to "Yu-Chuan Jack Li" in the corrected version of the paper.

In the originally published paper, the order of authors was listed as follows:

Hsuan-Chia Yang, Yu-Hsiang Wang, Hsuan-Chia Yang, Kuan-Jen Bai, Hsiao-Han Wang, Yu-Chuan (Jack) Li.

This has been corrected to:

Marvin Chia-Han Yeh, Yu-Hsiang Wang, Hsuan-Chia Yang, Kuan-Jen Bai, Hsiao-Han Wang, Yu-Chuan Jack Li.

In the originally published paper, the ORCID of author Hsuan-Chia Yang was incorrectly published as follows:

0000-0001-6710-4435

This has been corrected to:

0000-0001-9198-0697

The correction will appear in the online version of the paper on the JMIR Publications website on October 15, 2021, together with the publication of this correction notice. Because this was

made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 10.09.21; this is a non–peer-reviewed article;accepted 10.09.21; published 15.10.21.

Please cite as:
Yeh MCH, Wang YH, Yang HC, Bai KJ, Wang HH, Li YCJ
Correction: Artificial Intelligence–Based Prediction of Lung Cancer Risk Using Nonimaging Electronic Medical Records: Deep Learning Approach
J Med Internet Res 2021;23(10):e33519
URL: https://www.jmir.org/2021/10/e33519
doi:10.2196/33519
PMID:34653015

Corrigenda and Addenda

Correction: Engagement in an Interactive App for Symptom Self-Management during Treatment in Patients With Breast or Prostate Cancer: Mixed Methods Study

Marie-Therése Crafoord[1*], RN, BSc, MSc; Maria Fjell[1*], RN, BSc, OCN; Kay Sundberg[1], RN, PhD; Marie Nilsson[1], MSW, PhD; Ann Langius-Eklöf[1], RN, PhD

Division of Nursing, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden

[*]these authors contributed equally

Corresponding Author:
Marie-Therése Crafoord, RN, BSc, MSc
Division of Nursing
Department of Neurobiology, Care Sciences and Society
Karolinska Institutet
Alfred Nobels Allé 23
Stockholm, 14183
Sweden
Phone: 46 8524 837 49
Email: marie-therese.crafoord@ki.se

Related Article:

Correction of: https://www.jmir.org/2020/8/e17058

(J Med Internet Res 2021;23(10):e33140) doi:10.2196/33140

In "Engagement in an Interactive App for Symptom Self-Management during Treatment in Patients With Breast or Prostate Cancer: Mixed Methods Study" (J Med Internet Res 2020;22(8):e7058), 5 errors were noted.

In the originally published article, the following sentence was included in the Abstract under the "Results" section:

> *Among the patients treated for breast cancer, higher age predicted a higher total number of free text messages sent (P=.04).*

This sentence has been corrected to:

> *Among the patients treated for breast cancer, higher age predicted a lower total number of free text messages sent (P=.04).*

In the originally published article, the following sentence was included in the Abstract under the "Conclusions" section:

> *The predictive ability of demographic variables differed between patient groups, but higher age and a higher educational level predicted higher usage of specific app functions for both patient groups.*

This sentence has been corrected to:

> *The predictive ability of demographic variables differed between patient groups, but higher age and a higher educational level predicted usage of specific app functions for both patient groups.*

In the originally published article, the following sentence was included in the "Results" section under the paragraph "Patient Characteristics as Predictors of App Usage":

> *Higher age predicted a higher total number of free text messages sent (P=.04).*

This sentence has been corrected to:

> *Higher age predicted a lower total number of free text messages sent (P=.04).*

In the originally published article, the following sentence was included in the "Discussion" section under the paragraph "Principal Results":

> *The only variable tested that predicted usage for both groups was age; higher age predicted an increase of the total number of free text messages sent in the breast cancer group, while a higher age predicted more self-care advice views in the prostate cancer group.*

This sentence has been corrected to:

> *The only variable tested that predicted usage for both groups was age; higher age predicted a decrease of the total number of free text messages sent in the breast cancer group, while a higher age predicted more self-care advice views in the prostate cancer group.*

In the originally published article, the following sentence was included in the "Conclusions" section:

> *The predictive ability of demographic variables differed between patient groups, but higher age and a higher educational level predicted higher usage of specific app functions for both patient groups.*

This sentence has been corrected to:

> *The predictive ability of demographic variables differed between patient groups, but higher age and a higher educational level predicted the usage of specific app functions for both patient groups.*

The correction will appear in the online version of the paper on the JMIR Publications website on October 13, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 25.08.21; this is a non–peer-reviewed article;accepted 07.09.21; published 12.10.21.

Please cite as:
Crafoord MT, Fjell M, Sundberg K, Nilsson M, Langius-Eklöf A
Correction: Engagement in an Interactive App for Symptom Self-Management during Treatment in Patients With Breast or Prostate Cancer: Mixed Methods Study
J Med Internet Res 2021;23(10):e33140
URL: https://www.jmir.org/2021/10/e33140
doi:10.2196/33140
PMID:34637396

Corrigenda and Addenda

Correction: Characterization of Anorexia Nervosa on Social Media: Textual, Visual, Relational, Behavioral, and Demographical Analysis

Diana Ramírez-Cifuentes[1], MSc; Ana Freire[1,2], PhD; Ricardo Baeza-Yates[1,3], PhD; Nadia Sanz Lamora[4], MSc; Aida Álvarez[4,5,6], MD, PhD; Alexandre González-Rodríguez[4,5,6], MD, PhD; Meritxell Lozano Rochel[7], BSc; Roger Llobet Vives[7], MSc; Diego Alejandro Velazquez[8], MSc; Josep Maria Gonfaus[9], PhD; Jordi Gonzàlez[8], PhD

[1]Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain
[2]UPF Barcelona School of Management, Barcelona, Spain
[3]Institute for Experiential AI, Northeastern University, Boston, MA, United States
[4]Department of Mental Health, Centro de Investigación Biomédica en Red de Salud Mental, Parc Tauli University Hospital, Sabadell, Spain
[5]Parc Tauli Research and Innovation Institute, Sabadell, Spain
[6]Autonomous University of Barcelona, Bellaterra, Spain
[7]Fundación Instituto de Trastornos Alimentarios, Barcelona, Spain
[8]Computer Vision Center, Universitat Autonoma de Barcelona, Bellaterra, Spain
[9]Visual Tagging Services, Bellaterra, Spain

Corresponding Author:
Diana Ramírez-Cifuentes, MSc
Department of Information and Communication Technologies
Universitat Pompeu Fabra
Carrer de Tànger, 122-140
Barcelona, 08018
Spain
Phone: 34 643294184
Email: diana.ramirez@upf.edu

Related Article:

Correction of: https://www.jmir.org/2021/7/e25925

(J Med Internet Res 2021;23(10):e33447) doi:10.2196/33447

In "Characterization of Anorexia Nervosa on Social Media: Textual, Visual, Relational, Behavioral, and Demographical Analysis" (J Med Internet Res 2021; 23(7):e25925), the authors noted one error.

In the originally published paper, a reference was mistakenly placed in the following sentence in the Introduction – Background section:

> *EDs, such as AN, are strongly related to risk factors including perfectionistic traits, parenting style [2], and the existence of comorbid mood disorders such as depression, of which 33%-50% of anorexia patients experience.*

To address this issue and to avoid a misinterpretation of the conclusions of Reference 2, which analyses but does not state that there is a strong relation between AN and parenting style, the sentence has been corrected as follows:

> *EDs, such as AN, are related to risk factors including perfectionistic traits [2], and the existence of comorbid mood disorders such as depression, of which 33%-50% of anorexia patients experience.*

The correction will appear in the online version of the paper on the JMIR Publications website on October 12, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 08.09.21; this is a non–peer-reviewed article;accepted 06.10.21; published 12.10.21.

Please cite as:
Ramírez-Cifuentes D, Freire A, Baeza-Yates R, Sanz Lamora N, Álvarez A, González-Rodríguez A, Lozano Rochel M, Llobet Vives R, Velazquez DA, Gonfaus JM, Gonzàlez J
Correction: Characterization of Anorexia Nervosa on Social Media: Textual, Visual, Relational, Behavioral, and Demographical Analysis
J Med Internet Res 2021;23(10):e33447
URL: https://www.jmir.org/2021/10/e33447
doi:10.2196/33447
PMID:34637394

Original Paper

Public Sentiment and Discourse on Domestic Violence During the COVID-19 Pandemic in Australia: Analysis of Social Media Posts

Kim Usher[1], AM, RN, PhD; Joanne Durkin[1], MA, PhD; Sam Martin[2], PhD; Samantha Vanderslott[2], PhD; Cecilia Vindrola-Padros[3], PhD; Luke Usher[4]; Debra Jackson[5], AO, RN, PhD

[1]University of New England, Armidale, Australia

[2]Oxford Vaccine Group, University of Oxford, Oxford, United Kingdom

[3]University College London, London, United Kingdom

[4]Griffith University, Goldcoast, Australia

[5]University of Sydney, Sydney, Australia

Corresponding Author:
Joanne Durkin, MA, PhD
University of New England
Pat O'Shane Building
Armidale, 2350
Australia
Phone: 61 401096117
Email: jdurkin3@myune.edu.au

Abstract

Background: Measuring public response during COVID-19 is an important way of ensuring the suitability and effectiveness of epidemic response efforts. An analysis of social media provides an approximation of public sentiment during an emergency like the current pandemic. The measures introduced across the globe to help curtail the spread of the coronavirus have led to the development of a situation labeled as a "perfect storm," triggering a wave of domestic violence. As people use social media to communicate their experiences, analyzing public discourse and sentiment on social platforms offers a way to understand concerns and issues related to domestic violence during the COVID-19 pandemic.

Objective: This study was based on an analysis of public discourse and sentiment related to domestic violence during the stay-at-home periods of the COVID-19 pandemic in Australia in 2020. It aimed to understand the more personal self-reported experiences, emotions, and reactions toward domestic violence that were not always classified or collected by official public bodies during the pandemic.

Methods: We searched social media and news posts in Australia using key terms related to domestic violence and COVID-19 during 2020 via digital analytics tools to determine sentiments related to domestic violence during this period.

Results: The study showed that the use of sentiment and discourse analysis to assess social media data is useful in measuring the public expression of feelings and sharing of resources in relation to the otherwise personal experience of domestic violence. There were a total of 63,800 posts across social media and news media. Within these posts, our analysis found that domestic violence was mentioned an average of 179 times a day. There were 30,100 tweets, 31,700 news reports, 1500 blog posts, 548 forum posts, and 7 comments (posted on news and blog websites). Negative or neutral sentiment centered on the sharp rise in domestic violence during different lockdown periods of the 2020 pandemic, and neutral and positive sentiments centered on praise for efforts that raised awareness of domestic violence as well as the positive actions of domestic violence charities and support groups in their campaigns. There were calls for a positive and proactive handling (rather than a mishandling) of the pandemic, and results indicated a high level of public discontent related to the rising rates of domestic violence and the lack of services during the pandemic.

Conclusions: This study provided a timely understanding of public sentiment related to domestic violence during the COVID-19 lockdown periods in Australia using social media analysis. Social media represents an important avenue for the dissemination of information; posts can be widely dispersed and easily accessed by a range of different communities who are often difficult to reach. An improved understanding of these issues is important for future policy direction. Heightened awareness of this could help agencies tailor and target messaging to maximize impact.

(J Med Internet Res 2021;23(10):e29025) doi:10.2196/29025

KEYWORDS

COVID-19; domestic violence; social media; Twitter; sentiment analysis; discourse analysis; keyword analysis; pandemic; sentiment; public health; public expression

Introduction

COVID-19 has affected millions of people across the globe [1]. While the rates of COVID-19 in Australia have been low compared to the rest of the world, Australian people have experienced distress due to the nature of the infection and transmission routes, which have led to many societal and lifestyle changes associated with government attempts to contain the infection [2]. Social distancing regulations imposed to contain the spread of the infection have affected the social, psychological, and economic well-being of many Australians [3-5].

Studies have found that women experienced higher levels of psychological distress during the pandemic when compared to men [6-9]. Younger women in Australia reported higher levels of stress during 2020 compared to older women, with 1 in 4 women aged 25 to 31 years reporting being very or extremely stressed [10]. Reasons for these higher levels have been associated with caring roles and responsibilities [7], linked to homeschooling efforts, job insecurity and financial burden [10], and domestic violence [11,12].

Social media plays a significant role in the dissemination of health information [13], particularly during times of public health crisis [14], and can significantly bolster disaster management communication [15]. Communication during the COVID-19 pandemic has generated more reliance on online platforms such as Twitter, Facebook, and Instagram [16]. Social media can also accelerate the expression of feelings about public events [13,17-19] and social media platforms such as Twitter, blogs, and other platforms are ideal places to quickly receive news and express opinions in times of crisis such as the current pandemic [18,20-22]. As a real-time network, social media offers users the ability to communicate using both public and private messages. Since the beginning of the pandemic, people have used social media sites to express their opinions and share information about the COVID-19 pandemic and related issues. It has been suggested that reliance on social media platforms such as Twitter will continue to grow as long as social distancing measures are used by governments to contain the spread of the virus [23]. Numerous studies have already been conducted to understand the public response to issues related to COVID-19 [18,19,24-26], and social media provides researchers with an opportunity to study the role it plays in the current global health crisis.

The measures introduced across the globe to help curtail the spread of the coronavirus have led to the development of a situation labelled as a “perfect storm” to trigger a wave of domestic violence related to psychological and economic pressures, as well as negative coping mechanisms such as alcohol and other drug use [27,28]. In Australia, alcohol sales rose more than 36% as social distancing measures were implemented [29]. During this period of social isolation, reports of an increase in domestic or intimate partner violence have been heard internationally [27,30-33], reflected in a similar increase in calls to domestic violence hotlines [34]. Similar trends have been reported in previous pandemics [35,36]. In addition, there have also been reports of homicides associated with domestic violence since stay-at-home measures were introduced [37,38]. The pandemic also meant there were fewer opportunities for people experiencing domestic violence to seek help [5,39]. Previously, Twitter hashtags have been examined to determine the nature of domestic violence [18,40,41]. In this paper, we present the findings of a sample of geotagged Australian posts from across different social media platforms, as well as an analysis of users’ public emotional responses to domestic violence during the COVID-19 pandemic. The aim of this study was to understand more personal, self-reported experiences, emotions, and reactions toward domestic violence in Australia that were not always classified or collected by official public bodies during the pandemic.

Methods

Data Collection

An initial general Boolean search (Multimedia Appendix 1) was conducted to obtain a broad overview of the main domestic violence topics related to COVID-19. The research included an analysis of social media and news posts via the media monitoring software Meltwater [42] by using advanced Boolean search terms, agreed upon by 5 researchers (KU, JD, SM, SV, DJ), to collect an initial sample of posts related to domestic violence from the Australian region and during the first 6 months of the COVID-19 pandemic (January 1 to June 31, 2020). Posts were collected from January 1 to gather all mentions of the pandemic from the early stages of reports of COVID-19 (including the first reported case in Australia in late January 2020), and any pre-emptive discussion with regards to the effect of localized lockdowns on domestic home situations in Australia.

Overall, the initial search found a total of 137,300 posts that mentioned domestic violence in general across social media and news media. Within these posts, domestic violence was mentioned an average of 300 times a day. There were 90,200 tweets, with 40,300 news posts, 7000 blog posts, 250 forum posts, and 50 comments. Within this exploratory search, we found that with our simple Boolean, despite it being within the range of the start of the COVID-19 pandemic, 35% of all posts did not refer to the context of the pandemic. While this meant there was a need to refine the Boolean with more COVID-19–specific terms and hashtags, we managed to collect specific colloquialisms, slang, and Australian-specific terms linked with domestic violence as experienced in the region, which we then added to a more specific Boolean search. These included terms such as “DFV” (domestic family violence), “DV” (domestic violence), and “VAW” (violence against

women). Other terms specific to the Australian region included "domestic abuse," "family violence," "intimate partner violence," "gender violence," "spousal abuse," or "spousal violence"—which were found to be used online more often within the Australian region, than, for example, in the United Kingdom or the European Union.

Once an overview of the data was established, we conducted a second, more focused Boolean search (Multimedia Appendix 1), which included keywords and hashtags found in the initial data sample. Usage referred to English-language posts relating to domestic violence from the Australian region and included specific terms related to the first year of the COVID-19 pandemic (January 1 to December 31, 2020).

Data Analysis

While Meltwater, a standard media monitoring tool, was used to collect social media and news posts, we were aware of the difficulties of accurately assessing and classifying patterns of discourse, sentiment, and emoji-related analysis within the correct context or perspective of the topic under investigation [43]. We then conducted additional discourse and emoji analysis using InfraNodus, a text network analysis software [44,45], to measure themes and patterns occurring in discussions around domestic violence, and to analyze connections between subtopics that linked different types of conversations and word clusters together in order to identify similarities in behavior. Analysis of the sentiment of text and emojis shared within posts was based on the internal InfraNodus emoji sentiment lexicon [46] and sentiment analysis lexicon [47].

While these embedded tools were useful for a broad overview of sentiment, to obtain a better understanding of sentiment from the specific context of domestic violence, we drew on previous research [48] by creating a "manual sentiment framework" (Multimedia Appendix 2). We set up our own framework with definitions based on our analysis to "reannotate" a subsample of posts using this more specific, carefully constructed, qualitative framework. The sentiment of posts shared was measured in terms of differences in attitudes toward dealing with domestic violence within the context of the pandemic (Multimedia Appendix 2). Posts were classified as positive toward the handling of domestic violence issues if, for example, they agreed with official or government policies and actions around this topic. Posts were marked as negative if they contained negative attitudes or arguments against official or government policies and actions on domestic violence during lockdowns or during the pandemic in general, or shared bad experiences, or, for example, discouraged the following of official guidelines and actions due to concerns over the effectiveness of these measures. Posts were marked as neutral if they contained only a general statement, with no expression of sentiment or opinion.

Posts were then coded into predefined categories to create the final data set. Intercoder reliability comparison queries were assessed to ensure accuracy using the processes and scripts demonstrated by Kummervold et al [49], with an accuracy F-score of 0.76. Once we identified the key themes, we selected individual posts for textual analysis, drawing out specific issues for a qualitative interrogation of topics of concern. We draw on Lewis et al [50] for our approach to analysis, treating empirical materials not only as data that are freestanding but as an "empirical trigger" to guide us in describing and analyzing research topics. In addition, we employed constructivist grounded theory [51] in the interpretation of empirical material, following an iterative and reflexive process. Semantic discourse and topic analysis were used to understand frequently used keywords and topics of concern [52].

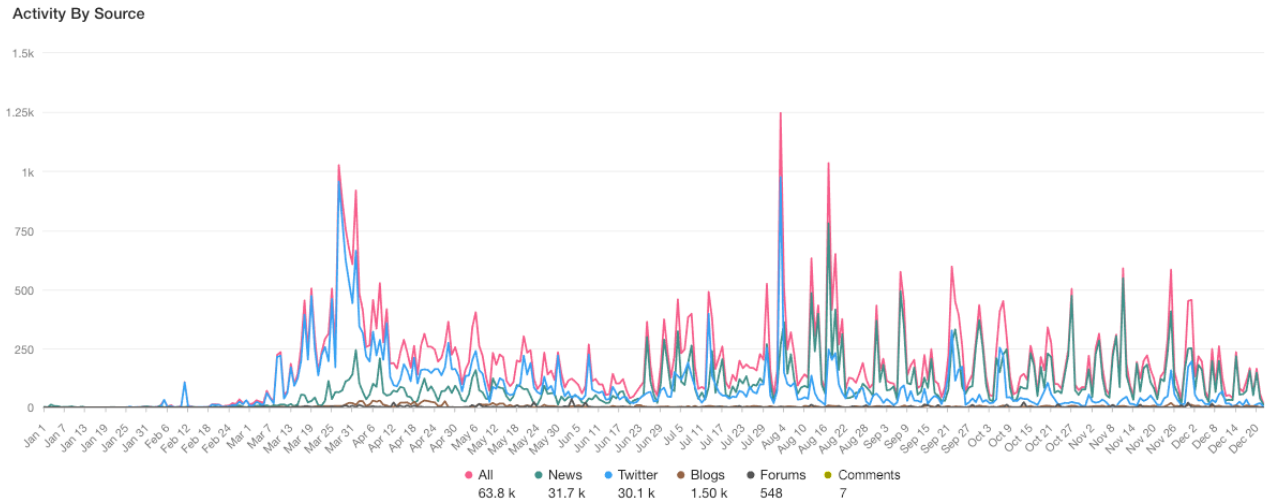
Ethics

Ethical issues related to internet research necessitate the need for researchers to carefully consider relevant guidelines to determine whether ethical approval and informed consent are required [53]. The research team discussed the proposed study with the Chair of Human Research Ethics Committee at the University of New England, Australia, who assessed the request and determined no formal ethics approval was required. While this study was beyond the scope of the human ethics committee, we adhered to the principles of ethics: beneficence, nonmaleficence, autonomy, and justice [54]. The appropriateness and benefit of conducting research during the pandemic was considered against the prospect of the results of this research benefiting a wider community [55]. We further considered the mitigation of the risk of harm to participants. In the case of this research, only publicly available tweets were included in the data set. Social media usernames were removed from the data samples, while within the manuscript, only summaries of tweets were used. No direct or easily traceable quotes have been included. All of these measures are in line with best practices [48,49,56,57]. Data were collected from and analyzed via secure encrypted servers via the Meltwater and InfraNodus platforms. Data were subsequently managed on a secure, password-protected drive only accessible by the research team.

Results

In the final data sample, overall, there were a total of 63,800 posts across social media and news media (Figure 1). Within these posts, our analysis found that domestic violence was mentioned an average of 179 times a day. There were 30,100 tweets, 31,700 news reports, 1500 blog posts, 548 forum posts, and 7 comments (posted on news and blog websites).

Figure 1. Social and mass media mentions of domestic violence in Australia (January 1, 2020, to December 22, 2020).



Demographics

Within the final data sample, it was found that only a smaller sample of 10,200 (out of 30,100) tweets had demographics data. This is indicative of the complexity of demographically and geographically tagged data [58], which we discuss further in the Limitations section. However, analysis of this smaller sample found that those who identified as female comprised 57.2% (n=5834) of the people posting about domestic violence in Australia during the 2020 pandemic, while those who identified as male accounted for 42.8% (n=4365). The age range of posters was as follows: 18-24 years (n=2029, 19.9%), 25-34 years (n=4386, 43%), 35-44 years (n=2519, 24.7%), and 45-54 years (n=948, 9.3%). While the main family status of posters was listed as parents (n=948, 91.1%), 6.3% (n=642) identified as married, with the rest identifying as senior or single.

On Twitter, the majority of Twitter mentions of domestic violence were retweets (n=17,600), followed by quoted (commented on) tweets (n=7130) and replies (n=1180). This conversation related to engagement among 11,800 Twitter users, centering around 4250 original tweets during this period (January 1 to December 22, 2020). Overall, based on the follower count of each of the users engaging with tweets, the potential reach and impressions of these conversations (possible number of people who may have read these messages) was 177 million.

Sentiment by Source

While retweeted and reposted news stories (without any additional commentary or human sentiment added) were classified as neutral across networks, the overall sentiment toward the topic of rates of domestic violence was negative at 38% (n=24,264), with just 7% (n=4469) of all posts across social media holding positive keywords. Forum discussions held the most negative sentiment at 50% (n=274), closely followed by reported news items with additional commentary (n=14,265, 45%), tweets (n=12,040, 40%), and blog posts (n=405, 27%).

Key Drivers of Sentiment

We found that a mixed range of sentiment was used in how domestic violence was discussed on social media and in news articles, as well as expressed in peoples' reactions to events as they retweeted/reposted or quoted original posts and articles discussing or describing worry or reactions to how domestic violence was handled in Australia during the pandemic and the various lockdowns.

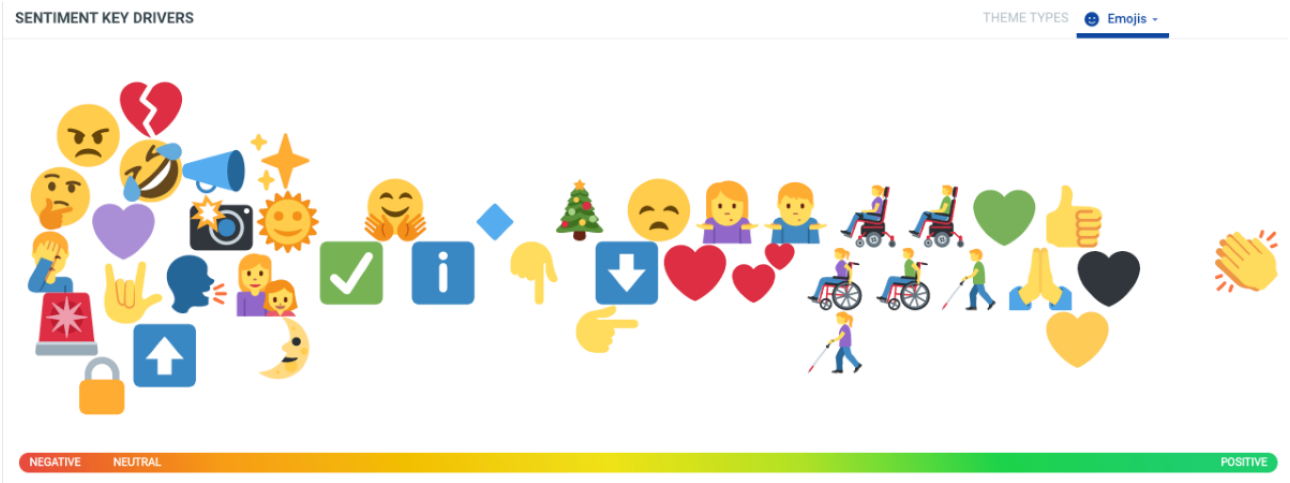
The majority of negative-to-neutral sentiment (Multimedia Appendix 2) centered on the sharp rise in domestic violence during different lockdown periods of the 2020 pandemic, with a focus on gender-based violence, sexism, and worries that a financial crisis and stress on overall services was driving up inequalities and creating a crisis of job insecurity and social pressures that were in turn fueling domestic violence. Neutral-to-positive sentiment centered on the praise of efforts raising awareness of domestic violence, a focus on both violence in immigrant families and stress on nonvisa versus visa holders, help with family-based violence, as well as the positive actions of domestic violence charities and support groups in their campaigns to help isolated victims in both mainstream and Aboriginal communities.

Further analysis of trending negative and ambiguous keywords (posts or articles with a mix of negative and positive terms) showed mixed sentiment, with calls for a positive and proactive handling (rather than a mishandling) of the pandemic. There were calls for a proactive and overdue review of family law, the use of family courts, an easing of the pressure of bankruptcies, and much better access to general practitioner surgeries and help via hospitals to make sure that the knock-on effects of issues that might lead to domestic and family violence could be better handled. More negative sentiment used the term "domestic violence" instead of family violence, with references to an ongoing crisis of domestic violence during various localized lockdowns around Australia, restrictions of access to needed help from the police and emergency services who were diverted elsewhere during the initial lockdowns, and an increased amount of pharmacies and frontline health care workers noticing more women with suspected domestic abuse injuries, as well as worry about pressures on mental health.

Emoji Analysis

The majority of negative emojis used on social media within discussions regarding domestic violence (Figure 2) involved symbols of anger, sadness, annoyance, and disbelief; reports of increased callouts to emergency services (siren emojis); a call for peace; and heartbreak at the rise in instances in family and gender-based violence (older and younger female emojis). The thinking and hand taking notes emojis were used to denote the need to think seriously about taking measures to protect the emotional and mental health well-being of those experiencing domestic and family violence. The sun, stars, and moon emojis conveyed the need to make services accessible to victims both during the day and night. Negative-to-neutral emojis were connected to the arrow and pointing finger emojis that were related to pointing toward links to articles, images, or other resources shared. The open-handed hug emoji was used to denote a show of support to those experiencing domestic violence. The sad or anxiety-face emoji was linked to worry about a surge in domestic violence around Christmas time (Christmas tree emoji), with offers of help, love, and support shared via links and messages (heart emojis). As part of the International Day of People with Disability, there was also a request to raise awareness around the plight of people with physical and visual disabilities (wheelchair and visibility cane emojis). Positive emotion was also represented by variations of different-colored heart emojis, as well as the clapping hands emoji, which symbolized both praise of a new $25 million United Nations emergency fund to tackle violence against women during the pandemic, as well as a celebration of the publication of a research paper that investigated risk and protective factors for family violence both during and after the COVID-19 lockdowns) [59]. While emojis were used across social media platforms (eg, tweets, blogs, forums, and other posts), which is to be expected, no emojis were found to be used within original news media articles.

Figure 2. Collection of top emojis used in relation to tweets mentioning domestic violence and the discourse behind them.



Hashtag Analysis

On social media and in news articles, outside of the main COVID-19 hashtags (#covid19 and #coronavirus or #lockdown), among the top 20 corelated hashtags regarding the topic of domestic violence in Australia were ones that focused specifically on the Australian experience (#covid19au or #covid19aus to represent COVID-19 Australia; #covid19vic or #covidvic to represent the experience of COVID-19 in the state of Victoria; #auspol to represent Australian politics; and #springst to represent news and discussion of politics in the Australian state of Victoria [the Victorian Parliament is located on Spring Street]). There were also hashtags linked to colloquialisms for domestic violence. These included #DV as a shorthand for "domestic violence," #DFV for "domestic family violence," as well as #VAW for "violence against women." The hashtag #thedrum was also used in relation to the Australian topical media/news company "ABC The Drum," which covered a host of topical pieces discussing the rise in domestic violence in Australia during the pandemic. More direct terms like #DomesticViolence and #DomesticAbuse were also popular, with #DomesticViolence being more prevalent (3700 mentions) than #DomesticAbuse (610 mentions).

Negative-to-ambiguous hashtags centered around posts discussing increased rates of stress, mental health, and femicides or deaths linked to substantial increases in domestic violence during lockdown, with specific mentions of New South Wales (#NSW, #nswbudget), as well as the feminist hashtag #MeToo, references to ongoing conversations in Australian politics discussing a need for changes in family law to better protect families experiencing trauma related to domestic family violence. Various corelated hashtags were also linked to wider campaigns focused on supporting victims, such as #ifyoucouldseewhatisee, #SexNotGender, #StaySafeStayOpen, #ChildPoverty, #GenderBasedViolence, #noexcuseforabuse, #ListentoVoices, and #Indigenous. More positive and ambiguous-to-positive hashtags used in campaigns included #LGBTIQ, #trans, #StopDV20, #WAFV2020, #16DaysOfActivism, and #justiceforall.

Discussion

Principal Findings

Government-regulated COVID-19 lockdowns occurred throughout 2020 and 2021 in Australia and internationally. Through the use of sentiment and discourse analysis, we identified negative or neutral sentiment centered on the sharp

rise in domestic violence during different lockdown periods in Australia. Neutral-to-positive sentiment centered around praising efforts to raise awareness of domestic violence, as well as the positive actions of domestic violence charities and support groups in their campaigns. We identified a high level of public discontent related to the rising rates of domestic violence and the lack of services during the pandemic.

Social media and news media are an important mechanism for discussing and forwarding information about domestic violence and available services [60]. They represent an important avenue for dissemination of information that can be widely dispersed and easily accessed by a range of different communities who are often difficult to reach. Increasingly, Twitter has been utilized to gain insights into public health outcomes, perspectives, and behaviors during the COVID-19 pandemic [61-64].

Our study examined public discussions and shared articles about domestic violence during the initial period of COVID-19 in Australia during 2020. The study contributed to our understanding of public sentiment about domestic violence. It indicated a high level of public discontent related to the rising rates of the violence and the lack of services during the pandemic. Previous studies are consistent with these findings [19,60]. The high rate of negative posts and articles related to domestic violence and COVID-19 supports recent reports and publications that have identified the increasing problems related to domestic violence during the COVID-19 pandemic stay-at-home orders [5,32,39,65-68]. The high rate of negative sentiments about domestic violence could have a negative impact on women experiencing domestic violence as women often feel blamed or blame themselves for the violence [69,70]. In this study, more negative sentiments were related to domestic violence than family violence. This may be related to an understanding of the term and indicates that family violence may be understood as inclusive of children as well as women. The language chosen for social media and news posts can offer insight into issues of what type of violence has been most prevalent and the emotional experiences and reactions to violence.

Emojis are useful for expressing ideas and sentiments on social media, and their use to communicate issues related to public health are not new [16]. In this study, they represented and symbolized the deeply emotional aspects of domestic violence and the vulnerability experienced by women. In a study by Al-Rawi et al [16], more negative emojis were used for women's concerns than those of men. Negative emojis in this study were used to express anger, sadness, annoyance, and disbelief, while open-handed hug emojis were used to offer support to people who discussed their personal experiences of domestic violence on social media. The sad or anxiety-face emoji was linked to worry about a surge in domestic violence around Christmas time (Christmas tree emoji), with offers of help, love, and support shared via links on social media (heart emojis). As part of the International Day of People with Disability, there was also a request to raise awareness around the plight of people with physical and visual disabilities (wheelchair and visibility cane emojis).

Within data tagged with demographic markers on Twitter, the majority of people posting and sharing articles about domestic violence in Australia during the 2020 pandemic were female (57.2%), which is not surprising given that women are more likely to be the target of domestic violence [5,18,19,41]. The age of the majority of users tweeting about domestic violence on Twitter specifically were between 25-34 years (43%), followed by those aged 35-44 years (24.7%). Younger women are more likely to be caregivers of children and older relatives, have the responsibility of managing the family budget, and have the added responsibility of homeschooling during the pandemic [11]. They also have greater access to and awareness of social media and technology in general.

Key hashtag drivers of sentiment included words such as "DV," "MeToo," "women," "law" "family law," "sexism," "trauma," "homelessness," "stress," "aged care," and "violence against women." Interestingly, ageist comments have been prevalent on social media during the COVID-19 pandemic, with posts implying that the lives of older adults are less valuable rising as quickly as 1 day after each news update on increases in COVID-19 infection and death rates or information related to COVID-19 risk factors [71]. The study reported a daily average of ageist tweets of 18% with the highest rate of almost 53% in March 2020. The content of those tweets ranged from suggesting that older people should be isolated to prevent the spread of COVID-19 to death jokes and ridicule targeted at older people.

Social and news media present a unique opportunity to investigate attitudes toward abuse and violence [60,64], and online forums provide an opportunity to offer support, engage in advocacy, and voice concerns or desires for social change [72,73], while also providing a central point for discussion of domestic violence prevention and promotion [60]. This information provides a useful window into the perspectives of persons expressing emotions related to domestic violence. Identifying periods of increased activity on Twitter is a useful way to identify changes in public opinion.

The rise in societal concerns evidenced by negative emotions on social media needs to be monitored regularly by public health professionals, who can release strategic information for consumers, health professionals, and government bodies so that urgent action can be taken during periods of high domestic violence–related activity.

Limitations

The study was conducted during a pandemic when everyday conditions deviated away from the norm, and use of social and news media was more amplified than in nonpandemic times. It should be noted that while the study was partially based on geotagged posts from across social media posts and articles from other (eg, news) media (in the Australian region), the limitations of geotagged tweets and posts should be kept in mind [74-76]. Drawing on the discussions of the reliability of geotagged research [58], first, while our data sample was quite large, it retained some bias, as users on social media are not representative of the general Australian population. The consequence is that the demographics indicated in this study do not perfectly mirror the larger Australian population. Second, not all users on social and other media platforms are required

to allow access to their geotagged location, so, while access was obtained for tens of thousands of geotagged posts, this is still a relatively small proportion of the total of nongeotagged posts across social and other media. However, despite these limitations, our results within this limited data sample have produced insights and metrics that align with existing qualitative trends and research with regards to domestic violence rates and policies within the COVID-19 pandemic [77,78]. Future research will need to add to unstructured social media and news data with analysis from interview or focus group data via traditional qualitative or digital ethnographic methods [52]. Further, domestic violence is a sensitive topic, and the way it is talked about on social media or online news posts may not fully represent all narratives. Hence, domestic violence posts on social media during COVID-19 can only represent the opinions and reactions of current geotagged social media users. In addition, the study focused on posts that included specific search terms and sentiment criteria; future studies would benefit from expanding the search terms used and the social and general media platforms included.

Conclusion

Social media and news media are important mechanisms for discussing and sharing information about domestic violence and available services. They represent an important avenue for the dissemination of information; information can be widely dispersed and easily accessed by a range of different communities who are often difficult to reach. The study showed that the use of sentiment and discourse analysis to assess social media and news media data is useful in measuring the general mood toward sensitive issues, the public expression of feelings, and sharing of resources in relation to the otherwise personal experience of domestic violence. Heightened awareness of this could help agencies tailor and target messaging to maximize impact.

Authors' Contributions

KU, DJ, and JD conceptualized the project, KU led the project; SM and SV retrieved and analyzed the data; and KU, DJ, JD, LU, and CV-P took part in the discussion on search terms, methodology, and the results. All authors contributed to the writing and annotation of the manuscript and approved the submission of the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
(A) Boolean search terms for domestic violence and (B) final Boolean (data sample: English-language posts originating within the Australian geographical region; January 1 to December 12, 2020).
[DOCX File , 14 KB - jmir_v23i10e29025_app1.docx]

Multimedia Appendix 2
Sentiment analysis of domestic violence posts in Australia: definition and context.
[DOCX File , 16 KB - jmir_v23i10e29025_app2.docx]

References

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases 2020 May;20(5):533-534. [doi: 10.1016/s1473-3099(20)30120-1]
2. Van Rheenen TE, Meyer D, Neill E, Phillipou A, Tan EJ, Toh WL, et al. Mental health status of individuals with a mood-disorder during the COVID-19 pandemic in Australia: Initial results from the COLLATE project. J Affect Disord 2020 Oct 01;275:69-77 [FREE Full text] [doi: 10.1016/j.jad.2020.06.037] [Medline: 32658826]
3. Usher K, Bhullar N, Jackson D. Life in the pandemic: Social isolation and mental health. J Clin Nurs 2020 Aug 06;29(15-16):2756-2757. [doi: 10.1111/jocn.15290] [Medline: 32250493]
4. Brooke J, Jackson D. Older people and COVID-19: Isolation, risk and ageism. J Clin Nurs 2020 Jul 05;29(13-14):2044-2046. [doi: 10.1111/jocn.15274] [Medline: 32239784]
5. Bradbury-Jones C, Isham L. The pandemic paradox: The consequences of COVID-19 on domestic violence. J Clin Nurs 2020 Jul 22;29(13-14):2047-2049 [FREE Full text] [doi: 10.1111/jocn.15296] [Medline: 32281158]
6. Stanton R, To QG, Khalesi S, Williams SL, Alley SJ, Thwaite TL, et al. Depression, Anxiety and Stress during COVID-19: Associations with Changes in Physical Activity, Sleep, Tobacco and Alcohol Use in Australian Adults. Int J Environ Res Public Health 2020 Jun 07;17(11):4065 [FREE Full text] [doi: 10.3390/ijerph17114065] [Medline: 32517294]
7. Rahman MA, Hoque N, Alif SM, Salehin M, Islam SMS, Banik B, et al. Factors associated with psychological distress, fear and coping strategies during the COVID-19 pandemic in Australia. Global Health 2020 Oct 08;16(1):95-15 [FREE Full text] [doi: 10.1186/s12992-020-00624-w] [Medline: 33032629]
8. French MT, Mortensen K, Timming AR. Psychological Distress and Coronavirus Fears During the Initial Phase of the COVID-19 Pandemic in the United States. J Ment Health Policy Econ 2020 Sep 01;23(3):93-100. [Medline: 32853158]

9. Gausman J, Langer A. Sex and Gender Disparities in the COVID-19 Pandemic. J Womens Health (Larchmt) 2020 Apr 01;29(4):465-466. [doi: 10.1089/jwh.2020.8472] [Medline: 32320331]
10. Loxton D, Forder P, Townsend N, Thomson C, Mishra G, Tooth L, et al. LSWH COVID-19 Survey 1: Report 1. Women's Health Australia. 2020 Apr 29. URL: https://alswh.org.au/post-outcomes/alswh-covid-19-survey-1-report-1/ [accessed 2021-09-15]
11. Dadi A, Gesesew H. Mental health for pregnant women and new mothers: why extra care is needed. The Conversation. 2020 Aug 16. URL: https://theconversation.com/mental-health-for-pregnant-women-and-new-mothers-why-extra-care-is-needed-143912 [accessed 2020-12-02]
12. Regenold N, Vindrola-Padros C. Gender Matters: A Gender Analysis of Healthcare Workers' Experiences during the First COVID-19 Pandemic Peak in England. Social Sciences 2021 Jan 27;10(2):43. [doi: 10.3390/socsci10020043]
13. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: Twitter and antibiotics. Am J Infect Control 2010 Apr;38(3):182-188 [FREE Full text] [doi: 10.1016/j.ajic.2009.11.004] [Medline: 20347636]
14. Hagen L, Keller T, Neely S, DePaula N, Robert-Cooperman C. Crisis Communications in the Age of Social Media. Social Science Computer Review 2017 Aug 21;36(5):523-541. [doi: 10.1177/0894439317721985]
15. Merchant RM, Elmer S, Lurie N. Integrating Social Media into Emergency-Preparedness Efforts. N Engl J Med 2011 Jul 28;365(4):289-291. [doi: 10.1056/nejmp1103591]
16. Al-Rawi A, Siddiqi M, Morgan R, Vandan N, Smith J, Wenham C. COVID-19 and the Gendered Use of Emojis on Twitter: Infodemiology Study. J Med Internet Res 2020 Nov 05;22(11):e21646 [FREE Full text] [doi: 10.2196/21646] [Medline: 33052871]
17. Rosenberg H, Syed S, Rezaie S. The Twitter pandemic: The critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. CJEM 2020 Jul 06;22(4):418-421 [FREE Full text] [doi: 10.1017/cem.2020.361] [Medline: 32248871]
18. Xue J, Chen J, Chen C, Hu R, Zhu T. The Hidden Pandemic of Family Violence During COVID-19: Unsupervised Learning of Tweets. J Med Internet Res 2020 Nov 06;22(11):e24361 [FREE Full text] [doi: 10.2196/24361] [Medline: 33108315]
19. Xue J, Chen J, Chen C, Zheng C, Li S, Zhu T. Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. PLoS One 2020 Sep 25;15(9):e0239441 [FREE Full text] [doi: 10.1371/journal.pone.0239441] [Medline: 32976519]
20. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
21. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS One 2011 May 04;6(5):e19467 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]
22. Park Y, Kim M, Seong K. Happy neighborhoods: Investigating neighborhood conditions and sentiments of a shrinking city with Twitter data. Growth and Change 2020 Nov 27;52(1):539-566. [doi: 10.1111/grow.12451]
23. Lopez C, Malolan V, Gallemore C. Understanding the perception of COVID-19 policies by mining a multilanguage Twitter dataset. arXiv Preprint posted online Mar 23, 2020 [FREE Full text]
24. Alvarez-Risco A, Mejia CR, Delgado-Zegarra J, Del-Aguila-Arcentales S, Arce-Esquivel AA, Valladares-Garrido MJ, et al. The Peru Approach against the COVID-19 Infodemic: Insights and Strategies. Am J Trop Med Hyg 2020 Aug;103(2):583-586 [FREE Full text] [doi: 10.4269/ajtmh.20-0536] [Medline: 32500853]
25. Mackey T, Purushothaman V, Li J, Shah N, Nali M, Bardier C, et al. Machine Learning to Detect Self-Reporting of Symptoms, Testing Access, and Recovery Associated With COVID-19 on Twitter: Retrospective Big Data Infoveillance Study. JMIR Public Health Surveill 2020 Jun 08;6(2):e19509 [FREE Full text] [doi: 10.2196/19509] [Medline: 32490846]
26. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
27. van Gelder N, Peterman A, Potts A, O'Donnell M, Thompson K, Shah N, GenderCOVID-19 working group. COVID-19: Reducing the risk of infection might increase the risk of intimate partner violence. EClinicalMedicine 2020 Apr;21:100348 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100348] [Medline: 32292900]
28. Usher K, Durkin J, Bhullar N. The COVID-19 pandemic and mental health impacts. Int J Ment Health Nurs 2020 Jun 10;29(3):315-318 [FREE Full text] [doi: 10.1111/inm.12726] [Medline: 32277578]
29. An early look at how the coronavirus is affecting household spending. Commonwealth Bank of Australia. 2020 Mar 30. URL: https://tinyurl.com/e7nx3jje [accessed 2020-04-10]
30. Campbell AM. An increasing risk of family violence during the Covid-19 pandemic: Strengthening community collaborations to save lives. Forensic Science International: Reports 2020 Dec;2:100089. [doi: 10.1016/j.fsir.2020.100089]
31. Peterman A, Potts A, O'Donnell M, Thompson K, Shah N, Oertelt-Prigione S, et al. Pandemics and Violence Against Women and Children. Center for Global Development. Washington DC: The Center for Global Development; 2020. URL: https://www.cgdev.org/sites/default/files/pandemics-and-vawg-april2.pdf [accessed 2021-09-15]
32. Allen-Ebrahimian B. China's Domestic Violence Epidemic. AXIOS. 2020. URL: https://www.axios.com/china-domestic-violencecoronavirus-quarantine-7b00c3ba-35bc-4d16-afdd-b76ecfb28882.html [accessed 2020-03-07]

33. Wagers SM. Domestic violence growing in wake of coronavirus outbreak. The Conversation. 2020. URL: https://theconversation.com/domestic-violence-growing-in-wake-of-coronavirus-outbreak-135598 [accessed 2020-04-08]
34. Taub A. A new COVID-19 crisis: domestic abuse rises worldwide. The New York Times. 2020 Apr 14. URL: https://www.nytimes.com/2020/04/06/world/coronavirus-domestic-violence.html [accessed 2021-09-15]
35. Adhiambo Onyango M, Resnick K, Davis A, Ramesh Shah R. Gender-Based Violence Among Adolescent Girls and Young Women: A Neglected Consequence of the West African Ebola Outbreak. In: Schwartz D, editor. Pregnant in the Time of Ebola. Switzerland: Springer; 2019:121-132.
36. Mittal S, Singh T. Gender-Based Violence During COVID-19 Pandemic: A Mini-Review. Front Glob Womens Health 2020 Sep 8;1:1-7. [doi: 10.3389/fgwh.2020.00004]
37. Kelly J, Morgan T. Coronavirus: Domestic abuse calls up 25% since lockdown, charity says. BBC News. 2020 Apr 6. URL: https://www.bbc.com/news/uk-52157620 [accessed 2020-04-12]
38. Knowels M. Coronavirus lockdown risks murder frenzy - PM sent warning on domestic abuse. Express. 2020 Apr 7. URL: https://www.express.co.uk/news/uk/1265912/coronavirus-lockdown-murder-warning-domestic-abuse-boris-johnson [accessed 2020-04-07]
39. Usher K, Bhullar N, Durkin J, Gyamfi N, Jackson D. Family violence and COVID-19: Increased vulnerability and reduced options for support. Int J Ment Health Nurs 2020 Aug 07;29(4):549-552 [FREE Full text] [doi: 10.1111/inm.12735] [Medline: 32314526]
40. Cravens JD, Whiting JB, Aamar RO. Why I Stayed/Left: An Analysis of Voices of Intimate Partner Violence on Social Media. Contemp Fam Ther 2015 Sep 25;37(4):372-385. [doi: 10.1007/s10591-015-9360-8]
41. Xue J, Macropol K, Jia Y, Zhu T, Gelles RJ. Harnessing big data for social justice: An exploration of violence against women‐related conversations on Twitter. Human Behav and Emerg Tech 2019 Jul 26;1(3):269-279. [doi: 10.1002/hbe2.160]
42. Social Media Monitoring. Meltwater. 2021. URL: https://www.meltwater.com/en/products/social-media-monitoring [accessed 2021-01-01]
43. van Atteveldt W, van der Velden MACG, Boukes M. The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms. Communication Methods and Measures 2021 Jan 28;15(2):121-140. [doi: 10.1080/19312458.2020.1869198]
44. Paranyushkin D. InfraNodus: Generating Insight Using Text Network Analysis. 2019 Presented at: WWW '19: The World Wide Web Conference; May 13-17; San Francisco, CA p. 3584-3589. [doi: 10.1145/3308558.3314123]
45. Paranyushkin D. Identifying the pathways for meaning circulation using text network analysis. Nodus Labs. 2011 Dec 25. URL: https://noduslabs.com/research/pathways-meaning-circulation-text-network-analysis/ [accessed 2021-09-27]
46. Kralj Novak P, Smailović J, Sluban B, Mozetič I. Sentiment of Emojis. PLoS One 2015;10(12):e0144296 [FREE Full text] [doi: 10.1371/journal.pone.0144296] [Medline: 26641093]
47. Nielsen FA. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. arXiv Preprint posted online Mar 15, 2011 [FREE Full text]
48. Martin S, Kilich E, Dada S, Kummervold PE, Denny C, Paterson P, et al. "Vaccines for pregnant women…?! Absurd" - Mapping maternal vaccination discourse and stance on social media over six months. Vaccine 2020 Sep 29;38(42):6627-6637. [doi: 10.1016/j.vaccine.2020.07.072] [Medline: 32788136]
49. Kummervold P, Martin S, Dada S, Kilich E. Categorising Vaccine Confidence with Transformer-Based Machine Learning Model: The Nuances of Vaccine Sentiment on Twitter. JMIR Preprints Preprint posted online Apr 13, 2021 [FREE Full text] [doi: 10.2196/29584]
50. Lewis J, Hughes J, Atkinson P. Relocation, realignment and standardisation: Circuits of translation in Huntington's disease. Soc Theory Health 2014 Oct 16;12(4):396-415. [doi: 10.1057/sth.2014.13]
51. Charmaz K. Shifting the grounds; Constructivist Grounded Theory Methods. In: Morse J, Bowers B, Charmaz K, Clakre J, Noerager Stern P, editors. Developing Grounded Theory: The Second Generation. Abingdon, Oxford: Routledge; 2009:127-154.
52. Vindrola-Padros C, Chisnall G, Cooper S, Dowrick A, Djellouli N, Symmons SM, et al. Carrying Out Rapid Qualitative Research During a Pandemic: Emerging Lessons From COVID-19. Qual Health Res 2020 Dec 31;30(14):2192-2204 [FREE Full text] [doi: 10.1177/1049732320951526] [Medline: 32865149]
53. Eysenbach G, Till JE. Ethical issues in qualitative research on internet communities. BMJ 2001 Nov 10;323(7321):1103-1105 [FREE Full text] [doi: 10.1136/bmj.323.7321.1103] [Medline: 11701577]
54. Varkey B. Principles of Clinical Ethics and Their Application to Practice. Medical Principles and Practice 2021 Jun 4;30(1):17-28 [FREE Full text] [doi: 10.1159/000509119] [Medline: 32498071]
55. Prichard J, Watters P, Krone T, Spiranovic C, Cockburn H. Social Media Sentiment Analysis: A New Empirical Tool for Assessing Public Opinion on Crime? Current Issues in Criminal Justice 2018 Nov 30;27(2):217-236. [doi: 10.1080/10345329.2015.12036042]
56. Ahmed W, Bath P, Demartini G. Using Twitter as a Data Source: An Overview of Ethical, Legal, and Methodological Challenges. Emerald Insight. 2017 Dec 17. URL: https://www.emerald.com/insight/content/doi/10.1108/S2398-601820180000002004/full/html [accessed 2021-09-27]

57. Ayers JW, Caputi TL, Nebeker C, Dredze M. Don't quote me: reverse identification of research participants in social media studies. NPJ Digital Medicine 2018 Aug 2;1(1):1-2 [FREE Full text] [doi: 10.1038/s41746-018-0036-2] [Medline: 31304312]
58. Xu P, Dredze M, Broniatowski DA. The Twitter Social Mobility Index: Measuring Social Distancing Practices With Geolocated Tweets. J Med Internet Res 2020 Dec 03;22(12):e21499 [FREE Full text] [doi: 10.2196/21499] [Medline: 33048823]
59. Spiranovic C, Hudson N, Winter R, Stanford S, Norris K, Bartkowiak-Theron I, et al. Navigating risk and protective factors for family violence during and after the COVID-19 'perfect storm'. Current Issues in Criminal Justice 2020 Dec 08;33(1):5-18. [doi: 10.1080/10345329.2020.1849933]
60. Dragiewicz M, Burgess J. Domestic Violence on #qanda: The "Man" Question in Live Twitter Discussion on the Australian Broadcasting Corporation's Q and A. Canadian Journal of Women and the Law 2016 Apr;28(1):211-229. [doi: 10.3138/cjwl.28.1.211]
61. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, et al. Global Sentiments Surrounding the COVID-19 Pandemic on Twitter: Analysis of Twitter Trends. JMIR Public Health Surveill 2020 May 22;6(2):e19447 [FREE Full text] [doi: 10.2196/19447] [Medline: 32412418]
62. Padala PR, Jendro AM, Padala KP. Conducting Clinical Research During the COVID-19 Pandemic: Investigator and Participant Perspectives. JMIR Public Health Surveill 2020 Apr 06;6(2):e18887 [FREE Full text] [doi: 10.2196/18887] [Medline: 32250281]
63. Chen E, Lerman K, Ferrara E. Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. JMIR Public Health Surveill 2020 May 29;6(2):e19273 [FREE Full text] [doi: 10.2196/19273] [Medline: 32427106]
64. Dubey AD. The Resurgence of Cyber Racism During the COVID-19 Pandemic and its Aftereffects: Analysis of Sentiments and Emotions in Tweets. JMIR Public Health Surveill 2020 Oct 15;6(4):e19833 [FREE Full text] [doi: 10.2196/19833] [Medline: 32936772]
65. A Perfect Storm: The impact of the Covid-19 pandemic on domestic abuse survivors and the services supporting them. Women's Aid. 2020 Aug 18. URL: https://www.womensaid.org.uk/a-perfect-storm-the-impact-of-the-covid-19-pandemic-on-domestic-abuse-survivors-and-the-services-supporting-them/ [accessed 2021-02-07]
66. Asunramu Z. The government's botched response to Covid-19 will set women back decades—and shows the need for more female leaders. Prospect Magazine. 2020 Jun 21. URL: https://www.prospectmagazine.co.uk/politics/government-coronavirus-covid-impact-on-women-domestic-violence-jobs-childcare [accessed 2020-10-02]
67. Chandan JS, Taylor J, Bradbury-Jones C, Nirantharakumar K, Kane E, Bandyopadhyay S. COVID-19: a public health approach to manage domestic violence is needed. The Lancet Public Health 2020 Jun;5(6):e309. [doi: 10.1016/s2468-2667(20)30112-2]
68. Das M, Das A, Mandal A. Examining the impact of lockdown (due to COVID-19) on Domestic Violence (DV): An evidences from India. Asian Journal of Psychiatry 2020 Dec;54:102335 [FREE Full text] [doi: 10.1016/j.ajp.2020.102335] [Medline: 32801078]
69. Martín-Fernández M, Gracia E, Lila M. Assessing victim-blaming attitudes in cases of intimate partner violence against women: Development and validation of the VB-IPVAW scale. Psychosocial Intervention 2018(27):133-143. [doi: 10.5093/pi2018a18]
70. Di Giacomo P, Cavallo A, Bagnasco A, Sartini M, Sasso L. Violence against women: knowledge, attitudes and beliefs of nurses and midwives. J of Clin Nursing 2017 Aug 14;26(15-16):2307-2316. [doi: 10.1111/jocn.13625] [Medline: 27805756]
71. Xiang X, Lu X, Halavanau A, Xue J, Sun Y, Lai PHL, et al. Modern Senicide in the Face of a Pandemic: An Examination of Public Discourse and Sentiment About Older Adults and COVID-19 Using Machine Learning. J Gerontol B Psychol Sci Soc Sci 2021 Mar 14;76(4):e190-e200 [FREE Full text] [doi: 10.1093/geronb/gbaa128] [Medline: 32785620]
72. Bogen KW, Bleiweiss K, Orchowski LM. Sexual violence is #NotOkay: Social reactions to disclosures of sexual victimization on twitter. Psychology of Violence 2019 Jan;9(1):127-137. [doi: 10.1037/vio0000192]
73. Homan CM, Schrading JN, Ptucha RW, Cerulli C, Ovesdotter Alm C. Quantitative Methods for Analyzing Intimate Partner Violence in Microblogs: Observational Study. J Med Internet Res 2020 Nov 19;22(11):e15347 [FREE Full text] [doi: 10.2196/15347] [Medline: 33211021]
74. Gore RJ, Diallo S, Padilla J. You Are What You Tweet: Connecting the Geographic Variation in America's Obesity Rate to Twitter Content. PLoS One 2015 Sep 2;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]
75. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018 Jun 14;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]
76. Smith A, Brenner J. Twitter use 2012. Pew Research Center. 2012 May 31. URL: https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/2012/PIP_Twitter_Use_2012.pdf [accessed 2021-09-27]

77. Usher K, Bradbury Jones C, Bhullar N, Durkin DJ, Gyamfi N, Fatema SR, et al. COVID-19 and family violence: Is this a perfect storm? Int J Ment Health Nurs 2021 Aug 18;30(4):1022-1032 [FREE Full text] [doi: 10.1111/inm.12876] [Medline: 34008291]
78. Piquero AR, Jennings WG, Jemison E, Kaukinen C, Knaul FM. Domestic violence during the COVID-19 pandemic - Evidence from a systematic review and meta-analysis. J Crim Justice 2021 May;74:101806. [doi: 10.1016/j.jcrimjus.2021.101806]

Edited by C Basch; submitted 23.03.21; peer-reviewed by R Gore, C Lynch; comments to author 19.05.21; revised version received 02.08.21; accepted 04.09.21; published 01.10.21.

Please cite as:
Usher K, Durkin J, Martin S, Vanderslott S, Vindrola-Padros C, Usher L, Jackson D
Public Sentiment and Discourse on Domestic Violence During the COVID-19 Pandemic in Australia: Analysis of Social Media Posts
J Med Internet Res 2021;23(10):e29025
URL: https://www.jmir.org/2021/10/e29025
doi:10.2196/29025
PMID:34519659

Original Paper

The National COVID Cohort Collaborative: Analyses of Original and Computationally Derived Electronic Health Record Data

Randi Foraker[1,2], MA, PhD; Aixia Guo[2], PhD; Jason Thomas[3], BS; Noa Zamstein[4], MSc, PhD; Philip RO Payne[1,2], PhD; Adam Wilcox[3], PhD; N3C Collaborative[5]

[1]Division of General Medical Sciences, School of Medicine, Washington University in St. Louis, St. Louis, MO, United States

[2]Institute for Informatics, School of Medicine, Washington University in St. Louis, St. Louis, MO, United States

[3]Department of Biomedical and Medical Education, School of Medicine, University of Washington, Seattle, WA, United States

[4]MDClone Ltd, Beer Sheva, Israel

[5]See Acknowlegments

Corresponding Author:
Randi Foraker, MA, PhD
Division of General Medical Sciences
School of Medicine
Washington University in St. Louis
600 S. Taylor Avenue, Suite 102
Campus Box 8102
St. Louis, MO, 63110
United States
Phone: 1 314 273 2211
Fax: 1 314 273 1390
Email: randi.foraker@wustl.edu

Abstract

Background: Computationally derived ("synthetic") data can enable the creation and analysis of clinical, laboratory, and diagnostic data as if they were the original electronic health record data. Synthetic data can support data sharing to answer critical research questions to address the COVID-19 pandemic.

Objective: We aim to compare the results from analyses of synthetic data to those from original data and assess the strengths and limitations of leveraging computationally derived data for research purposes.

Methods: We used the National COVID Cohort Collaborative's instance of MDClone, a big data platform with data-synthesizing capabilities (MDClone Ltd). We downloaded electronic health record data from 34 National COVID Cohort Collaborative institutional partners and tested three use cases, including (1) exploring the distributions of key features of the COVID-19–positive cohort; (2) training and testing predictive models for assessing the risk of admission among these patients; and (3) determining geospatial and temporal COVID-19–related measures and outcomes, and constructing their epidemic curves. We compared the results from synthetic data to those from original data using traditional statistics, machine learning approaches, and temporal and spatial representations of the data.

Results: For each use case, the results of the synthetic data analyses successfully mimicked those of the original data such that the distributions of the data were similar and the predictive models demonstrated comparable performance. Although the synthetic and original data yielded overall nearly the same results, there were exceptions that included an odds ratio on either side of the null in multivariable analyses (0.97 vs 1.01) and differences in the magnitude of epidemic curves constructed for zip codes with low population counts.

Conclusions: This paper presents the results of each use case and outlines key considerations for the use of synthetic data, examining their role in collaborative research for faster insights.

(J Med Internet Res 2021;23(10):e30697) doi:10.2196/30697

KEYWORDS

synthetic data; protected health information; COVID-19; electronic health records and systems; data analysis

Introduction

COVID-19 presents data and knowledge sharing challenges [1]. Clinical data exist at individual institutions; however, these data are rarely shared with external entities. Big data from multiple institutions allow for more comprehensive analyses, particularly for characterizing rare outcomes [2,3]. In response to this need, the National COVID Cohort Collaborative (N3C), an open science community, was formed to ingest and harmonize COVID-19 data from institutions across the United States [4]. The N3C sought a solution to preserve the privacy and confidentiality of these clinical data while enabling their broad dissemination [5-7] and partnered with MDClone (Beer Sheva, Israel) to computationally derive "synthetic" N3C data and support the rapid advancement of population health insights [8].

Since synthetic derivatives of data can enable privacy-preserving data downloads and accelerate discovery, these data assets can potentially be of great utility to the N3C and the broader informatics community. Our synthetic data validation workstream was established to assist the N3C community in better understanding the utility of synthetic data for research purposes. Our previous work demonstrated statistical equivalency between original and computationally derived data sets from a local instance of MDClone [9]. We also used synthetic data sets exclusively to apply machine learning to predict decompensation in heart failure [10]. Others have demonstrated repeatedly generated synthetic data sets from MDClone produced stable results that were similar to the original data [11].

However, the performance of MDClone—the comparison of original to synthetic data—has not been validated using data comprising multiple sources such as those originating across health systems. To that end, we tested three use cases, including (1) exploring the distributions of key features of the COVID-19–positive cohort; (2) training and testing predictive models for assessing the risk of admission among these patients; and (3) determining geospatial and temporal COVID-19–related measures and outcomes, and constructing their epidemic curves. We analyzed data for each use case using original and synthetic data. We conducted analyses using traditional statistics, machine learning approaches, and temporal and spatial representations of the data. Here we present the results of these analyses and describe the strengths and limitations of using synthetic data for research.

Methods

Overview

The Clinical and Translational Science Award Program developed N3C in response to a need for integrating, harmonizing, and democratizing individual-level COVID-19 data [4]. The N3C established a secure data enclave to store data and conduct collaborative analytics. The subsequent analyses resulted from a synthetic data pilot designed to evaluate the utility of computationally derived data for the N3C community. Synthetic data generation represents an emerging technology that can support population health research at scale. As described in more detail elsewhere [9], MDClone uses a computational derivation approach. Briefly, novel data whose features are queried independently for each distinct use case are produced in a multidimensional space that adheres to the statistical properties of the original source data. MDClone censors categorical values that are unique to few patients by removing the value and replacing it with the word "censored" in the computationally derived data set. Extreme numerical values also do not appear in the synthetic data set. Together, these approaches ensure that outliers in the original data set will not be identifiable in the synthetic data derivative.

All analyses were conducted using original data and computationally derived data, respectively, which allowed us to compare the results of analyses and assess the strengths and limitations of leveraging synthetic data for COVID-19 insights. All statistical analyses on the original and synthetic data sets were done outside of MDClone on the Palantir Foundry Analytic Platform (Palantir Technologies).

Analyses were conducted using Python (3.6.10l Python Software Foundation). We obtained institutional review board approval from our institutions for these analyses, in addition to completing data use agreements and requests with the National Center for Advancing Translational Sciences at the National Institutes of Health.

Use Case 1: Exploring the Distributions of Key Features of the COVID-19–Positive Cohort

The goal of this use case was to evaluate whether synthetic data had similar distributions of demographic and clinical characteristics among the COVID-19–positive cohort as compared to original data. Key characteristics (n=15) of the COVID-19–positive cohort were extracted from MDClone to compare distributions between the synthetic and the original data. The 15 features included age, gender, race, patients' state of residence, institution, median household income, BMI, number of days between testing positive and hospital admission (if hospitalized), diagnosis of diabetes, dyspnea, chronic kidney disease (CKD), fever, cough, and in-hospital mortality. We calculated mean and SD for continuous variables, and counts and proportions for categorical variables.

Use Case 2: Training and Testing Predictive Models for Assessing the Risk of Admission Among COVID-19–Positive Patients

The goal of this use case was to evaluate whether synthetic data would perform similarly when training and testing predictive machine learning models on synthetic data as compared to training and testing the models on original data. We included 230,703 patients who tested positive for COVID-19. Features for predictive modeling included 11 variables: age, gender, race, median household income, BMI, minimum oxygen saturation, diabetes, dyspnea, CKD, fever, and cough. These variables were chosen because of initial data suggesting their significant impact on COVID-19 outcomes.

We calculated odds ratios (ORs) and 95% CIs for admission within 14 days of a COVID-19 diagnosis by univariate logistic regression (LR) and multivariable LR using synthetic and original data, respectively. We then developed two widely used

machine learning models, random forest (RF) and LR, to predict admission within 14 days of a COVID-19 diagnosis based on the 11 features. We randomly split the cohort into training (80%) and testing (20%) data. The models were trained on the 80% subset of the data and then tested on the remaining 20%. We used a variety of metrics, including accuracy, precision, recall, F1-score, area under the receiver operating characteristic (ROC) curve, and precision-recall curves to evaluate model performance. Each model was trained and evaluated on the synthetic data set, the results of which were then compared against a model trained and evaluated on the original data.

Use Case 3: Determining Geospatial and Temporal COVID-19–Related Measures and Outcomes, and Constructing Their Epidemiologic Curves

The purpose of this analysis was to assess concordance of geospatial and temporal relationships between the synthetic and original data to make the data actionable and interpretable according to geography and time. Our data sets (original: n=1,854,968 tests; synthetic: n=1,854,950 tests) were event-based with each row representing a patient's first COVID-19 test result. The data sets included the following variables: source partner with which the patient was affiliated; lab test result (negative/positive); lab test date and time (reference time point for data generation); age at confirmed lab test result; admission start date (days from reference if admission occurred within ±7 days of COVID-19–positive test result); admission length of stay (in days); death (yes/null) during admission; patient's state of residence; patient's 5-digit zip code; and median household income, percent of residents under the poverty line, percent without health insurance, and total population by zip code.

On both the synthetic and original data sets, we calculated the aggregate count, 7-day midpoint moving average, and 7-day slope (count – count 6 days prior) per day for positive tests. We then plotted epidemic curves (Plotly version 4.14.1, Plotly Technologies Inc) for positive tests with synthetic and original data overlaid in the same figure. To test for significant differences or equivalence between the synthetic and original data epidemic curves, the paired two-sided *t* test (scipy version 1.5.3, stats.ttest_rel) and two-sided Wilcoxon signed rank test (scipy version 1.5.3, stats.wilcoxon) were run for each metric (count, 7-day moving average, and 7-day slope) treating the counts for individual dates as pairs.

Next, we calculated the differences in the mean, SD, median, IQR, and missingness of zip code–level social determinants of health (SDOH) variables within the original data set. We then compared these original data SDOH values for unique zip codes in the original data that were censored versus uncensored in the synthetic data. We defined censored zip codes as those present within the original data set that could not be matched (n=11,222) within the synthetic data set either due to not being present or being labeled as *censored* within the synthetic data set. We defined uncensored zip codes as present within both the synthetic data and original data (n=5819).

Ethics

This study was reviewed and approved by the Washington University in St. Louis' and the University of Washington's institutional review boards.

Results

Use Case 1: Exploring the Distributions of Key Features of the COVID-19–Positive Cohort

The MDClone synthetic data process generated 230,650 participants, compared to 230,703 in the original data. Demographic and clinical variables comparing synthetic and original data sets are displayed in Table 1. The mean age from both data sources was the same (mean 41.6, SD 20.4 years; Table 1). Approximately 47% of patients were male and 53% were White in both data sources. The values of all means and SDs (or counts and percentages) were the same or very similar between original and synthetic data. Table 1 shows that the distribution of demographic and clinical variables was similar between original and synthetic populations.

Table 1. Comparison of patient characteristics of available demographic and clinical variables: original vs synthetic data.

	Original data (n=230,703)	Synthetic data (n=230,650)
Age (years), mean (SD)	41.6 (20.4)	41.6 (20.4)
Gender (male), n (%)	108,194 (46.9)	107,892 (46.8)
Race, n (%)		
White	121,706 (52.8)	121,564 (52.7)
Black	40,930 (17.7)	40,824 (17.7)
Asian	5203 (2.3)	5117 (2.2)
Other/unknown	62,864 (27.2)	62,733 (27.2)
Top 5 most prevalent states, n (%)		
1	29,875 (12.9)	28,617 (12.4)
2	21,191 (9.2)	20,671 (9.0)
3	21,045 (9.1)	20,319 (9.0)
4	18,006 (7.8)	16,998 (7.4)
5	14,391 (6.2)	13,840 (6.0)
Top 5 most prevalent institutions, n (%)		
1	33,413 (14.5)	32,743 (14.2)
2	24,533 (10.6)	23,986 (10.4)
3	15,578 (6.8)	15,065 (6.5)
4	11,870 (5.1)	11,255 (4.9)
5	11,354 (4.9)	10,850 (4.7)
Household income (US $), median (IQR)	56,738 (45,214, 71,250)	56,662 (45,223, 71,029)
BMI, mean (SD)	30.3 (8.4)	30.3 (8.2)
Admission start date (days from reference), mean (SD)	2.1 (3.3)	2.0 (3.2)
Minimum oxygen saturation, mean (SD)	90.9 (10.1)	91.0 (9.7)
Diabetes, n (%)	31,942 (13.8)	31,929 (13.8)
Dyspnea, n (%)	20,867 (9.0)	20,826 (9.0)
Chronic kidney disease, n (%)	11,225 (4.9)	11,194 (4.9)
Fever, n (%)	30,210 (13.1)	30,200 (13.1)
Cough, n (%)	39,703 (17.2)	39,689 (17.2)
Deceased, n (%)	1133 (0.5)	1008 (0.4)

Use Case 2: Training and Testing Predictive Models for Assessing the Risk of Admission Among COVID-19–Positive Patients

Features (n=11) used for prediction included age, gender, race, median household income, BMI, minimum oxygen saturation, diagnosis of diabetes, dyspnea, CKD, fever, and cough. Table 2 shows the OR for admission and for each of the 11 variables by univariable LR yielded by original and synthetic data sources, respectively. The comparison of ORs between original and synthetic data sources show that the values for all 11 features were the same or similar. For example, the OR for admission by age from the original data was 1.04 (95% CI 1.04-1.04), which was the same as that obtained from synthetic data.

Table 2. Logistic regression for admission: original vs synthetic data.

	Univariate LR[a], OR[b] (95% CI)		Multivariable LR, OR (95% CI)	
	Original data	Synthetic data	Original data	Synthetic data
Age	1.04 (1.04-1.04)	1.04 (1.04-1.04)	1.00 (1.00-1.00)	1.00 (1.00-1.00)
Male gender	1.20 (1.16-1.24)	1.14 (1.10-1.17)	1.11 (0.99-1.23)	1.03 (0.93-1.15)
Black race	2.15 (2.07-2.22)	2.09 (2.02-2.17)	0.99 (0.87-1.12)	0.93 (0.82-1.06)
Median household income	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)
BMI	1.02 (1.01-1.02)	1.02 (1.01-1.02)	0.97 (0.97-0.98)	1.01 (1.00-1.02)
Minimum oxygen saturation	0.97 (0.96-0.97)	0.97 (0.96-0.97)	0.97 (0.97-0.98)	0.97 (0.97-0.98)
Diabetes	6.14 (5.94-6.34)	6.15 (5.95-6.36)	1.45 (1.29-1.62)	1.46 (1.30-1.63)
Dyspnea	4.79 (4.62-4.97)	4.79 (4.61-4.97)	1.23 (1.09-1.38)	1.25 (1.11-1.41)
Chronic kidney disease	7.20 (6.89-7.52)	7.17 (6.87-7.49)	1.23 (1.07-1.42)	1.26 (1.09-1.45)
Fever	2.62 (2.52-2.71)	2.62 (2.53-2.72)	1.44 (1.29-1.61)	1.45 (1.30-1.62)
Cough	1.38 (1.33-1.43)	1.38 (1.32-1.43)	1.50 (1.32-1.70)	1.45 (1.28-1.65)

[a]LR: logistic regression.

[b]OR: odds ratio.

The comparison of ORs between original and synthetic data sources shows that the multivariable LR yielded the same or similar results. For example, the OR for admission by Black race from the original data was 0.99 (95% CI 0.87-1.12), which was similar to that obtained from synthetic data (OR 0.93, 95% CI 0.82-1.06). Of note, the ORs that corresponded to a one-unit increase in BMI were on either side of the null (0.97 vs 1.01).

The machine learning models that were trained and tested on original data and then trained and tested on synthetic data used the same 11 features. Figure 1 shows the comparison of model prediction performance using original and synthetic data, respectively. We found the RF model achieved an under the ROC curve of 0.814 (0.816 by LR) using original data, and 0.812 (0.815 by LR) using synthetic data (Figure 1 A and C). Meanwhile, the RF model achieved an average precision of 0.298 (0.286 by LR) with original data and 0.308 (0.278 by LR) with synthetic data (Figure 1 B and D).

Figure 2 shows additional metrics for the evaluation of model performance. We observed the same or similar patterns by accuracy, specificity, precision, sensitivity, and F1-score when comparing models that were trained and tested on original data as compared to those trained and tested on synthetic data.

Figure 3 shows the feature importance according to RF (Figure 3 A) and LR models (Figure 3 B) using original (magenta) and synthetic (blue) data. Both the RF and LR models' demonstrated that features such as age, income, and minimum oxygen saturation were high-ranking informative features.

Figure 1. Prediction performance for the two models by receiver operating characteristic curves (A, C) and precision-recall curves (B, D) by using original and synthetic data. Results for the RF model are in the first row (A, B); the second row (C, D) is for LR. AUC: area under the curve; LR: logistic regression; RF: random forest.

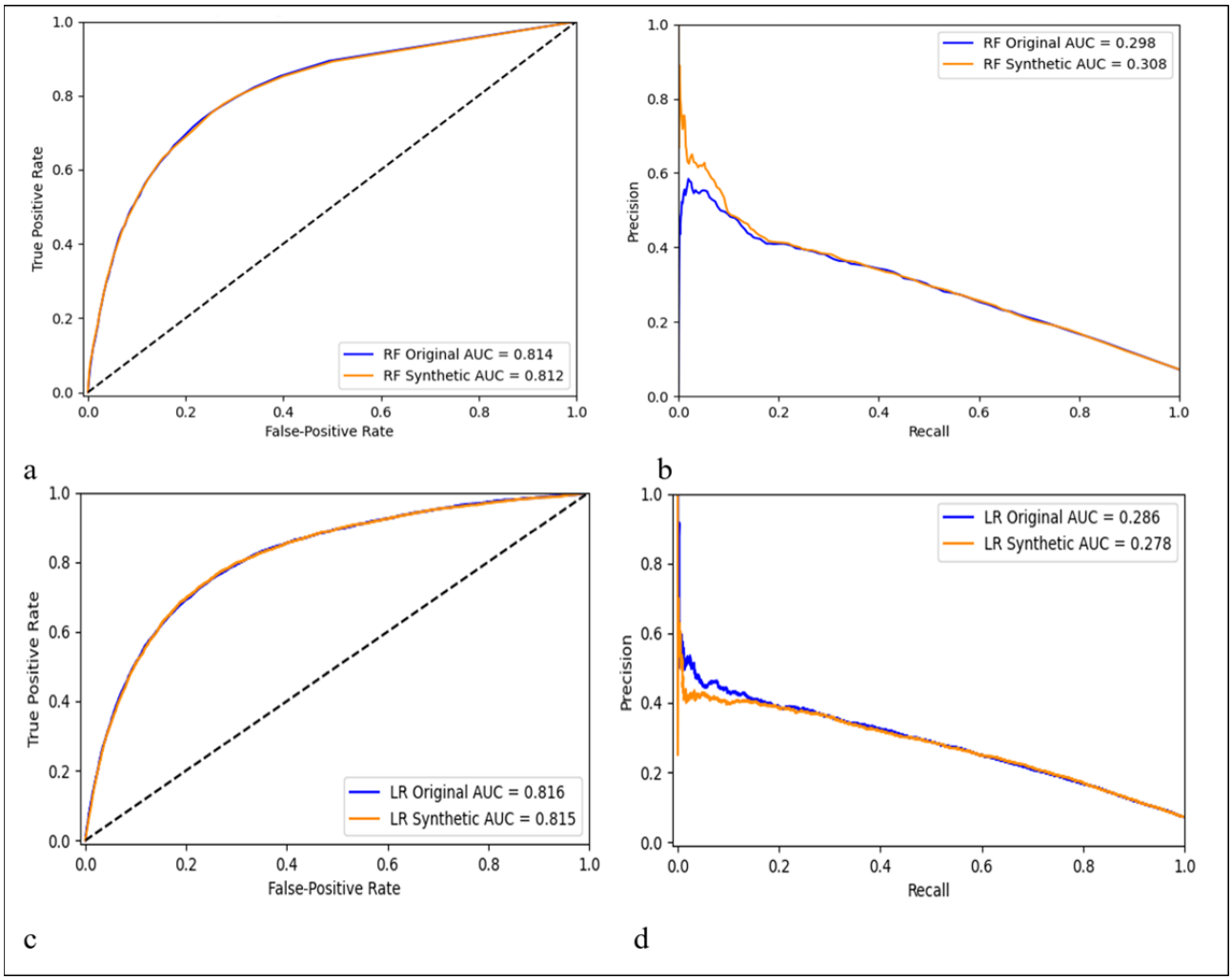


Figure 2. Model performance metrics from original (green) and synthetic (gold) data by accuracy, specificity, precision, sensitivity, and F1-score: RF model (A) and LR model (B). LR: logistic regression; RF: random forest.

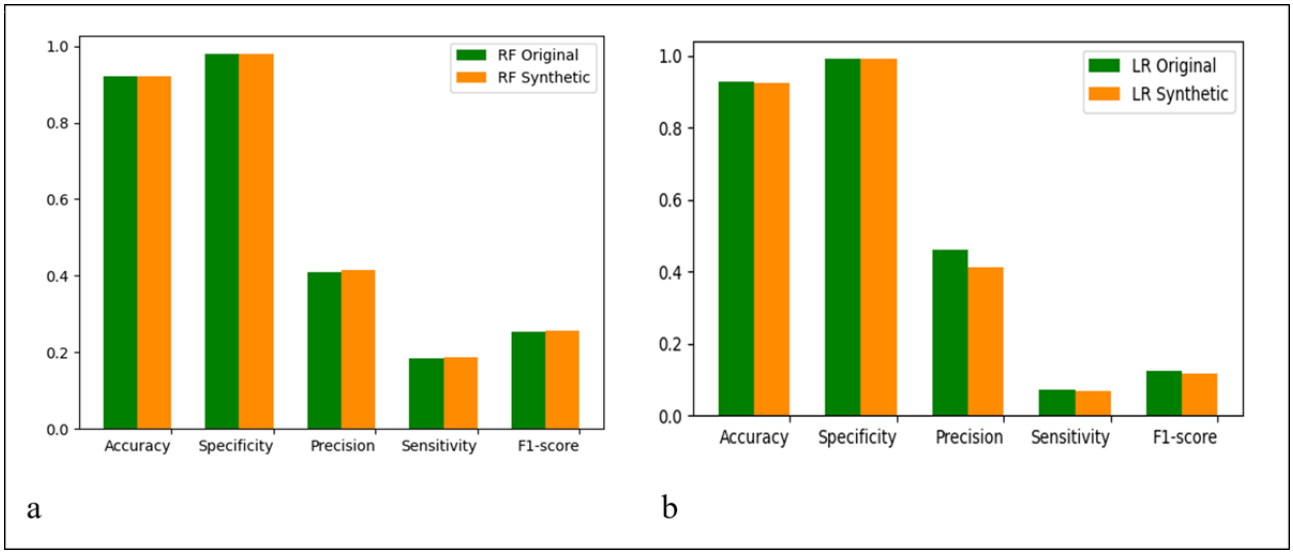
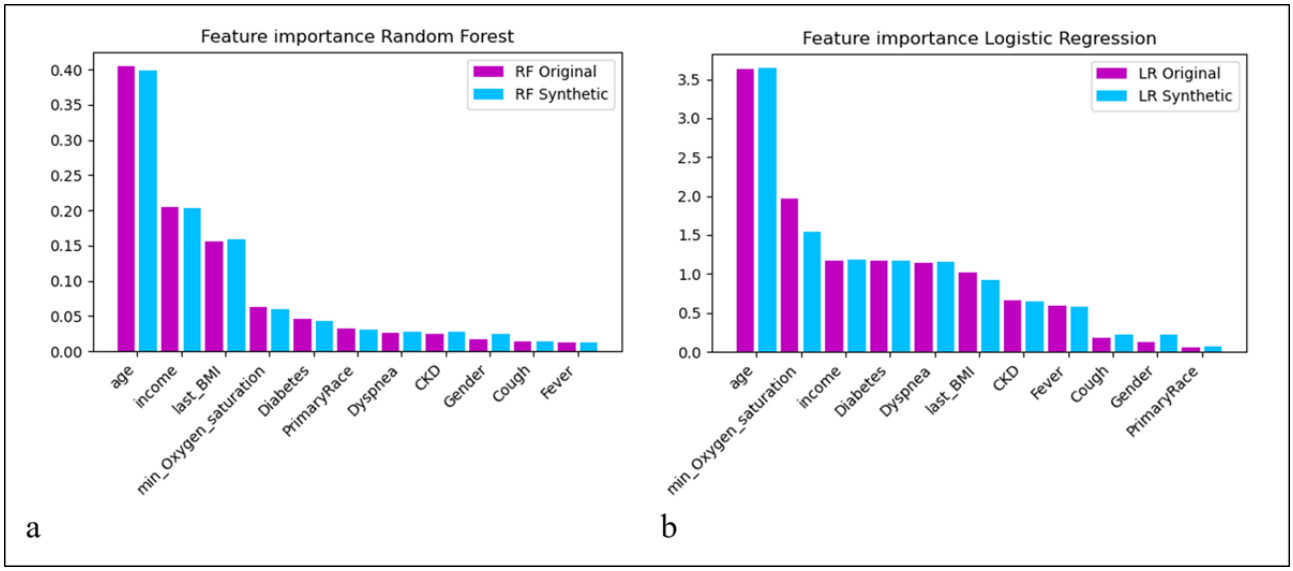


Figure 3. Feature importance for the 11 variables in RF (a) and LR (b) models: original vs synthetic data. CKD: chronic kidney disease; LR: logistic regression; RF: random forest.



Use Case 3: Determining the Zip Code–Level Distributions of COVID-19–Related Outcomes and Calculating Their Epidemiologic Curves

A graphical comparison of the epidemic curves for aggregate positive tests (cases) between the synthetic and original data is shown in Figure 4. Pairwise statistics for the epidemic curve metrics are shown in Table 3; no significant differences were found between the synthetic and original data epidemic curves across all metrics (Wilcoxon signed rank test *P* value range .50-.90; Student paired *t* test *P* value range .996-.998).

Compared to censored zip codes, uncensored zip codes had a higher median household income, a lower percentage of residents under the poverty line, a lower percentage of patients without health insurance, a higher total population, and fewer missing values for all four SDOH. Total population and data missingness were the two greatest differences between uncensored and censored zip codes. Uncensored zip codes had a 74% higher median total population and had approximately 70% fewer missing SDOH values than censored zip codes (Table 4).

Figure 4. Original data (light blue) and synthetic data (light red), with their overlap (purple).

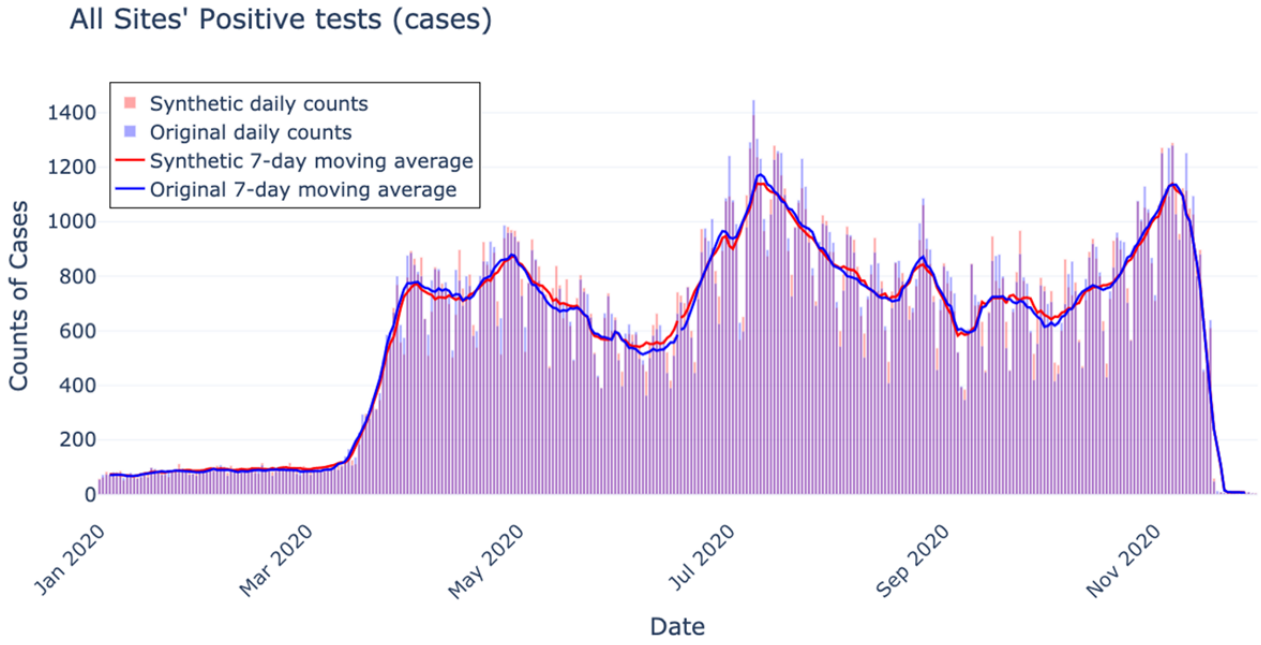


Table 3. Epidemic curves aggregate cases' paired statistical tests, comparing original to synthetic data.

Metric	Date range	Wilcoxon result	Wilcoxon *P* value	*t* statistic	*t* test *P* value
Counts	335	26,288	.50	–0.002	>.99
7-day moving average	329	26,005	.78	–0.006	>.99
7-day slope	329	25,788.5	.90	–0.002	>.99

Table 4. SDOH values for zip codes that were uncensored (n=5819) compared to censored (n=11,222) zip codes.

SDOH[a] and censored status	Mean	SD	Median	IQR	% missing
Median household income (US $)					
Uncensored	63,536	26,755	57,352	28,692	3.28
Censored	60,544	26,549	54,358	27,067	10.98
Difference (%)	+2992 (4.9)	+206 (0.8)	+2994 (5.5)	+1625 (6.0)	–7.70 (70.1)
Percent under poverty line					
Uncensored	12.89	8.74	10.80	10.40	2.92
Censored	13.87	10.15	11.60	11.50	9.12
Difference (%)	–0.98 (7.1)	–1.41 (13.9)	–0.80 (6.9)	–1.10 (9.6)	–6.20 (68.0)
Percent without health insurance					
Uncensored	8.52	5.09	7.50	6.50	2.84
Censored	9.65	7.09	8.10	8.00	9.00
Difference (%)	–1.13 (11.7)	–2.00 (28.2)	–0.60 (7.4)	–1.50 (18.8)	–6.16 (68.4)
Total population					
Uncensored	17,363	16,128	12,263	23,172	2.73
Censored	14,540	17,317	7048	21,436	8.69
Difference (%)	+2823 (19.4)	–1189 (6.9)	+5215 (74.0)	+1736 (8.1)	–5.96 (68.6)

[a]SDOH: social determinants of health.

Discussion

Principal Findings

Our main findings demonstrated that computationally derived data had the same or similar statistical output as the original data sets, with the caveat that zip codes with a lower population had data suppressed/censored for privacy reasons more often than zip codes with a higher population. In each use case, the results of the analyses appear sufficiently similar between the synthetic derivative and the original data across the various methods used to assess similarity (means, medians, *P*>.05, overlapping CI, etc) to draw the same conclusions with the exception of one OR on either side of the null in multivariable analyses (0.97 vs 1.01). In several instances, the results were exactly the same and rarely were there statistically significant differences between data sets.

Small sample sizes, missing values, and high dimensionality can all adversely affect the data synthesis process and the precision and interpretability of original data. Our geospatial analysis shows that zip codes that are censored to protect patient privacy have a lower population, which will likely make using these computationally derived data to study rural populations more challenging. Additionally, the lower original data quality found within censored zip codes—seen in greater SDOH missingness—as compared to uncensored may indicate broader data quality issues in rural zip code data. Such issues may pose a further challenge to data synthesis.

This was the first validation of computationally derived data using the N3C data. Our study adds to the growing literature of synthetic data validation in the following ways. First, our study is the first assessment of N3C synthetic data utility and has been conducted prior to the broad dissemination of N3C synthetic data. Thus, our study provides insight to the validity of N3C synthetic data prior to its dissemination for use by the broader N3C community. Second, our results from use case 3 support the temporal validity of these computationally derived data as an alternative to date-shifting when privacy must be protected yet temporality maintained.

For these descriptive and quantitative analyses, the synthetic data appear to produce similar patterns and results compared to the original data, except for in the context of high missingness. We acknowledge that these use cases may not represent all possible ways in which the synthetic data may be used by the N3C community and thus validation should continue. In addition to continuously validating these data for different use cases and analytic methods, we seek to explore the performance of other commercial systems in the N3C community and their approaches

to synthetic data generation and the privacy-preserving aspects of each approach.

We also suggest that the synthetic data can be used by researchers for hypothesis generation to then be validated later on original data. Another potential use case that could be valuable to the N3C community, which we do not explicitly test here, is the potential for synthetic data to be used for software engineering projects that seek to develop digital health tools for combating the COVID-19 pandemic. Computationally derived data that are faithful to the original data could be used to develop and test such tools.

Limitations

For these analyses, we compared the data statistically and did not conduct privacy evaluations of the synthetic data that will be a focus of future investigations. We used a *P* value threshold of .05 to maintain simplicity of presenting results from multiple use cases. We acknowledge that such thresholds would (and should) vary by use case and specifically by the amount of error a researcher is willing to tolerate given the context of the research question. We also acknowledge that other statistical tests such as equivalence testing could be suitable to assess the equivalence of computationally derived data to original data. However, the threshold for equivalence will yet again depend on the use case.

Our geospatial and temporal analysis was limited in scope. Our work is ongoing, and future analyses will assess validity of other measures (eg, tests, admissions, deaths, or positivity) over time—both in aggregate and at the zip code level—in greater detail.

Conclusions

We conclude that the potential for leveraging synthetic data for the conduct of COVID-19 research in N3C is substantial. We expect that the use of synthetic data will accelerate the conduct of data-driven research studies across the community, as it will allow the N3C to overcome data sharing barriers and rapidly create COVID-19 analytic insights [4]. Future directions for this work include developing and validating additional clinical risk prediction models, using a larger repertoire of analytic methods, conducting geospatial and temporal analyses in greater detail and at the zip code level, and evaluating additional strengths and limitations of computationally derived data for research [1].

Clinical Relevance Statement

Data synthesis platforms like MDClone are expected to enhance the N3C community's ability to use clinical data for faster COVID-19 insights and reduce barriers to data access by multiple stakeholders.

Acknowledgments

The analyses described in this publication were conducted with data or tools accessed through the National Center for Advancing Translational Sciences (NCATS) National COVID Cohort Collaborative (N3C) Data Enclave [12] and supported by NCATS U24 TR002306. This research was possible because of the patients whose information is included within the data from participating organizations [13] and the organizations and scientists [14] who have contributed to the ongoing development of this community resource [4].

The N3C data transfer to NCATS is performed under a Johns Hopkins University Reliance Protocol IRB00249128 or individual site agreements with the National Institutes of Health (NIH). The N3C Data Enclave is managed under the authority of the NIH; information can be found at [15].

We gratefully acknowledge contributions from the following N3C core teams (leads noted with *):

- Principal investigators: Melissa A Haendel*, Christopher G Chute*, Kenneth R Gersing, and Anita Walden

- Workstream, subgroup, and administrative leaders: Melissa A Haendel*, Tellen D Bennett, Christopher G Chute, David A Eichmann, Justin Guinney, Warren A Kibbe, Hongfang Liu, Philip RO Payne, Emily R Pfaff, Peter N Robinson, Joel H Saltz, Heidi Spratt, Justin Starren, Christine Suver, Adam B Wilcox, Andrew E Williams, and Chunlei Wu

- Key liaisons at data partner sites

- Regulatory staff at data partner sites

- Individuals at the sites who are responsible for creating the data sets and submitting data to N3C

- Data ingest and harmonization team: Christopher G Chute*, Emily R Pfaff*, Davera Gabriel, Stephanie S Hong, Kristin Kostka, Harold P Lehmann, Richard A Moffitt, Michele Morris, Matvey B Palchuk, Xiaohan Tanner Zhang, and Richard L Zhu

- Phenotype team (individuals who create the scripts that the sites use to submit their data, based on the COVID-19 and long COVID definitions): Emily R Pfaff*, Benjamin Amor, Mark M Bissell, Marshall Clark, Andrew T Girvin, Stephanie S Hong, Kristin Kostka, Adam M Lee, Robert T Miller, Michele Morris, Matvey B Palchuk, and Kellie M Walters

- Project management and operations team: Anita Walden*, Yooree Chae, Connor Cook, Alexandra Dest, Racquel R Dietz, Thomas Dillon, Patricia A Francis, Rafael Fuentes, Alexis Graves, Julie A McMurry, Andrew J Neumann, Shawn T O'Neil, Usman Sheikh, Andréa M Volz, and Elizabeth Zampino

- Partners from the NIH and other federal agencies: Christopher P Austin*, Kenneth R Gersing*, Samuel Bozzette, Mariam Deacy, Nicole Garbarini, Michael G Kurilla, Sam G Michael, Joni L Rutter, and Meredith Temple-O'Connor

- Analytics team (individuals who build the Enclave infrastructure, help create code sets, variables, and help domain teams and project teams with their data sets): Benjamin Amor*, Mark M Bissell, Katie Rebecca Bradwell, Andrew T Girvin, Amin Manna, and Nabeel Qureshi

- Publication committee management team: Mary Morrison Saltz*, Christine Suver*, Christopher G Chute, Melissa A Haendel, Julie A McMurry, Andréa M Volz, and Anita Walden

- Publication committee review team: Carolyn Bramante, Jeremy Richard Harper, Wenndy Hernandez, Farrukh M Koraishy, Federico Mariona, Saidulu Mattapally, Amit Saha, and Satyanarayana Vedula

- Synthetic data domain team: Yujuan Fu, Nisha Mathews, and Ofer Mendelevitch

Data was provided from the following institutions: Stony Brook University (U24TR002306); University of Oklahoma Health Sciences Center (U54GM104938: Oklahoma Clinical and Translational Science Institute); West Virginia University (U54GM104942: West Virginia Clinical and Translational Science Institute); University of Mississippi Medical Center (U54GM115428: Mississippi Center for Clinical and Translational Research); University of Nebraska Medical Center (U54GM115458: Great Plains IDeA-Clinical & Translational Research); Maine Medical Center (U54GM115516: Northern New England Clinical & Translational Research Network); Wake Forest University Health Sciences (UL1TR001420: Wake Forest Clinical and Translational Science Institute); Northwestern University at Chicago (UL1TR001422: Northwestern University Clinical and Translational Science Institute); University of Cincinnati (UL1TR001425: Center for Clinical and Translational Science and Training); The University of Texas Medical Branch at Galveston (UL1TR001439: The Institute for Translational Sciences); Medical University of South Carolina (UL1TR001450: South Carolina Clinical & Translational Research Institute); University of Massachusetts Medical School Worcester (UL1TR001453: The UMass Center for Clinical and Translational Science); University of Southern California (UL1TR001855: The Southern California Clinical and Translational Science Institute); Columbia University Irving Medical Center (UL1TR001873: Irving Institute for Clinical and Translational Research); George Washington Children's Research Institute (UL1TR001876: Clinical and Translational Science Institute at Children's National); University of Kentucky (UL1TR001998: UK Center for Clinical and Translational Science); University of Rochester (UL1TR002001: UR Clinical & Translational Science Institute); University of Illinois at Chicago (UL1TR002003: UIC Center for Clinical and Translational Science); Penn State Health Milton S. Hershey Medical Center (UL1TR002014: Penn State Clinical and Translational Science Institute); The University of Michigan at Ann Arbor (UL1TR002240: Michigan Institute for Clinical and Health Research); Vanderbilt University Medical Center (UL1TR002243: Vanderbilt Institute for Clinical and Translational Research); University of Washington (UL1TR002319: Institute of Translational Health Sciences); Washington University in St. Louis (UL1TR002345: Institute of Clinical and Translational Sciences); Oregon Health & Science University (UL1TR002369: Oregon Clinical and Translational Research Institute); University of Wisconsin-Madison (UL1TR002373: UW Institute for Clinical and Translational Research); Rush University Medical Center (UL1TR002389: The Institute for Translational Medicine [ITM]); The University of Chicago (UL1TR002389: ITM); University of North Carolina at Chapel Hill (UL1TR002489: North Carolina Translational and Clinical Science Institute); University of Minnesota (UL1TR002494: Clinical and Translational Science Institute); Children's Hospital Colorado (UL1TR002535: Colorado Clinical and Translational Sciences Institute); The University of Iowa (UL1TR002537: Institute for Clinical and Translational Science); The University of Utah (UL1TR002538: Uhealth Center for Clinical and Translational Science); Tufts Medical Center (UL1TR002544: Tufts Clinical and Translational Science Institute); Duke University (UL1TR002553: Duke Clinical and Translational Science Institute); Virginia Commonwealth University (UL1TR002649: C. Kenneth and Dianne Wright Center for Clinical and Translational Research); The Ohio State University (UL1TR002733: Center for Clinical and Translational Science); The University of Miami Leonard M. Miller School of Medicine (UL1TR002736: University of Miami Clinical and Translational Science Institute); University of Virginia (UL1TR003015: iTHRIVL Integrated Translational Health Research Institute of Virginia, Carilion Clinic (UL1TR003015: iTHRIVL Integrated Translational Health Research Institute of Virginia); University of Alabama at Birmingham (UL1TR003096: Center for Clinical and Translational Science); Johns Hopkins University (UL1TR003098: Johns Hopkins Institute for Clinical and Translational Research); University of Arkansas for Medical Sciences (UL1TR003107: UAMS Translational Research Institute); Nemours (U54GM104941: Delaware CTR ACCEL Program); University Medical Center New Orleans (U54GM104940: Louisiana Clinical and Translational Science Center); University of Colorado Denver, Anschutz Medical Campus (UL1TR002535: Colorado Clinical and Translational Sciences Institute); Mayo Clinic Rochester (UL1TR002377: Mayo Clinic Center for Clinical and Translational Science); Tulane University (UL1TR003096: Center for Clinical and Translational Science); Loyola University Medical Center (UL1TR002389: ITM); Advocate Health Care Network (UL1TR002389: ITM); and OCHIN (INV-018455: Bill and Melinda Gates Foundation grant to Sage Bionetworks).

The following are additional data partners who have signed the Data Transfer Agreement and data release pending: The Rockefeller University (UL1TR001866: Center for Clinical and Translational Science); The Scripps Research Institute (UL1TR002550: Scripps Research Translational Institute); University of Texas Health Science Center at San Antonio (UL1TR002645: Institute for Integration of Medicine and Science); The University of Texas Health Science Center at Houston (UL1TR003167: Center

for Clinical and Translational Sciences); NorthShore University HealthSystem (UL1TR002389: ITM); Yale New Haven Hospital (UL1TR001863: Yale Center for Clinical Investigation); Emory University (UL1TR002378: Georgia Clinical and Translational Science Alliance); Weill Medical College of Cornell University (UL1TR002384: Weill Cornell Medicine Clinical and Translational Science Center); Montefiore Medical Center (UL1TR002556: Institute for Clinical and Translational Research at Einstein and Montefiore); Medical College of Wisconsin (UL1TR001436: Clinical and Translational Science Institute of Southeast Wisconsin); University of New Mexico Health Sciences Center (UL1TR001449: University of New Mexico Clinical and Translational Science Center); George Washington University (UL1TR001876: Clinical and Translational Science Institute at Children's National); Stanford University (UL1TR003142: Spectrum: The Stanford Center for Clinical and Translational Research and Education); Regenstrief Institute (UL1TR002529: Indiana Clinical and Translational Science Institute); Cincinnati Children's Hospital Medical Center (UL1TR001425: Center for Clinical and Translational Science and Training); Boston University Medical Campus (UL1TR001430: Boston University Clinical and Translational Science Institute); The State University of New York at Buffalo (UL1TR001412: Clinical and Translational Science Institute); Aurora Health Care (UL1TR002373: Wisconsin Network For Health Research); Brown University (U54GM115677: Advance Clinical Translational Research); Rutgers, The State University of New Jersey (UL1TR003017: New Jersey Alliance for Clinical and Translational Science); Loyola University Chicago (UL1TR002389: ITM); New York University (UL1TR001445: Langone Health's Clinical and Translational Science Institute); Children's Hospital of Philadelphia (UL1TR001878: Institute for Translational Medicine and Therapeutics); University of Kansas Medical Center (UL1TR002366: Frontiers: University of Kansas Clinical and Translational Science Institute); Massachusetts General Brigham (UL1TR002541: Harvard Catalyst); Icahn School of Medicine at Mount Sinai (UL1TR001433: ConduITS Institute for Translational Sciences); Ochsner Medical Center (U54GM104940: Louisiana Clinical and Translational Science Center); HonorHealth (None, voluntary); University of California, Irvine (UL1TR001414: The UC Irvine Institute for Clinical and Translational Science); University of California, San Diego (UL1TR001442: Altman Clinical and Translational Research Institute); University of California, Davis (UL1TR001860: UCDavis Health Clinical and Translational Science Center); University of California, San Francisco (UL1TR001872: UCSF Clinical and Translational Science Institute); University of California, Los Angeles (UL1TR001881: UCLA Clinical Translational Science Institute); University of Vermont (U54GM115516: Northern New England Clinical & Translational Research Network); and Arkansas Children's Hospital (UL1TR003107: UAMS Translational Research Institute).

Authors' Contributions

RF, JT, and AW contributed to the study conception and design; RF contributed to the generation of the data; AG and JT contributed to data analysis; authors RF, JT, AW, PROP, and NZ contributed to the writing and editing of the paper.

Conflicts of Interest

NZ is employed by MDClone Ltd. All other authors declare no competing interests.

References

1. Foraker RE, Lai AM, Kannampallil TG, Woeltje KF, Trolard AM, Payne PRO. Transmission dynamics: data sharing in the COVID-19 era. Learn Health Syst 2020 Jun 28:e10235. [doi: 10.1002/lrh2.10235] [Medline: 32838037]
2. Foraker R, Mann DL, Payne PRO. Are synthetic data derivatives the future of translational medicine? JACC Basic Transl Sci 2018 Oct;3(5):716-718 [FREE Full text] [doi: 10.1016/j.jacbts.2018.08.007] [Medline: 30456342]
3. Nair S, Hsu D, Celi LA. Challenges and opportunities in secondary analyses of electronic health record data. In: Secondary Analysis of Electronic Health Records. Cham: Springer; 2016:17-26.
4. Haendel MA, Chute CG, Bennett TD, Eichmann DA, Guinney J, Kibbe WA, N3C Consortium. The National COVID Cohort Collaborative (N3C): rationale, design, infrastructure, and deployment. J Am Med Inform Assoc 2021 Mar 01;28(3):427-443 [FREE Full text] [doi: 10.1093/jamia/ocaa196] [Medline: 32805036]
5. Federal Policy for the Protection of Human Subjects ('Common Rule'). US Department of Health and Human Services. URL: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html [accessed 2019-03-11]
6. The HIPAA Privacy Rule. US Department of Health and Human Services. URL: https://tinyurl.com/rnhxnbf5 [accessed 2019-03-11]
7. Miller AR, Tucker C. Privacy protection and technology diffusion: the case of electronic medical records. Manage Sci 2009 Jul;55(7):1077-1093. [doi: 10.1287/mnsc.1090.1014]
8. Crawford R, Bishop M, Bhumiratana B, Clark L, Levitt K. Sanitization models and their limitations. In: Proceedings of the 2006 workshop on New security paradigms. 2006 Presented at: NSPW '06; September 19-22, 2006; Germany. [doi: 10.1145/1278940.1278948]
9. Foraker RE, Yu SC, Gupta A, Michelson AP, Pineda Soto JA, Colvin R, et al. Spot the difference: comparing results of analyses from real patient data and synthetic derivatives. JAMIA Open 2020 Dec;3(4):557-566 [FREE Full text] [doi: 10.1093/jamiaopen/ooaa060] [Medline: 33623891]

10. Guo A, Foraker RE, MacGregor RM, Masood FM, Cupps BP, Pasque MK. The use of synthetic electronic health record data and deep learning to improve timing of high-risk heart failure surgical intervention by predicting proximity to catastrophic decompensation. Front Digital Health 2020 Dec 7;2:576945. [doi: 10.3389/fdgth.2020.576945]
11. Reiner Benaim A, Almog R, Gorelik Y, Hochberg I, Nassar L, Mashiach T, et al. Analyzing medical research results based on synthetic data and their relation to real data results: systematic comparison from five observational studies. JMIR Med Inform 2020 Feb 20;8(2):e16492 [FREE Full text] [doi: 10.2196/16492] [Medline: 32130148]
12. Enclave essentials. N3C. URL: https://covid.cd2h.org/enclave [accessed 2021-06-24]
13. Data Transfer Agreement Signatories. National Center for Advancing Translational Sciences. URL: https://covid.cd2h.org/dtas [accessed 2021-06-24]
14. DUA Signatories. N3C. URL: https://covid.cd2h.org/duas [accessed 2021-06-24]
15. National COVID Cohort Collaborative Forms and Resources. National Center for Advancing Translational Sciences. URL: https://ncats.nih.gov/n3c/resources [accessed 2021-06-24]

Abbreviations

CKD: chronic kidney disease
ITM: Institute for Translational Medicine
LR: logistic regression
NCATS: National Center for Advancing Translational Sciences
NIH: National Institutes of Health
N3C: National COVID Cohort Collaborative
OR: odds ratio
RF: random forest
ROC: receiver operating characteristic
SDOH: social determinants of health

Edited by C Basch; submitted 03.06.21; peer-reviewed by M Salimi, S Molani; comments to author 03.08.21; revised version received 24.08.21; accepted 12.09.21; published 04.10.21.

Please cite as:
Foraker R, Guo A, Thomas J, Zamstein N, Payne PRO, Wilcox A, N3C Collaborative
The National COVID Cohort Collaborative: Analyses of Original and Computationally Derived Electronic Health Record Data
J Med Internet Res 2021;23(10):e30697
URL: https://www.jmir.org/2021/10/e30697
doi:10.2196/30697
PMID:34559671

Viewpoint

Positioning Digital Tracing Applications in the Management of the COVID-19 Pandemic in France

Marion Albouy-Llaty[1,2,3], MD, PhD; Caroline Martin[4,5]; Daniel Benamouzig[6,7], PhD; Eric Bothorel[8]; Gilles Munier[9], MD; Catherine Simonin[10]; Jean-Louis Guéant[11,12], MD, PhD; Emmanuel Rusch[13,14], MD, PhD

[1]French Public Health Society, Paris, France
[2]Clinical Investigation Centre of Poitiers University Hospital 1402, Inserm, Poitiers, France
[3]Faculty of Medicine and Pharmacy, University of Poitiers, Poitiers, France
[4]National Consultative Ethics Committee for Health and Life Sciences, Paris, France
[5]Conseil d'Etat, Paris, France
[6]French Covid-19 Scientific Council, Paris, France
[7]Sciences Po, Center for the Sociology of Organizations, Paris, France
[8]Assemblée Nationale (French Parliament), Paris, France
[9]National Council of the Physicians's College, Paris, France
[10]France-Assos Santé, Paris, France
[11]National Council of Medical Biology, Paris, France
[12]University of Lorraine, Inserm Unit 1256, Nancy, France
[13]National Health Conference, Paris, France
[14]Unit Education, Ethics and Health, University of Tours, Tours, France

Corresponding Author:
Marion Albouy-Llaty, MD, PhD
Faculty of Medicine and Pharmacy
University of Poitiers
6 rue de la Milétrie
Poitiers, 86021
France
Phone: 33 0669660192
Email: marion.albouy@univ-poitiers.fr

Abstract

To combat the COVID-19 pandemic, many European countries have developed a public health strategy involving the use of digital contact tracing (DCT) applications to improve timely tracking and contact tracing of COVID-19 cases. France's independent COVID-19 Control and Society Connection Council (CCL) was established by law in May 2020 to issue advice and recommendations on the national epidemic digital systems. In this paper, we present the recommendations by the CCL, with the objective to increase the uptake and utility of French DCT applications. As the country's most vulnerable population has been subjected to greater virus exposure, a stronger impact of the lockdown, and less access to preventive and health care services, the CCL is particularly aware of health inequalities. The French DCT app TousAntiCovid had been downloaded by 13.6 million users (ie, 20% of the French population) in March 2021. To promote the use of DCT apps, the CCL has recommended that communication about the app's individual and collective objectives be increased. The CCL has also recommended the introduction of clear, simple, accessible, incentivizing, noncoercive information within the digital tools. In addition, the CCL has recommended improving public health policies to address the needs of the underprivileged. The CCL calls for promoting population empowerment with the use of digital tools, improving public health culture for decision-makers dealing with health determinants, taking social considerations into account, and incorporating community participation.

(J Med Internet Res 2021;23(10):e27301) doi:10.2196/27301

KEYWORDS

COVID-19 pandemic; digital contact tracing applications; health inequalities; Europe; health promotion

Introduction

To combat the first wave of the COVID-19 pandemic, many European countries instituted a strict lockdown coupled with reverse-transcription polymerase chain reaction (RT-PCR) testing, which led to a relative decline in case fatality [1]. These countries also developed a public health strategy involving the use of digital contact tracing (DCT) applications to improve timely tracking and contact tracing [2,3]. Despite these measures, the second epidemic wave occurred in autumn 2020, more rapidly and with a higher severity than expected, leading to a second lockdown in many countries. France developed a national "Test, Trace and Isolate" strategy, but did not achieve its goal of less than 5000 confirmed COVID-19 cases per day by December 15, 2020. As of that date, incidence had increased, and a curfew went into effect first. Several weeks later, in facing a third wave, a lockdown was reinstituted by many countries.

Recommendations for DCT Use in France

France's independent COVID-19 Control and Society Connection Council (CCL) was established by law in May 2020 in order to issue advice and recommendations on the national epidemic digital systems. The CCL comprises 13 members, including representatives from the parliament and civil society, patient organizations, jurists, and academics from different disciplines (see Multimedia Appendix 1 [4]). Considering the characteristics of European DCT apps and their potential, we present the CCL's recommendations, with the objective to increase the uptake and utility of French DCT apps. As the country's most vulnerable population has been subjected to greater virus exposure, a stronger impact of the lockdown, and less access to preventive and health care services, the CCL is particularly aware of prevailing health inequalities.

Findings

Several mobile apps for contact tracing and information on public health policy have been developed in Europe (Table 1). These apps should respond to scientifically valid and time-bound ethical guidelines. Insufficient privacy protection could erode trust in the government and public health services [5,6]. Unfortunately, adoption of these apps has been below expectations [7]; although, adoption has been high in Finland (45%), it has been limited elsewhere (24% in Ireland, 21% in Switzerland, 19% in Germany, and 13% in Italy [8]; see Table 1). The French DCT *StopCovid* app was based on Bluetooth-based exposure notification. Conceived as a silent app (ie, without interactivity), it was replaced on October 22, 2020, by a conversational app renamed *TousAntiCovid*. Among 66 million inhabitants, only 1.5 million (2.3%) had downloaded the *StopCovid* app and, as of March 2021, this number increased to 13.6 million (ie, 20% of the French population). Moreover, positive test declarations (n=183,377), such as notifications of a contact (n=107,465) have been weak. The poor results of the StopCovid app stemmed from the lack of attractiveness and prevention information, a fear of insufficient privacy protection, Bluetooth connection problems, and the incompatibility of apps with older smartphones.

To promote the use of the apps, the CCL recommended increasing the communication about the individual and collective objectives of the *TousAntiCovid* app. The CCL also recommended the introduction of clear, simple, accessible, incentivizing, noncoercive information in the digital tool. Indeed, the messages aimed at emphasizing risk are less effective than those encouraging self- and collective efficacy [9]. Moreover, to stimulate DCT uptake, apps should enhance perceived benefits [10]. In addition, the CCL recommended improving public health policies to address the needs of the underprivileged population [11]. It bears mentioning that apps are effective only if users are comfortable with digital tools; technical problems are among the main reasons for failure to download [12]. It is also worth noting that the COVID-19 pandemic has accelerated digital health uptake: 49% of the French population used their first health digital tool during the first lockdown [13], whereas 16% of French population never connect to internet [14].

The CCL believes that the national "Test, Trace and Isolate" strategy needed consistency—testing without effective tracing and efficient isolation of positive cases makes little, if any, sense. In Europe and France, apps have progressively incorporated an isolation module (questions about needs during isolation). Their development should be aimed at measuring isolation prescription and adherence. Apps could help families to stay connected and maintain links to resources that support their physical and mental well-being [15]. Despite contrasting strategies of lockdown, no dramatic difference in the magnitude of the second epidemic wave seems to have been observed among these countries (Table 1). For persons with positive test results, the strategy adopted in most countries included a mandatory 7-day isolation until newer variants of SARS-CoV-2 were discovered [16].

Table 1. European data on COVID-19 mortality, government stringency responses, and digital contact tracing (DCT) applications as of December 15, 2020.

European country	Population	Unstandardized COVID-19 mortality (per million inhabitants)	Government response stringency index[a] [17]	DCT app name	Data destruction apps data[b]	Technology	Technology	Uptake rate (%)
Finland	5,540,720	83.2	44.91	Koronavilkku	Yes	Bluetooth	Google/Apple	45
Denmark	5,792,202	164.0	45.37	Smittestopp	Yes	Bluetooth	Google/Apple	__[c]
Switzerland	8,654,622	643.8	46.30	SwissCovid	Yes	Bluetooth	Google/Apple DP3T[d]	21
Estonia	1,326,535	116.1	48.15	Estonia's App	No	Bluetooth	Google/Apple	—
Norway	5,421,241	72.5	52.78	Smittestopp	Yes	Bluetooth and location	Other	—
United Kingdom	67,886,011	948.7	54.17	NHSCovid19	Yes	Bluetooth	Google/Apple	—
Netherlands	17,134,872	587.9	56.48	Coronamelder	Yes	Bluetooth	Google/Apple	—
Bulgaria	6,948,445	840.2	57.41	ViruSafe	Yes	Location	Other	—
Belgium	11,589,623	1557.8	60.19	Coronoalert	Yes	Bluetooth	Google/Apple DP3T	—
Germany	83,783,942	268.2	67.59	Corona-Warn-App	Yes	Bluetooth	Google/Apple	19
Cyprus	1,207,359	67.9	69.44	CovTracer	Yes	Location	Other	—
Poland	37,846,611	615.9	71.30	ProteGO	Yes	Bluetooth	Google/Apple	—
Spain	46,754,778	1026.9	71.30	RadarCOVID	Yes	Bluetooth	Google/Apple DP3T	—
Ireland	4,937,786	430.6	72.22	Covid Tracker	Yes	Bluetooth	Google/Apple	24
Hungary	9,660,351	749.1	72.22	VirusRadar	Yes	Bluetooth	Other	—
France	65,273,511	887.2	75.00	TousAntiCovid	Yes	Bluetooth	Other	19
Portugal	10,196,709	554.0	77.78	Stay away Covid	Yes	Bluetooth	Google/Apple DP3T	—
Italy	60,461,826	1075.2	79.63	Immuni	No	Bluetooth	Google/Apple	13
Austria	9,006,398	500.8	82.41[e]	Stopp Corona	Yes	Bluetooth	Google/Apple	—

[a]The Government Stringency Index is a composite measure based on 9 response indicators, including school closures, workplace closures, and travel bans, rescaled to a value ranging from 0 to 100 (100=strictest).

[b]Data were extracted several days after their acquisition.

[c]Not available.

[d]DP3T: decentralized privacy-preserving proximity tracing.

[e]Data as of December 8, 2020.

Discussion

There are several challenges in motivating populations to download DCT apps, even though it is a promising tool. Between European policies of isolation, which are either strict and instituted by national regulations with severe fines or based on incentives, the CCL recommends an intermediate position with national guidelines accompanied by strong incentive supports, ensuring good adherence from the population and avoiding the weakening of the "testing and contact tracing" facets of the national strategy. However, apps will not by themselves resolve the problem of isolation nonadherence, particularly among persons who are outside the health care system. The CCL recommends adaptations of the informational process for these persons. With regard to those for whom pragmatic isolation is difficult, because of home overcrowding, family composition, or work needs, the CCL recommends a prosocial approach with community officers who ensure that sufficient supplies (eg, essentials such as food, medication, and childcare resources) are provided [18].

To conclude, the CCL calls for promoting empowerment of the population with digital tools, improving public health culture for decision-makers dealing with health determinants, taking social considerations into account, and incorporating community participation [19].

Conflicts of Interest

All authors are members of the French COVID-19 Control and Society Connection Council (CCL).

Multimedia Appendix 1
Description of the French COVID-19 Control and Society Connection Council.
[DOCX File , 23 KB - jmir_v23i10e27301_app1.docx]

References

1. Pachetti M, Marini B, Giudici F, Benedetti F, Angeletti S, Ciccozzi M, et al. Impact of lockdown on Covid-19 case fatality rate and viral mutations spread in 7 countries in Europe and North America. J Transl Med 2020 Sep 02;18(1):338 [FREE Full text] [doi: 10.1186/s12967-020-02501-x] [Medline: 32878627]
2. Whitelaw S, Mamas MA, Topol E, Van Spall HGC. Applications of digital technology in COVID-19 pandemic planning and response. The Lancet Digital Health 2020 Aug;2(8):e435-e440 [FREE Full text] [doi: 10.1016/S2589-7500(20)30142-4]
3. Murray CJL, Alamro NMS, Hwang H, Lee U. Digital public health and COVID-19. The Lancet Public Health 2020 Sep;5(9):e469-e470. [doi: 10.1016/s2468-2667(20)30187-0]
4. Decree no 2020-572 of May 15, 2020 relating to the Covid-19 Control and Liaison Committee. Legifrance. 2020 May 17. URL: https://www.legifrance.gouv.fr/loda/id/JORFTEXT000041884917/2020-12-16/ [accessed 2021-08-04]
5. Bengio Y, Janda R, Yu YW, Ippolito D, Jarvie M, Pilat D, et al. The need for privacy with public digital contact tracing during the COVID-19 pandemic. Lancet Digit Health 2020 Jul;2(7):e342-e344 [FREE Full text] [doi: 10.1016/S2589-7500(20)30133-3] [Medline: 32835192]
6. Morley J, Cowls J, Taddeo M, Floridi L. Ethical guidelines for COVID-19 tracing apps. Nature 2020 Jun;582(7810):29-31. [doi: 10.1038/d41586-020-01578-0] [Medline: 32467596]
7. Jonker M, de Bekker-Grob E, Veldwijk J, Goossens L, Bour S, Rutten-Van Mölken M. COVID-19 contact tracing apps: predicted uptake in the Netherlands based on a discrete choice experiment. JMIR Mhealth Uhealth 2020 Oct 09;8(10):e20741 [FREE Full text] [doi: 10.2196/20741] [Medline: 32795998]
8. Blasimme A, Vayena E. What's next for COVID-19 apps? governance and oversight. Science 2020 Nov 13;370(6518):760-762. [doi: 10.1126/science.abd9006] [Medline: 33184192]
9. Heffner J, Vives M, FeldmanHall O. Emotional responses to prosocial messages increase willingness to self-isolate during the COVID-19 pandemic. Pers Individ Dif 2021 Feb 15;170:110420 [FREE Full text] [doi: 10.1016/j.paid.2020.110420] [Medline: 33082614]
10. Walrave M, Waeterloos C, Ponnet K. Adoption of a contact tracing app for containing COVID-19: a health belief model approach. JMIR Public Health Surveill 2020 Sep 01;6(3):e20572 [FREE Full text] [doi: 10.2196/20572] [Medline: 32755882]
11. Watts G. COVID-19 and the digital divide in the UK. Lancet Digit Health 2020 Aug;2(8):e395-e396 [FREE Full text] [doi: 10.1016/S2589-7500(20)30169-2] [Medline: 32835198]
12. Thomas R, Michaleff ZA, Greenwood H, Abukmail E, Glasziou P. Concerns and misconceptions about the Australian government's COVIDSafe app: cross-sectional survey study. JMIR Public Health Surveill 2020 Nov 04;6(4):e23081 [FREE Full text] [doi: 10.2196/23081] [Medline: 33048826]
13. Conclusion of the Assises Citoyennes du Numérique en Santé: the French are ready for digital health if it is built for and with them. DSIH. 2020 Nov 20. URL: https://www.dsih.fr/article/3986/conclusion-des-assises-citoyennes-du-numerique-en-sante-les-francais-sont-prets-pour-le-numerique-en-sante-s-il-est-construit-pour-et-avec-eux.html [accessed 2020-12-08]
14. Beauchamps M. First results of the Capacity survey. Marsouin. 2017 Mar 23. URL: https://www.marsouin.org/article953.html [accessed 2021-08-04]
15. Banskota S, Healy M, Goldberg EM. 15 smartphone apps for older adults to use while in isolation during the COVID-19 pandemic. West J Emerg Med 2020 Apr 14;21(3):514-525 [FREE Full text] [doi: 10.5811/westjem.2020.4.47372] [Medline: 32302279]
16. Atlani-Duault L, Lina B, Malvy D, Yazdanpanah Y, Chauvin F, Delfraissy J. COVID-19: France grapples with the pragmatics of isolation. The Lancet Public Health 2020 Nov;5(11):e573-e574. [doi: 10.1016/s2468-2667(20)30235-8]
17. Hale T, Angrist N, Goldszmidt R, Kira B, Petherick A, Phillips T, et al. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). Nat Hum Behav 2021 Apr 08;5(4):529-538. [doi: 10.1038/s41562-021-01079-8] [Medline: 33686204]
18. Webster RK, Brooks SK, Smith LE, Woodland L, Wessely S, Rubin GJ. How to improve adherence with quarantine: rapid review of the evidence. Public Health 2020 May;182:163-169 [FREE Full text] [doi: 10.1016/j.puhe.2020.03.007] [Medline: 32334182]
19. Marston C, Renedo A, Miles S. Community participation is crucial in a pandemic. The Lancet 2020 May;395(10238):1676-1678. [doi: 10.1016/s0140-6736(20)31054-0]

Abbreviations

CCL: COVID-19 Control and Society Connection Council
DCT: digital contact tracing
RT-PCR: reverse-transcription polymerase chain reaction

Edited by C Basch; submitted 20.01.21; peer-reviewed by S Six, E Goldberg; comments to author 02.03.21; revised version received 31.03.21; accepted 31.03.21; published 07.10.21.

Please cite as:
Albouy-Llaty M, Martin C, Benamouzig D, Bothorel E, Munier G, Simonin C, Guéant JL, Rusch E
Positioning Digital Tracing Applications in the Management of the COVID-19 Pandemic in France
J Med Internet Res 2021;23(10):e27301
URL: https://www.jmir.org/2021/10/e27301
doi:10.2196/27301
PMID:34313588

Original Paper

Expanding Video Consultation Services at Pace and Scale in Scotland During the COVID-19 Pandemic: National Mixed Methods Case Study

Joseph Wherton[1*], PhD; Trisha Greenhalgh[1*], PhD; Sara E Shaw[1], PhD

Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Trisha Greenhalgh, PhD
Nuffield Department of Primary Care Health Sciences
University of Oxford
Radcliffe Primary Care Building
Woodstock Rd
Oxford, OX2 6GG
United Kingdom
Phone: 44 1865 617855
Email: trish.greenhalgh@phc.ox.ac.uk

Abstract

Background: Scotland—a country of 5.5 million people—has a rugged geography with many outlying islands, creating access challenges for many citizens. The government has long sought to mitigate these through a range of measures including an ambitious technology-enabled care program. A strategy to develop a nationwide video consultation service began in 2017. Our mixed methods evaluation was commissioned in mid-2019 and extended to cover the pandemic response in 2020.

Objective: To draw lessons from a national evaluation of the introduction, spread, and scale-up of Scotland's video consultation services both before and during the pandemic.

Methods: Data sources comprised 223 interviews (with patients, staff, technology providers, and policymakers), 60 hours of ethnographic observation (including in-person visits to remote settings), patient and staff satisfaction surveys (n=20,349), professional and public engagement questionnaires (n=5400), uptake statistics, and local and national documents. Fieldwork during the pandemic was of necessity conducted remotely. Data were analyzed thematically and theorized using the Planning and Evaluating Remote Consultation Services (PERCS) framework which considers multiple influences interacting dynamically and unfolding over time.

Results: By the time the pandemic hit, there had been considerable investment in material and technological infrastructure, staff training, and professional and public engagement. Scotland was thus uniquely well placed to expand its video consultation services at pace and scale. Within 4 months (March-June 2020), the number of video consultations increased from about 330 to 17,000 per week nationally. While not everything went smoothly, video was used for a much wider range of clinical problems, vastly extending the prepandemic focus on outpatient monitoring of chronic stable conditions. The technology was generally considered dependable and easy to use. In most cases (14,677/18,817, 78%), patients reported no technical problems during their postconsultation survey. Health care organizations' general innovativeness and digital maturity had a strong bearing on their ability to introduce, routinize, and expand video consultation services.

Conclusions: The national-level groundwork before the pandemic allowed many services to rapidly extend the use of video consultations during the pandemic, supported by a strong strategic vision, a well-resourced quality improvement model, dependable technology, and multiple opportunities for staff to try out the video option. Scotland provides an important national case study from which other countries may learn.

(J Med Internet Res 2021;23(10):e31374) doi:10.2196/31374

KEYWORDS

technology-enabled care; video consultations; quality improvement; COVID-19; PERCS framework

Introduction

Background

The first documented video-mediated medical consultations in health care were conducted in the 1950s via closed-circuit television [1]. While telephone consultations have long been offered in both primary and secondary care, until the COVID-19 pandemic only a tiny fraction of clinicians had ever conducted a consultation by video and an even smaller fraction of patients had received care this way [2-7]. The video consultation is thus of academic interest as an example of a promising service-level innovation that has taken decades to catch on. We initially flagged the pandemic as an "opportunity in a crisis" for giving video consulting the push it needed [8]. In this paper, we consider how one country's efforts fared.

Prepandemic research on the acceptability, effectiveness, and cost-effectiveness of video consultations seemed to convey a positive message but was potentially misleading. Numerous research trials (generally small in size, parochial in setting, and led by a local enthusiast), in which a digitally confident and low-risk sample of patients selected from a much more diverse clinic population was randomized to continuing their usual outpatient care or trying the video option, usually showed that the latter group did no worse clinically and were no less satisfied than the former, and that costs (when measured) were similar [9-14]. Almost all such studies were underpowered to test their central hypothesis.

In published research trials of video consultations, the service was usually available only as part of the trial and discontinued thereafter, so the challenges of embedding it in business-as-usual were never addressed. Any conclusion that video is effective, acceptable, and safe was therefore naïve and premature. To "work" in the real world, video consultations require new infrastructure, new technologies, new clinical and support roles, new organizational routines, new approaches to privacy and information governance, new clinical techniques (eg, for remote examinations and safeguarding conversations), and new payment and reimbursement frameworks—not to mention measures to improve the digital skills of staff and patients and mitigate digital exclusion [15-20]. Given these complexities, it is small wonder that efforts to rollout video consultation services even following a successful trial or pilot study progressed slowly or not at all [18,21].

Early in the pandemic, an overview pointed out that despite the technical potential of remote forms of consulting to help address infection control,"[m]ost countries … lack a regulatory framework to authorize, integrate, and reimburse telemedicine services, including in emergency and outbreak situations" [22] (p1). But as it turned out, the global emergency gave governments reason enough to cut red tape—that is, remove regulatory blocks to rapid purchase and use of newly developed technologies, especially bespoke software for supporting video consultations [23]. During the first wave of the pandemic, most countries saw a rapid reduction in face-to-face medical consultations and an increase in remote ones in both primary and secondary care [3,24]. Such shifts were part of a system-wide response in what has been termed the world's first "digital pandemic," which included technologies to support outbreak monitoring and management, triaging and severity assessment, ordering and documentation, secure messaging, real-time data analytics, fast-track research trials, global knowledge sharing, and living systematic reviews [25-27]. But as Gkeredakis et al [28] (p2) have observed, while the pandemic provided fertile soil for rapid growth of new technologies, "the shifts in digital technology use to cope with the COVID-19 crisis are fast-paced, dramatic and not well understood," and successful embedding and use of novel solutions is "contingent upon the openness, distributedness, recombinability, re-programmability, and accessibility of digital technologies".

In these early months of the pandemic, politicians, policymakers, and the lay press across different countries and regions all emphasized the role of new modalities such as video and e-consultations in this unprecedented service change [29-33]. But despite this focus on the novel, the reality in most countries was that most remote consultations—especially in primary care—occurred using the old-fashioned telephone [34-38].

An important question for researchers is why, even in the context of an unprecedented global emergency, establishing and sustaining video consultations as business-as-usual in a mainstream health service has proved such a stubborn challenge. Rather than analyze failures, we have chosen to consider the case of Scotland, UK, as—broadly speaking—a success both before the pandemic (when steady incremental progress was being made to introduce the video option region by region and service by service) and during it (when a rapid and dramatic increase in video consultations was achieved at pace and scale).

In telling Scotland's story, we draw inspiration from a theoretical approach called appreciative inquiry, a form of action research or action evaluation which explicitly seeks to highlight, learn from, and reinforce the positive—things that went right, preconditions that helped, people who made a difference, and so on—while also identifying and learning from less successful aspects of the case [39]. Appreciative inquiry involves building collaborative researcher–practitioner relationships, systematically identifying "the best of what is," using creativity and experimentation to try to improve things further, and seeking to extend and replicate positive mechanisms and outcomes across the system.

The aim of this study was to draw lessons from an in-depth study of one country that could inform video consultation services and policy decisions more widely. In the remainder of this paper, we first give a historical background to the Scottish case. In the "Methods" section, we describe our aims, study setup, research questions, theoretical framework (Planning and Evaluating Remote Consultation Services [PERCS]), data sources, and methodological approach for our national evaluation of video consultation services both pre- and peripandemic. We then describe our findings, structured along the 7 domains of the PERCS framework. Finally, we discuss the Scottish case in the context of the wider literature, highlighting learning points for other countries.

Scotland: A National Case Study of Technology-Enabled Care

Scotland (population 5.5 million) is 1 of the 4 jurisdictions in the United Kingdom with a land area only 40% smaller than England (population 56 million). Much of it is mountainous and rugged with lakes (known as lochs), rivers, and offshore islands. Scotland thus shares some of the geographical challenges of remote Scandinavian regions, with some journeys involving a combination of land, water, and air. Scotland's health service is organized separately from those of England, Wales, and Northern Ireland. Health and care services are mainly delivered by 14 territorial health boards and are underpinned by a strong public-sector ethos which emphasizes professionally led quality improvement and reducing inequalities [40]. Scotland resisted the purchaser–provider split, which has existed in England and Wales since 1991 and created additional hurdles for introducing new technologies [41].

There has long been a strategic intent in Scotland to support and extend remote consulting options, including via video. This was advocated from 2008 through Scotland's eHealth Strategy [42,43] and extended in a more recent Digital Health and Care Strategy [44]. In all these documents, remote care is framed as a means to improve citizens' access to services and, ultimately, to improve outcomes and potentially reduce inequalities. In 2014, the Scottish Government established the Technology-Enabled Care (TEC) Programme to drive the widespread adoption of technology to support self-management of illness (eg, self-monitoring of long-term conditions) as well as improve access to professional care, partly in response to the perceived need for service transformation in the context of rising demand for both health and social care. The TEC Programme aims to support local deployment as well as strengthening national technical and support infrastructure.

The video consulting workstream of the TEC Programme was seen as enabling pooling of expertise and provision across the country to ensure a high-quality patient experience. Initially, this involved various pilot studies which used different video technologies, including Cisco Jabber and Polycom devices, before the TEC team decided in 2015 to introduce a more bespoke product (Attend Anywhere), described below. Based on the success of a pilot co-design and quality improvement program in one health board (Highland) in 2017, the video consulting service using the Attend Anywhere platform was branded nationally as "Near Me" (a name chosen by a patient). In November 2018 the TEC Programme launched a £1.6 million (US $2.3 million) "scale-up challenge," to support wider rollout across all health boards. By February 2020, all 14 health boards and the Golden Jubilee National Hospital (the main tertiary referral center based in Glasgow) were enrolled in the program.

Even before the pandemic, the Near Me video service had been adopted by about 180 services, spanning 35 different clinical specialties, albeit at different levels of implementation. But actual use of video within most of these services remained relatively low, with many clinicians describing their use of it as "ad hoc" rather than business-as-usual. Nevertheless, the scale-up effort continued to progress and established a strong national profile, steadily working through regulatory, infrastructural, and operational challenges. The rationale for scaling up Near Me was initially reducing patient travel and improving access and service efficiency [45]. Video consulting was generally—but not universally—seen as enhancing the existing face-to-face services, rather than replacing them. But in the context of the pandemic, emphasis shifted to infection control and the maintenance of core services.

In March 2020, when the COVID-19 outbreak reached Scotland, most routine and nonurgent care in both primary and secondary care was halted. Space and staffing were repurposed to support the pandemic response. Rollout of the Near Me video service was accelerated via a 12-week scale-up plan, led by a rapidly assembled national implementation team within the existing TEC Programme. Staff were drafted in from across Healthcare Improvement Scotland (a Special NHS Board in Scotland with a remit to help implement health care priorities), the Scottish Access Collaborative (a government program to sustainably improve waiting times for non-emergency procedures), and the Care Inspectorate (a regulatory body for social work and social care services in Scotland). They prepared guidance and resources for deployment of video consultations across a range of health and care settings and built links with other key government departments—for example, with the Primary Care Division which covered general practice. National-level groundwork and strategic planning over the previous 2 years to create technical infrastructure, service readiness, and positive attitudes helped services transform, at pace and scale, to a remote-first mode of operating as the pandemic took hold.

Following this 12-week scale-up, an engagement exercise was undertaken with various service teams to consolidate implementation plans for the video service going forward. A key element of this was the need to streamline and coordinate activities that had been undertaken in the immediate response, as well as leverage learning and resources developed. More broadly, health boards were required to produce recovery plans for clinical services during and beyond the pandemic, as part of the Scottish Government Re-mobilise, Recover, Re-design framework [46].

Evaluation and system learning were built into the Scottish Government's strategy both before and during the pandemic. Below, we describe the methods and findings of a commissioned evaluation.

Methods

Aim and Set Up of the Study

This commissioned evaluation aimed to highlight—at individual, organizational, and system level—how clinical and nonclinical staff adapted their practices and systems to accommodate and optimize the use of remote consultations both before and in response to the pandemic, and to inform policy going forward. As noted above, our *research* aim was to draw generalizable learning from an in-depth analysis of this case.

The initial evaluation contract was awarded in July 2019 by competitive tender following a public call for proposals. A follow-up contract was awarded in June 2020 using COVID-19 emergency procurement regulations. Ethics approval was

obtained from London – Camberwell St Giles Research Ethics Committee (ref 19/LO/0550) and the NHS Research Scotland Permissions Coordinating Centre. A small advisory group was set up within the Scottish government to oversee the project. Progress of the second phase of the evaluation was also monitored by an external advisory group with wide stakeholder representation and a lay chair established to oversee a number of rapid-response research studies on remote care occurring during the pandemic. Fieldwork was conducted by JW and TG.

Research Questions

Our research questions were, in relation to this national case study:

- What were the multiple interacting influences (clinical, social, technical, organizational, regulatory, and so on) on the uptake, implementation, scale-up, acceptability, effectiveness, and appropriateness of video consultations?
- What was the impact of the COVID-19 pandemic on the scale-up effort?
- What can we learn from this case about the kind of knowledge, capabilities, and infrastructures needed to support the introduction and use of video consultations in different parts of a public-sector health service?

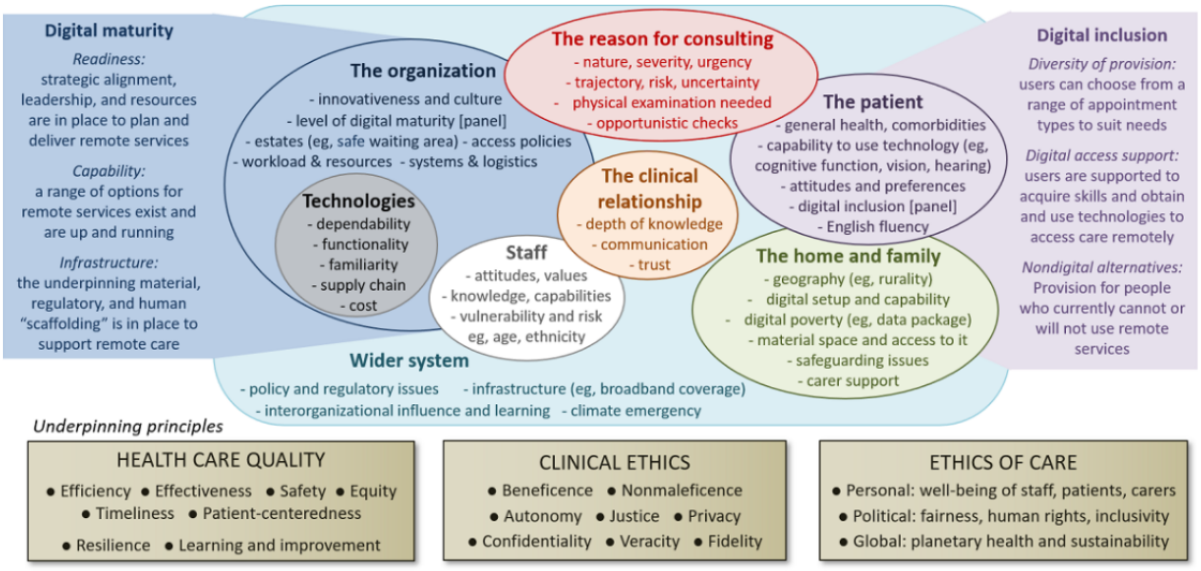
Study Design

This was a mixed methods naturalistic case study, using an "n of 1" hermeneutic approach drawing on the theoretical work of Flyvbjerg [47], Stake [48], Tsoukas [49], and Cooperrider et al [39]. These authors emphasize the use of narrative methods and rich description to produce a unique account of the case for its own sake. They warn against imposing a rigid analytic framework, producing abstracted models, or getting drawn into disjunctive theorizing (dividing the data into formal themes and categories which are then separately theorized, resulting in a neat but reductive account) [50]. Rather, as Tsoukas explains, the free-text narrative form is used to produce *conjunctive* theorizing—that is, producing an account which weaves multiple themes and influences together in a way that conveys the complexity, historical emergence, and inherent messiness of the case and draws attention to the interdependencies between different aspects of it. Appreciative inquiry applies this methodology to largely successful cases to gain insights particularly—though not exclusively—from what went well [39].

Theoretical Framework

Development and refinement of the PERCS framework (Figure 1) are described in detail elsewhere [51]. PERCS—which is specific to remote consultations—is an adaptation of a previous framework (nonadoption and abandonment by individuals, and challenges to scale-up, spread and sustainability [NASSS]) of technological innovation in health and care [52], which in turn built earlier work on diffusion of innovations in health care [53]. The domains of the PERCS framework are explained in Multimedia Appendix 1.

Figure 1. Planning and Evaluating Remote Consultation Services (PERCS) framework.



Data Collection Procedure

Data were collected in 2 phases, before and during the COVID-19 pandemic. The periods and data sources for the 2 phases are presented in Table 1.

Table 1. Data sources for the 2 phases of the evaluation.

Periods and data sources	Phase 1 (before the pandemic)	Phase 2 (in pandemic)	Total
Period of data collection	July 2019 to February 2020	July to October 2020	12 months
Number of health boards included	8[a]	8[b]	10
Ethnographic observation	60 hours in 11 clinical sites	No ethnography possible due to the pandemic	60 hours
Interviews	140, mostly conducted face to face	83 conducted remotely (mostly by video)	223 (36 were interviewed in both phases)
Interview participant characteristics	Doctors (n=29), nurses (n=18), allied health professionals (n=22), health support workers (n=3), managers (n=18), admin/IT[c] staff (n=17), patients/carers (n=21); national stakeholders (n=12)	Doctors (n=30), nurses (n=5), allied health professionals (n=13), managers (n=11), admin/IT staff (n=7), national stakeholders (n=17)	Doctors (n=59), nurses (n=23), allied health professionals (n=35), managers (n=29), admin/IT staff (n=24), patients/carers (n=21), national stakeholders (n=29), health support workers (n=3)
Documents	National (eg, on technology-enabled care strategy) and local (eg, protocols)	Relating to pandemic response (eg, remobilization and recovery plans)	N/A[d]
User experience surveys conducted online after consultation	Patients (n=679), staff (n=755)	Patients (n=18,915)	20,349
Patient and public engagement survey	N/A	Patients/public (n=4197), staff (n=1203)	5400
Uptake statistics for the Near Me service, by health board and clinical specialty	January to December 2019	January to September 2020	21 months

[a]Health Boards included Forth Valley, Highland, Golden Jubilee, Grampian, Greater Glasgow and Clyde, Lothian, Orkney, Western Isles.

[b]Health Boards included Grampian, Greater Glasgow and Clyde, Dumfries and Galloway, Highland, Forth Valley, Fife, Orkney, Western Isles.

[c]IT: information technology.

[d]N/A: not applicable.

In sum, our data set comprised 223 interviews with patients, staff, technology providers, and policymakers (further details on participant characteristics are provided in Multimedia Appendix 2); 60 hours of ethnographic observation (including visits to remote settings); local and national documents; and process data such as uptake statistics, patient and staff satisfaction surveys; and patient enablement scores. Participants for interviews were identified in 1 of 3 ways: direct contact (eg, clinicians or managers recommended to us by the Scottish Government or who were listed as having a strategic role); indirect contact ("snowballing" from interviewees by asking them to recommend someone else); and social media (via a Twitter call).

Fieldwork before the pandemic occurred in person but during the pandemic was of necessity conducted remotely. To aid conjunctive theorizing, interviews were conversational in style and lasted between 15 and 60 minutes. Local and national stakeholders were invited to speak about their efforts to develop and scale-up the service. Patients and health and care staff were asked to talk about their experience of video consulting (or why they had chosen not to use this medium). When interviewees talked in the abstract about problems and challenges, we asked them to describe specific examples of these. Qualitative data collection was conducted within a subsample of health boards purposefully selected to explore variation in geography (urban, rural, islands) and progress in the implementation and uptake of video consultations before and during the pandemic.

In addition to qualitative interviews and fieldwork, analysis was informed by evaluation data captured nationally by the TEC team in both phases of the study. A short online survey was completed by patients and clinicians immediately after each video consultation. In phase 1, the patient survey questions focused on experience and perception of technology (eg, usability, call quality). In phase 2, we added a 6-question validated patient enablement instrument to assess perceived quality and usefulness of the clinical aspects of the consultation [54]. The online staff surveys captured their experience of, and perceived satisfaction with, the consultation. In total, there were 19,594 patient responses and 755 clinician responses to the surveys.

A public engagement exercise conducted by the TEC team during the pandemic explored perspectives on the mainstreamed use of video consultations during and beyond the pandemic. This included online and paper questionnaires with members of the public (n=4197) and care professionals (n=1203), which were disseminated through a range of national and community networks.

Data on the uptake and use of the Near Me service were captured nationally during Phase 1 and Phase 2, spanning 12 months (January-December 2019) and 9 months (January-September 2020), respectively. The activity data were captured through the Attend Anywhere platform, as opposed to NHS systems, so it was not possible to establish the proportion of video in relation

to other appointment types (ie, face-to-face and telephone appointments).

Data Management and Analysis

Data were pseudonymized by giving each participant a different name. A spreadsheet containing real names and pseudonyms was stored securely in accordance with General Data Protection Regulations (GDPR). Interviews were anonymized and stored securely and selective sections transcribed. Interviews were not fully transcribed, partly for resource reasons and partly to avoid loss of overview, because salient issues were often captured succinctly in field notes. We returned to the audiotape to obtain a verbatim record where needed. We organized and gained initial familiarity with our qualitative data by organizing field notes and interview notes into an Excel spreadsheet to identify emerging themes. Each row represented an interviewee and each column represented a thematic category. We then considered each thematic category in turn, along with interactions and interdependencies. Following this familiarization phase, we undertook a more theoretically driven analysis using the PERCS framework described above so as to highlight how multiple influences interacted dynamically and unfolded over time.

Quantitative data were used to illustrate and affirm the narrative and inform ongoing data collection and analysis. Uptake and use across different health boards and specialties (and how these changed over time) were used to provide a national picture on the pace and scale of rollout, and highlight areas of focus for ongoing fieldwork and interviews (eg, to explore difference in use across settings and specialties). Survey and questionnaire data were analyzed using descriptive statistics in order to provide a national-level account and explore perspectives across different specialties and regions.

Results

Overview

Phase 1 (before the pandemic) of the evaluation generated over 300 pages of interview transcript, field notes, and document excerpts as well as raw quantitative data on staff and patient experience. At the time, our analysis (described in the official report [55]) focused mainly on the question of effectiveness, cost-effectiveness, and sustainability of the service in the context of Scotland's general policy priorities. Our formal evaluation report of the pandemic period is also available online [56].

In the prepandemic phase, we were particularly struck by 3 things. The first was the focus on region-by-region quality improvement. As described above, there had been long-established strategic drivers for Near Me in Scotland, with strong national policy support for such systems to reduce the human, financial, and environmental burden of travel. Commencing 2017, in one of Scotland's largest geographic regions (Highland), systematic efforts had been made to work collaboratively with clinicians and patients to implement a video consultation service. By 2019, this work was well underway in Highland, led not by a technical expert but by a clinician (pharmacist) who was well-regarded regionally and had previously been trained in—and personally inspired by—the system-wide approach to quality improvement promoted by the US Institute of Medicine [57]. Building on these developments, work commenced in one other health board (Grampian) to mainstream Near Me, and small-scale implementation began in the other health boards across Scotland. Thus, Near Me had, from the outset, an "organic" (locally grown and locally owned) ethos and a sense that it was being developed to improve access, reduce inequalities, and help save the planet.

The second striking feature of the service in late 2019 was the emergence of different service models reflecting Scotland's remote geography. At that time video consultations were rarely used in primary care, as almost every citizen lived fairly close to a local general practice and doctors in remote areas were generally happy to do home visits to those unable to travel to surgery. Video was largely a secondary and tertiary care service taken up in particular by remote regions. We identified 3 different models of use:

- Hub-home, in which the clinician connects from the clinic (hub) to the patient at home (or some other location via a personal device);
- Dyadic hub-spoke, in which the clinician in a specialist clinic (hub) connects to the patient in a remote health or care site ("spoke"—typically, an unstaffed kiosk equipped with a self-service video screen and connection);
- Triadic hub-spoke, in which the clinician in a specialist clinic connects to the patient in a remote health or care site with an additional staff member present.

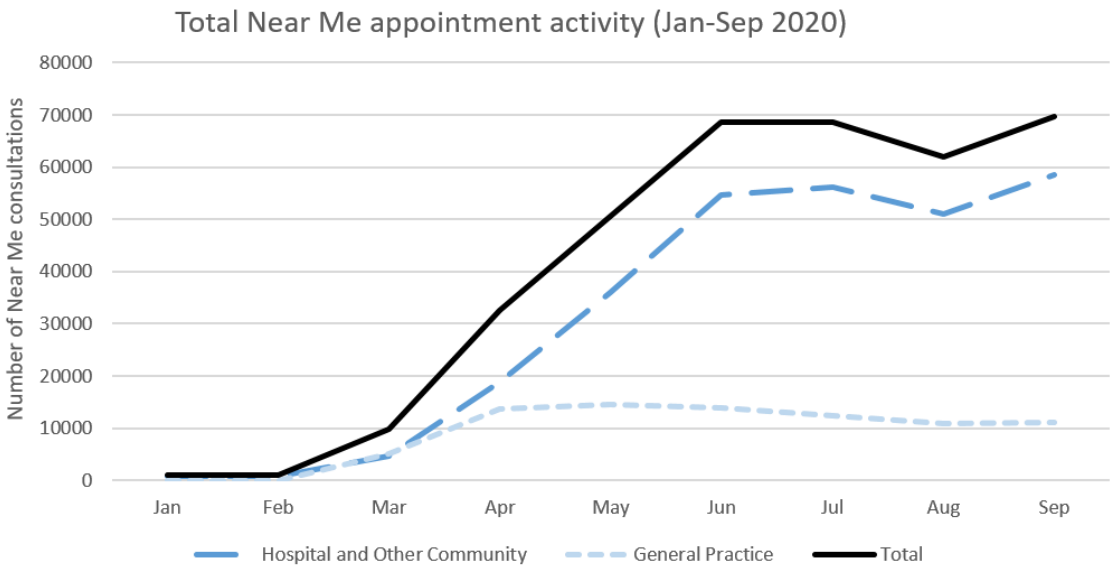
Contemporary images of video consultations generally depict some variant of the hub-home model (eg, doctor in clinic connecting directly to patient at home). This was extremely rare in our data set. Most video consultations in the prepandemic phase were organizationally far more complex—involving a triadic hub-spoke model in which a specialist in a secondary or tertiary care center connected with a remote hospital, primary care clinic, or care home and the patient received both technical and clinical support from a staff member such as a nurse, general practitioner (GP), or health care support worker. As we describe below, this unusual and resource-intensive arrangement produced challenges both at the time and—even more so—once pandemic-related infection control measures were imposed.

The third striking aspect of the Scottish Near Me service was its asymmetric development, driven in some places by particular local enthusiasts and thwarted in other places by lack of them. Often, a video service had been established serendipitously—for example, as a specialist consultant who moved away sought to keep some clinical contact with their patients—and maintained through strong working relationships between key members of staff. This patchwork nature of video consultation services had advantages and disadvantages. On the one hand, it reflected the Scottish Government's enabling (rather than command and control) approach, in which professionals could be creative and locally adaptive. On the other hand, the lack of a centrally mandated policy meant that despite Scotland's relative success, *most* services still offered few video appointments and many offered none at all.

When we returned—using virtual methodologies—to explore the response to the pandemic, our data confirmed a rapid

expansion of the service (Figure 2). In 2019, just under 7000 Near Me consultations had been conducted nationally (134 per week on average). In the months preceding the rapid scale-up (January-February 2020), there were approximately 230 video consultations per week. Between March and June 2020, the number of video appointments increased 50-fold, from about 330 to 17,000 appointments per week nationally, and over 50 clinical specialties introduced video consultations for the first time. Other forms of remote consulting were used (eg, telephone and emailing or uploading of photographs). Unfortunately, relative proportions of these different modes could not be accurately captured for further analysis.

Figure 2. Growth of video consultations before and during the pandemic. The graph shows the total number of video consultations for general practice, hospital, and other community services.



Although the pandemic saw a significant shift in the use of Near Me at a national level, the extent of this change varied across care settings. For instance, while many general practices introduced the Near Me service model, most used it infrequently and ad hoc, so that general practice as a whole accounted for only 22.55% (81,822/362,828) of all video consultations in Scotland. Among hospital and community specialties, the services accounting for most video activity were psychiatry, psychology and community mental health (36.41% of all hospital and community care activity, 94,876/260,547), physiotherapy (8.79%, 22,909/260,547), and pediatrics (7.81%, 20,354/260,547).

In the analysis below, we present findings from each domain of the PERCS framework. Interview quotes are provided by domain in Multimedia Appendix 3.

The Reason for Consulting

Prior to the pandemic, almost all new consultations in secondary care were face to face, in order to establish a clinical relationship, undertake a full physical examination, and conduct baseline assessments (eg, a standardized severity score for rheumatoid arthritis). Video consultations were used primarily for routine follow-up of chronic, stable conditions, especially to convey test results and affirm that the patient remained asymptomatic (see quote Q1a in Multimedia Appendix 3).

Other services with high use of video consultations before the pandemic included physiotherapy, speech and language therapy, pediatrics (for nonacute conditions such as gait abnormality), community mental health follow-up (eg, for patients with depression), and postoperative follow-up. In many such situations, clinicians saw value in “eyeballing” the patient (ie, a visual overview, albeit via video) to assess their general health. Some remote physical examinations were possible, especially by experienced practitioners (Q1b in Multimedia Appendix 3).

During the pandemic, use of video was extended to embrace a much wider range of clinical conditions and reasons for consulting. This was driven by a concerted national effort to maintain the provision of priority clinical specialties (eg, oncology, maternity, mental health), as well as local teams seeking to maintain some service provision across a range of different specialties. Most of these required no formal physical examination. Video was noted to be very useful in psychiatry, psychology, counseling, respiratory medicine, and speech and language therapy, where visual, nonverbal communication was important and personal protective equipment might interfere (Q1c in Multimedia Appendix 3). Some clinicians talked of the need to *feel* a lesion as well as see it (eg, palpating regional lymph nodes in cases of possible cancer), which precluded video examination.

Some conditions for which a visual examination was crucial were considered unsuitable for video consultations. For skin lesions, for example, the variable quality of the video image transmitted from a patient’s device was not always adequate to make a confident diagnosis; instead, patients were often encouraged to obtain and upload a high-quality still photograph taken in good light. This approach also allowed the image to be stored and, where necessary, sent on for a specialist opinion. Most acute ophthalmology consultations required slit-lamp examination or fundoscopy (use of high-intensity light and magnification to see inside the eye) even when there was an evident lesion on the eye (and especially when there was not). While video was used for remote examinations by

ophthalmologists, patients still needed to be physically co-present with the optometrist in order to capture magnified images of the eye through specialist equipment. In sexual health, use of video and still images was limited by legal and regulatory restrictions on transmission of intimate images (and by practitioners' discomfort about exchanging such images).

Both before and during the pandemic, video consultations were widely viewed as clinically less appropriate for poorly defined and less predictable conditions, rare conditions (ie, those with which the clinician and system were unfamiliar), life-changing diagnoses, planning of major interventions (eg, bone marrow transplant), unstable or unpredictable psychiatric conditions (eg, severe emotional trauma, psychosis), or when the patient would need to attend in person anyway to have tests or collect medication. The threshold for offering such patients a remote consultation changed, however, as the risk–benefit balance was dramatically altered by the pandemic. In the context of pandemic restrictions, GPs considered telephone adequate for most problems in known patients, as they felt they rarely needed to rely on visual assessment or physical examination. But video was sometimes considered crucial for visual assessment to exclude rare but potentially serious acute problems such as infection (Q1d in Multimedia Appendix 3).

Primary care is traditionally the "risk sink" of a health service: most new problems are low risk and self-limiting, and potentially serious symptoms or signs can be observed and referred on as needed. During the pandemic, GPs talked of the difficulty in managing risk without the option of bringing the patient in for a face-to-face examination, especially because secondary care colleagues did not always trust an assessment made by telephone. One experienced GP felt they had picked up a possible malignancy purely from the history, so referred the patient on the basis of that, but the patient was "triaged" at the secondary care end and the appointment refused (Q1e in Multimedia Appendix 3).

The Patient

Our prepandemic evaluation identified numerous comorbidities and preexisting conditions which affected patients' ability to use the video technology. These included temporary or permanent confusion or other cognitive limitation, visual impairment, or serious anxieties about the technology (including body image concerns about being seen, or seeing themselves, on video display). But more commonly, the barrier to using video was the patient's general level of debility (Q2a in Multimedia Appendix 3).

Many patients with deafness and partial hearing loss found remote consultations by telephone impossible, while those able to lip-read or use the chat function often welcomed the option of video. Professional interpreters using the British Sign Language were available for some via remote triadic consultations, depending on the availability of interpreters and organizational structures in place to coordinate this. During the pandemic, patients with hearing impairment—and staff—were greatly handicapped by the requirement to use face coverings in face-to-face settings, making a video consultation a more attractive option (Q2b in Multimedia Appendix 3).

Interpreting services for patients who spoke limited English were rarely available by video before the pandemic, due to logistical challenges of bringing the interpreter into a 3-way call. However, remote interpreting services gained impetus during the pandemic, when resources were directed into meeting the technical and logistical challenges (Multimedia Appendix 3, Q2c and 2d).

Before and during the pandemic, lack of familiarity and low digital literacy explained some patients' reluctance to use video even in the absence of a relevant disability or co-existing condition. While the pandemic provided impetus for upskilling, it was difficult for clinicians or support staff to estimate a patient's likely capability prior to the consultation, and making decisions based on age, gender, or social stereotypes felt unprofessional (Q2e in Multimedia Appendix 3). Older patients and those assumed to be unfamiliar with digital technology (eg, those in manual and outdoor occupations) were often offered telephone rather than video.

For patients able to use video technologies, their attitudes and preferences toward video, telephone, and face-to-face consultations before the pandemic typically centered on the trade-offs between home and work commitments, travel and transport access, the nature of the clinical problem, a desire (or not) to establish or strengthen a personal relationship with the clinician, and sheer convenience (Q2f and 2g in Multimedia Appendix 3).

More prosaically, patients were not always aware of the video option (some informants commented that this needed flagging when the appointment is booked).

During the pandemic, concerns about infection risk became over-riding for most, leading many people to try video consulting for the first time. Findings from the public engagement survey indicated a high level of support for continuing remote consulting beyond the pandemic, partly because people were now familiar with this medium and partly due to on-going anxieties about risk of infection.

The Home and Family

Our prepandemic visits to Scotland highlighted the very remote settings in which some people lived. Transport links were sometimes few and unreliable (eg, ferry and plane cancellations due to the weather), and staff as well as patients spent many hours (and sometimes whole days) traveling to and from clinics. Even before the pandemic, many people described how video consultations had transformed their lives simply by saving travel time. But connecting from home was far from an instant option for some patients. Many homes in remote areas were small, of basic construction, and had limited privacy. In some communities (especially certain inner-city areas), there were high levels of poverty and deprivation. In the public engagement survey, respondents' reported "very significant" barriers to uptake including lack of access to an appropriate device (23.04%, 924/4010), poor internet connectivity (29.92%, 1200/4011), restrictions on mobile data packages (17.07%, 685/4012), and lack of private space at home (19.61%, 786/4009).

Requirements for physical distancing and managing risk of infection during the pandemic required a shift from the triadic hub-spoke model of video consulting to a hub-home model, for which neither the patients nor the service were fully prepared. Failed attempts at video consultations to patients affected by digital poverty (eg, no smartphone, no webcam, limited data package) were common and frustrating (Q3a in Multimedia Appendix 3). Video consulting to home could also mean an inadequate material space (not everyone had a desk or table for example; some consulted from their cars, their bed, or even the bathroom) and potential distractions (especially from children either present in the room or unsupervised somewhere off camera). Some interviewees expressed concerns about the possibility of a patient's abusive partner listening in.

Patients with low digital literacy or confidence sometimes benefited from on-hand carer support for video consultations; those lacking such support were often limited to telephone. Carer support included setting up and troubleshooting the video link, adjusting the camera angle to facilitate a remote physical examination, preparing and overseeing a child's appointment (while enabling rather than interfering with the direct clinician–child interaction), and assisting with translation or communication (Q3b in Multimedia Appendix 3).

As we have shown previously, supporting a relative's remote clinical examination can be emotionally as well as technically challenging, because it may involve complex negotiations between carer and patient about the balance between assistance and autonomy [58]. During the pandemic, shielding and physical isolation measures limited the availability of carer support. The quality of video consultations with care home residents also depended on care workers' varying technical knowledge and skill.

The Clinical Relationship

An established clinical relationship, based on previous face-to-face encounters, made clinicians and patients more relaxed about video consultations and allowed clinicians to tolerate the higher levels of uncertainty associated with this medium. Prior to the pandemic, most clinicians liked to have an initial face-to-face consultation to establish rapport and confirm suitability of video for follow-up appointments. Initially, the default in all services was for first assessments to be done in person, but the pandemic required many new referrals to be assessed via video. While these posed challenges (chiefly around technical connectivity), many informants told stories of how such interactions had gone surprisingly well.

Mental health specialties in particular considered video important for communicating (Q3c in Multimedia Appendix 3). A recurring theme in our data was the importance of a high-quality technical connection for establishing and building deep therapeutic rapport.

Perceptions on how remote consulting altered the relationship and interaction between patients and clinicians were nuanced and highly contingent upon clinicians' interaction styles, perceived value of tactile information and facial expression, and the clinical context of the encounter. Some emphasized the therapeutic value of the in-person physical examination and regretted the loss of such contact during the pandemic (Q3d in Multimedia Appendix 3). Others saw video as a way of circumventing the interpersonal barriers created by facemasks (Q3e in Multimedia Appendix 3).

Video was occasionally advocated for more paternalistic reasons. One medical consultant, for example, considered that teenagers did not take telephone consultations seriously and were likely to "pay attention" more if they were seen by video.

The Technology

Attend Anywhere, the technology used for Near Me across Scotland, is an internet browser–based video technology that can be accessed by a staff member on a work computer or a member of the public using their own device. One defining feature is its "inbound" workflow, which seeks to emulate how patients physically attend their appointments. For example, a single button on a website (or consistent weblink address on an appointment letter) offers a one-stop "virtual front door" for patients. On clicking that link, the patient enters a "virtual waiting room" (potentially managed by a live receptionist), before being invited into the clinician's virtual consulting room. Because Attend Anywhere does not require the downloading of software or creation of user accounts, it is easier for patients to use securely. The system has also been designed specifically for health and care, with a strong information governance model that was reviewed and endorsed nationally. This helped avoid information governance restrictions which prevail in many health care organizations.

In our prepandemic evaluation, the Attend Anywhere technology was generally considered by staff and patients to be dependable and to produce high-quality video and audio. In most cases (503/662, 75.98%), patients reported no technical problems during their postconsultation survey. Of those reporting technical problems, the issues mainly related to internet connection and audio–video quality (eg, moments of sound loss, lack of synchronization between video and audio), as opposed to complete technical failure or usability issues. Staff and patients told us that ease of use was partly due to the well-designed software, also partly because many services had invested in high-quality peripherals such as screens and noise-cancelling microphones. Additional data from postconsultation surveys are provided in Multimedia Appendix 4.

During the pandemic, 2 problems occurred. First, the shift to hub-home care models (see above) meant that the connection came to depend heavily on patient connectivity and device (and sometimes also on the home connections of homeworking clinicians), leading to loss of video or audio connection or awkward lag (Q4a in Multimedia Appendix 3). Second, the software platform initially came under significant strain due to an unprecedented increase in volume of consultations, resulting in periods of poor service reliability. These problems were dealt with promptly by the technology supplier by removing bottlenecks from the underlying application and increasing server capacity.

The video connection also depended on other technical systems, particularly for accessing the virtual waiting area. For example, text-messaging systems through the electronic booking systems

to provide the patient with the URL were prone to error during rapid rollout, sometimes sending incorrect links to patients.

When the video connection failed, telephone was used—fairly unproblematically—as a backup.

Despite these challenges, patient survey responses remained generally positive. Most (14,677/18,817, 78%) reported no technical problems. Of the remainder, most problems were similar to those encountered in the prepandemic evaluation (eg, audio or video quality).

The functionality of Attend Anywhere was considered good by most interviewees. Clinicians particularly liked the virtual waiting room and the option for screensharing (Q4b in Multimedia Appendix 3). This functionality, however, required a reasonable screen size and was of limited use if the patient was using a smartphone or small tablet device.

While Attend Anywhere was unfamiliar to most patients, the "inbound" workflow with a single point of entry and virtual waiting area made sense to patients because (they told us) using the technology "felt like" going into a clinic and physically sitting in a waiting room—a finding that others have also observed [59]. There were occasional glitches such as when a patient, offered several waiting rooms, selected the wrong one (Q4c in Multimedia Appendix 3). In the larger services, an actual live receptionist would meet and greet the patient on the video call and transfer the patient to the correct virtual waiting room.

A national policy decision to provide Attend Anywhere to all NHS organizations, alongside financial investment in the model in 2018, strengthened the organizational incentives to expand use of the technology prior to the pandemic. At the start of the pandemic, central procurement of Attend Anywhere was further extended by 2 years. Significant challenges were faced in resourcing laptops, video cameras, and audio equipment in the face of national shortages and disrupted supply chains during the pandemic. But because Attend Anywhere is an encrypted browser-based technology, some staff could make use of personal devices to run video consultations.

Staff

Staff attitudes toward video consulting varied considerably, especially before the pandemic. Most clinicians we spoke to who had used Near Me had positive things to say about it, describing it as a significant way of improving patient access and experience by reducing the need for travel, providing faster and more direct access to specialists, and helping overcome reluctance to visit clinical spaces. Some clinicians, who used video never or rarely before the pandemic, depicted such services as unprofessional or unsafe (Q5a in Multimedia Appendix 3). But interviews during the pandemic found that many had changed their perspective (Q5b in Multimedia Appendix 3).

Many clinicians talked about still being on a learning curve about when to offer the remote option and how to conduct such consultations effectively, including adapting ways of interacting with patients to take account of the physical and symbolic differences of the virtual environment (Q5c in Multimedia Appendix 3).

The shift to video was not universally welcomed. Some staff felt its continued use beyond the pandemic was a retrograde step because it was less professionally fulfilling (Q5d in Multimedia Appendix 3).

Before the pandemic, the use of video consulting was almost never spoken of in relation to staff well-being (with the exception of saving travel time). During the pandemic, video consultations were seen, on the one hand, as protecting staff—especially vulnerable ones—from risk of infection. On the other hand, they were described as more cognitively demanding and tiring than face-to-face ones. In a few cases, the clinician described becoming unwell during a video consultation (Q5e in Multimedia Appendix 3).

The Organization

The general innovativeness and digital maturity of health care organizations had a strong bearing on their ability to introduce, routinize, expand, and evaluate their video consultation service. In this regard, the nationwide effort by the Scottish Government to strengthen digital infrastructure over the previous 10 years was evident. Many, though not all, organizations had good broadband connection, adequate hardware, and sufficient numbers of trained staff to implement the technology. In many specialties, the equipment had been installed but had not been routinely used until pandemic pressures created an impetus.

Developments before the pandemic revealed the importance of equipment setup and availability. This included dual screens, high-quality cameras and noise-cancelling microphones and speakers, as well as specialist equipment (eg, high-quality audio headsets for speech and language therapists, high-magnification cameras for dermatologists). Much work had gone into these details alongside the rollout of Near Me in 2018-19. However, during the early stages of the pandemic, demand for this kit soon exceeded supply, especially because infection control protocols prevented sharing between staff. Some ran short and had to rely on the phone.

Video consulting relied on other technical systems, such as electronic booking and secure asynchronous communication channels with patients (eg, texting, email). Digital maturity, in this regard, was the extent to which standardized processes had been established for the smooth running and reconfiguration of appointment schedules to accommodate different modalities. The rechanneling of IT and outpatient resources during the pandemic helped address these challenges, but staff in some settings described various glitches, such as patients entering incorrect virtual waiting areas, due to rapid restructuring of administrative workflows and systems (Q6a in Multimedia Appendix 3).

Prepandemic infrastructure strengthening through the TEC Programme had mainly focused on outpatient hospital sites, which were able to scale-up quickly as the pandemic hit. By contrast, general practice services had had little interest in video consultations before the pandemic and had to be rapidly set up in early 2020. Because of shielding, and also because clinic space was repurposed (eg, for seeing potentially infected

patients), many general practice staff worked from home. Despite input from mobile IT teams to install the necessary equipment, this remote working did not always go smoothly (Q6b in Multimedia Appendix 3).

A major challenge across all sites was establishing adequate space and equipment for a video consultation. During the pandemic, staff worked pragmatically and adaptively, with technology to hand. Particularly for the larger city hospitals, moves toward hot-desking and shared office space were not conducive to the expansion of video services, because open-plan working encroaches on privacy and may require the clinician to wear a mask.

Some staff described logistical barriers to establishing and running a remote consultation, notably requesting and obtaining blood test results from the patient's local primary care practice, transmitting a prescription to the patient's local pharmacy when the patient was not there to collect it in person, providing patients with printed information sheets, or obtaining written consent (Q6c in Multimedia Appendix 3).

Many of these logistical issues required a redistribution of resources across the system (eg, additional staff were needed to run a more complex appointment system and virtual waiting area). The expansion of video appointments also required new ways of working and sharing data across departments and with patients (eg, respiratory services purchased pulse oximeters for patients to use at home). While these changes were initially developed (and resourced) as an interim "workaround" measure, there is both enthusiasm and concern for sustainability of these practices.

Before the pandemic, efforts to introduce and use remote services often stalled because of staff shortages (especially when senior clinicians were replaced by a series of short-term locums) or general lack of resources. Attempts by IT and service managers to set up remote services and embed them in business-as-usual were dependent on clinicians who were willing to join the change effort, use the technology, and consider working in a different way. Such individuals were relatively rare (described by one interviewee as the "keenies" but perhaps more formally classified as "innovators" or "early adopters" [53]). The rechanneling of local resources during the pandemic, alongside a lull in routine activity in some specialties, provided clinicians with the opportunity to try out and adapt new ways of working.

We found in our prepandemic evaluation that both dyadic and (even more so) triadic hub-spoke models raised logistical challenges and required various kinds of double-handling (eg, appointments needed to be made, rooms booked, and staff members made available, at both the hub and the spoke site). In some remote sites, there was much redundancy (eg, staff were allocated to a hub clinic for a whole morning but only 1 or 2 patients were seen); in others, a lone staff member had to juggle multiple roles (Q6d in Multimedia Appendix 3).

Because much of this double-handled activity related to outpatient consultations or cold surgery, it was stalled during the pandemic. It is unclear how, as video consultation services expand beyond the pandemic, this issue will be resolved.

All the organizations we studied were committed to a policy of inclusion. Service teams were encouraged to ensure that new remote models did not disadvantage people in relation to service access and allowed patients to exercise choice where clinically appropriate.

During the pandemic, patient choice was heavily constrained by infection control protocols, a measure that created huge challenges for ensuring equity of access. Some of our respondents were keen that video should not be the default option for everyone going forward (Q6e in Multimedia Appendix 3).

The Wider System

The Scottish Government's longstanding commitment to using technologies to achieve high-quality, accessible, and equitable care and contributing to a low-carbon future created an important national-level context for the introduction and mainstreaming of video consultations. While remote areas had limited or no broadband access, this was improving as a result of a policy push for connectivity. However, there was no strong tradition of digital communication (some outlying islands, for example, had only had broadband outside the largest town for a few years, so not everyone owned, or was comfortable using, a smartphone). When the pandemic hit, the Near Me service was immediately mandated across the country, allowing rapid and consistent implementation locally and regionally (Q7a in Multimedia Appendix 3).

Also important was engagement of professional bodies such as Royal Colleges, who endorsed the TEC Programme's vision and guidance documents from an early stage. Proactive communication between government and professional bodies ensured that frontline clinicians believed that the changes were professionally endorsed and led rather than imposed by central government. In addition, the TEC team worked with NHS National Services Scotland (a public body that provides national strategic support to NHS services) to review and monitor network bandwidth capacity for the rapid expansion of video, and sought data protection approvals at a national level to provide confidence and continuity across local organizations.

There was a concerted effort at national and local level to support collaborative learning and interorganizational support. This included shared learning within specialties at national level (supported by NHS Education Scotland) and board-level collaboration between departments about local processes (Q7b in Multimedia Appendix 3). As well as supporting knowledge sharing across the 14 health boards, the Scottish TEC team engaged with national and regional leads across England and Wales to facilitate local rollout shortly prior, and during, the pandemic. This included the sharing of training resources, patient facing materials, and governance documentation. The English and Welsh NHS Trusts were also hosted temporarily on the Scottish Attend Anywhere platform to help them off the ground until separate platforms could be established, at some risk to their own system integrity.

Discussion

Principal Findings

This is a mixed methods case study of the development and pandemic-driven scale-up of video consultation services across Scotland. Using the PERCS framework, we have mapped a complex data set of qualitative and quantitative findings to explain multiple interacting domains of influence. Before the pandemic, a national program to extend a service that had been successful in local pilots was already underway, driven by an ethos of collaborative quality improvement, reducing inequalities, and achieving cross-government low-carbon goals. By the time the pandemic hit, there had been considerable investment in material and technological infrastructure, staff training, and professional and public engagement. Scotland was thus uniquely well placed to expand its video consultation services at pace and scale, resulting in a dramatic increase in number of services using video. While not everything went smoothly, video consultations became available as business-as-usual for a much wider range of clinical problems, vastly extending the prepandemic focus on outpatient monitoring of chronic stable conditions.

Strengths and Limitations

The great strength of this study is that, somewhat serendipitously, we had built a good working relationship with the Scottish Government and many regional implementation teams just before the pandemic hit, and were able to mobilize quickly to undertake a second phase of the evaluation. This meant that—perhaps uniquely in any country—we obtained both pre- and peripandemic data of various kinds. We were also undertaking other research on remote consultations across the UK and developing the PERCS framework, which proved useful for explaining and organizing multiple streams of data.

The limitations of this study are threefold. First, pandemic restrictions meant that we could undertake no ethnographic work in phase 2, and our data collection more generally was affected by the unprecedented pressures on NHS staff (who, for example, had had little time to reflect individually or collectively on what was happening). This also raises potential sample biases toward more technically literate participants (ie, to conduct interviews by video), as well as those individuals with the time available to speak with us. We sought to mitigate these issues by offering phone as well as video interviews and adapting interview schedules to meet individual circumstances. Second, our positive and now longstanding working relationship with national- and regional-level stakeholders may have led us to view their change efforts in a positive light, though in other large-scale evaluations we have had equally positive relations with stakeholders but produced less positive reports [21]. Third, the pace, scale, and scope of the evaluation did not allow us to produce an economic component to explore the costs and cost-effectiveness of the video option in different circumstances and settings.

Comparison With Prior Work

Scotland's story of scaling up video consultations resonates with what we know of other countries that had a relatively advanced infrastructure for telehealth—for example, Australia, where a small (and possibly skewed) survey found that up to 60% of health professional respondents had consulted by video during the pandemic, aided by a slackening of regulatory restrictions and more flexible reimbursement [35]. In Norway, the relative proportion of remote GP appointments increased from approximately 3% (before the pandemic) to almost 60% during the initial lockdown [60]. This move was encouraged by the Norwegian Ministry of Health and Care services and incentivized by temporary modifications to reimbursement systems. Clinician surveys revealed that, while such shift raised new possibilities for video consulting in the longer term, important clinical, technical, and operational challenges remain [60]. In New Zealand, the Royal College of General Practitioners urged all members to switch to remote (video, phone, email) consultations with a goal of reducing in-person visits by 70% within 48 hours of the national lockdown [61]. This rapid response was aided by a NZ $20 (US $13.8) million government commitment to increase telehealth capacity, training webinars through the National Telehealth Resource Center, and temporary relaxation of electronic prescribing rules [62]. While nationwide uptake of this service model is yet to be comprehensively reviewed [63], preliminary research indicates longstanding potential; pending further investments; and closer attention to IT infrastructure, regulatory, and accessibility considerations [64,65]. We suspect there will be useful comparisons between New Zealand and Scotland, possibly using the PERCS framework.

Basu et al [27] have sought to capture the different perspectives of the International Medical Informatics Association Telehealth Working Group, to broadly explore the role of telehealth in 10 different countries during the pandemic. The authors present 6 themes which align broadly with our Scottish case study, namely, strategic (policy decisions and legal changes); operational (increasing capacity and delivery by building skills and resources at pace and scale); regulatory (including pandemic-related unofficial workarounds with unregulated products); changes in attitudes and uptake; public engagement; and training and education. The authors use the World Health Organization Health System model to emphasize the sociotechnical nature of these changes. Other system-focused frameworks such as i-PARIHS (integrated Promoting Action on Research Implementation in Health Services) take a similar though not identical approach [66].

In the language of system innovation, the pandemic was what Van de Ven [67] would call an "environmental shock"—something that generates uncertainty, puts organizations under stress, and requires an urgent adaptive response. A weak system is highly vulnerable to such shocks but a resilient one that is able to adapt can sometimes use the shock as an impetus to innovation and thereby become better able to weather the next shock [68]. The litmus test for Scotland is perhaps not the impressive expansion of video consulting during the pandemic under emergency measures (unregulated telehealth) but the extent to which the positive elements of this expansion will be retained and mainstreamed once such measures are rolled back (regulated telehealth) [27]. It is to Scotland's credit that both its technological and human

infrastructure were sufficiently resilient to respond in ways that could often be routinized within the existing system.

Gkeredakis et al [28] apply 3 perspectives to shed light on the varied uses of digital technology, and associated tensions, during the COVID-19 crisis: *opportunity* for accelerated innovation and removal of barriers to experimentation; *disruption* to organizational and occupational practices, generating new dependencies and risks; and *exposure* of vulnerabilities in both people and infrastructures that have previously gone unnoticed, such as physical work spaces, IT networks, and key workers [28]. Our findings illustrated each of these themes.

Conclusion

Scotland's national-level groundwork before the pandemic allowed many services to transform, at pace and scale, to a video-first mode of operating. Key contributors included the "burning platform" of the pandemic, national strategic vision, a well-resourced quality improvement model based around communities of practice and system learning, dependable technology, and multiple opportunities for staff to try out the video option.

We anticipate that sustaining video as the new normal will depend on multiple issues such as digital infrastructure, human and financial resources (distributed fairly across the system), training (including digital literacy and teleconsulting skills), workforce (including extent to which video can help compensate for staff shortages), data security (including overcoming the tendency for regulations to be overly restrictive), and research into remote clinical examinations [25,69].

Evidence that remote care contributes significantly to greener health services is currently limited (one study suggests that a substantial reduction in carbon footprint could be made [70]), but we and others are continuing to study this important factor. Further research is also recommended into different service models. We flagged, for example, that the triadic hub-spoke model is somewhat labor intensive and unlikely to be scalable, whereas hub-home is limited by patients' digital and material setup. Other models have also been described [71].

Scotland provides an important national case study from which other countries may learn. We invite others to apply the PERCS framework to their own case studies and propose refinements to it.

Acknowledgments

We thank the research participants including patients, health care staff, and wider stakeholders who all gave generously of their time while working under considerable pressure. Fieldwork for this study was funded by 2 research contracts commissioned by the Scottish Government from TG and JW in 2019 and 2020. Additional funding for TG, SES, and JW's salaries over the course of the study was provided by the UK National Institute for Health Research Oxford Biomedical Research Centre (BRC-1215-20008), Economic and Social Research Council COVID-19 Emergency Fund (ES/V010069/1), and Health Foundation and Wellcome Trust (WT104830MA). Funders played no part in the data collection, analysis, or preparation of the paper except for correction of factual errors such as dates on a near-final draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Planning and Evaluating Remote Consultation Services (PERCS) domain explanations.
[DOCX File , 26 KB - jmir_v23i10e31374_app1.docx]

Multimedia Appendix 2
Interview participant characteristics.
[DOCX File , 18 KB - jmir_v23i10e31374_app2.docx]

Multimedia Appendix 3
Qualitative quotes.
[DOCX File , 22 KB - jmir_v23i10e31374_app3.docx]

Multimedia Appendix 4
Postconsultation survey data.
[DOCX File , 21 KB - jmir_v23i10e31374_app4.docx]

References

1. Brown FW. Rural telepsychiatry. Psychiatr Serv 1998 Jul;49(7):963-964. [doi: 10.1176/ps.49.7.963] [Medline: 9661235]
2. Zanaboni P, Wootton R. Adoption of routine telemedicine in Norwegian hospitals: progress over 5 years. BMC Health Serv Res 2016 Sep 20;16:496 [FREE Full text] [doi: 10.1186/s12913-016-1743-5] [Medline: 27644324]

3. Hughes G, Wood G, Shaw S. Video Consultations in the NHS: Summary Findings from a UK-Wide Survey. London, UK: Health Foundation; 2020 Nov. URL: https://www.phc.ox.ac.uk/files/reports-and-working-papers/vc-survey-and-report-final-draft-for-website.pdf/view [accessed 2021-09-27]
4. Bhatta R, Aryal K, Ellingsen G. Opportunities and challenges of a rural-telemedicine program in Nepal. J Nepal Health Res Counc 2015;13(30):149-153. [Medline: 26744201]
5. Wade VA, Taylor AD, Kidd MR, Carati C. Transitioning a home telehealth project into a sustainable, large-scale service: a qualitative study. BMC Health Serv Res 2016 May 16;16(1):183 [FREE Full text] [doi: 10.1186/s12913-016-1436-0] [Medline: 27185041]
6. Alami H, Gagnon M, Wootton R, Fortin J, Zanaboni P. Exploring factors associated with the uneven utilization of telemedicine in Norway: a mixed methods study. BMC Med Inform Decis Mak 2017 Dec 28;17(1):180 [FREE Full text] [doi: 10.1186/s12911-017-0576-4] [Medline: 29282048]
7. Brooks E, Manson SM, Bair B, Dailey N, Shore JH. The diffusion of telehealth in rural American Indian communities: a retrospective survey of key stakeholders. Telemed J E Health 2012 Jan;18(1):60-66 [FREE Full text] [doi: 10.1089/tmj.2011.0076] [Medline: 22082106]
8. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
9. Chongmelaxme B, Lee S, Dhippayom T, Saokaew S, Chaiyakunapruk N, Dilokthornsakul P. The effects of telemedicine on asthma control and patients' quality of life in adults: a systematic review and meta-analysis. J Allergy Clin Immunol Pract 2019 Jan;7(1):199-216.e11. [doi: 10.1016/j.jaip.2018.07.015] [Medline: 30055283]
10. Gentry MT, Lapid MI, Rummans TA. Geriatric telepsychiatry: systematic review and policy considerations. Am J Geriatr Psychiatry 2019 Feb;27(2):109-127. [doi: 10.1016/j.jagp.2018.10.009] [Medline: 30416025]
11. Hong Y, Lee SH. Effectiveness of tele-monitoring by patient severity and intervention type in chronic obstructive pulmonary disease patients: A systematic review and meta-analysis. Int J Nurs Stud 2019 Apr;92:1-15. [doi: 10.1016/j.ijnurstu.2018.12.006] [Medline: 30690162]
12. Kelson J, Rollin A, Ridout B, Campbell A. Internet-delivered acceptance and commitment therapy for anxiety treatment: systematic review. J Med Internet Res 2019 Jan 29;21(1):e12530 [FREE Full text] [doi: 10.2196/12530] [Medline: 30694201]
13. McFarland S, Coufopolous A, Lycett D. The effect of telehealth versus usual care for home-care patients with long-term conditions: A systematic review, meta-analysis and qualitative synthesis. J Telemed Telecare 2021 Feb;27(2):69-87. [doi: 10.1177/1357633X19862956] [Medline: 31394973]
14. Thiyagarajan A, Grant C, Griffiths F, Atherton H. Exploring patients' and clinicians' experiences of video consultations in primary care: a systematic scoping review. BJGP Open 2020;4(1):bjgpopen20X101020 [FREE Full text] [doi: 10.3399/bjgpopen20X101020] [Medline: 32184212]
15. Wherton J, Shaw S, Papoutsi C, Seuren L, Greenhalgh T. Guidance on the introduction and use of video consultations during COVID-19: important lessons from qualitative research. leader 2020 May 18;4(3):120-123. [doi: 10.1136/leader-2020-000262]
16. Donaghy E, Atherton H, Hammersley V, McNeilly H, Bikker A, Robbins L, et al. Acceptability, benefits, and challenges of video consulting: a qualitative study in primary care. Br J Gen Pract 2019 Sep;69(686):e586-e594 [FREE Full text] [doi: 10.3399/bjgp19X704141] [Medline: 31160368]
17. Odendaal WA, Anstey Watkins J, Leon N, Goudge J, Griffiths F, Tomlinson M, et al. Health workers' perceptions and experiences of using mHealth technologies to deliver primary healthcare services: a qualitative evidence synthesis. Cochrane Database Syst Rev 2020 Mar 26;3:CD011942 [FREE Full text] [doi: 10.1002/14651858.CD011942.pub2] [Medline: 32216074]
18. Greenhalgh T, Wherton J, Shaw S, Papoutsi C, Vijayaraghavan S, Stones R. Infrastructure revisited: an ethnographic case study of how health information infrastructure shapes and constrains technological innovation. J Med Internet Res 2019 Dec 19;21(12):e16093 [FREE Full text] [doi: 10.2196/16093] [Medline: 31855184]
19. Atherton H, Brant H, Ziebland S, Bikker A, Campbell J, Gibson A, et al. Alternatives to the face-to-face consultation in general practice: focused ethnographic case study. Br J Gen Pract 2018 Apr;68(669):e293-e300 [FREE Full text] [doi: 10.3399/bjgp18X694853] [Medline: 29378697]
20. Brant H, Atherton H, Ziebland S, McKinstry B, Campbell JL, Salisbury C. Using alternatives to face-to-face consultations: a survey of prevalence and attitudes in general practice. Br J Gen Pract 2016 Jul;66(648):e460-e466 [FREE Full text] [doi: 10.3399/bjgp16X685597] [Medline: 27215571]
21. Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, Hanson P, et al. Advantages and limitations of virtual online consultations in a NHS acute trust: the VOCAL mixed-methods study. Health Serv Deliv Res 2018 Jun;6(21):1-136. [doi: 10.3310/hsdr06210] [Medline: 29939519]
22. Ohannessian R, Duong TA, Odone A. Global telemedicine implementation and integration within health systems to fight the COVID-19 pandemic: a call to action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]
23. NHS Digital. What we did in 2019-20: COVID-19 response. NHS Digital Annual Report and Accounts 2019 to 2020. London: NHS Digital; 2020 Aug 27. URL: https://tinyurl.com/28due5rd [accessed 2021-09-27]

24. Wong MYZ, Gunasekeran DV, Nusinovici S, Sabanayagam C, Yeo KK, Cheng C, et al. Telehealth demand trends during the COVID-19 pandemic in the top 50 most affected countries: infodemiological evaluation. JMIR Public Health Surveill 2021 Feb 19;7(2):e24445 [FREE Full text] [doi: 10.2196/24445] [Medline: 33605883]
25. Thomas EE, Haydon HM, Mehrotra A, Caffery LJ, Snoswell CL, Banbury A, et al. Building on the momentum: Sustaining telehealth beyond COVID-19. J Telemed Telecare 2020 Sep 26:1357633X20960638. [doi: 10.1177/1357633X20960638] [Medline: 32985380]
26. Reeves JJ, Hollandsworth HM, Torriani FJ, Taplitz R, Abeles S, Tai-Seale M, et al. Rapid response to COVID-19: health informatics support for outbreak management in an academic health system. J Am Med Inform Assoc 2020 Jun 01;27(6):853-859 [FREE Full text] [doi: 10.1093/jamia/ocaa037] [Medline: 32208481]
27. Basu A, Kuziemsky C, de Araújo Novaes M, Kleber A, Sales F, Al-Shorbaji N, et al. Telehealth and the COVID-19 pandemic: international perspectives and a health systems framework for telehealth implementation to support critical response. Yearb Med Inform 2021 Aug;30(1):126-133 [FREE Full text] [doi: 10.1055/s-0041-1726484] [Medline: 33882598]
28. Gkeredakis M, Lifshitz-Assaf H, Barrett M. Crisis as opportunity, disruption and exposure: Exploring emergent responses to crisis through digital technology. Information and Organization 2021 Mar 01;31(1):100344. [doi: 10.1016/j.infoandorg.2021.100344]
29. Hancock M. The Future of Health care (Speech, 30th July). London: UK Government; 2020 Jun 30. URL: https://www.gov.uk/government/speeches/the-future-of-healthcare [accessed 2021-09-27]
30. Mroz G, Papoutsi C, Rushforth A, Greenhalgh T. Changing media depictions of remote consulting in COVID-19: analysis of UK newspapers. Br J Gen Pract 2021 Jan;71(702):e1-e9 [FREE Full text] [doi: 10.3399/BJGP.2020.0967] [Medline: 33318086]
31. Ministry of Health and Family Welfare. A Big Win for Digital India: Health Ministry's 'eSanjeevani' Telemedicine Service Records 2 Lakh Tele-consultations. New Delhi, India: Press Information Bureau, Government of India; 2020 Aug 19. URL: https://pib.gov.in/PressReleasePage.aspx?PRID=1646913 [accessed 2021-09-27]
32. Davis A. COVID-19: Acknowledging the need to adopt digital health in New Zealand. Healthcare IT News. 2020 May 22. URL: https://www.healthcareitnews.com/news/anz/covid-19-acknowledging-need-adopt-digital-health-new-zealand [accessed 2021-09-27]
33. Webster P. Virtual health care in the era of COVID-19. Lancet 2020 Apr 11;395(10231):1180-1181 [FREE Full text] [doi: 10.1016/S0140-6736(20)30818-7] [Medline: 32278374]
34. Murphy M, Scott LJ, Salisbury C, Turner A, Scott A, Denholm R, et al. Implementation of remote consulting in UK primary care following the COVID-19 pandemic: a mixed-methods longitudinal study. Br J Gen Pract 2021;71(704):e166-e177 [FREE Full text] [doi: 10.3399/BJGP.2020.0948] [Medline: 33558332]
35. Taylor A, Caffery LJ, Gesesew HA, King A, Bassal A, Ford K, et al. How Australian health care services adapted to telehealth during the COVID-19 pandemic: a survey of telehealth professionals. Front Public Health 2021;9:648009 [FREE Full text] [doi: 10.3389/fpubh.2021.648009] [Medline: 33718325]
36. Snoswell CL, Caffery LJ, Haydon HM, Thomas EE, Smith AC. Telehealth uptake in general practice as a result of the coronavirus (COVID-19) pandemic. Aust Health Rev 2020 Sep;44(5):737-740. [doi: 10.1071/AH20183] [Medline: 32853536]
37. Reddy A, Gunnink E, Deeds SA, Hagan SL, Heyworth L, Mattras TF, et al. A rapid mobilization of 'virtual' primary care services in response to COVID-19 at Veterans Health Administration. Healthc (Amst) 2020 Dec;8(4):100464 [FREE Full text] [doi: 10.1016/j.hjdsi.2020.100464] [Medline: 32992109]
38. Spelman JF, Brienza R, Walsh RF, Drost P, Schwartz AR, Kravetz JD, et al. A model for rapid transition to virtual care, VA Connecticut primary care response to COVID-19. J Gen Intern Med 2020 Oct;35(10):3073-3076 [FREE Full text] [doi: 10.1007/s11606-020-06041-4] [Medline: 32705471]
39. Cooperrider D, Whitney D, Stavros J. The Appreciative Inquiry Handbook: For Leaders of Change. Oakland, CA: Berrett-Koehler Publishers; 2008.
40. Trust N. Learning from Scotland's NHS. British Journal of Healthcare Assistants 2017 Aug 02;11(8):407-411. [doi: 10.12968/bjha.2017.11.8.407]
41. Rosen R, Mays N. The impact of the UK NHS purchaser-provider split on the 'rational' introduction of new medical technologies. Health Policy 1998 Feb;43(2):103-123. [doi: 10.1016/s0168-8510(97)00091-2] [Medline: 10177613]
42. Scottish Government. eHealth Strategy 2008-2011. Edinburgh: Government of Scotland; 2008. URL: https://www.webarchive.org.uk/wayback/archive/20170701232915/http://www.gov.scot/Publications/2008/08/27103130/13 [accessed 2021-09-27]
43. Scottish Government. eHealth Strategy 2011-2017 (Revised November 2010). Edinburgh: Government of Scotland; 2012. URL: https://www.gov.scot/publications/ehealth-strategy-2011-2017-revised-july-2012-include-sixth-strategic-aim/ [accessed 2021-09-27]
44. Scottish G. Scotland's Digital Health and Care Strategy: Enabling, Connecting and Empowering. Edinburgh: Government of Scotland; 2018. URL: https://www.gov.scot/publications/scotlands-digital-health-care-strategy-enabling-connecting-empowering/ [accessed 2021-09-27]
45. Scottish Government. Programme for Government. Edinburgh: Government of Scotland; 2019. URL: https://www.gov.scot/publications/protecting-scotlands-future-governments-programme-scotland-2019-20/pages/6/ [accessed 2021-09-27]

46. Scottish Government. Re-mobilise, Recover, Re-design: The Framework for NHS Scotland. Edinburgh: Government of Scotland; 2020. URL: https://www.gov.scot/publications/re-mobilise-recover-re-design-framework-nhs-scotland/ [accessed 2021-09-27]
47. Flyvbjerg B. Five misunderstandings about case-study research. Qualitative Inquiry 2016 Jun 29;12(2):219-245. [doi: 10.1177/1077800405284363]
48. Stake R. Case studies. In: Denzin NK, Lincoln YS, editors. Strategies of Qualitative Inquiry (2nd Ed). Thousand Oaks, CA: Sage; 2003:134-164.
49. Tsoukas H. Craving for Generality and Small-N Studies: A Wittgensteinian Approach towards the Epistemology of the Particular in Organization and Management Theory. In: Tsoukas H, editor. Philosophical Organization Theory. Oxford: Oxford University Press; 2009:384-407.
50. Tsoukas H. Don't simplify, complexify: from disjunctive to conjunctive theorizing in organization and management studies. Jour. of Manage. Stud 2016 Jul 11;54(2):132-153. [doi: 10.1111/joms.12219]
51. Greenhalgh T, Rosen R, Shaw SE, Byng R, Faulkner S, Finlay T, et al. Planning and evaluating remote consultation services: a new conceptual framework incorporating complexity and practical ethics. Front. Digit. Health 2021 Aug 13;3:726095. [doi: 10.3389/fdgth.2021.726095]
52. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
53. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
54. Howie JG, Heaney DJ, Maxwell M, Walker JJ. A comparison of a Patient Enablement Instrument (PEI) against two established satisfaction scales as an outcome measure of primary care consultations. Fam Pract 1998 Apr;15(2):165-171. [doi: 10.1093/fampra/15.2.165] [Medline: 9613486]
55. Wherton J, Greenhalgh T. Evaluation of the Attend Anywhere/Near Me Video Consulting Service in Scotland. Edinburgh: Scottish Government; 2019. URL: https://www.gov.scot/publications/evaluation-attend-anywhere-near-video-consulting-service-scotland-2019-20-main-report/pages/2/; [accessed 2021-09-27]
56. Wherton J, Greenhalgh T. Evaluation of the Near Me Video Consulting Service in Scotland During COVID-19. Edinburgh: Scottish Government; 2020. URL: https://tinyurl.com/djzs56k [accessed 2021-09-27]
57. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academies Press; 2001.
58. Seuren LM, Wherton J, Greenhalgh T, Cameron D, A'Court C, Shaw SE. Physical examinations via video for patients with heart failure: qualitative study using conversation analysis. J Med Internet Res 2020 Feb 20;22(2):e16694 [FREE Full text] [doi: 10.2196/16694] [Medline: 32130133]
59. Gunasekeran D, Liu Z, Tan W, Koh J, Cheong C, Tan L, et al. Evaluating safety and efficacy of follow-up for patients with abdominal pain using video consultation (SAVED Study): randomized controlled trial. J Med Internet Res 2020 Jun 15;22(6):e17417 [FREE Full text] [doi: 10.2196/17417] [Medline: 32459637]
60. Johnsen TM, Norberg BL, Kristiansen E, Zanaboni P, Austad B, Krogh FH, et al. Suitability of video consultations during the COVID-19 pandemic lockdown: cross-sectional survey among Norwegian general practitioners. J Med Internet Res 2021 Feb 08;23(2):e26433 [FREE Full text] [doi: 10.2196/26433] [Medline: 33465037]
61. McBeth R. GPs Move to Digital First Health Informatics New Zealand. 2020. URL: https://www.hinz.org.nz/news/495292/GPs-move-to-digital-first-approach.htm [accessed 2021-09-27]
62. New Zealand Government. Extension of Temporary Prescription Waivers During COVID-19. Auckland, New Zealand: Government of New Zealand; 2021 Sep 14. URL: https://www.health.govt.nz/system/files/documents/pages/advisory-extension-temporary-prescription-waivers-during-covid-15sep21.pdf [accessed 2021-09-27]
63. Day K, Hunter I, Rao V, Russell G, Roskvist R, Moir F, et al. Survey protocol for exploring video and phone use in Aotearoa New Zealand general practice: considerations for future telehealth. BMJ Health Care Inform 2021 Mar 22;28(1):e100309 [FREE Full text] [doi: 10.1136/bmjhci-2020-100309] [Medline: 33753326]
64. Al-Busaidi I, Martin M. The transition to a "virtual practice" in primary care during the COVID-19 pandemic: experience from one medical centre in New Zealand. N Z Med J 2020 Aug 21;133(1520):91-98. [Medline: 32994597]
65. Imlach F, McKinlay E, Middleton L, Kennedy J, Pledger M, Russell L, et al. Telehealth consultations in general practice during a pandemic lockdown: survey and interviews on patient experiences and preferences. BMC Fam Pract 2020 Dec 13;21(1):269 [FREE Full text] [doi: 10.1186/s12875-020-01336-1] [Medline: 33308161]
66. Harvey G, Kitson A. PARIHS revisited: from heuristic to integrated framework for the successful implementation of knowledge into practice. Implement Sci 2016 Mar 10;11:33 [FREE Full text] [doi: 10.1186/s13012-016-0398-2] [Medline: 27013464]
67. Van de Ven A. The Innovation Journey. Oxford, UK: Oxford University Press; 2008.

68. Linden E. Pandemics and environmental shocks: What aviation managers should learn from COVID-19 for long-term planning. J Air Transp Manag 2021 Jan;90:101944 [FREE Full text] [doi: 10.1016/j.jairtraman.2020.101944] [Medline: 33071486]
69. Kichloo A, Albosta M, Dettloff K, Wani F, El-Amir Z, Singh J, et al. Telemedicine, the current COVID-19 pandemic and the future: a narrative review and perspectives moving forward in the USA. Fam Med Community Health 2020 Aug;8(3):e000530 [FREE Full text] [doi: 10.1136/fmch-2020-000530] [Medline: 32816942]
70. Holmner, Ebi KL, Lazuardi L, Nilsson M. Carbon footprint of telemedicine solutions--unexplored opportunity for reducing carbon emissions in the health sector. PLoS One 2014;9(9):e105040 [FREE Full text] [doi: 10.1371/journal.pone.0105040] [Medline: 25188322]
71. Tham Y, Husain R, Teo KYC, Tan ACS, Chew ACY, Ting DS, et al. New digital models of care in ophthalmology, during and beyond the COVID-19 pandemic. Br J Ophthalmol 2021 Mar 22 (forthcoming). [doi: 10.1136/bjophthalmol-2020-317683] [Medline: 33753407]

Abbreviations

GDPR: General Data Protection Regulations
i-PARIHS: integrated Promoting Action on Research Implementation in Health Services
IT: information technology
NASSS: nonadoption and abandonment by individuals, and challenges to spread, scale-up and sustainability
NHS: National Health Service (UK)
PERCS: Planning and Evaluating Remote Consultation Services
TEC: technology-enabled care

Edited by C Basch; submitted 18.06.21; peer-reviewed by P Murphie, S Hall, M Mbwogge, D Gunasekeran; comments to author 12.07.21; revised version received 20.07.21; accepted 06.08.21; published 07.10.21.

Please cite as:
Wherton J, Greenhalgh T, Shaw SE
Expanding Video Consultation Services at Pace and Scale in Scotland During the COVID-19 Pandemic: National Mixed Methods Case Study
J Med Internet Res 2021;23(10):e31374
URL: https://www.jmir.org/2021/10/e31374
doi:10.2196/31374
PMID:34516389

Original Paper

International Changes in COVID-19 Clinical Trajectories Across 315 Hospitals and 6 Countries: Retrospective Cohort Study

Griffin M Weber[1*], MD, PhD; Harrison G Zhang[1*], HSDG[‡]; Sehi L'Yi[1*], PhD; Clara-Lea Bonzel[1], MSc; Chuan Hong[1], PhD; Paul Avillach[1], MD, PhD; Alba Gutiérrez-Sacristán[1], PhD; Nathan P Palmer[1], PhD; Amelia Li Min Tan[1], BSc, PhD; Xuan Wang[1], PhD; William Yuan[1], PhD; Nils Gehlenborg[1], PhD; Anna Alloni[2], PhD; Danilo F Amendola[3], MSc; Antonio Bellasi[4], MD, PhD; Riccardo Bellazzi[5], MS, PhD; Michele Beraghi[6], MS; Mauro Bucalo[2], MS; Luca Chiovato[7], MD, PhD; Kelly Cho[8], PhD; Arianna Dagliati[9], MS, PhD; Hossein Estiri[10], PhD; Robert W Follett[11], BS; Noelia García Barrio[12], MS; David A Hanauer[13], MS, MD; Darren W Henderson[14], BS; Yuk-Lam Ho[8], MPH; John H Holmes[15,16], MS, PhD; Meghan R Hutch[17], BS; Ramakanth Kavuluru[18], PhD; Katie Kirchoff[19], MSHI; Jeffrey G Klann[10], MEng, PhD; Ashok K Krishnamurthy[20], PhD; Trang T Le[15], PhD; Molei Liu[21], MSc; Ne Hooi Will Loh[22], MBBS; Sara Lozano-Zahonero[23], PhD; Yuan Luo[17], PhD; Sarah Maidlow[24], AA; Adeline Makoudjou[23], MD; Alberto Malovini[25], PhD; Marcelo Roberto Martins[26], MSc; Bertrand Moal[27], MD, PhD; Michele Morris[28], BA; Danielle L Mowery[15], PhD; Shawn N Murphy[29], MD, PhD; Antoine Neuraz[30], MD, PhD; Kee Yuan Ngiam[31], MBBS, MRCS, MMed; Marina P Okoshi[32], MD, PhD; Gilbert S Omenn[33], MD, PhD; Lav P Patel[34], MS; Miguel Pedrera Jiménez[12], MS; Robson A Prudente[32], PhD; Malarkodi Jebathilagam Samayamuthu[28], MD; Fernando J Sanz Vidorreta[11], BS; Emily R Schriver[35], MS; Petra Schubert[8], MPH; Pablo Serrano Balazote[12], MS, MD; Byorn WL Tan[36], MBBS; Suzana E Tanni[32], MD, PhD; Valentina Tibollo[25], MS; Shyam Visweswaran[28], MD, PhD; Kavishwar B Wagholikar[10], MBBS, PhD; Zongqi Xia[37], MD, PhD; Daniela Zöller[23], PhD; The Consortium For Clinical Characterization Of COVID-19 By EHR (4CE)[39]; Isaac S Kohane[1], MD, PhD; Tianxi Cai[1], ScD[‡]; Andrew M South[38], MS, MD; Gabriel A Brat[1], MPH, MD

[1]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, United States

[2]BIOMERIS (BIOMedical Research Informatics Solutions), Pavia, Italy

[3]Clinical Research Unit, Botucatu Medical School, São Paulo State University, Botucatu, Brazil

[4]Division of Nephrology, Department of Medicine, Ente Ospedaliero Cantonale, Lugano, Switzerland

[5]Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy

[6]Information Technology Department, Azienda Socio-Sanitaria Territoriale di Pavia, Pavia, Italy

[7]Unit of Internal Medicine and Endocrinology, Istituti Clinici Scientifici Maugeri SpA SB IRCCS, Pavia, Italy

[8]Massachusetts Veterans Epidemiology Research and Information Center, Veterans Affairs Boston Healthcare System, Boston, MA, United States

[9]Department of Electrical Computer and Biomedical Engineering, University of Pavia, Pavia, Italy

[10]Department of Medicine, Massachusetts General Hospital, Boston, MA, United States

[11]Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States

[12]Health Informatics, Hospital Universitario 12 de Octubre, Madrid, Spain

[13]Department of Learning Health Sciences, University of Michigan Medical School, Ann Arbor, MI, United States

[14]Department of Biomedical Informatics, University of Kentucky, Lexington, KY, United States

[15]Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States

[16]Institute for Biomedical Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States

[17]Department of Preventive Medicine, Northwestern University, Chicago, IL, United States

[18]Institute for Biomedical Informatics, University of Kentucky, Lexington, KY, United States

[19]Medical University of South Carolina, Charleston, SC, United States

[20]Department of Computer Science, Renaissance Computing Institute, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[21]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, United States

[22]Department of Anaesthesia, National University Health System, Singapore, Singapore

[23]Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany

[24]Michigan Institute for Clinical & Health Research Informatics, University of Michigan, Ann Arbor, MI, United States

[25]Laboratory of Informatics and Systems Engineering for Clinical Research, Istituti Clinici Scientifici Maugeri SpA SB IRCCS, Pavia, Italy
[26]Clinical Hospital of Botucatu Medical School, São Paulo State University, Botucatu, Brazil
[27]Informatique et archivistique médicales unit, Bordeaux University Hospital, Bordeaux, France
[28]Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, United States
[29]Department of Neurology, Massachusetts General Hospital, Boston, MA, United States
[30]Department of Biomedical Informatics, Hôpital Necker-Enfants Malade, Assistance Publique Hôpitaux de Paris, University of Paris, Paris, France
[31]Department of Biomedical Informatics, Institute for Digital Medicine, National University Health System, Singapore, Singapore
[32]Internal Medicine Department, Botucatu Medical School, São Paulo State University, Botucatu, Brazil
[33]Department of Computational Medicine & Bioinformatics, Internal Medicine, Human Genetics, and Public Health, University of Michigan, Ann Arbor, MI, United States
[34]Division of Medical Informatics, Department of Internal Medicine, University of Kansas Medical Center, Kansas City, KS, United States
[35]Data Analytics Center, University of Pennsylvania Health System, Philadelphia, PA, United States
[36]Department of Medicine, National University Health System, Singapore, Singapore
[37]Department of Neurology, University of Pittsburgh, Pittsburgh, PA, United States
[38]Section of Nephrology, Department of Pediatrics, Brenner Children's Hospital, Wake Forest School of Medicine, Winston Salem, NC, United States
[39]see Authors' Contributions
[‡]The Consortium for Clinical Characterization of COVID-19 by EHR (4CE)
[*]these authors contributed equally

Corresponding Author:
Tianxi Cai, ScD
Department of Biomedical Informatics
Harvard Medical School
10 Shattuck St
Boston, MA, 02115
United States
Phone: 1 6175488520
Email: tcai@hsph.harvard.edu

Abstract

Background: Many countries have experienced 2 predominant waves of COVID-19–related hospitalizations. Comparing the clinical trajectories of patients hospitalized in separate waves of the pandemic enables further understanding of the evolving epidemiology, pathophysiology, and health care dynamics of the COVID-19 pandemic.

Objective: In this retrospective cohort study, we analyzed electronic health record (EHR) data from patients with SARS-CoV-2 infections hospitalized in participating health care systems representing 315 hospitals across 6 countries. We compared hospitalization rates, severe COVID-19 risk, and mean laboratory values between patients hospitalized during the first and second waves of the pandemic.

Methods: Using a federated approach, each participating health care system extracted patient-level clinical data on their first and second wave cohorts and submitted aggregated data to the central site. Data quality control steps were adopted at the central site to correct for implausible values and harmonize units. Statistical analyses were performed by computing individual health care system effect sizes and synthesizing these using random effect meta-analyses to account for heterogeneity. We focused the laboratory analysis on C-reactive protein (CRP), ferritin, fibrinogen, procalcitonin, D-dimer, and creatinine based on their reported associations with severe COVID-19.

Results: Data were available for 79,613 patients, of which 32,467 were hospitalized in the first wave and 47,146 in the second wave. The prevalence of male patients and patients aged 50 to 69 years decreased significantly between the first and second waves. Patients hospitalized in the second wave had a 9.9% reduction in the risk of severe COVID-19 compared to patients hospitalized in the first wave (95% CI 8.5%-11.3%). Demographic subgroup analyses indicated that patients aged 26 to 49 years and 50 to 69 years; male and female patients; and black patients had significantly lower risk for severe disease in the second wave than in the first wave. At admission, the mean values of CRP were significantly lower in the second wave than in the first wave. On the seventh hospital day, the mean values of CRP, ferritin, fibrinogen, and procalcitonin were significantly lower in the second wave than in the first wave. In general, countries exhibited variable changes in laboratory testing rates from the first to the second wave. At admission, there was a significantly higher testing rate for D-dimer in France, Germany, and Spain.

Conclusions: Patients hospitalized in the second wave were at significantly lower risk for severe COVID-19. This corresponded to mean laboratory values in the second wave that were more likely to be in typical physiological ranges on the seventh hospital day compared to the first wave. Our federated approach demonstrated the feasibility and power of harmonizing heterogeneous

EHR data from multiple international health care systems to rapidly conduct large-scale studies to characterize how COVID-19 clinical trajectories evolve.

(J Med Internet Res 2021;23(10):e31400) doi:10.2196/31400

KEYWORDS

SARS-CoV-2; electronic health records; federated study; retrospective cohort study; meta-analysis; COVID-19; severe COVID-19; laboratory trajectory

Introduction

From January 2020 to June 2021, the COVID-19 pandemic has resulted in over 170 million confirmed cases of SARS-CoV-2 infection and 3.7 million confirmed deaths worldwide [1]. Similar to previous viral pandemics, the resurgence in SARS-CoV-2 infections and subsequent hospitalizations since the first documented outbreaks have been characterized by a series of "waves." To date, there have been reports of at least two waves in numerous countries, including an initial one in the Spring of 2020 and a resurgence of cases in the Summer and Fall of 2020 [1-10]. A limited number of single-center studies have reported differences in laboratory values, demographic composition, and disease management between patients with COVID-19 admitted in the first and second waves [6,8,11,12]. Thus, there is substantial interest in comparing the clinical trajectories of patients with SARS-CoV-2 who were hospitalized across different waves of the pandemic to better understand the rapidly evolving epidemiology, pathophysiology, and health care dynamics of the COVID-19 pandemic. This may further inform health care workers, policymakers, and public health experts on how to anticipate potential additional waves due to SARS-CoV-2 variants [13].

Single-center studies are limited in scope, power, and generalizability, and there is a need for robust multicenter analyses using multinational cohorts that compare first and second wave patient characteristics. The goal of this study was to use a federated electronic health record (EHR)-based approach to examine international temporal trends in the clinical trajectories of patients hospitalized with SARS-CoV-2 across 6 countries obtained from contributing health care systems in the Consortium for Clinical Characterization of COVID-19 by EHR (4CE) [14], an international research collaborative of more than 300 hospitals across 7 countries that collects patient-level EHR data to study the epidemiology and clinical course of COVID-19. We collected data from 26 participating international health care systems covering 79,613 hospitalized patients with SARS-CoV-2 to study changes in (1) hospitalization rates across calendar time; (2) risk of developing severe COVID-19; and (3) mean laboratory values and laboratory testing rates between the first and second waves. We stratified severity risk analyses by country and demographic subgroups.

Methods

Description of the Federated Approach: Participating Health Care Systems, Local Data Collection, and Central Data Aggregation

Our analyses were performed on EHR data collected from 315 hospitals (affiliated with 26 regional health care systems) across the following 6 countries: Brazil, France, Germany, Italy, Spain, and the United States [14,15]. In the United States, we grouped the 170 Veterans Affairs (VA) hospitals into 5 regional health care systems [16]. See Table 1 for details about participating health care systems and Figure 1 for a map of participating health care systems.

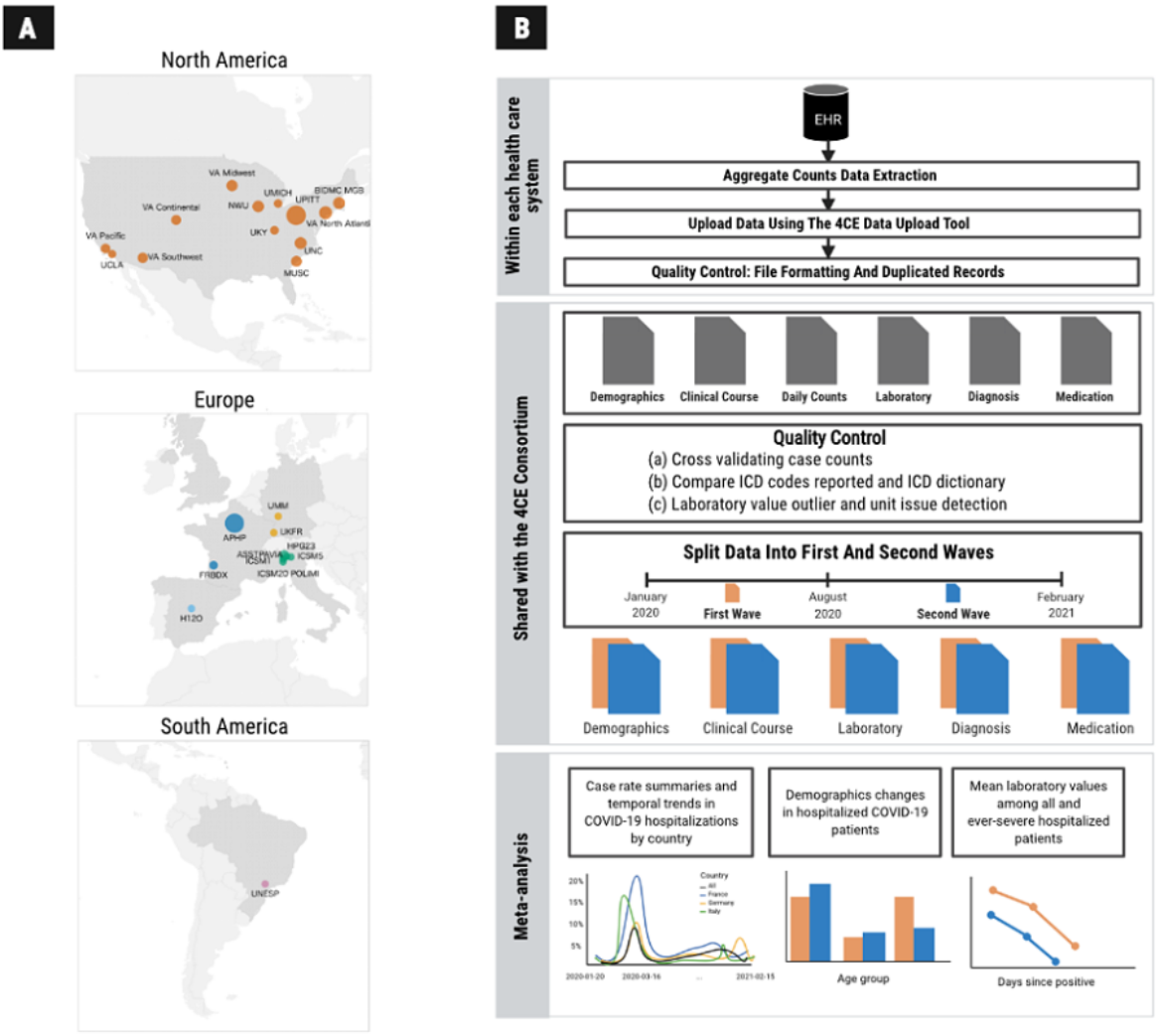
Table 1. Participating health care systems, metadata on the number of hospitals and beds, and hospitalization date used to define the first and second wave cohorts.

Health care system	Country	Hospitals, n	Beds, n	Inpatient discharges/year, n	First wave date range	Second wave date range	First wave sample size, n	Second wave sample size, n
Assistance Publique - Hôpitaux de Paris	France	39	20,098	1,375,538	January 29, 2020, to August 10, 2020	August 11, 2020, to November 06, 2020	9827	4584
Azienda Socio-Sanitaria Territoriale della provincia di Pavia	Italy	7	958	29,103	February 28, 2020, to April 30, 2020	May 01, 2020, to February 15, 2021	945	1543
Beth Israel Deaconess Medical Center	United States	1	673	40,752	March 23, 2020, to September 02, 2020	September 03, 2020, to February 22, 2021	685	585
Bordeaux University Hospital	France	3	2676	130,033	January 23, 2020, to July 31, 2020	August 01, 2020, to November 07, 2020	331	439
Hospital Universitario 12 de Octubre	Spain	1	1256	45,035	March 01, 2020, to July 20, 2020	July 21, 2020, to February 28, 2021	2369	3730
Azienda Socio-Sanitaria Territoriale Papa Giovanni XXIII Bergamo	Italy	1	1080	45,000	February 25, 2020, to May 21, 2020	May 22, 2020, to November 13, 2020	1533	371
Istituto Clinico Scientifico Maugeri Pavia Hospital	Italy	1	426	8616	February 29, 2020, to May 08 2020	May 09, 2020, to November 16, 2020	113	58
Istituto Clinico Scientifico Maugeri Milano Hospital	Italy	1	200	2432	February 21, 2020, to May 08, 2020	May 09, 2020, to November 16, 2020	38	119
Istituto Clinico Scientifico Maugeri Lumezzane/Brescia Hospitals	Italy	1	149	1296	March 11, 2020, to May 08, 2020	May 09, 2020, to November 16, 2020	111	21
Mass General Brigham (Partners Healthcare)	United States	10	3418	163,521	March 11, 2020, to July 31, 2020	August 01, 2020, to February 28, 2021	2736	1735
Medical University of South Carolina	United States	8	1600	55,664	March 12, 2020, to May 25, 2020	May 26, 2020, to November 15, 2020	127	1482
Northwestern University	United States	10	2234	103,279	March 05, 2020, to July 31, 2020	August 01, 2020, to December 31, 2020	2313	3567
Policlinico di Milano	Italy	1	900	40,000	February 25, 2020, to August 01, 2020	August 02, 2020, to November 13, 2020	612	304
Medical Center, University of Freiburg	Germany	1	1660	71,500	March 13, 2020, to July 31, 2020	August 01, 2020, to February 28, 2021	186	490
University of California, LA	United States	2	786	40,526	March 10, 2020, to August 03, 2020	August 04, 2020, to November 13, 2020	425	151
University of Kentucky	United States	3	881	45,714	March 18, 2020, to July 07, 2020	July 08, 2020, to November 06, 2020	113	352
University of Michigan	United States	3	1000	49,008	March 09, 2020, to July 31, 2020	August 01, 2020, to February 28, 2021	745	1619
University Medicine Mannheim	Germany	1	1352	50,748	March 18, 2020, to August 03, 2020	August 04, 2020, to January 23, 2021	81	497

Health care system	Country	Hospitals, n	Beds, n	Inpatient discharges/year, n	First wave date range	Second wave date range	First wave sample size, n	Second wave sample size, n
University of North Carolina at Chapel Hill	United States	11	3095	52,000	March 14, 2020, to June 05, 2020	June 06, 2020, to October 30, 2020	458	1525
Universidade Estadual Julio de Mesquita Filho	Brazil	1	490	28,167	April 01, 2020, to July 31, 2020	August 01, 2020, to February 28, 2021	171	425
University of Pittsburgh	United States	39	8085	369,300	March 13, 2020, to July 31, 2020	August 01, 2020, to February 28, 2021	685	5021
Veteran Affairs North Atlantic	United States	49	3594	151,075	March 01, 2020, to July 31, 2020	August 01, 2020, to February 04, 2021	1949	2984
Veteran Affairs Southwest	United States	29	3115	156,315	March 01, 2020, to July 31, 2020	August 01, 2020, to February 04, 2021	1679	4071
Veteran Affairs Midwest	United States	39	2686	145,468	March 01, 2020, to July 31, 2020	August 01, 2020, to February 04, 2021	1544	4617
Veteran Affairs Continental	United States	24	2110	113,260	March 01, 2020, to July 31, 2020	August 01, 2020, to February 04, 2021	1497	3495
Veteran Affairs Pacific	United States	29	2296	114,569	March 01, 2020, to July 31, 2020	August 01, 2020, to February 04, 2021	1194	3361
Total	N/A[a]	315	66,818	3,427,919	N/A	N/A	32,467	47,146

[a]N/A: not applicable.

Figure 1. (A) World map with health care systems. (B) Schematic of the federated electronic health record (EHR)-based study involving health care systems from 6 countries. 4CE: Consortium for Clinical Characterization of COVID-19 by EHR; ICD: International Classification of Diseases.



Similar to our previous 4CE studies, we distributed a SQL database script to each of the contributing health care systems, which they ran on their patient-level EHR data to generate aggregate counts and statistics about their patient cohorts after gaining institutional review board approval [14,15,17]. Health care systems then uploaded their aggregate data via a central 4CE data upload tool. Aggregate data included hospital admission summaries over calendar time; dates of positive SARS-CoV-2 reverse transcription polymerase chain reaction tests; demographic counts for age, sex, and race groups; and daily trajectories of laboratory test values. Race data were only reported in participating health care systems from the United States and only included categories for black and white patients given the considerable heterogeneity present in race coding systems across health care system EHRs. A schematic of our workflow is presented in Figure 1, and further details of collected data are reported in Multimedia Appendix 1.

In order to ensure high-quality EHR data across health care systems and countries, extensive data quality control was performed. When sites uploaded the data using the 4CE data upload tool, an initial online quality control verified that all comma separated value (CSV) files were under the standard format, which included verification of the file and column names, column orders, data types, code values and ranges, and the absence of duplicated records. This step was crucial in ensuring proper downstream statistical analysis. At the central site, additional quality control steps were completed on all submitted data. These steps included cross-validating the consistency of the total case counts, checking that there are no negative values in patient counts, and verifying that no data types were missing. We also checked for consistency between the 3-digit International Classification of Diseases (ICD) codes and the ICD dictionary. To assess the general consistency of the laboratory data from each site with data observed from all health care systems and to detect outliers, we plotted laboratory values across time with all sites overlaid on each other. Using these plots, we were also able to check if certain laboratory values from specific health care systems were systematically lower or higher than for other sites, which usually indicated unit errors. If a health care system presented any quality control

issues, the central site directly contacted its corresponding informaticians to resolve them.

Cohort Identification

Our study included all patients hospitalized up to February 28, 2021, at participating 4CE health care systems with an admission date between 7 days before to 14 days after the date of the first positive SARS-CoV-2 reverse transcription polymerase chain reaction test result. We chose this time window in an effort to mitigate selection bias by identifying hospitalized patients who may have tested positive for SARS-CoV-2 before or after being admitted to a hospital. We further defined the first admission date within this −7 to +14-day time window as the index date, and "days since admission" and "hospital day" are referenced to this index date. EHR data were available for 79,613 patients.

We partitioned patients into first- and second-wave cohorts according to their index date. Although different regions had slightly varying temporal trajectories in COVID-19–related hospitalizations, our data indicated 2 predominant waves of hospitalizations, which we used to partition patients as follows: a first wave from January 1 to July 31, 2020, and a second wave from August 1, 2020, to February 28, 2021. Relatively few admissions occurred between July and August 2020 across all health care systems. Table 1 defines health care system-specific hospitalization date criteria for the first and second waves.

We further categorized patients as "ever-severe" using the validated 4CE COVID-19 severity algorithm that allows us to determine whether patients, at any time during their hospitalization, progressed to severe disease, regardless of their recovery from COVID-19 [18]. The algorithm leverages a set of EHR data elements to define severe COVID-19 including (1) laboratory tests for partial pressure of carbon dioxide or partial pressure of oxygen; (2) ordered medications for sedatives and anesthetics; (3) diagnosis codes for acute respiratory distress syndrome or ventilator-associated pneumonia; and (4) procedures such as endotracheal tube insertion and invasive mechanical ventilation [18].

Statistical Analysis

Centralized random effect meta-analyses were performed to summarize individual health care system effect sizes. To account for heterogeneity between health care systems, we harmonized effect sizes using DerSimonian and Laird random effect meta-analysis [19]. Weights assigned to health care system effect sizes during meta-analysis were kept constant between corresponding first and second cohort analyses to facilitate effective comparisons between waves. All statistical analyses were performed using R software version 4.0.2 (R Foundation for Statistical Computing).

We estimated the intensity rate of hospitalizations over time within each participating health care system and averaged at the country level. Within each health care system, the intensity rate for a given calendar date was estimated as the proportion of patients in the cohort who were hospitalized on that date. We further summarized the prevalence of demographic subgroups in the first and second waves. We excluded the VA health care systems only when estimating the prevalence of demographic subgroups in our cohort due to their unique demographic profiles [20]. We report the prevalence of demographic subgroups including the VA health care systems in Multimedia Appendix 2.

We then estimated the absolute risk of severe COVID-19 in the first and second waves and the relative risk (RR) of severe COVID-19 in the second wave compared with the first wave. Within each health care system and over a set time period of interest, the absolute risk was estimated as the proportion of patients who ever developed severe disease among all patients in the corresponding cohort. We stratified these analyses by country and demographic subgroups. Analyses of absolute risk and RR for severe COVID-19 included all participating health care systems.

We then compared standardized mean laboratory test values stratified by disease severity at days 0, 1, and 7 to investigate changes in laboratory trajectories between the 2 waves. We focused on the following 6 blood laboratory values associated with worse outcomes and severe disease in patients with COVID-19: C-reactive protein (CRP), ferritin, fibrinogen, procalcitonin, D-dimer, and creatinine [21-28]. To facilitate effective comparisons, we defined standardized laboratory values as relative to each laboratory test's harmonized value on the index date in the first wave. We also summarized the proportion of all and ever-severe patients having each laboratory test at days 0 to 14 stratified by country to examine any changes in clinical practice regarding laboratory testing.

Ethics Approval

All study sites were responsible for and obtained ethics approval, as needed, from the appropriate ethics committee at their institutions.

Data Sharing Statement

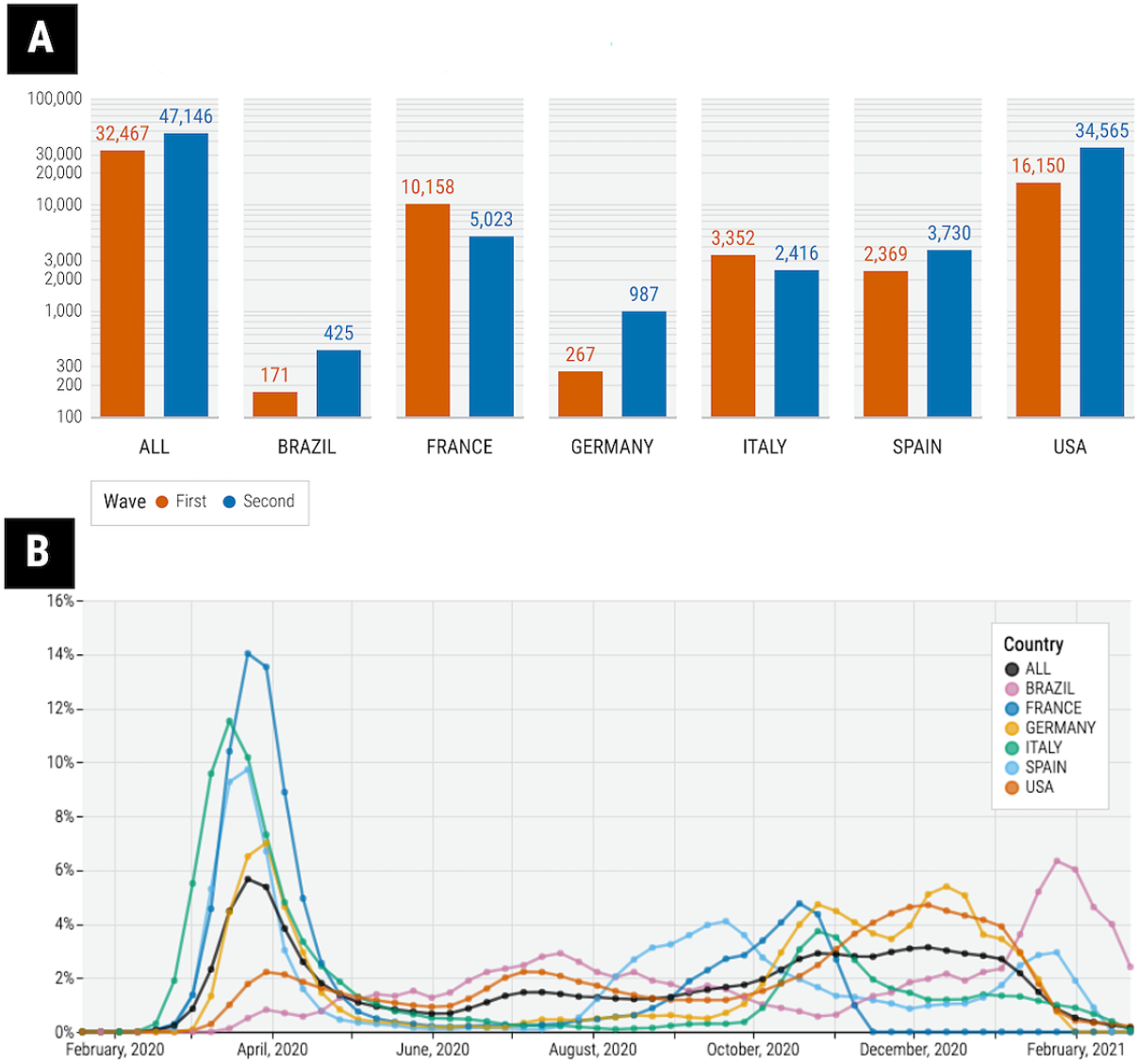
Deidentified aggregate data were provided by sites for this study. The 4CE provides samples of deidentified data collected by the consortium and some corresponding visualizations on the consortium website [29].

Results

Characteristics of the Study Population and Trends in Hospitalization

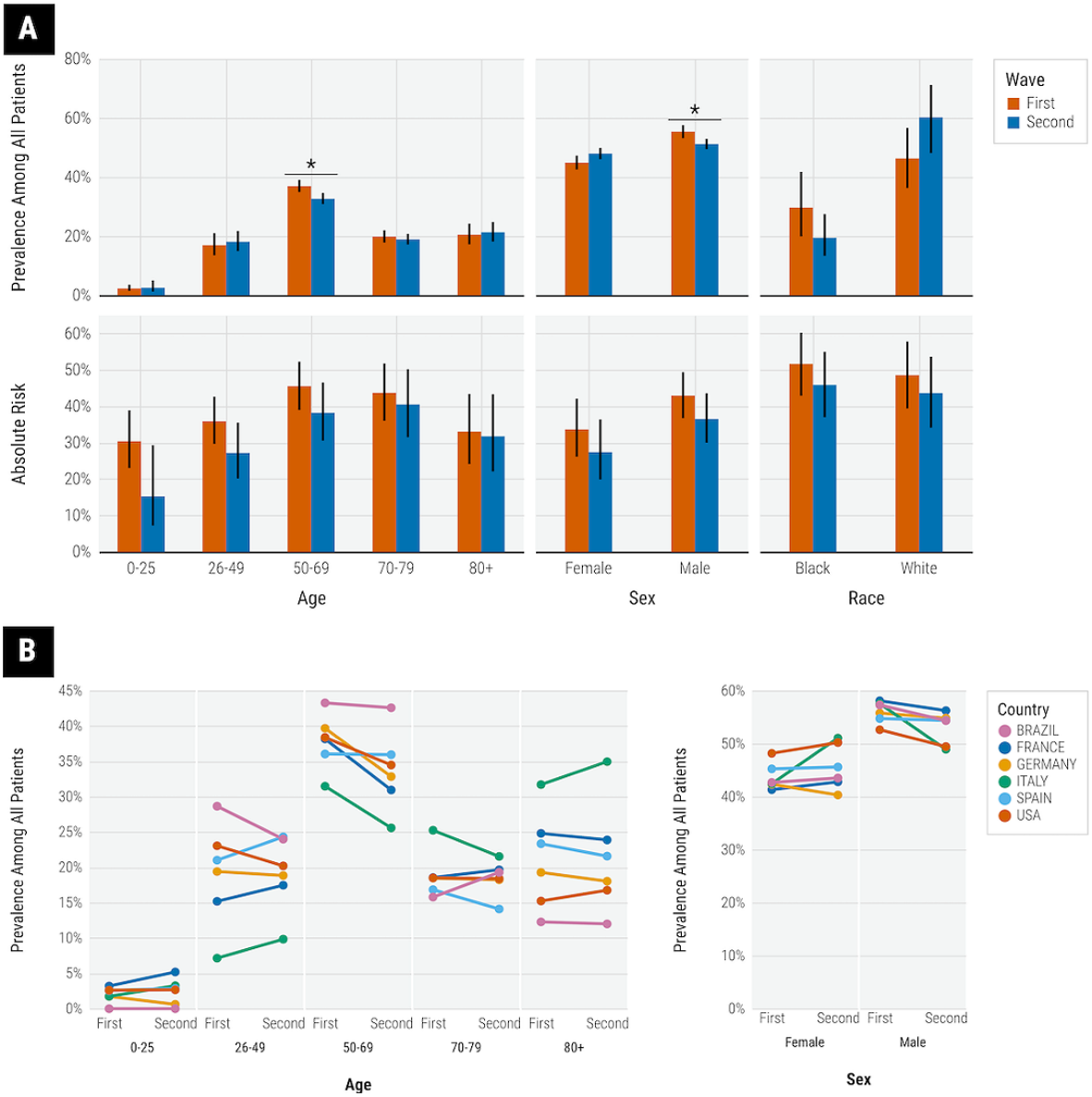
In the study population of 79,613 hospitalized patients with SARS-CoV-2, 32,467 were hospitalized during the first wave and 47,146 were hospitalized during the second wave. In this cohort, the United States represented the country with the most hospitalizations. As seen in Figure 2B, hospitalization rates generally peaked in March-April of 2020 and again in the final months of 2020 across all 6 countries.

Figure 2. (A) Total hospitalizations in the cohort between the first and second waves. (B) The intensity rate of hospitalizations over time by country.



We report the prevalence of demographic subgroups in Figure 3. Overall, there was a higher prevalence of male and older age patients in both waves. The prevalence of patients aged 50 to 69 years decreased significantly from the first wave (37.1%, 95% CI 35.0%-39.3%) to the second wave (32.3%, 95% CI 30.5%-34.2%). The prevalence of male patients also decreased from the first wave (55.3%, 95% CI 53.1%-57.5%) to the second wave (50.9%, 95% CI 49.1%-52.6%). There were no statistically significant changes in the prevalence of other age or sex groups for the entire cohort. At the country level, we observed that in Spain, the prevalence of patients aged 26 to 49 years increased significantly from 21.0% (95% CI 19.4%-22.7%) in the first wave to 24.3% (95% CI 23.0%-25.7%) in the second wave, while the prevalence of patients aged 70 to 79 years decreased significantly from 16.8% (95% CI 15.4%-18.4%) in the first wave to 14.1% (95% CI 13.0%-15.3%) in the second wave. In the United States, the prevalence of white patients increased (first wave: 46.3%, 95% CI 36.3%-56.6% vs second wave: 60.2%, 95% CI 48.1%-71.1%), while the prevalence of black patients decreased (first wave: 29.7%, 95% CI 20.0%-41.8% vs second wave: 19.4%, 95% CI 13.3%-27.5%), although both results did not reach statistical significance.

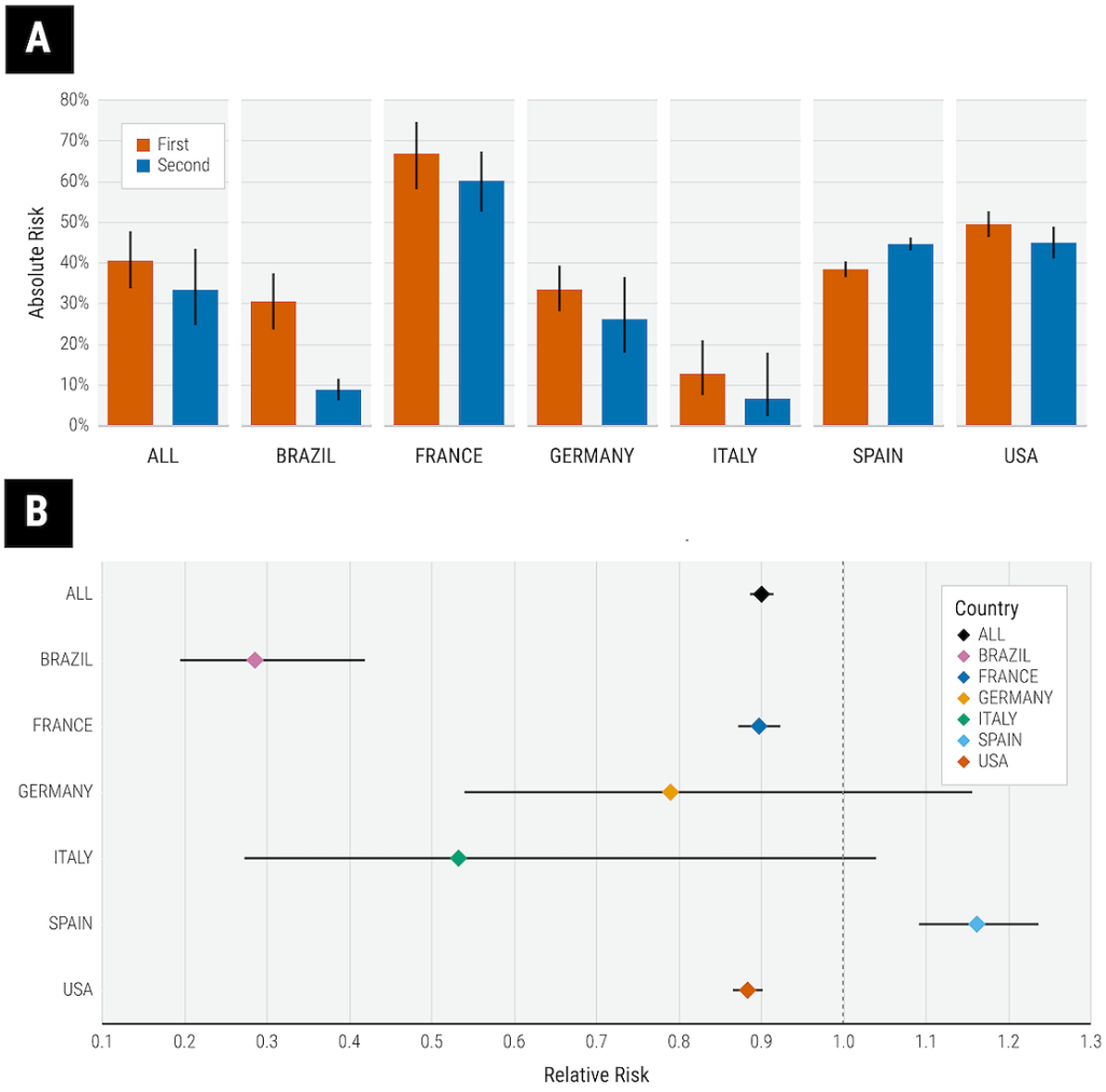
Figure 3. (A) Prevalence of demographic subgroups in the first and second waves for the entire cohort and estimated absolute risk for severe COVID-19 by demographic subgroup. (B) Country-level demographics of all patients by wave. Error bars indicate 95% CIs. *Significant difference by nonoverlapping 95% CIs.



Risk of Severe Disease in the First and Second Waves

We report the absolute risk and RR for severe COVID-19 stratified by country in Figure 4. The absolute risk of severe disease was 0.40 (95% CI 0.34-0.48) in the first wave and 0.33 (95% CI 0.25-0.43) in the second wave. The absolute risk for severe disease varied significantly across countries in both waves (first wave vs second wave) as follows: Brazil, 30.1% vs 8.7%; France, 66.7% vs 60.1%; Germany, 33.3% vs 25.8%; Italy, 12.6% vs 6.5%; Spain, 38.3% vs 44.5%; United States, 49.4% vs 44.8%. The RR of severe disease in the second wave compared to the first wave was more comparable across Brazil (RR 0.29, 95% CI 0.20-0.42), France (RR 0.90, 95% CI 0.87-0.92), and the United States (RR 0.88, 95% CI 0.87-0.90). The observed RR was 0.78 (95% CI 0.52-1.17) in Germany and 0.53 (95% CI 0.27-1.04) in Italy, but the reduction in risk was not statistically significant. In contrast, patients in Spain (RR 1.16, 95% CI 1.09-1.24) had a slightly higher risk of severe COVID-19 in the second wave than in the first wave.

Figure 4. (A) Absolute risk for severe COVID-19 in the first and second waves. (B) Relative risk of severe COVID-19 in the second compared to the first wave stratified by country. Error bars indicate 95% CIs.



We report the absolute risk and RR for severe COVID-19 stratified by demographic subgroups in Figure 3A and Table 2, respectively. Across demographic subgroups in the second wave, there were significant reductions in risk among patients aged 26 to 49 years and 50 to 69 years; male and female patients; and black patients. RR effect sizes were generally comparable between sexes and between races. The reduction in risk in the second wave was slightly greater for younger age groups than for older age groups.

Table 2. Relative risk of severe disease in the second wave compared to the first wave stratified by demographic subgroups and by country.

Demographic group	All countries, RR[a] (95% CI)	Brazil, RR (95% CI)	France, RR (95% CI)	Germany, RR (95% CI)	Italy, RR (95% CI)	Spain, RR (95% CI)	United States, RR (95% CI)
Age (years)							
0-25	0.75 (0.56-1.02)	N/A[b]	1.50 (1.08-2.09)[c]	N/A	N/A	N/A	0.80 (0.59-1.09)
26-49	0.77 (0.63-0.94)[c]	0.31 (0.16-0.61)[c]	0.86 (0.57-1.31)	1.1 (0.51-2.45)	0.24 (0.07-0.78)[c]	1.08 (0.90-1.29)	0.81 (0.69-0.94)[c]
50-69	0.84 (0.72-0.97)[c]	0.23 (0.12-0.43)[c]	0.95 (0.89-1.01)	0.88 (0.56-1.39)	0.50 (0.17-1.46)	1.12 (1.02-1.23)[c]	0.88 (0.79-0.98)[c]
70-79	0.91 (0.80-1.02)	0.26 (0.12-0.60)[c]	1.00 (0.95-1.06)	0.84 (0.53-1.34)	1.38 (0.97-1.97)	1.16 (1.03-1.32)[c]	0.87 (0.76-0.99)[c]
≥80	1.01 (0.87-1.17)	0.62 (0.11-3.34)	0.93 (0.77-1.12)	0.87 (0.37-1.96)	1.13 (0.56-2.28)	1.49 (1.31-1.69)[c]	0.97 (0.83-1.15)
Sex							
Female	0.84 (0.73-0.96)[c]	0.22 (0.12-0.42)[c]	0.88 (0.84-0.92)[c]	0.59 (0.13-2.71)	0.87 (0.49-1.53)	1.13 (1.02-1.25)[c]	0.86 (0.76-0.98)[c]
Male	0.85 (0.76-0.95)[c]	0.32 (0.19-0.53)	0.93 (0.90-0.96)[c]	0.58 (0.25-1.34)	0.61 (0.29-1.25)	1.18 (1.10-1.28)[c]	0.89 (0.81-0.98)*[c]
Race							
Black	0.89 (0.81-0.98)[c]	N/A	N/A	N/A	N/A	N/A	0.89 (0.81-0.98)[c]
White	0.91 (0.80-1.03)	N/A	N/A	N/A	N/A	N/A	0.91 (0.80-1.03)

[a]RR: relative risk.

[b]N/A: not applicable; no patients reported in specific demographic subgroups for certain countries.

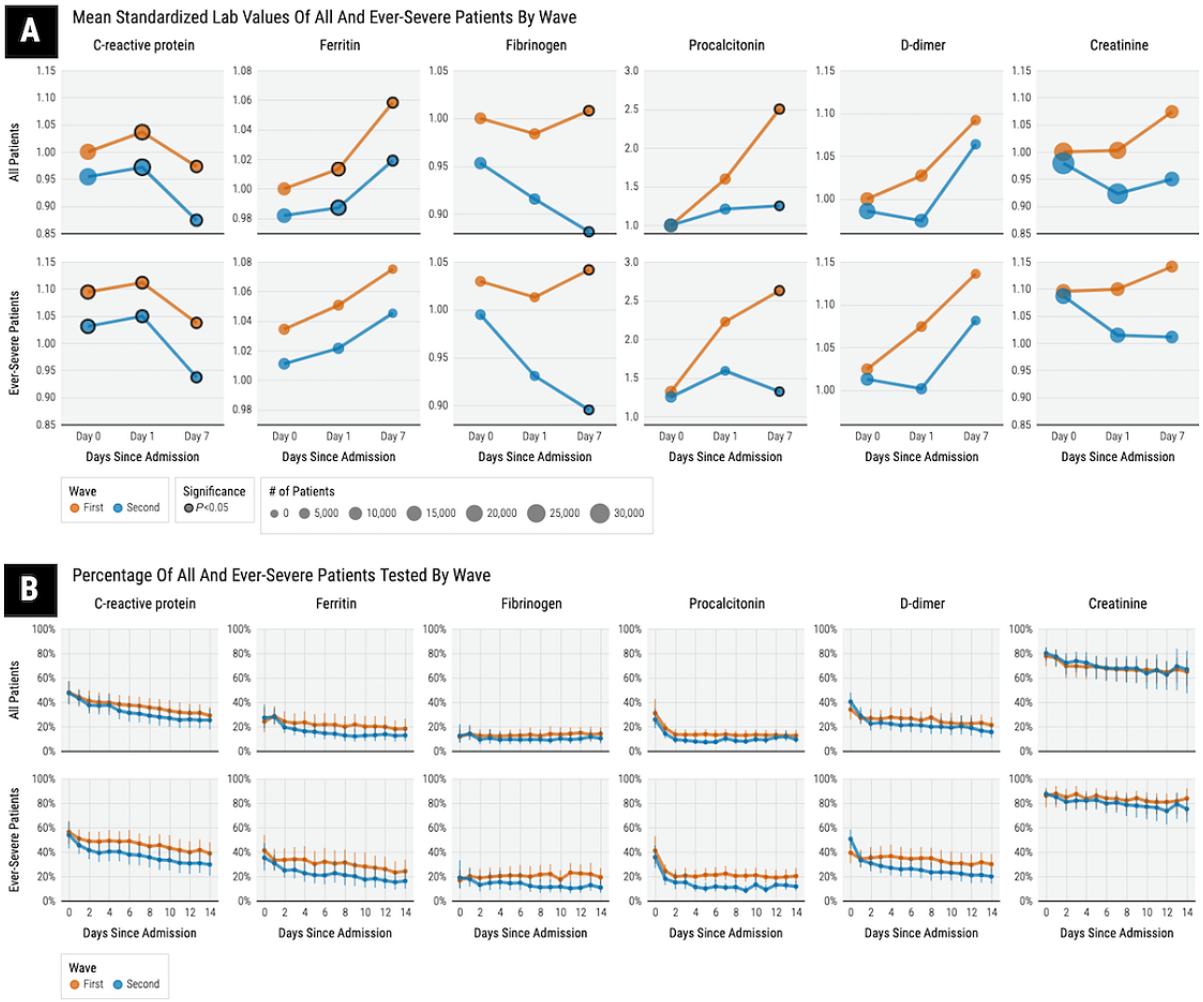
[c]Statistically significant.

Change in Mean Laboratory Values and Laboratory Testing Rates

We report standardized mean laboratory values in the first and the second waves at days 0, 1, and 7 since the index date of admission for CRP, ferritin, fibrinogen, procalcitonin, D-dimer, and creatinine in Figure 5. Among all patients, we observed significantly lower mean CRP values throughout the first week of hospitalization on days 0, 1, and 7 in the second wave than in the first wave. All other mean laboratory values on day 0 were not significantly different between the first and second waves. At day 7, we further observed that the mean values of ferritin, fibrinogen, and procalcitonin were all significantly lower in the second wave than in the first wave.

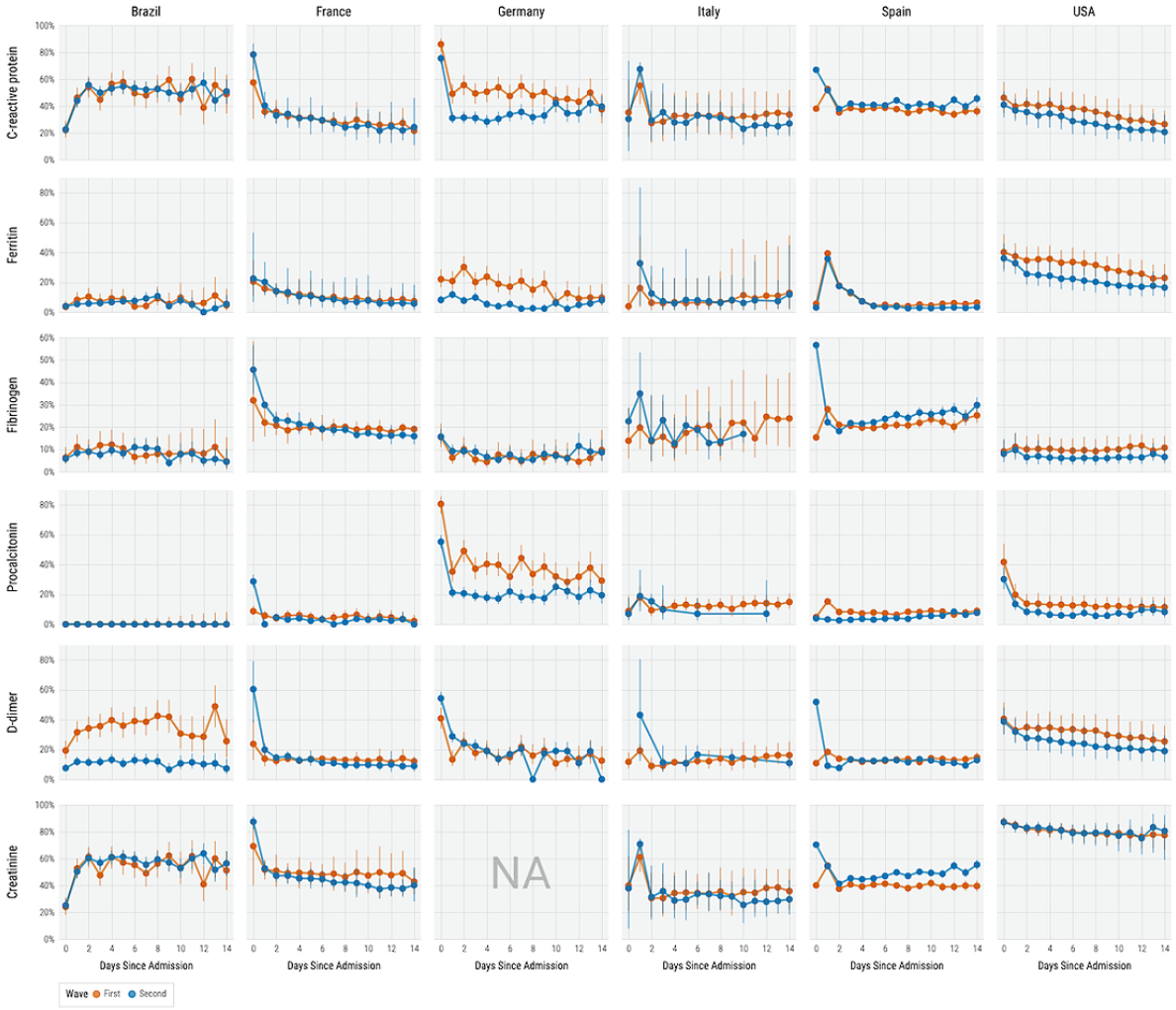
Among patients with severe disease, we observed similar results with regard to comparing mean laboratory values between the first and second waves. Mean CRP values during the first week of hospitalization on days 0, 1, and 7 were significantly lower in the second wave than in the first wave. Mean fibrinogen and procalcitonin values were significantly lower on day 7 in the second wave than in the first wave.

Figure 5. Standardized mean laboratory values (A) and corresponding laboratory testing rates (B) among all patients and those with severe disease in the first and second waves. Error bars indicate 95% CIs.



When comparing the overall laboratory testing rates during hospitalization, as reported in Figure 5, we observed that overall testing rates among all patients for procalcitonin were significantly lower across hospitalization days in the second wave than in the first wave. Overall procalcitonin testing rates among patients with severe disease were similarly significantly lower across hospitalization days in the second wave than in the first wave. There were no other significant changes in overall laboratory testing rates. We report laboratory testing rates within each country in Figure 6; creatinine laboratory data were not available from participating health care systems in Germany. Laboratory testing rates among countries varied significantly between the first and second waves. European countries exhibited the most changes in testing rates at admission (day 0). In the second wave, there was a significant increase in the D-dimer testing rate in France, Germany, and Spain, and in the CRP, creatinine, and fibrinogen testing rates in Spain. In Germany, there were significant decreases in the testing rates for CRP, ferritin, and procalcitonin at admission. In Brazil, there was a significant decrease in the testing rate for D-dimer at admission. In contrast, the United States did not have any significant changes in laboratory testing rates at admission. In all countries, except Brazil and Spain, laboratory testing rates in the second wave were generally lower during the second week of hospitalization. In Spain, second hospital week testing rates for CRP, fibrinogen, and creatinine were higher in the second wave than in the first wave.

Figure 6. Laboratory testing rates across hospitalization days in each country. Error bars indicate 95% CIs. Laboratory data for creatinine were unavailable for health care systems in Germany. NA: Not Available.



Discussion

Principal Findings

In this large EHR-based study, we employed a federated approach to rapidly aggregate and harmonize clinical data across 315 international hospitals from 6 countries that included 79,613 hospitalized patients with SARS-CoV-2 to offer insights on the evolving clinical trajectory of COVID-19 across the first and second waves. We found that patients hospitalized in the second wave were at significantly lower risk for severe COVID-19, corresponding to lower mean laboratory values for several inflammatory markers during the first week of hospitalization in the second wave than in the first wave.

In this study, we capitalized on the availability of real-world EHR data from participating international health care systems within the 4CE to capture pertinent clinical characteristics effectively and accurately. Despite the high heterogeneity in the health systems, we were able to rigorously perform quality checks across all health care centers using a multidisciplinary team approach that engages statisticians, informaticians, and clinicians. Additionally, the multinational nature of our data allowed us to identify country-level variations in temporal trends, as well as distinguish different clinical phenotypes and trajectories in the second wave compared with the first wave of the pandemic.

Notably, our federated approach demonstrated several advantages over methods where hospitals transfer patient-level data to a central repository [30]. By keeping data local within hospitals, we avoided privacy concerns and regulatory barriers that often delay multisite research studies. By comparing the results from different hospitals, rather than treating all the data as a single combined data set, we could identify outliers that suggested data quality problems. Finally, by having local data experts at each site running the database queries, we could leverage their help in addressing these data quality concerns [31,32]. Although more sophisticated analyses, such as machine learning models or robust multivariable models that can adjust for multiple sources of various types of bias, are harder to implement in a federated architecture than in a central repository, this study demonstrated that our federated approach

enabled us to obtain early clinical insights into the evolving pandemic and helped us gain confidence in the data.

Our data demonstrated a near uniform peak in country-specific hospitalizations in the first wave and a variable peak in country-specific hospitalizations in the second wave, reflecting country-specific patterns in the resurgence of COVID-19–related hospitalizations that were consistent with international tracking sites [1,9,10]. In all countries, except Brazil and the United States, the second wave peak was characterized by a lower intensity compared to the first wave peak, possibly reflecting the effects of successful COVID-19 mitigation measures implemented after the first wave [33,34].

Consistent with previous single-country studies, we observed that patients hospitalized in the second wave had an overall lower risk for severe COVID-19 than patients hospitalized in the first wave [6,12,35,36]. On further stratifying our analyses by country, we were able to observe that patients in Spain instead had a significantly higher risk for severe disease in the second wave, contrary to what we observed in France and the United States. Although the reasons for the increased risk of severe COVID-19 during Spain's second wave are unclear and likely complex, it is consistent with international tracking sites that indicated increased mortality rates and health care resource use in Spain's second wave, and it reflects the importance of being able to identify country-specific variations in our data [37]. We further note that the data from Spain originated from 1 hospital and were likely subject to some forms of bias. Even when stratifying by demographic subgroups, we observed similar patterns indicating reduced risk for severe COVID-19 in the second wave, particularly among patients aged 26 to 49 and among black patients. However, while the risk of severe disease for the entire population was lower in the second wave for patients aged 26 to 49, country-specific results demonstrated possible collider bias and small-sample bias in that the individual estimates for Brazil, France, Germany, and Spain were not statistically significant and were imprecise with wide CIs. This is likely due in large part to the smaller sample size of this age group in each country and reflects the need for multicenter studies to improve power. While one might expect to see a less severe disease later in the course of the pandemic due to changes in patient populations over time, improved clinical care, and greater utilization of health care resources compared to the beginning of the pandemic, it is unclear why we did not observe similar relationships for other patient groups. These possible discrepancies are likely due to a variety of different factors, as noted above, as well as sources of bias in our data. Further investigation into these country-specific demographic differences in the change in severe disease risk over the course of the pandemic is warranted and is ongoing in the 4CE.

Our observations comparing laboratory values between the first and second waves support our finding that patients hospitalized during the second wave compared to the first wave had a lower risk for severe disease. We found that mean laboratory values in the second wave exhibited considerable improvement toward typical physiological values compared to the first wave, especially those of inflammatory markers. In particular, the mean values of the positive acute phase reactant CRP were lower across the first week of hospitalization in the second wave than in the first wave, while the mean values of the positive acute phase reactants ferritin, fibrinogen, and procalcitonin were lower at day 7 [38]. This indicates that, on average, patients hospitalized during the second wave may have had less overall systemic inflammation at admission and had improved inflammatory states during the first week of hospitalization in comparison to patients admitted in the first wave [39-41]. Considering that there were no new major effective pharmacologic therapies for patients with COVID-19 introduced between the first and second waves, these general patterns may be reflective of a less vulnerable patient population in the second wave, as well as improved general clinical management strategies of COVID-19 in the second wave [42-50]. Ongoing 4CE analyses are further investigating these findings.

We further observed variations in laboratory testing rates among countries between the first and second waves. These changes in laboratory testing rates at admission may be reflective of greater understanding of COVID-19 pathophysiology and clinical trajectories leading to changes in clinical protocols. For example, there was a significant increase in the testing rate of D-dimer at admission in France, Germany, and Spain in the second wave. This particular change in clinical practice may have been driven in part by the growing literature supporting the association of high D-dimer values with worse outcomes in COVID-19 and the possibility of using D-dimer to clinically classify and evaluate the prognosis of COVID-19 patients [51-55]. Further, although there were no significant changes in laboratory testing at admission in the United States, we observed that testing rates across hospitalization days were generally higher than in other European countries regardless of wave. Future investigations are warranted to infer why we observed these patterns.

Study Limitations

We acknowledge several limitations for this EHR-based observational cohort study. This study was limited to patients who were admitted to a hospital, either because they experienced more severe illness or because they had other possibly biasing conditions; as with many EHR-based studies, we were unable to ascertain the precise reasons for admission. Similar to other EHR-based studies, we were not able to validate if patients were hospitalized due to COVID-19 or happened to have a positive test when admitted for an unrelated medical condition. Thus, we could not completely mitigate selection bias or misclassification bias in our cohort identification. Due to the limited scope of the extracted aggregate data, we could not effectively control for patient-level potentially confounding variables such as comorbidities, medication use (both prior to and during hospitalization), and other societal and environmental factors, all of which can induce many types of biases [56]. Data pertaining to certain countries, most notably Brazil and Germany, may have been subject to small sample bias. Furthermore, mean laboratory values at later days of hospitalization were subject to censoring (transfer, discharge, and death) and thus dropout bias, so we could not effectively compare mean laboratory values within a single wave at different timepoints. However, we believe that facilitating comparisons at identical timepoints between different waves is not subject to as much dropout bias. In an effort to provide

information regarding the nature of censoring that existed in the data, we report in Multimedia Appendix 3 the proportion of patients who were alive and remained in the hospital across hospitalization days for each country by wave. Further, considering the aforementioned limitations, we took special caution to make conclusions that were mostly descriptive in nature. In the future, we hope to disaggregate EHR data to the patient level in order to adjust for many of these biases, if possible, under institutional review board approval.

Conclusions

For assessing the evolving epidemiology, pathophysiology, and health care dynamics of the COVID-19 pandemic, we leveraged EHR data in a large international cohort of hospitalized patients with SARS-CoV-2 to rapidly characterize the clinical course of patients admitted to hospital during the first 2 major waves of the pandemic. We were able to characterize changes in hospitalization rates, demographic characteristics, severity risk, and mean laboratory values using data from 79,613 patients across 315 health care systems in 6 countries. Our study's federated approach demonstrates the feasibility and power of leveraging real-world EHR data from multiple countries to support our understanding of evolving pandemics such as COVID-19.

Acknowledgments

GMW is supported by National Institutes of Health (NIH)/National Center for Advancing Translational Sciences (NCATS) UL1TR002541 and UL1TR000005, NIH/National Library of Medicine (NLM) R01LM013345, and NIH/National Human Genome Research Institute (NHGRI) 3U01HG008685-05S2. NG is supported by NIH/NLM T15LM007092. DAH is supported by NCATS UL1TR002240. RK is supported by Clinical and Translational Science Awards (CTSA) Award UL1TR001998. KK is supported by NIH/NCATS UL1TR001450. DLM is supported by NIH/NCATS UL1TR001878. SNM is supported by NIH/NCATS 5UL1TR001857-05 and NIH/NHGRI 5R01HG009174-04. GSO is supported by NIH P30ES017885 and U24CA210967. LPP is supported by CTSA Award UL1TR002366. JS is supported by NIH/NCATS UL1TR001881. SV is supported by NIH/NLM R01LM012095 and NIH/NCATS UL1TR001857. ZX is supported by National Institute of Neurological Disorders and Stroke (NINDS) R01NS098023. AMS is supported by NIH/National Heart, Lung, and Blood Institute (NHLBI) K23HL148394, L40HL148910, and NIH/NCATS UL1TR001420.

Authors' Contributions

All authors made contributions to conception and design; acquisition, analysis, and interpretation of data; drafting the article or revising it critically for important intellectual content; and final approval of the version to be published. The Consortium for Clinical Characterization of COVID-19 by EHR (4CE) consists of: James R Aaron, Giuseppe Agapito, Adem Albayrak, Mario Alessiani, Anna Alloni, Danilo F Amendola, Li LLJ Anthony, Bruce J Aronow, Fatima Ashraf, Andrew Atz, Paul Avillach, Paula S Azevedo, James Balshi, Brett K Beaulieu-Jones, Douglas S Bell, Antonio Bellasi, Riccardo Bellazzi, Vincent Benoit, Michele Beraghi, José Luis Bernal Sobrino, Mélodie Bernaux, Romain Bey, Alvar Blanco Martínez, Martin Boeker, Clara-Lea Bonzel, John Booth, Silvano Bosari, Florence T Bourgeois, Robert L Bradford, Gabriel A Brat, Stéphane Bréant, Carlos Tadeu Breda Neto, Nicholas W Brown, William A Bryant, Mauro Bucalo, Anita Burgun, Tianxi Cai, Mario Cannataro, Aldo Carmona, Charlotte Caucheteux, Julien Champ, Jin Chen, Krista Chen, Luca Chiovato, Lorenzo Chiudinelli, Kelly Cho, James J Cimino, Tiago K Colicchio, Sylvie Cormont, Sébastien Cossin, Jean B Craig, Juan Luis Cruz Bermúdez, Jaime Cruz Rojo, Arianna Dagliati, Mohamad Daniar, Christel Daniel, Priyam Das, Anahita Davoudi, Batsal Devkota, Julien Dubiel, Loic Esteve, Hossein Estiri, Shirley Fan, Robert W Follett, Thomas Ganslandt, Noelia García Barrio, Lana X Garmire, Nils Gehlenborg, Alon Geva, Tobias Gradinger, Alexandre Gramfort, Romain Griffier, Nicolas Griffon, Olivier Grisel, Alba Gutiérrez-Sacristán, David A Hanauer, Christian Haverkamp, Bing He, Darren W Henderson, Martin Hilka, Yuk-Lam Ho, John H Holmes, Chuan Hong, Petar Horki, Kenneth M Huling, Meghan R Hutch, Richard W Issitt, Anne Sophie Jannot, Vianney Jouhet, Mark S Keller, Chris J Kennedy, Katie Kirchoff, Jeffrey G Klann, Isaac S Kohane, Ian D Krantz, Detlef Kraska, Ashok K Krishnamurthy, Sehi L'Yi, Trang T Le, Judith Leblanc, Andressa RR Leite, Guillaume Lemaitre, Leslie Lenert, Damien Leprovost, Molei Liu, Ne Hooi Will Loh, Qi Long, Sara Lozano-Zahonero, Yuan Luo, Kristine E Lynch, Sadiqa Mahmood, Sarah Maidlow, Adeline Makoudjou, Alberto Malovini, Kenneth D Mandl, Chengsheng Mao, Anupama Maram, Patricia Martel, Marcelo R Martins, Aaron J Masino, Maria Mazzitelli, Arthur Mensch, Marianna Milano, Marcos F Minicucci, Bertrand Moal, Jason H Moore, Cinta Moraleda, Jeffrey S Morris, Michele Morris, Karyn L Moshal, Sajad Mousavi, Danielle L Mowery, Douglas A Murad, Shawn N Murphy, Thomas P Naughton, Antoine Neuraz, Kee Yuan Ngiam, Wanjiku FM Njoroge, James B Norman, Jihad Obeid, Marina P Okoshi, Karen L Olson, Gilbert S Omenn, Nina Orlova, Brian D Ostasiewski, Nathan P Palmer, Nicolas Paris, Lav P Patel, Miguel Pedrera Jimenez, Emily R Pfaff, Danielle Pillion, Hans U Prokosch, Robson A Prudente, Víctor Quirós González, Rachel B Ramoni, Maryna Raskin, Siegbert Rieg, Gustavo Roig Domínguez, Pablo Rojo, Carlos Sáez, Elisa Salamanca, Malarkodi J Samayamuthu, L Nelson Sanchez-Pinto, Arnaud Sandrin, Nandhini Santhanam, Janaina CC Santos, Javier W Sanz, Maria Savino, Emily R Schriver, Petra Schubert, Juergen Schuettler, Luigia Scudeller, Neil J Sebire, Pablo Serrano Balazote, Patricia Serre, Arnaud Serret-Larmande, Zahra Shakeri, Domenick Silvio, Piotr Sliz, Jiyeon Son, Charles Sonday, Andrew M South, Anastasia Spiridou, Amelia LM Tan, Bryce WQ Tan, Byorn WL Tan, Suzana E Tanni, Deanne M Taylor, Ana I Terriza Torres, Valentina Tibollo, Patric Tippmann, Carlo Torti, Enrico M Trecarichi, Yi-Ju Tseng, Andrew K Vallejos, Gael Varoquaux, Margaret E Vella,

Guillaume Verdy, Jill-Jênn Vie, Shyam Visweswaran, Michele Vitacca, Kavishwar B Wagholikar, Lemuel R Waitman, Xuan Wang, Demian Wassermann, Griffin M Weber, Zongqi Xia, Xin Xiong, Nadir Yehya, William Yuan, Alberto Zambelli, Harrison G Zhang, Daniel Zoeller, and Chiara Zucco.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Descriptions of CSV files generated at participating health care systems.
[DOCX File , 13 KB - jmir_v23i10e31400_app1.docx]

Multimedia Appendix 2
Prevalence of demographic subgroups in the first and second waves for the entire cohort and by country, and estimated absolute risk for severe COVID-19 by demographic subgroups inclusive of 5 Veterans Affairs health care systems. Error bars indicate 95% CIs. *Significant difference by nonoverlapping 95% CIs.
[PNG File , 297 KB - jmir_v23i10e31400_app2.png]

Multimedia Appendix 3
Proportion of patients who were alive and still in the hospital across hospitalization days.
[PNG File , 199 KB - jmir_v23i10e31400_app3.png]

References

1. COVID-19 Map - Johns Hopkins Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/map.html [accessed 2021-05-03]
2. Devi S. COVID-19 resurgence in Iran. The Lancet 2020 Jun 20;395(10241):1896 [FREE Full text] [doi: 10.1016/S0140-6736(20)31407-0] [Medline: 32563363]
3. Looi M. Covid-19: Is a second wave hitting Europe? BMJ 2020 Oct 28;371:m4113. [doi: 10.1136/bmj.m4113] [Medline: 33115704]
4. Shimizu K, Wharton G, Sakamoto H, Mossialos E. Resurgence of covid-19 in Japan. BMJ 2020 Aug 18;370:m3221. [doi: 10.1136/bmj.m3221] [Medline: 32816821]
5. Sabino EC, Buss LF, Carvalho MPS, Prete CA, Crispim MAE, Fraiji NA, et al. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. The Lancet 2021 Feb 06;397(10273):452-455 [FREE Full text] [doi: 10.1016/S0140-6736(21)00183-5] [Medline: 33515491]
6. Saito S, Asai Y, Matsunaga N, Hayakawa K, Terada M, Ohtsu H, et al. First and second COVID-19 waves in Japan: A comparison of disease severity and characteristics. J Infect 2021 Apr;82(4):84-123 [FREE Full text] [doi: 10.1016/j.jinf.2020.10.033] [Medline: 33152376]
7. Win A. Rapid rise of COVID-19 second wave in Myanmar and implications for the Western Pacific region. QJM 2020 Dec 01;113(12):856-857 [FREE Full text] [doi: 10.1093/qjmed/hcaa290] [Medline: 33095232]
8. Vahidy FS, Drews AL, Masud FN, Schwartz RL, Askary BB, Boom ML, et al. Characteristics and Outcomes of COVID-19 Patients During Initial Peak and Resurgence in the Houston Metropolitan Area. JAMA 2020 Sep 08;324(10):998-1000 [FREE Full text] [doi: 10.1001/jama.2020.15301] [Medline: 32789492]
9. Cumulative deaths. IHME COVID-19 Projections. URL: https://covid19.healthdata.org/united-states-of-america?view=cumulative-deaths&tab=trend [accessed 2021-06-17]
10. Coronavirus (COVID-19) Deaths. Our World in Data. URL: https://ourworldindata.org/covid-deaths [accessed 2021-06-17]
11. Mollinedo-Gajate I, Villar-Álvarez F, Zambrano-Chacón M, Núñez-García L, de la Dueña-Muñoz L, López-Chang C, et al. First and Second Waves of Coronavirus Disease 2019 in Madrid, Spain: Clinical Characteristics and Hematological Risk Factors Associated With Critical/Fatal Illness. Crit Care Explor 2021 Feb;3(2):e0346 [FREE Full text] [doi: 10.1097/CCE.0000000000000346] [Medline: 33634266]
12. Contou D, Fraissé M, Pajot O, Tirolien J, Mentec H, Plantefève G. Comparison between first and second wave among critically ill COVID-19 patients admitted to a French ICU: no prognostic improvement during the second wave? Crit Care 2021 Jan 04;25(1):3 [FREE Full text] [doi: 10.1186/s13054-020-03449-6] [Medline: 33397421]
13. Long SW, Olsen RJ, Christensen PA, Bernard DW, Davis JJ, Shukla M, et al. Molecular Architecture of Early Dissemination and Massive Second Wave of the SARS-CoV-2 Virus in a Major Metropolitan Area. mBio 2020 Oct 30;11(6):e02707-e02720 [FREE Full text] [doi: 10.1128/mBio.02707-20] [Medline: 33127862]
14. Brat GA, Weber GM, Gehlenborg N, Avillach P, Palmer NP, Chiovato L, et al. International electronic health record-derived COVID-19 clinical course profiles: the 4CE consortium. NPJ Digit Med 2020;3:109 [FREE Full text] [doi: 10.1038/s41746-020-00308-0] [Medline: 32864472]

15. Weber GM, Hong C, Palmer NP, Avillach P, Murphy SN, Gutiérrez-Sacristán A, Consortium for Clinical Characterization of COVID-19 by EHR (4CE), et al. International Comparisons of Harmonized Laboratory Value Trajectories to Predict Severe COVID-19: Leveraging the 4CE Collaborative Across 342 Hospitals and 6 Countries: A Retrospective Cohort Study. medRxiv 2021 Feb 05 [FREE Full text] [doi: 10.1101/2020.12.16.20247684] [Medline: 33564777]
16. Jones AL, Pettey WBP, Carter ME, Brignone E, Redd A, Suo Y, et al. Regional Variations in Documentation of Sexual Trauma Concepts in Electronic Medical Records in the United States Veterans Health Administration. AMIA Annu Symp Proc 2019;2019:514-522 [FREE Full text] [Medline: 32308845]
17. Le TT, Gutiérrez-Sacristán A, Son J, Hong C, South AM, Beaulieu-Jones BK, et al. Multinational Prevalence of Neurological Phenotypes in Patients Hospitalized with COVID-19. medRxiv 2021 Jan 29 [FREE Full text] [doi: 10.1101/2021.01.27.21249817] [Medline: 33655281]
18. Klann JG, Estiri H, Weber GM, Moal B, Avillach P, Hong C, Consortium for Clinical Characterization of COVID-19 by EHR (4CE) (CONSORTIA AUTHOR), et al. Validation of an internationally derived patient severity phenotype to support COVID-19 analytics from electronic health record data. J Am Med Inform Assoc 2021 Jul 14;28(7):1411-1420 [FREE Full text] [doi: 10.1093/jamia/ocab018] [Medline: 33566082]
19. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986 Sep;7(3):177-188. [doi: 10.1016/0197-2456(86)90046-2] [Medline: 3802833]
20. National Center for Veterans Analysis and Statistics. United States Department of Veterans Affairs. URL: https://www.va.gov/vetdata/veteran_population.asp [accessed 2021-06-07]
21. Huang I, Pranata R, Lim MA, Oehadian A, Alisjahbana B. C-reactive protein, procalcitonin, D-dimer, and ferritin in severe coronavirus disease-2019: a meta-analysis. Ther Adv Respir Dis 2020;14:1753466620937175 [FREE Full text] [doi: 10.1177/1753466620937175] [Medline: 32615866]
22. Carubbi F, Salvati L, Alunno A, Maggi F, Borghi E, Mariani R, et al. Ferritin is associated with the severity of lung involvement but not with worse prognosis in patients with COVID-19: data from two Italian COVID-19 units. Sci Rep 2021 Mar 01;11(1):4863 [FREE Full text] [doi: 10.1038/s41598-021-83831-8] [Medline: 33649408]
23. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
24. Smilowitz NR, Kunichoff D, Garshick M, Shah B, Pillinger M, Hochman JS, et al. C-reactive protein and clinical outcomes in patients with COVID-19. Eur Heart J 2021 Jun 14;42(23):2270-2279 [FREE Full text] [doi: 10.1093/eurheartj/ehaa1103] [Medline: 33448289]
25. Shi S, Nie B, Chen X, Cai Q, Lin C, Zhao G, et al. Clinical and laboratory characteristics of severe and non-severe patients with COVID-19: A retrospective cohort study in China. J Clin Lab Anal 2021 Jan;35(1):e23692 [FREE Full text] [doi: 10.1002/jcla.23692] [Medline: 33389777]
26. Bi X, Su Z, Yan H, Du J, Wang J, Chen L, et al. Prediction of severe illness due to COVID-19 based on an analysis of initial Fibrinogen to Albumin Ratio and Platelet count. Platelets 2020 Jul 03;31(5):674-679 [FREE Full text] [doi: 10.1080/09537104.2020.1760230] [Medline: 32367765]
27. Statsenko Y, Al Zahmi F, Habuza T, Gorkom KN, Zaki N. Prediction of COVID-19 severity using laboratory findings on admission: informative values, thresholds, ML model performance. BMJ Open 2021 Feb 26;11(2):e044500 [FREE Full text] [doi: 10.1136/bmjopen-2020-044500] [Medline: 33637550]
28. Wang F, Hou H, Wang T, Luo Y, Tang G, Wu S, et al. Establishing a model for predicting the outcome of COVID-19 based on combination of laboratory tests. Travel Med Infect Dis 2020;36:101782 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101782] [Medline: 32526372]
29. Consortium for Clinical Characterization of COVID-19 by EHR. URL: https://covidclinical.net/ [accessed 2021-09-23]
30. Haendel MA, Chute CG, Bennett TD, Eichmann DA, Guinney J, Kibbe WA, N3C Consortium. The National COVID Cohort Collaborative (N3C): Rationale, design, infrastructure, and deployment. J Am Med Inform Assoc 2021 Mar 01;28(3):427-443 [FREE Full text] [doi: 10.1093/jamia/ocaa196] [Medline: 32805036]
31. Wallace S, Myles PS. Solving the challenges of large multicenter trials in anesthesia. HSR Proc Intensive Care Cardiovasc Anesth 2009;1(3):46-53 [FREE Full text] [Medline: 23439979]
32. de Jonge JC, Reinink H, Colam B, Alpers I, Ciccone A, Csiba L, et al. Regulatory delays in a multinational clinical stroke trial. Eur Stroke J 2021 Jun 30;6(2):120-127 [FREE Full text] [doi: 10.1177/23969873211004845] [Medline: 34414286]
33. Haug N, Geyrhofer L, Londei A, Dervic E, Desvars-Larrive A, Loreto V, et al. Ranking the effectiveness of worldwide COVID-19 government interventions. Nat Hum Behav 2020 Dec;4(12):1303-1312. [doi: 10.1038/s41562-020-01009-0] [Medline: 33199859]
34. Tam K, Walker N, Moreno J. Effect of mitigation measures on the spreading of COVID-19 in hard-hit states in the U.S. PLoS One 2020;15(11):e0240877 [FREE Full text] [doi: 10.1371/journal.pone.0240877] [Medline: 33141823]
35. Docherty AB, Mulholland RH, Lone NI, Cheyne CP, De Angelis D, Diaz-Ordaz K, et al. Changes in in-hospital mortality in the first wave of COVID-19: a multicentre prospective observational cohort study using the WHO Clinical Characterisation Protocol UK. The Lancet Respiratory Medicine 2021 Jul 14;9(7):773-785 [FREE Full text] [doi: 10.1016/S2213-2600(21)00175-2] [Medline: 34000238]

36. Karagiannidis C, Windisch W, McAuley DF, Welte T, Busse R. Major differences in ICU admissions during the first and second COVID-19 wave in Germany. The Lancet Respiratory Medicine 2021 May;9(5):e47-e48 [FREE Full text] [doi: 10.1016/S2213-2600(21)00101-6] [Medline: 33684356]
37. Daily infections and testing. IHME COVID-19 Projections. URL: https://covid19.healthdata.org/spain?view=infections-testing&tab=trend&test=infections [accessed 2021-06-17]
38. Gulhar R, Ashraf M, Jialal I. Physiology, Acute Phase Reactants. In: StatPearls [Internet]. Treasure Island, FL: StatPearls Publishing; 2021.
39. Nehring SM, Goyal A, Bansal P, Patel BC. C Reactive Protein. In: StatPearls [Internet]. Treasure Island, FL: StatPearls Publishing; 2021.
40. Knovich MA, Storey JA, Coffman LG, Torti SV, Torti FM. Ferritin for the clinician. Blood Rev 2009 May;23(3):95-104 [FREE Full text] [doi: 10.1016/j.blre.2008.08.001] [Medline: 18835072]
41. Kell DB, Pretorius E. Serum ferritin is an important inflammatory disease marker, as it is mainly a leakage product from damaged cells. Metallomics 2014 Apr;6(4):748-773. [doi: 10.1039/c3mt00347g] [Medline: 24549403]
42. Sanders JM, Monogue ML, Jodlowski TZ, Cutrell JB. Pharmacologic Treatments for Coronavirus Disease 2019 (COVID-19): A Review. JAMA 2020 May 12;323(18):1824-1836. [doi: 10.1001/jama.2020.6019] [Medline: 32282022]
43. Sheahan TP, Sims AC, Leist SR, Schäfer A, Won J, Brown AJ, et al. Comparative therapeutic efficacy of remdesivir and combination lopinavir, ritonavir, and interferon beta against MERS-CoV. Nat Commun 2020 Jan 10;11(1):222 [FREE Full text] [doi: 10.1038/s41467-019-13940-6] [Medline: 31924756]
44. Gautret P, Lagier J, Parola P, Hoang VT, Meddeb L, Mailhe M, et al. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. International Journal of Antimicrobial Agents 2020 Jul;56(1):105949 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.105949] [Medline: 32205204]
45. Spinner CD, Gottlieb RL, Criner GJ, Arribas López JR, Cattelan AM, Soriano Viladomiu A, GS-US-540-5774 Investigators. Effect of Remdesivir vs Standard Care on Clinical Status at 11 Days in Patients With Moderate COVID-19: A Randomized Clinical Trial. JAMA 2020 Sep 15;324(11):1048-1057 [FREE Full text] [doi: 10.1001/jama.2020.16349] [Medline: 32821939]
46. Wang Y, Zhang D, Du G, Du R, Zhao J, Jin Y, et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. The Lancet 2020 May 16;395(10236):1569-1578 [FREE Full text] [doi: 10.1016/S0140-6736(20)31022-9] [Medline: 32423584]
47. Boulware DR, Pullen MF, Bangdiwala AS, Pastick KA, Lofgren SM, Okafor EC, et al. A Randomized Trial of Hydroxychloroquine as Postexposure Prophylaxis for Covid-19. N Engl J Med 2020 Aug 06;383(6):517-525 [FREE Full text] [doi: 10.1056/NEJMoa2016638] [Medline: 32492293]
48. Mitjà O, Corbacho-Monné M, Ubals M, Alemany A, Suñer C, Tebé C, BCN-PEP-CoV2 Research Group. A Cluster-Randomized Trial of Hydroxychloroquine for Prevention of Covid-19. N Engl J Med 2021 Feb 04;384(5):417-427 [FREE Full text] [doi: 10.1056/NEJMoa2021801] [Medline: 33289973]
49. Self WH, Semler MW, Leither LM, Casey JD, Angus DC, Brower RG, National Heart, Lung,Blood Institute PETAL Clinical Trials Network, et al. Effect of Hydroxychloroquine on Clinical Status at 14 Days in Hospitalized Patients With COVID-19: A Randomized Clinical Trial. JAMA 2020 Dec 01;324(21):2165-2176 [FREE Full text] [doi: 10.1001/jama.2020.22240] [Medline: 33165621]
50. WHO Solidarity Trial Consortium, Pan H, Peto R, Henao-Restrepo A, Preziosi M, Sathiyamoorthy V, et al. Repurposed Antiviral Drugs for Covid-19 - Interim WHO Solidarity Trial Results. N Engl J Med 2021 Feb 11;384(6):497-511 [FREE Full text] [doi: 10.1056/NEJMoa2023184] [Medline: 33264556]
51. Yao Y, Cao J, Wang Q, Shi Q, Liu K, Luo Z, et al. D-dimer as a biomarker for disease severity and mortality in COVID-19 patients: a case control study. J Intensive Care 2020;8:49 [FREE Full text] [doi: 10.1186/s40560-020-00466-z] [Medline: 32665858]
52. Düz ME, Balcı A, Menekşe E. D-dimer levels and COVID-19 severity: Systematic Review and Meta-Analysis. Tuberk Toraks 2020 Dec;68(4):353-360 [FREE Full text] [doi: 10.5578/tt.70351] [Medline: 33448732]
53. Yu H, Qin C, Chen M, Wang W, Tian D. D-dimer level is associated with the severity of COVID-19. Thromb Res 2020 Nov;195:219-225 [FREE Full text] [doi: 10.1016/j.thromres.2020.07.047] [Medline: 32777639]
54. Berger JS, Kunichoff D, Adhikari S, Ahuja T, Amoroso N, Aphinyanaphongs Y, et al. Prevalence and Outcomes of D-Dimer Elevation in Hospitalized Patients With COVID-19. Arterioscler Thromb Vasc Biol 2020 Oct;40(10):2539-2547 [FREE Full text] [doi: 10.1161/ATVBAHA.120.314872] [Medline: 32840379]
55. He X, Yao F, Chen J, Wang Y, Fang X, Lin X, et al. The poor prognosis and influencing factors of high D-dimer levels for COVID-19 patients. Sci Rep 2021 Jan 19;11(1):1830 [FREE Full text] [doi: 10.1038/s41598-021-81300-w] [Medline: 33469072]
56. Cohen JB, D'Agostino McGowan L, Jensen ET, Rigdon J, South AM. Evaluating sources of bias in observational studies of angiotensin-converting enzyme inhibitor/angiotensin II receptor blocker use during COVID-19: beyond confounding. J Hypertens 2021 Apr 01;39(4):795-805 [FREE Full text] [doi: 10.1097/HJH.0000000000002706] [Medline: 33186321]

Abbreviations

4CE: Consortium for Clinical Characterization of COVID-19 by EHR
CRP: C-reactive protein
EHR: electronic health record
ICD: International Classification of Diseases
RR: relative risk
VA: Veterans Affairs

Edited by C Basch; submitted 23.06.21; peer-reviewed by J Lee, M Dehghani, KM Kuo; comments to author 28.07.21; revised version received 02.09.21; accepted 02.09.21; published 11.10.21.

Please cite as:
Weber GM, Zhang HG, L'Yi S, Bonzel CL, Hong C, Avillach P, Gutiérrez-Sacristán A, Palmer NP, Tan ALM, Wang X, Yuan W, Gehlenborg N, Alloni A, Amendola DF, Bellasi A, Bellazzi R, Beraghi M, Bucalo M, Chiovato L, Cho K, Dagliati A, Estiri H, Follett RW, García Barrio N, Hanauer DA, Henderson DW, Ho YL, Holmes JH, Hutch MR, Kavuluru R, Kirchoff K, Klann JG, Krishnamurthy AK, Le TT, Liu M, Loh NHW, Lozano-Zahonero S, Luo Y, Maidlow S, Makoudjou A, Malovini A, Martins MR, Moal B, Morris M, Mowery DL, Murphy SN, Neuraz A, Ngiam KY, Okoshi MP, Omenn GS, Patel LP, Pedrera Jiménez M, Prudente RA, Samayamuthu MJ, Sanz Vidorreta FJ, Schriver ER, Schubert P, Serrano Balazote P, Tan BWL, Tanni SE, Tibollo V, Visweswaran S, Wagholikar KB, Xia Z, Zöller D, The Consortium For Clinical Characterization Of COVID-19 By EHR (4CE), Kohane IS, Cai T, South AM, Brat GA
International Changes in COVID-19 Clinical Trajectories Across 315 Hospitals and 6 Countries: Retrospective Cohort Study
J Med Internet Res 2021;23(10):e31400
URL: https://www.jmir.org/2021/10/e31400
doi:10.2196/31400
PMID:34533459

Original Paper

Intervening on Trust in Science to Reduce Belief in COVID-19 Misinformation and Increase COVID-19 Preventive Behavioral Intentions: Randomized Controlled Trial

Jon Agley[1], MPH, PhD; Yunyu Xiao[2], MPhil, PhD; Esi E Thompson[3], PhD; Xiwei Chen[4], MS; Lilian Golzarri-Arroyo[4], MS

[1]Prevention Insights, Department of Applied Health Science, School of Public Health Bloomington, Indiana University Bloomington, Bloomington, IN, United States

[2]Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, United States

[3]Indiana University Media School, Indiana University Bloomington, Bloomington, IN, United States

[4]Biostatistics Consulting Center, School of Public Health Bloomington, Indiana University Bloomington, Bloomington, IN, United States

Corresponding Author:
Jon Agley, MPH, PhD
Prevention Insights, Department of Applied Health Science
School of Public Health Bloomington
Indiana University Bloomington
809 E. 9th St.
Bloomington, IN, 47404
United States
Phone: 1 812 855 3123
Email: jagley@indiana.edu

Abstract

Background: Trust in science meaningfully contributes to our understanding of people's belief in misinformation and their intentions to take actions to prevent COVID-19. However, no experimental research has sought to intervene on this variable to develop a scalable response to the COVID-19 infodemic.

Objective: Our study examined whether brief exposure to an infographic about the scientific process might increase trust in science and thereby affect belief in misinformation and intention to take preventive actions for COVID-19.

Methods: This two-arm, parallel-group, randomized controlled trial aimed to recruit a US representative sample of 1000 adults by age, race/ethnicity, and gender using the Prolific platform. Participants were randomly assigned to view either an intervention infographic about the scientific process or a control infographic. The intervention infographic was designed through a separate pilot study. Primary outcomes were trust in science, COVID-19 narrative belief profile, and COVID-19 preventive behavioral intentions. We also collected 12 covariates and incorporated them into all analyses. All outcomes were collected using web-based assessment.

Results: From January 22, 2021 to January 24, 2021, 1017 participants completed the study. The intervention slightly improved trust in science (difference-in-difference 0.03, SE 0.01, t_{1000}=2.16, *P*=.031). No direct intervention effect was observed on belief profile membership, but there was some evidence of an indirect intervention effect mediated by trust in science (adjusted odds ratio 1.06, SE 0.03, 95% CI 1.00-1.12, *z*=2.01, *P*=.045) on membership in the "scientific" profile compared with the others. No direct nor indirect effects on preventive behaviors were observed.

Conclusions: Briefly viewing an infographic about science appeared to cause a small aggregate increase in trust in science, which may have, in turn, reduced the believability of COVID-19 misinformation. The effect sizes were small but commensurate with our 60-second, highly scalable intervention approach. Researchers should study the potential for truthful messaging about how science works to serve as misinformation inoculation and test how best to do so.

Trial Registration: NCT04557241; https://clinicaltrials.gov/ct2/show/NCT04557241

International Registered Report Identifier (IRRID): RR2-10.2196/24383

(J Med Internet Res 2021;23(10):e32425) doi:10.2196/32425

KEYWORDS

infodemic; misinformation; trust in science; COVID-19; RCT; randomized controlled trial

Introduction

Background

The COVID-19 pandemic has been accompanied by a substantive, pervasive outpouring of misinformation about the disease [1] that can be described as an infodemic [2]. Concerns about this infodemic were raised by members of the science community almost immediately, and steps were taken to develop a research agenda [3], as misinformation about COVID-19 has taken many forms and been amplified across numerous types of media [4-6]. Anecdotal stories about behaviors and consequences associated with COVID-19 misinformation can be readily identified (as we have done in small measure in our prior work [7,8]), and some scholars have documented negative outcomes of COVID-19 misinformation [9,10]. Though Greene and Murphy [11] recently found "surprisingly little" experimental research examining the effects of misinformation on behavior, their study found that even brief, single exposures to COVID-19 misinformation may nudge (have a small effect on) some behavioral intentions. Loomba et al [12] similarly found evidence for decreases in COVID-19 vaccination intention due to exposure to misinformation.

Unfortunately, far from tapering more than a year into the pandemic, the volume of misinformation has remained high; representatives of multiple organizations, including the World Health Organization and US Food and Drug Administration, recently warned that misinformation poses a global concern and may drive pandemic-related harms [13]. It is clearly incumbent on researchers to develop a thorough understanding of COVID-19 misinformation and to establish evidence-based mitigation tools.

Belief in COVID-19 Misinformation Clusters Is Associated With Trust in Science

In May 2020, in response to growing concern about the COVID-19 infodemic, our team conducted one of the first studies of COVID-19 misinformation believability and the factors associated therewith [7]. We examined 5 brief narrative statements ranging from clearly false (eg, 5G transmission of COVID-19) to likely misinformed or improbable but not impossible (eg, purposeful laboratory development as a weapon) to a statement reflecting the scientific consensus at the time (eg, zoonotic origin). Using latent profile analysis, we identified 4 belief profiles into which it was possible to classify participants. Members of the largest profile (70.15% of the sample) reported high believability for a statement about the zoonotic origin and much lower believability for the misinformed statements. Members of the other, smaller profiles did not disbelieve the zoonotic statement, but tended to report higher believability for misinformation. In other words, findings suggested the existence of a large "scientific" or science-consistent group and multiple smaller groups that found misinformation believable to various degrees.

Then, we found that—controlling for race/ethnicity, gender, age, and education level—trust in science and scientists, a scale variable computed from 21 Likert-type questions of the Trust in Science and Scientists Inventory [14], was strongly associated with belief profile membership, with greater trust being associated with considerably higher odds of belonging to the "scientific profile." The magnitude of adjusted odds for trust substantially exceeded that of other variables hypothesized to be associated with profile membership (political orientation and religious commitment) that were simultaneously analyzed [7].

Based on our findings and research described subsequently, we speculated that the strong association between COVID-19 narrative belief profile and trust in science might mean that (1) if a brief, inexpensive intervention could increase trust in science, it might possibly (2) affect individuals' COVID-19 narrative belief profile membership. We also wondered whether this effect, mediated by belief profile, might (3) influence behavioral intentions to undertake COVID-19 preventive behaviors. Much of our rationale for these ideas is laid out in the published protocol for the present study [8]. Here, we present a brief explanation outlining why we have focused on trust and how this study fits among current COVID-19 misinformation interventions.

Trust in Science May be an Effective Intervention Target for Misinformation Prevention

Theoretical Basis for Focusing on Trust

Trust is highly complex [14]. Often, "we know by trusting what others tell us" [15]. This is the case because there are many things about which we cannot produce our own knowledge, but there are often experts who *do* have that capability. Here, we posit that beliefs about COVID-19 are linked to rational epistemic trust, the idea that it is reasonable to believe statements made by experts. This might be expressed by the principle, "If [person] has good reasons to believe that [scientist] has good reasons to believe [a finding], then [person] has good reasons to believe that [finding]" [16].

Importantly, though, the prior formulation only pertains to claims about research findings, which make assertions about reality with varying degrees of certainty (eg, face mask use can reduce community transmission of COVID-19 [17]). *Recommendations* from scientists or experts, which may be *based on* scientific claims, instead suggest what people can do to achieve a specific outcome, and so they appeal to a different form of trust. For example, the manuscript reporting that face mask use could reduce community transmission also stated, "face mask use should be as nearly universal as possible" [17]. However, one can be logically consistent and believe both that face mask use can prevent community spread (eg, trust the findings of the research study) and that face mask use should not be universal (eg, not thinking that preventing community spread is important or that it is less important than other interests such as social identity [18]). An alternate formulation that pertains to trusting recommendations might be, "If [person] has

good reasons to believe [scientist] has good reasons to believe that a certain action is in [person's] interest, then [person] has good reasons to believe that they should perform that action" [19].

These concepts can help illustrate how the Trust in Science and Scientists Inventory [14] might be associated with COVID-19 preventive behavioral intentions or belief in COVID-19 misinformation. For example, the item "when scientists change their mind about a scientific idea, it diminishes my trust in their work" might reflect the conditions in which a person trusts a scientific finding, or the item "today's scientists will sacrifice the well-being of others to advance their research" could inform our understanding of how a person perceives recommendations from a scientist.

Evidentiary Basis for Focusing on Trust

In addition to our own identification of associations between trust in science and belief in misinformation [7], studies conducted early in the pandemic found that willingness to abide by COVID-19 preventive guidelines was directly associated with trust in science and risk perception [20-22], the former of which also served as a mediator [22] or moderator [23] for other characteristics such as political conservatism. Trust in science has also been associated with intention to get vaccinated for COVID-19 [21,24]. However, we have not located any prior studies examining misinformation as a mediator in any such relationships.

Extant Interventions to Reduce the Influence of COVID-19 (and Related) Misinformation

One prominent approach to addressing misinformation is debunking (eg, fact checking). Despite some initial concerns, fact checking appears unlikely to backfire [25], and a recent Dutch randomized controlled trial demonstrated that debunking messaging can reduce endorsement of myths about vaccines and influenza [26]. At the same time, fact checking still suffers from scalability issues and a variety of other nuanced concerns and effects [27]. For example, on its face, the time and effort needed to prepare and disseminate a specific piece of misinformation are typically less than the time and effort spent debunking it. Thus, fact checking is likely a useful but not sufficient response amid rapid proliferation of misinformation. Other issues also add complexity to debunking, such as the recent finding that social media users who reported correcting others about COVID-19 online were also more likely to endorse misperceptions about the disease [28].

A promising additional approach is prebunking (eg, inoculation) to confer resistance to the potential influence of misinformation before it is encountered [29-31]. Somewhat related are recent, robust studies suggesting that the likelihood of sharing fake news can be reduced by interventions to reduce inattention and encourage a focus on accuracy [32,33]. Further, in one study, active (online game) and passive (infographic) prebunking interventions targeting misinformation and fake news improved participants' ability to identify misinformation about COVID-19, and the active condition also reduced willingness to share it [34].

Study Aims and Hypotheses

Rationale

In our prior study of COVID-19 misinformation, around 70% of respondents were classified as belonging to the "scientific profile," and classification therein was strongly associated with trust in science. Such an association is also supported both by theoretical and scientific literature. Separately, research on COVID-19 misinformation has suggested the value of scalable, universal prophylaxis that can support people in resisting the influence of misinformation. Therefore, our current study combines those ideas to examine an inoculation approach to COVID-19 misinformation using trust in science as a scalable intervention target.

In this preregistered, randomized controlled trial, we examined the effects of a brief prophylactic intervention (viewing a single infographic about the scientific process for at least 60 seconds). The study had 3 aims with corresponding hypotheses, which we copied verbatim from the study protocol [8] here for narrative clarity.

Aim 1

We aim to assess the effect of a brief informational infographic about the scientific process on trust in science. We hypothesize that exposure to such an intervention will have a moderate, positive effect on trust in science.

Aim 2

We aim to assess the effect of a brief informational infographic about the scientific process on the likelihood of believing scientifically implausible narratives about COVID-19. We hypothesize that exposure to such an intervention will have a small, negative effect on the likelihood of believing implausible narratives, as evidenced by profile membership, and that this will be partly mediated by trust in science.

Aim 3

We aim to assess the effect of a brief informational infographic about the scientific process on behavioral intentions to engage in recommended COVID-19 [nonpharmaceutical preventive behaviors (NPBs)]. We hypothesize that exposure to such an intervention will have a small, positive effect on behavioral intentions to engage in recommended COVID-19 NPBs that will be partly mediated by misinformation profile membership.

Methods

Study Design and Participants

This study of COVID-19 misinformation prophylaxis was a single-stage, two-arm, parallel-group, randomized superiority trial with a 1:1 allocation ratio. Participants were a US-based nationally representative population sample by age, sex, race, and ethnicity recruited using the online data collection platform Prolific [35]. Participants were eligible for this study if they were aged 18 years or older and were selected by Prolific to be part of the nationally representative sample. Prior to randomization, evidence-based quality control mechanisms to manage virtual private network usage, automated responses, dishonest respondents, and inattentive respondents were

implemented [36], and participants were considered ineligible if they failed any of these steps. Replacements were drawn in a manner that preserved the representative nature of the sample. All participants provided digitally signed informed consent according to the protocol approved by the Indiana University Institutional Review Board. This study was preregistered with Clinicaltrials.gov (NCT04557241), and the protocol was published in full before any data collection [8].

Randomization and Masking

After providing sociodemographic information and passing quality control checks, participants were randomly assigned to 1 of 2 study arms: (1) a control group that viewed an infographic about how hunting dogs point at targets (Figure 1) or (2) an intervention group that viewed an infographic about purchasing butter and margarine at the grocery store that was intended to highlight how scientific recommendations change along with newly available evidence (Figure 2). The same artist designed both infographics.

Figure 1. Control infographic.

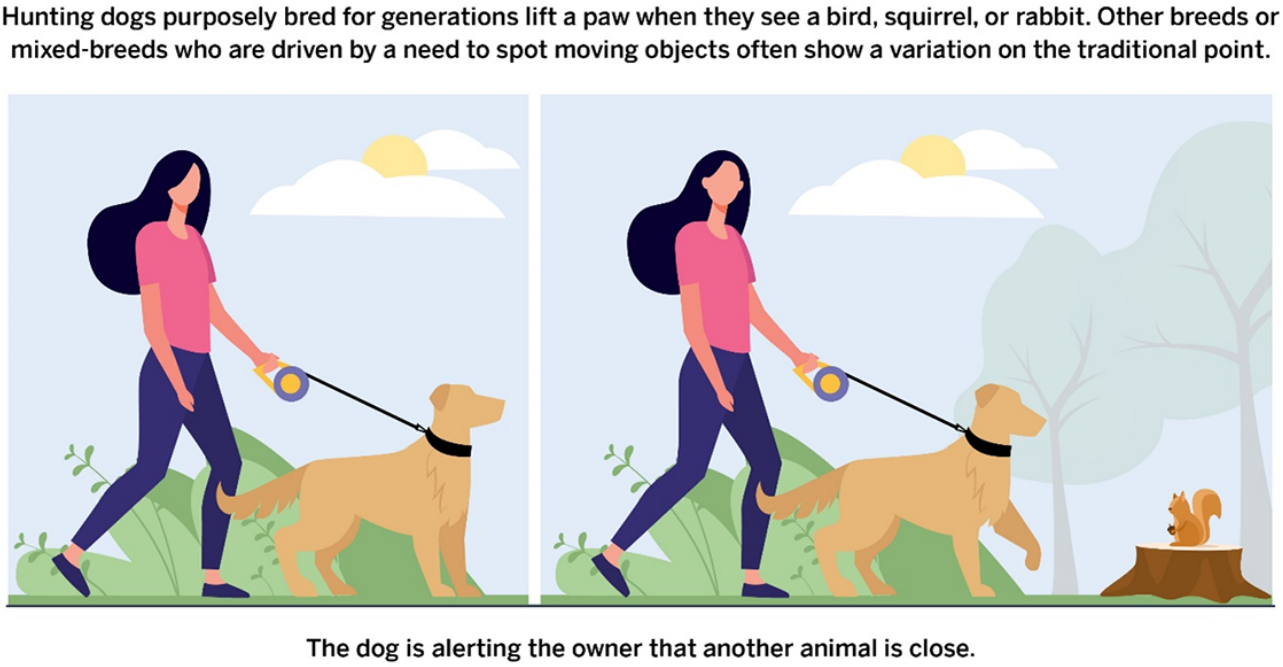
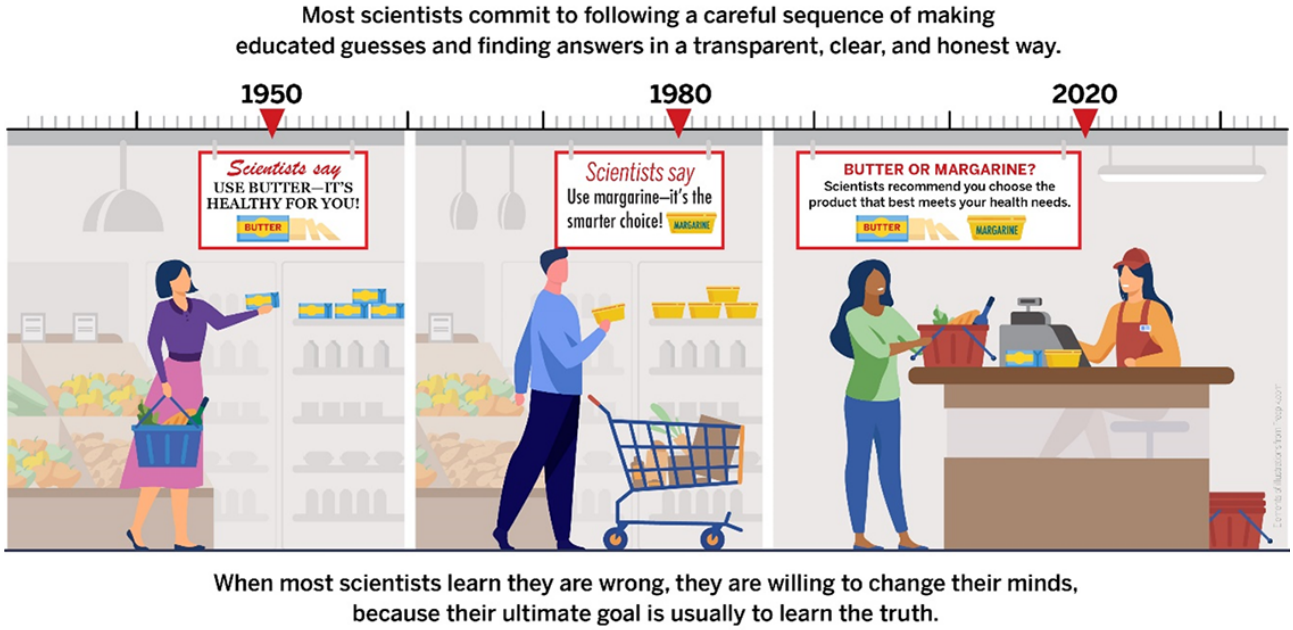


Figure 2. Intervention infographic.



Enrollment was managed by Prolific, entirely independent of the study team. Enrolled subjects accessed a link to Qualtrics (QualtricsXM, Seattle, WA) to participate in the study. Eligible participants were randomized to study arms using the randomizer procedure in Qualtrics with a 1:1 allocation ratio, ensuring no involvement by study personnel. To prevent expectancy biases, study hypotheses and intentions were masked to participants. The summary statement indicated only that "we are interested in understanding how people perceive and think about messages and images."

Procedures

As prespecified [8], the intervention infographic (Figure 2) was iteratively developed using a multistage pilot procedure prior to study initiation. The results of that procedure, which included a randomized pilot comparison between 5 potential infographics, are described in a separate publication [37]. Participants who entered the Qualtrics survey completed sociodemographic items (see the Covariates section) and quality control checks, the latter of which are described in detail in a separate methodological paper [36].

Eligible participants who passed quality control checks were randomized (no indication of this was provided to participants) and then proceeded to the Trust in Science and Scientists Inventory [14]. Then, the program proceeded to display either the control or intervention infographic for the participant to view for a minimum of 60 seconds. To ensure maximum visibility of the infographic for participants on multiple platforms, the Lightbox script [38] was integrated into Qualtrics to allow participants to manually enlarge and reduce images. Following the intervention, participants were asked about the believability of 7 statements about COVID-19 and then were asked about 7 behavioral intentions based on recommendations by the US Centers for Disease Control and Prevention (CDC). These measures were described in detail in the protocol [8], and their use in this study is described in the Outcomes section. Finally, participants completed the Trust in Science Inventory a second time.

Outcomes

This study had 3 primary prespecified outcome measures corresponding with 3 aims.

Aim One

Aim One investigated the effect of the intervention on participants' trust in science and scientists. That construct was measured using the 21-item Trust in Science Inventory [14], which produces a composite score ranging from 1 (low trust) to 5 (high trust). Items in this inventory use Likert-type responses to statements like, "When scientists change their mind about a scientific idea, it diminishes my trust in their work," and "Scientists will protect each other even when they are wrong."

Aim Two

Aim Two investigated the effect of the intervention on participants' classification into misinformation believability profiles [8]. To compute these profiles, participants were asked how believable they found 7 different statements about COVID-19, with responses ranging from 1 (extremely unbelievable) to 7 (extremely believable). Four of these statements were used in our prior research [7] and were derived from an early list of COVID-19 misinformation [5]:

> *The rollout of 5G cellphone networks caused the spread of COVID-19.*
>
> *Bill Gates caused (or helped cause) the spread of COVID-19 in order to expand his vaccination programs.*
>
> *COVID-19 was developed as a military weapon (by China, the United States, or some other country).*
>
> *The number of deaths from COVID-19 has been exaggerated as a way to restrict liberties in the United States.*

A fifth statement referenced the explanation that is currently considered most plausible by much of the scientific community [39]:

> *SARS-Cov-2, the virus that causes COVID-19, likely originated in animals [like bats] and then spread to humans.*

Finally, 2 additional misinformed statements about face masks were added for this study [6,40,41]:

> *Wearing a face mask for COVID-19 prevention can cause oxygen deficiency or carbon dioxide intoxication.*
>
> *Face masks are probably not helpful in reducing COVID-19 spread in a community.*

Statistical and logical classification of participants into latent profiles based on the believability of misinformation was demonstrated in our prior research [7]. However, the current study occurred 8 months later than the original study and included new statements about face masks. Thus, profiles were computed based on the data from this study without prespecifying the existence of any classes (see the Statistical Analysis section). Then, all participants were assigned a numeric variable corresponding with their latent profile membership.

Aim Three

Aim Three targeted the intervention's effect on participants' behavioral intentions to engage in the COVID-19 preventive behaviors recommended by the CDC at the time of study administration [8,42]. Questions were based on structured measurement of intentions using the Theory of Planned Behavior [43], with response options ranging from 1 (unlikely) to 7 (likely). We prespecified 6 intentions in the protocol:

> *Wash your hands often (or use a hand sanitizer that contains at least 60% alcohol).*
>
> *Avoid close contact (stay at least 6 feet from other people).*
>
> *Cover your mouth and nose with a mask when around others.*
>
> *Cover coughs and sneezes.*
>
> *Clean and disinfect frequently touched surfaces daily.*
>
> *Monitor your health daily.*

Intention to get vaccinated was not prespecified in the protocol but was added as the seventh behavioral intention prior to administration in response to availability of vaccination for some US residents.

As planned, overall preventive behavioral intentions were assessed using exploratory factor analysis (see the Statistical Analysis section) to determine the number of factors present and then by computing mean scores for each factor to serve as outcomes. Intention to get vaccinated was analyzed as an isolated outcome of interest in a separate study [44] but was also included as a preventive behavior in this study's factor analysis.

Because they already had received at least one shot of the vaccine, 49 participants were not asked to respond to the question about intention to get vaccinated for COVID-19; data for those individuals were imputed as a 7 (likely). Sensitivity analyses were performed without imputing data for those 49 participants, which led to similar results and conclusions. Therefore, imputed results were used in analyses throughout the manuscript.

Covariates

Additional measures were added as covariates for analysis, as prespecified, including political orientation and religious commitment [7,45]; race, gender, age, and education level; whether the participant had been diagnosed with (or believed they had) COVID-19 [46]; perceived severity of contracting COVID-19 and perceived ability to avoid contracting COVID-19 [47]; and normative belief about friends' and family's avoidance of crowded areas [48].

Due to evolving circumstances in the United States during this study, a question about COVID-19 vaccination status was added after the protocol was published. It read, "Vaccines to prevent COVID-19 have been approved by the Food and Drug Administration for use in the United States. The vaccines will be available to different people at different times. Did you already get a COVID-19 vaccine (at least one shot)?"

Statistical Analysis

We planned to recruit 1000 participants, which would allow detection of small differences (Cohen d=0.18) with 80% power and would be sufficient for both types of planned analysis, linear mixed models (LMM) and path analyses [8].

Aim One

The primary outcome for Aim One, the effect of the infographic intervention on trust in science, was analyzed using an LMM controlling for all covariates (see the Outcomes section) with a random intercept for the individual participant. The interaction between study condition (intervention/control) and time (pre/postintervention) was estimated using contrasts to obtain the difference-in-difference using Kenward-Roger degrees of freedom approximation.

Aim Two

For the first component of this aim, we examined believability profiles for narrative statements about COVID-19 using latent profile analysis. To select the number of classes, we reviewed the Akaike information criterion (AIC), Bayesian information criterion (BIC) and adjusted BIC, class size, entropy, and results from the Vuong-Lo-Mendell-Rubin likelihood ratio test (LMR) to examine improvements in model fit for *k* versus *k-1* classes.

Next, we assigned a "profile" value to each participant based on the profile to which they most closely belonged. That variable was used as an outcome in the prespecified path analysis for this aim, which investigated adjusted odds of being a member of a less-scientific profile by examining the direct effect of the intervention and the indirect effect of the intervention mediated by trust in science, controlling for all other covariates. Finally, we presented results in parallel, treating profile as a multinomial variable (single model) and treating it as a dummy variable (one model per identified profile).

To elucidate other potentially interesting connections between the study variables, we conducted an exploratory, unplanned multivariate logistic regression analysis using profile membership as the outcome variable. All other variables served as dependent predictors except pre-intervention trust in science and having a professional diagnosis of COVID-19, which were highly associated with postintervention trust and believing one had been infected by COVID-19, respectively.

Aim Three

To determine the format of the outcome variable for this aim, we first conducted exploratory factor analysis (maximum likelihood with varimax rotation) to decide whether it was appropriate to treat the behavioral intentions regarding preventive behaviors as a monotonic scale [8]. Identification of a solution was based on assessment of eigenvalues, parallel analysis, factor loadings, and 2-dimensional spatial inspection. The computed scale variable(s) were used as outcomes in the prespecified path analyses for this aim, which computed the direct effect of the intervention on intentions to perform preventive behaviors and the indirect effect of the intervention thereon mediated by misinformation belief profile, controlling for all other covariates.

We computed exploratory path analyses to assess the influence of trust in science on preventive behaviors, with a mediation pathway through believability profile membership, with other variables serving as covariates. These analyses were for informative purposes only and were not used to generate any causal inferences.

Role of the Funding Source

The funders of the study had no role in data collection, analysis, interpretation, or writing of the report. As reported in the protocol, grant reviewers made suggestions to improve study rigor that were incorporated prior to study initiation. Grant reviews were published alongside the protocol [8].

Results

Sample Characteristics

A representative panel of 1000 paid US respondents by gender, age, and race/ethnicity was solicited from Prolific on January 22, 2021 [35]. In total, 1077 Prolific panel members accepted the survey on the Prolific platform and accessed the Qualtrics

study platform through January 24, 2021. The additional 77 cases included those who declined to participate after reading the study information sheet (n=2); who were rejected for failing a quality check (n=23); who exited the study (eg, closed their internet browser) prior to the intervention, most often immediately following a failed quality check (n=35); and who successfully completed the study but for unknown reasons did not request payment from Prolific (n=17). The latter 17 cases were retained for analysis in the arm to which they were randomly assigned, but random assignment beyond 1000 participants did not adhere to a 1:1 allocation ratio. The 3 cases who did not provide complete data for trust in science were excluded listwise from analyses except the latent profile computation for Aim 2. Thus, the final sample included 511 individuals randomized to the intervention arm and 503 individuals randomized to the control arm (Figure 3; Table 1).

Figure 3. CONSORT flow diagram.

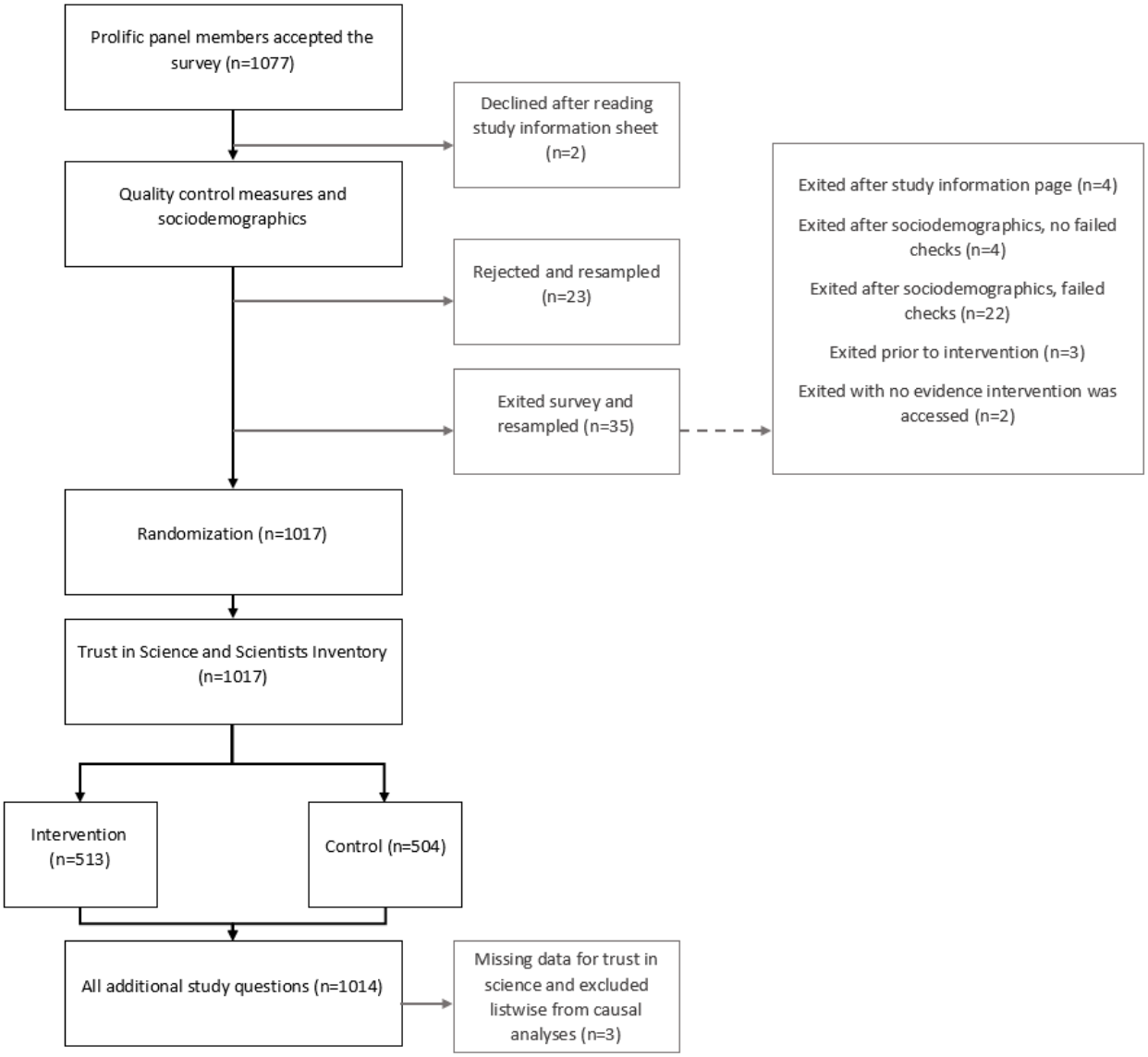


Table 1. Sample characteristics by study arm.

Variable	Intervention (n=511)	Control (n=503)
Gender, n (%)		
Male	251 (49.1)	238 (47.3)
Female	254 (49.7)	261 (51.9)
Nonbinary	3 (0.6)	3 (0.6)
Transgender	3 (0.6)	1 (0.2)
Race, n (%)		
White	388 (75.9)	394 (78.3)
Black or African American	74 (14.5)	58 (11.5)
American Indian or Alaska Native	3 (0.6)	2 (0.4)
Asian	35 (6.8)	37 (7.4)
Native Hawaiian or Pacific Islander	0 (0.0)	1 (0.2)
Other	11 (2.2)	11 (2.2)
Hispanic or Latino/[a] (Yes), n (%)	28 (5.5)	35 (7.0)
COVID-19 diagnosis from a professional, n (%)		
Yes	15 (2.9)	27 (5.4)
No/unsure	496 (97.1)	476 (94.5)
Religious commitment (1=low to 10=high), mean (SD)	4.15 (3.45)	4.06 (3.35)
Political orientation (1=liberal to 10=conservative), mean (SD)	4.27 (2.78)	4.15 (2.71)
Vaccination intention[a] (1=unlikely to 7=likely), mean (SD)	5.48 (2.14)	5.50 (2.10)
Seriousness of contracting COVID-19 (1=not at all serious to 10=very serious), mean (SD)	6.60 (2.72)	6.41 (2.61)
Confidence avoiding COVID-19 (1=not very confident to 5=very confident), mean (SD)	3.25 (0.98)	3.27 (0.97)
Family/friends COVID-19 avoidance (1=strongly disagree to 7=strongly agree), mean (SD)	5.56 (1.51)	5.72 (1.40)
Age (years), mean (SD)	45.50 (16.61)	45.28 (16.19)

[a]Data do not include imputed values of "7" for vaccinated individuals.

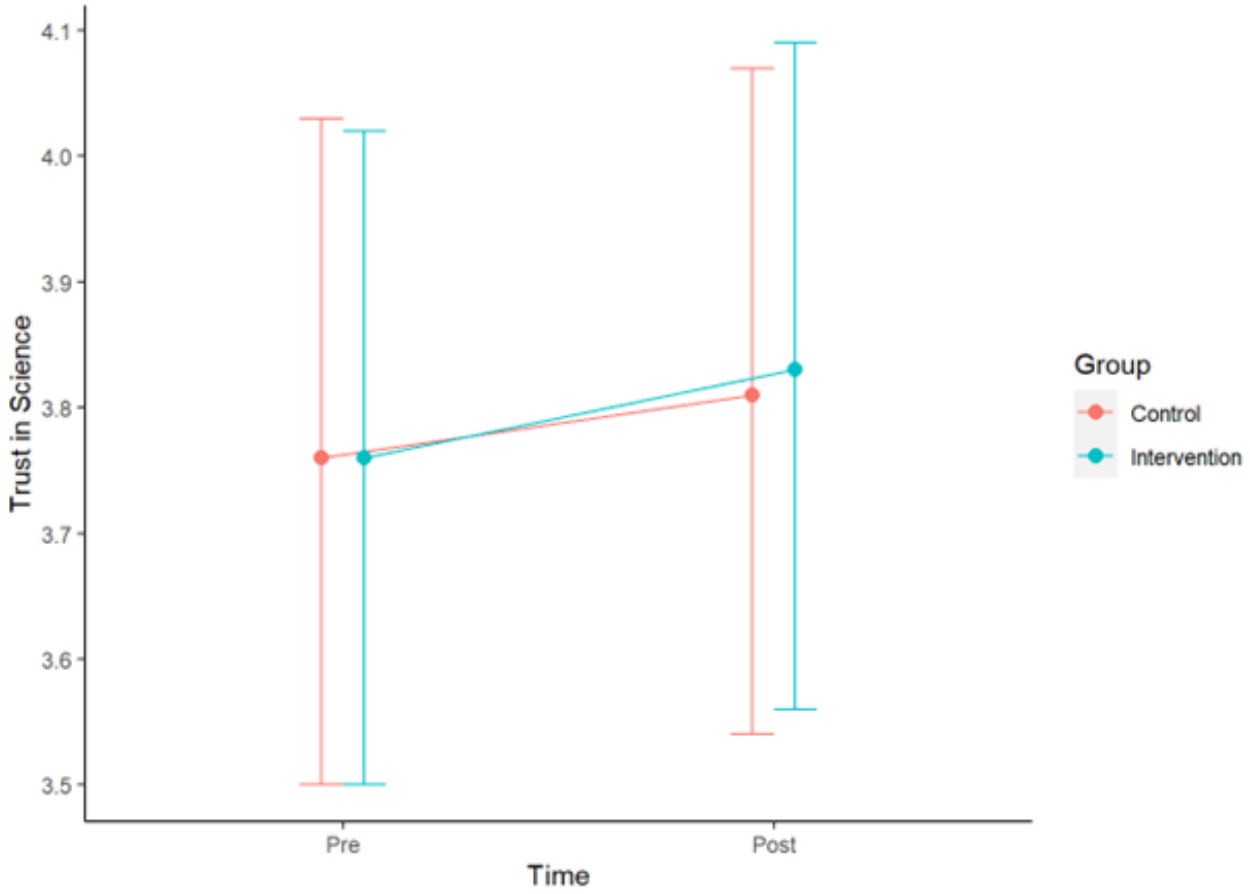
Aim One (Primary Outcome)

We hypothesized that exposure to the infographic intervention would have a moderate, positive effect on trust in science. This hypothesis was partly upheld. Our difference-in-difference analysis suggested that, controlling for all covariates, viewing the intervention infographic had a small, positive effect (0.03, SE 0.01, t_{1000}=2.16, *P*=.031) on trust in science (Table 2; Figure 4). Additional details from the model as well as parceled analytic code are available in Multimedia Appendix 1.

Table 2. Contrast estimates for aim one.

Contrast	Estimate	SE	df	*t*	*P*
Control pre vs intervention pre	0.00	0.03	1048.17	0.10	>.999
Control pre vs control post	–0.04	0.01	1000.00	–5.11	<.001
Intervention pre vs intervention post	–0.07	0.01	1000.00	–8.23	<.001
Control post vs intervention post	–0.02	0.03	1048.17	–0.69	.90
Difference-in-difference (control pre-post) vs (intervention pre-post)	0.03	0.01	1000.00	2.16	.031

Figure 4. Trust in science scores and 95% CIs.



Aim Two (Primary Outcome)

Computation of the Outcome Variable (Narrative Believability Profiles)

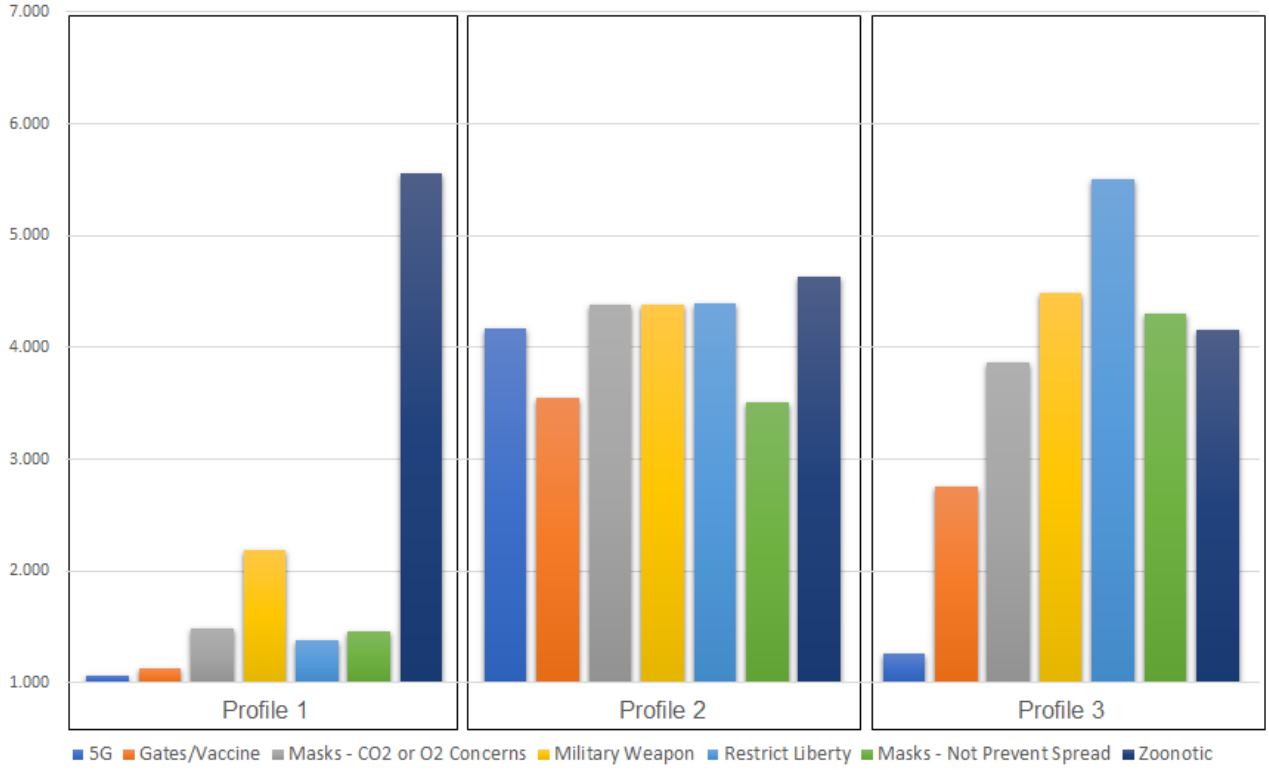
Based on fit statistics (see Multimedia Appendix 1 for analytic code), we selected a 3-class model for use in this study. The primary metrics used to make this decision were the LMR test, entropy, and correspondence with extant data and prior studies. Table 3 and Figure 5 demonstrate the believability of narrative statements across each of the 3 profiles.

Table 3. Standardized means for latent profiles of narrative believability[a] for aim two.

Statement	Profile One (828/1017, 81.42%)	Profile Two (42/1017, 4.13%)	Profile Three (147/1017, 14.45%)
5G	1.06	4.17	1.26
Gates/vaccine	1.14	3.54	2.75
Masks—CO2 or O2 concerns	1.49	4.38	3.86
Military weapon	2.19	4.38	4.48
Restrict liberty	1.39	4.39	5.51
Masks—not prevent spread	1.47	3.51	4.31
Zoonotic	5.55	4.63	4.15

[a]Believability scores ranged from 1 (Extremely unbelievable) to 7 (Extremely believable).

Figure 5. Believability of narrative statements by latent profile. Believability scores range from 1 (Extremely unbelievable) to 7 (Extremely believable).



Profile One (828/1017, 81.42%), the largest class, was most likely to believe the zoonotic narrative (mean 5.55) and found most other narratives to be extremely unbelievable (mean <1.50), with the exception of the military weapon narrative (mean 2.19).

Profile Two (42/1017, 4.13%) was the smallest class and considered all the narratives to be moderately plausible, within a narrow band of believability scores (mean >3.50 and <4.65).

Profile Three (147/1017, 14.45%) reported differential believability across narratives. Members reported that the 5G theory (mean 1.26) and Bill Gates/vaccine narrative (mean 2.75) were extremely or mostly unbelievable. The misinformed idea that face masks can cause carbon dioxide intoxication or oxygen deficiency was perceived to be somewhat more believable (mean 3.86), as were the scientifically implausible statements that masks are not helpful in reducing COVID-19 spread (mean 4.31) or that COVID-19 was developed as a military weapon (mean 4.48). Believability of the zoonotic narrative also fell within this range (mean 4.15). For this profile, the most believable narrative was that the number of deaths from COVID-19 was exaggerated as a way to restrict liberties in the United States (mean 5.51).

Impact of the Intervention on Profile Membership

We hypothesized that exposure to the intervention would have a small, negative effect on the likelihood of belonging to a profile that believed misinformed or implausible narratives and that it would be partially mediated by trust in science. This hypothesis was partly upheld, as there was no evidence of a direct effect, but some evidence of a mediated effect (Table 4; Figure 6).

Table 4. Path analysis of the effects of the intervention on the believability profile.

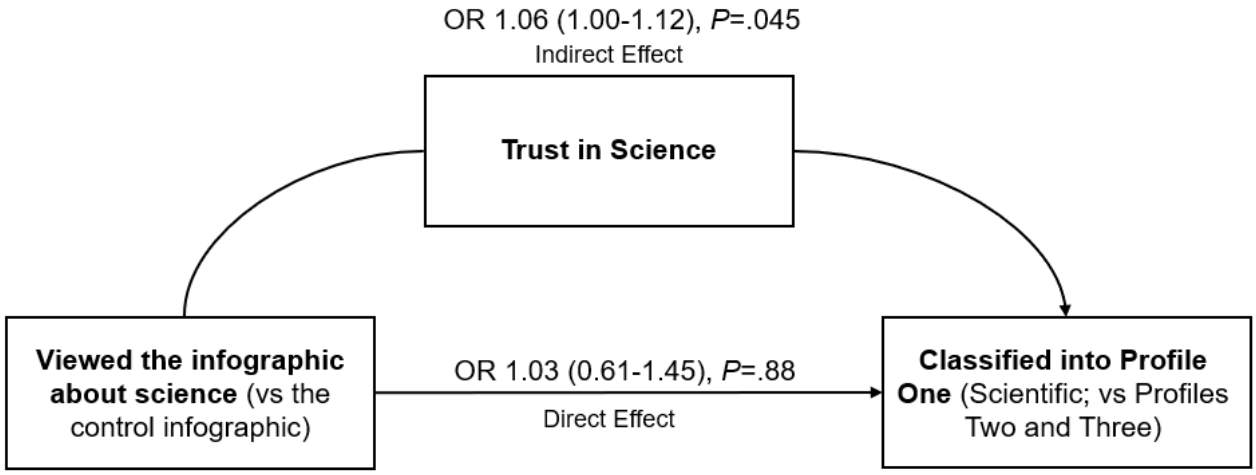
Dependent variables	Odds ratio	SE	Lower CI	Upper CI	z	*P* value	AIC[a]
Multinomial analysis: Profile Two[b]							416.88
Direct effect	0.96	0.35	0.28	1.64	–0.10	.92	
Indirect effect	0.92	0.04	0.84	1.00	–1.95	.051	
Total effect	0.89	0.32	0.25	1.52	–0.34	.74	
Multinomial analysis: Profile Three[b]							
Direct effect	0.98	0.22	0.54	1.41	–0.11	.91	
Indirect effect	0.95	0.02	0.90	1.00	–1.82	.07	
Total effect	0.93	0.21	0.52	1.35	–0.32	.75	
Binomial analysis: Profile One[c]							
Direct effect	1.03	0.21	0.61	1.45	0.16	.88	214.38
Indirect effect	1.06	0.03	1.00	1.12	2.01	.045	
Total effect	1.10	0.23	0.65	1.54	0.44	.66	
Binomial analysis: Profile Two[c]							
Direct effect	1.00	0.35	0.31	1.70	0.01	.99	–147.42
Indirect effect	0.93	0.04	0.86	1.00	–1.84	.07	
Total effect	0.94	0.33	0.28	1.59	–0.19	.85	
Binomial analysis: Profile Three[c]							
Direct effect	1.00	0.22	0.56	1.43	0.00	1.00	146.46
Indirect effect	0.97	0.02	0.93	1.01	–1.56	.12	
Total effect	0.97	0.22	0.55	1.39	–0.15	.88	

[a]AIC: Akaike information criterion.

[b]Reference is Profile One.

[c]Each profile is a dummy variable.

Figure 6. Influence of the intervention on the likelihood of being classified in Profile One, adjusted for age, gender, race, vaccination status, political orientation, perceived severity, perceived susceptibility, family behavior, prior diagnosis, prior infection, and pre-intervention trust. OR: odds ratio.



In the multinomial analysis, controlling for all covariates, the direct effect of viewing the intervention on belonging to Profile Two (versus Profile One) was nonsignificant (adjusted odds ratio [AOR] 0.96, SE 0.35, 95% CI 0.28-1.64, *z*=–0.10, *P*=.92), but there was some evidence of a marginal indirect effect mediated by trust in science (AOR 0.92, SE 0.04, 95% CI 0.84-1.00, *z*=–1.95, *P*=.051). Results for Profile 3 (versus Profile One) were similar, with a nonsignificant direct effect (AOR 0.98, SE 0.22, 95% CI 0.54-1.41, *z*=–0.11, *P*=.91) and limited evidence of a marginal indirect effect mediated by trust in

science (AOR 0.95, SE 0.02, 95% CI 0.90-1.00, z=–1.82, P=.07).

To support disambiguation of the indirect effect, we also conducted binomial path analyses using each profile as a dummy variable. The direct effect of viewing the intervention on belonging to Profile One was nonsignificant (AOR 1.03, SE 0.21, 95% CI 0.61-1.45, z=0.16, P=.88), but there was evidence of a small indirect effect mediated by trust in science (AOR 1.06, SE 0.03, 95% CI 1.00-1.12, z=2.01, P=.045). Full output from the model and analytic code are available in Multimedia Appendix 1.

Aim Three (Primary Outcome)

Computation of the Outcome Variable (COVID-19 Preventive Behaviors)

Exploratory factor analysis did not clearly indicate whether the 7 preventive behaviors formed a monotonic or 2-factor scale. Discrimination based on eigenvalues favored a 2-factor solution, which cumulatively explained 46% of the variance (χ^2_8=124.1, P<.001), while parallel analysis favored a 1-factor solution, which explained 37% of the variance (χ^2_{14}=357.1, P<.001). Conceptually, both approaches were logical.

In the 2-factor solution, handwashing, cleaning and disinfecting surfaces daily, and monitoring one's health daily cleanly loaded on factor 1, while avoiding close contact, covering one's mouth and nose with a mask when around others, and getting vaccinated for COVID-19 loaded on factor 2, with covering coughs and sneezes loading weakly on both factors, but more strongly (0.41) on factor 1. The 95% CIs for the Cronbach alpha were 0.68-0.73 for factor 1 and 0.64-0.71 for factor 2. In the 1-factor solution, variable loadings ranged from 0.48 to 0.71, and the 95% CIs for the Cronbach alpha was 0.74-0.79.

As prespecified [8], the factor analysis guided further analyses for this aim. Given the conceptual complexity, we opted to complete separate analyses for both 1-factor and 2-factor preventive behavior solutions and to interpret them in tandem.

Impact of the Intervention on Behavioral Intentions

We hypothesized that exposure to the intervention would have a small, positive effect on behavioral intentions that would be partially mediated by believability profile membership (Figure 7). Controlling for all covariates, we found no evidence that the intervention affected behavioral intentions to engage in COVID-19 preventive behavior using either a 1-factor or 2-factor outcome (Tables 5, 6, and 7).

Figure 7. Hypothesized causal pathway of the intervention (not supported), adjusted for age, gender, race, vaccination status, political orientation, perceived severity, perceived susceptibility, family behavior, prior diagnosis, prior infection, and pre-intervention trust.

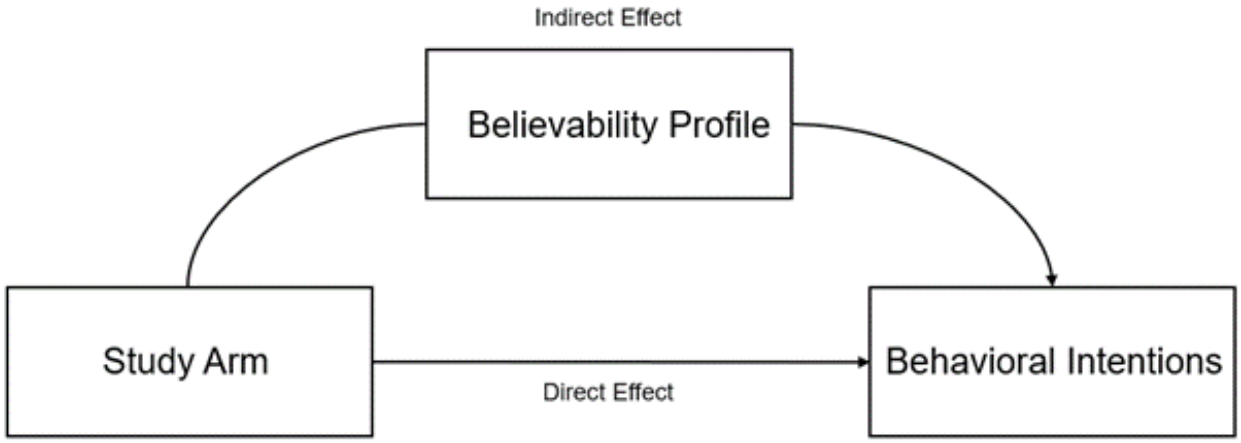


Table 5. Path analysis of effects of the intervention on preventive behaviors (1-factor solution).

Mediators	Coefficient	SE	Lower CI	Upper CI	z	*P* value	AIC[a]
Multinomial analysis: Profile Two[b]							
Direct effect	–0.04	0.04	–0.12	0.05	–0.85	.39	3018.13
Indirect effect	0.05	0.22	–0.38	0.49	0.23	.82	
Total effect	0.01	0.23	–0.43	0.46	0.06	.95	
Multinomial analysis: Profile Three[b]							
Direct effect	–0.04	0.04	–0.12	0.05	–0.85	.39	
Indirect effect	0.04	0.15	–0.26	0.33	0.24	.82	
Total effect	0.00	0.16	–0.30	0.30	0.00	1.00	
Binomial analysis: Profile One[c]							
Direct effect	–0.04	0.04	–0.12	0.05	–0.86	.39	2814.37
Indirect effect	0.05	0.13	–0.22	0.31	0.35	.73	
Total effect	0.01	0.14	–0.27	0.29	0.07	.94	
Binomial analysis: Profile Two[c]							
Direct effect	–0.03	0.05	–0.12	0.06	–0.70	.48	2528.63
Indirect effect	0.01	0.15	–0.28	0.31	0.10	.92	
Total effect	–0.02	0.16	–0.32	0.29	–0.11	.91	
Binomial analysis: Profile Three[c]							
Direct effect	–0.03	0.04	–0.12	0.05	–0.76	.45	2763.44
Indirect effect	0.01	0.13	–0.24	0.27	0.08	.93	
Total effect	–0.02	0.14	–0.29	0.25	–0.17	.87	

[a]AIC: Akaike information criterion.

[b]Reference is Profile One.

[c]Each profile is a dummy variable.

Table 6. Multinomial path analysis of the effects of the intervention on preventive behaviors (2-factor solution), in which the reference is Profile One.

Mediator	Coefficient	SE	Lower CI	Upper CI	z	*P* value	AIC[a]
Factor 1[b]as the dependent variable: Profile Two							3376.55
Direct effect	–0.03	0.05	–0.13	0.07	–0.62	.53	
Indirect effect	0.03	0.13	–0.22	0.28	0.23	.82	
Total effect	0.00	0.14	–0.27	0.26	–0.02	.98	
Factor 1[b] as the dependent variable: Profile Three							
Direct effect	–0.03	0.05	–0.13	0.07	–0.62	.53	
Indirect effect	0.02	0.07	–0.12	0.15	0.24	.81	
Total effect	–0.02	0.09	–0.18	0.15	–0.18	.86	
Factor 2[c] as the dependent variable: Profile Two							3440.75
Direct effect	–0.04	0.05	–0.15	0.06	–0.81	.42	
Indirect effect	0.08	0.35	–0.61	0.77	0.23	.82	
Total effect	0.04	0.35	–0.66	0.73	0.11	.92	
Factor 2[c] as the dependent variable: Profile Three							
Direct effect	–0.04	0.05	–0.15	0.06	–0.81	.42	
Indirect effect	0.06	0.26	–0.44	0.57	0.24	.81	
Total effect	0.02	0.26	–0.49	0.53	0.07	.94	

[a]AIC: Akaike information criterion.

[b]Preventive behaviors 1, 4, 5, and 6.

[c]Preventive behaviors 2, 3, and 7.

Table 7. Binomial path analysis of the effects of the intervention on preventive behaviors (2-factor solution), in which each profile is a dummy variable.

Mediator	Coefficient	SE	Lower CI	Upper CI	z	*P* value	AIC[a]
Factor 1[b] as the dependent variable: Profile One							
Direct effect	–0.03	0.05	–0.13	0.07	–0.62	.53	3172.83
Indirect effect	0.02	0.06	–0.10	0.15	0.35	.73	
Total effect	–0.01	0.08	–0.17	0.15	–0.12	.91	
Factor 1[b] as the dependent variable: Profile Two							
Direct effect	–0.03	0.05	–0.13	0.07	–0.57	.57	2820.14
Indirect effect	0.01	0.09	–0.17	0.19	0.09	.92	
Total effect	–0.02	0.11	–0.23	0.19	–0.20	.84	
Factor 1[b] as the dependent variable: Profile Three							
Direct effect	–0.03	0.05	–0.13	0.07	–0.58	.56	3100.05
Indirect effect	0.00	0.06	–0.11	0.12	0.08	.93	
Total effect	–0.03	0.08	–0.18	0.13	–0.33	.74	
Factor 2[c] as the dependent variable: Profile One							
Direct effect	–0.04	0.05	–0.15	0.06	–0.81	.42	3237.87
Indirect effect	0.08	0.23	–0.36	0.52	0.35	.73	
Total effect	0.04	0.23	–0.42	0.49	0.15	.88	
Factor 2[c] as the dependent variable: Profile Two							
Direct effect	–0.03	0.06	–0.15	0.08	–0.59	.56	3020.59
Indirect effect	0.02	0.22	–0.42	0.46	0.10	.92	
Total effect	–0.01	0.23	–0.47	0.44	–0.05	.96	
Factor 2[c] as the dependent variable: Profile Three							
Direct effect	–0.04	0.05	–0.15	0.07	–0.69	.49	3203.54
Indirect effect	0.02	0.23	–0.42	0.46	0.08	.93	
Total effect	–0.02	0.23	–0.48	0.44	–0.08	.94	

[a]AIC: Akaike information criterion.

[b]Preventive behaviors 1, 4, 5, and 6.

[c]Preventive behaviors 2, 3, and 7.

Secondary Outcomes

Exploratory Multivariate Logistic Regression on Profile Membership

Independent of the path analysis, we investigated what factors were associated with classification in each of the 3 belief profiles. Each 1-point movement toward political conservatism on a 10-point scale was associated with 1.39 adjusted odds of belonging to Profile Three versus Profile One (χ^2_1=40.52, 95% CI 1.25-1.53, *P*<.001), and each 1-point increase in perceived severity of COVID-19 was associated with 0.82 adjusted odds of belonging to Profile Three versus Profile One (χ^2_1=19.04, 95% CI 0.75-0.90, *P*<.001). Each additional year of age was associated with a slight decrease in adjusted odds (0.972) of belonging to Profile Two versus Profile One (χ^2_1=4.65, 95% CI 0.95-1.00, *P*=.031).

Finally, each 1-point increase in trust in science was associated with substantially lower adjusted odds of belonging to Profile Three (0.21) or Profile Two (0.14) compared with Profile One (Profile Three: χ^2_1=55.57, 95% CI 0.14-0.31, *P*<.001; Profile Two: χ^2_1=39.20, 95% CI 0.08-0.26, *P*<.001). See Multimedia Appendix 1 and Multimedia Appendix 2 for all outputs and code.

Exploratory Path Analysis of the Association Between Trust in Science and Preventive Behavioral Intentions

We computed exploratory path analyses to assess the influence of trust in science on preventive behaviors, with a mediation pathway through believability profile membership. In the 1-factor preventive behavior model treating believability Profile One as a dummy variable, there was a significant direct effect (0.46, SE 0.11, 95% CI 0.24-0.69, *z*=4.05, *P*<.001) of trust in science on preventive behavior, as well as a significant indirect effect mediated by believability profile (1.28, SE 0.36, 95% CI

0.57-1.99, z=3.53, P<.001). Similar outcomes were observed for the 2-factor model (see Multimedia Appendix 1 and Multimedia Appendix 2). These specific findings were correlational, not causal.

Discussion

This study provides preliminary evidence and a proof-of-concept for using infographics that truthfully address underlying reasons why a person might not trust science or scientists to (1) improve trust in science and (2) provide inoculation against COVID-19 misinformation. However, observed effects were small, as expected for a short, passive, and inexpensive intervention. Much remains to be learned in this area of research. Here, we discuss the main findings separately by study aim, provide additional interpretation of exploratory analyses, and make recommendations for future work.

Aim One

Principal Finding

This study found that viewing an infographic designed to truthfully address underlying reasons why a person might not trust science or scientists [37] once, for a minimum of 60 seconds, caused a small aggregate increase in participants' overall trust in science.

Interpretation

The scale used in this study measured trust in science as a composite from 21 questions to yield a score from 1 to 5. We posit that the small difference-in-difference improvement estimate (+0.03) is meaningful due to the simplicity of the intervention and the ease with which such an intervention could be deployed to large numbers of people. Especially given recent research indicating that aggregate social trust in science may affect variables like vaccine confidence beyond individual-level trust [49], we believe this finding merits replication research. At the same time, though trust in science is a worthy concept to study in and of itself, we were particularly interested in the degree to which changing trust might affect misinformation or behavioral intentions. This aim simply established that it is plausible that a brief, single exposure to an infographic can improve trust in science.

Aim Two

Principal Finding

We found some evidence that viewing an infographic designed to truthfully address underlying reasons why a person might not trust science or scientists [37] once, for a minimum of 60 seconds, may have had a very small indirect effect on belief in COVID-19 misinformation.

Interpretation

There is ongoing discussion among methodologists and metascientists as to how to interpret mediation effects in terms of causal attribution, especially when the direct and total effects are nonsignificant. In general, we can think of a direct effect as proposing that "*X* is regularly followed by *Y*," while an indirect effect suggests "*Y* if and only if *A*" [50]. Here, *X* is "viewing the infographic about science," and *Y* is "belonging to the scientific latent belief profile," while *A* is "an increase in trust in science and scientists." We did not find any evidence of a direct effect or total effect—that is, viewing our intervention infographic was not regularly followed by increased likelihood of belonging to the scientific profile, nor was it sufficient for establishing increased likelihood of belonging to that profile. However, the mediation effect (in this case, OR 1.06) suggests that viewing the intervention resulted in increased likelihood of belonging to the scientific profile *if and only if* viewing the intervention also increased trust in science to a sufficient degree. In conjunction with the main finding from Aim One, this supports the plausibility of the intervention functioning in this manner.

Because this study used a randomized, controlled experimental design and included numerous covariates, endogeneity bias was not a substantive concern in interpretation [51,52]. Further, the sample was drawn to be nationally representative of the US population by age, sex, race, and ethnicity [35], attenuating but not eliminating concerns about generalizability [53]. Given the complexity of the topic, further investigation of this relationship as well as experimental replication are both needed before drawing any sort of definitive conclusion.

That noted, we encourage such research to be undertaken with some urgency. This work has meaningful, practical application if the findings hold true. While fact checking can reduce belief in misinformation, it is not likely feasible to respond to the amount and variability that is produced, even for a single topic like COVID-19 [27]. Further, the type and nature of misinformation can rapidly shift in unexpected ways. For example, during revision of this manuscript, the Mississippi State Department of Health issued a warning that 70% of recent calls to their Poison Control Center were related to ingesting livestock or animal formulations of ivermectin [54]. The comparative advantage of the type of trust-mediated prophylaxis proposed and tested here (versus debunking or fact checking) is that it does not require addressing each new misinformed claim. In that sense, such interventions affecting believability of misinformation might work well alongside interventions to reduce the likelihood of sharing it [32,33].

Aim 3

Principal Finding

Viewing an infographic designed to truthfully address underlying reasons why a person might not trust science or scientists [37] once, for a minimum of 60 seconds, did not significantly affect behavioral intentions to engage in COVID-19 preventive behaviors.

Interpretation

We hypothesize, but cannot be certain, that this null finding emerged because this specific infographic addressed a component of rational epistemic trust (eg, why it makes sense to trust scientific findings even when they change over time) [16] but did not address recommendation trust (eg, did not offer any reassurance that scientists make recommendations that are in the best interests of others) [19]. Our exploratory analyses did indicate a significant, strong association between trust in science and COVID-19 preventive behavioral intentions,

indicating the plausibility that a strategic effect targeting that recommendation trust might affect such intentions. It is also possible, of course, that behavioral intentions are sufficiently complex and difficult to change that a single infographic viewing, regardless of content, would not "move the needle" [55].

Exploratory Findings

This study also identified potentially valuable information about how beliefs about COVID-19 may cluster. In May 2020, we identified a single latent profile that endorsed the zoonotic narrative and generally found other narratives unbelievable and 3 profiles that believed misinformed narratives to varying degrees but also believed the zoonotic narrative [7]. The present study was conducted in a different information ecosystem (around 8 months later) and included 2 additional narratives about face masks.

Two major findings about profiles were consistent between our studies: There was a single profile endorsing the zoonotic narrative and generally disbelieving other narratives, and no narrative profile rejected the zoonotic explanation. However, there were only 2 nonscientific profiles in this study rather than 3, and interpretation of their meaning was clearer than in the original study. The smallest profile (Profile Two) found all narratives to be at least somewhat believable. In contrast, Profile Three was comparatively less likely to endorse narratives that were subjectively less political in the United States (eg, 5G, Gates/vaccine) and more likely to believe other narratives (eg, restrict liberty, masks don't prevent spread). It is unclear whether this difference was due to the addition of the face mask narratives, a change in the information ecosystem, the use of a nationally representative sample, or a different reason altogether.

Notably, in our unplanned regression analysis, trust in science remained the most substantive predictor of profile membership, as in May 2020. However, unlike our previous study, in which political orientation was not associated with profile membership, here we found that conservative political orientation was associated with classification in Profile Three versus Profile One, but not with classification in Profile Two. Along with the profile analysis itself, this suggests the possibility of 2 "typologies" of misinformation belief, 1 that is apolitical and may believe even scientifically impossible narratives (eg, finding all narratives to be plausible) and 1 that is associated with political orientation and that believes misinformation somewhat selectively, applying an alternate decision heuristic in determining what is plausible.

Limitations and Future Directions

This study investigated multiple outcomes and so there was some increased risk of Type 1 error. For this reason, we interpreted the outcomes cautiously and recommend replication prior to any definitive determination about these findings. At the same time, the primary outcomes were prespecified and were assessed using a limited number of models. A limitation specific to the third aim is that behavioral intentions are not behaviors, so this study should not be interpreted to assess the effect of the intervention on actual behavior. In addition, we opted to limit the allowable content in the intervention. As we note in our pilot study [37], we very purposefully used messaging about science and scientists that we believed to be truthful. Our intention specifically was not to "manipulate" trust in science but rather to determine whether exposure to an easily digested, truthful accounting had a causal effect.

There were numerous decisions made in the course of developing the single image used as the intervention in this study, as well as the structure of the intervention. Given this proof-of-concept, there is much room to explore alternative approaches, including, but not limited to, investigating whether a brief video would be more efficacious than a static image, the art style or amount of wording matters, embedding the image as an ad in social media (eg, repeated natural exposures) over a period of time affects the results, and comparison to real negative messages about science would produce similar results to this study, which used an active placebo about dogs.

Acknowledgments

The infographics and art were produced by Ms. Amanda Goehlert, a Designer on the Creative Team at Indiana University Studios.

This study was funded by the Indiana Clinical and Translational Sciences Institute funded, in part by Award Number UL1TR002529 from the National Institutes of Health, National Center for Advancing Translational Sciences, Clinical and Translational Sciences Award. The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

Authors' Contributions

JA was the lead investigator and authored the first draft of the paper. JA, YX, EET, and LG contributed to the design, prepared the protocol and analytic plan, and conducted the randomized pilot test. EET and AG (see Acknowledgments) led the process to develop infographics from our working documents. XC and YX conducted statistical analyses with support from LG and JA, all of whom had access to the data and take responsibility for the integrity of the data and the analytic outcomes. JA, YX, and EET secured funding for the project. All authors contributed to the manuscript and had joint responsibility for the decision to submit for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
All analytic code and instructions for its use to replicate results and generate additional tables.
[DOCX File , 34 KB - jmir_v23i10e32425_app1.docx]

Multimedia Appendix 2
CSV datasets used with Appendix 1.
[ZIP File (Zip Archive), 99 KB - jmir_v23i10e32425_app2.zip]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1170 KB - jmir_v23i10e32425_app3.pdf]

References

1. Tangcharoensathien V, Calleja N, Nguyen T, Purnat T, D'Agostino M, Garcia-Saiso S, et al. Framework for managing the COVID-19 infodemic: Methods and results of an online, crowdsourced WHO Technical Consultation. J Med Internet Res 2020 Jun 26;22(6):e19659 [FREE Full text] [doi: 10.2196/19659] [Medline: 32558655]
2. Eysenbach G. Infodemiology: the epidemiology of (mis)information. The American Journal of Medicine 2002 Dec;113(9):763-765. [doi: 10.1016/S0002-9343(02)01473-0]
3. Calleja N, AbdAllah A, Abad N, Ahmed N, Albarracin D, Altieri E, et al. A Public health research agenda for managing infodemics: Methods and results of the first WHO Infodemiology Conference. JMIR Infodemiology 2021 Sep 15;1(1):e30979 [FREE Full text] [doi: 10.2196/30979] [Medline: 34604708]
4. Brennen JS, Simon FM, Howard PN, Nielsen RK. Types, sources, and claims of COVID-19 misinformation. Reuters Institute for the Study of Journalism.: University of Oxford; 2020 Apr 07. URL: https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation [accessed 2021-10-09]
5. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, et al. The COVID-19 social media infodemic. Sci Rep 2020 Oct 06;10(1):16598 [FREE Full text] [doi: 10.1038/s41598-020-73510-5] [Medline: 33024152]
6. Lynas M. COVID: Top 10 current conspiracy theories. Alliance for Science. 2020 Apr 20. URL: https://allianceforscience.cornell.edu/blog/2020/04/covid-top-10-current-conspiracy-theories/ [accessed 2021-10-09]
7. Agley J, Xiao Y. Misinformation about COVID-19: evidence for differential latent profiles and a strong association with trust in science. BMC Public Health 2021 Jan 07;21(1):89 [FREE Full text] [doi: 10.1186/s12889-020-10103-x] [Medline: 33413219]
8. Agley J, Xiao Y, Thompson EE, Golzarri-Arroyo L. COVID-19 misinformation prophylaxis: Protocol for a randomized trial of a brief informational intervention. JMIR Res Protoc 2020 Dec 7;9(12):e24383. [doi: 10.2196/24383]
9. Enders AM, Uscinski JE, Klofstad C, Stoler J. The different forms of COVID-19 misinformation and their consequences. HKS Misinfo Review 2020 Nov 12:1. [doi: 10.37016/mr-2020-48]
10. Sallam M, Dababseh D, Yaseen A, Al-Haidar A, Taim D, Eid H, et al. COVID-19 misinformation: Mere harmless delusions or much more? A knowledge and attitude cross-sectional study among the general public residing in Jordan. PLoS One 2020 Dec 3;15(12):e0243264 [FREE Full text] [doi: 10.1371/journal.pone.0243264] [Medline: 33270783]
11. Greene CM, Murphy G. Quantifying the effects of fake news on behavior: Evidence from a study of COVID-19 misinformation. J Exp Psychol Appl 2021 Jun 10:1 [FREE Full text] [doi: 10.1037/xap0000371] [Medline: 34110860]
12. Loomba S, de Figueiredo A, Piatek SJ, de Graaf K, Larson HJ. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nat Hum Behav 2021 Mar 05;5(3):337-348. [doi: 10.1038/s41562-021-01056-1] [Medline: 33547453]
13. Mendez R. WHO says Covid misinformation is a major factor driving pandemic around the world. CNBC. 2021 Aug 24. URL: https://www.cnbc.com/2021/08/24/who-says-covid-misinformation-is-a-major-factor-driving-pandemic-around-the-world.html [accessed 2021-10-09]
14. Nadelson L, Jorcyk C, Yang D, Jarratt Smith M, Matson S, Cornell K, et al. I just don't trust them: The development and validation of an assessment instrument to measure trust in science and scientists. Sch Sci Math 2014 Jan 19;114(2):76-86. [doi: 10.1111/ssm.12051]
15. Faulkner P, Simpson T, editors. The philosophy of trust. Oxford, England: Oxford University Press; 2017.
16. Hardwig J. The role of trust in knowledge. The Journal of Philosophy 1991 Dec;88(12):693. [doi: 10.2307/2027007]
17. Eikenberry SE, Mancuso M, Iboi E, Phan T, Eikenberry K, Kuang Y, et al. To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. Infect Dis Model 2020;5:293-308 [FREE Full text] [doi: 10.1016/j.idm.2020.04.001] [Medline: 32355904]
18. Powdthavee N, Riyanto YE, Wong ECL, Yeo JXW, Chan QY. When face masks signal social identity: Explaining the deep face-mask divide during the COVID-19 pandemic. PLoS One 2021 Jun 10;16(6):e0253195 [FREE Full text] [doi: 10.1371/journal.pone.0253195] [Medline: 34111233]

19. Bennett M. Should I do as I'm told? Trust, experts, and COVID-19. Kennedy Institute of Ethics Journal 2020;30(3-4):243-263. [doi: 10.1353/ken.2020.0014]
20. Dohle S, Wingen T, Schreiber M. Acceptance and adoption of protective measures during the COVID-19 pandemic: The role of trust in politics and trust in science. Soc. Psychol. Bull 2020 Dec 23;15(4):1. [doi: 10.32872/spb.4315]
21. Hromatko I, Tonković M, Vranic A. Trust in science, perceived vulnerability to disease, and adherence to pharmacological and non-pharmacological COVID-19 recommendations. Front Psychol 2021 Apr 30;12:664554 [FREE Full text] [doi: 10.3389/fpsyg.2021.664554] [Medline: 33995222]
22. Plohl N, Musil B. Modeling compliance with COVID-19 prevention guidelines: the critical role of trust in science. Psychol Health Med 2021 Jan 01;26(1):1-12. [doi: 10.1080/13548506.2020.1772988] [Medline: 32479113]
23. Koetke J, Schumann K, Porter T. Trust in science increases conservative support for social distancing. Group Processes & Intergroup Relations 2021 May 31;24(4):680-697. [doi: 10.1177/1368430220985918]
24. Attwell K, Lake J, Sneddon J, Gerrans P, Blyth C, Lee J. Converting the maybes: Crucial for a successful COVID-19 vaccination strategy. PLoS One 2021 Jan 20;16(1):e0245907 [FREE Full text] [doi: 10.1371/journal.pone.0245907] [Medline: 33471821]
25. Wood T, Porter E. The elusive backfire effect: Mass attitudes' steadfast factual adherence. Polit Behav 2018 Jan 16;41(1):135-163. [doi: 10.1007/s11109-018-9443-y]
26. Yousuf H, van der Linden S, Bredius L, Ted van Essen GA, Sweep G, Preminger Z, et al. A media intervention applying debunking versus non-debunking content to combat vaccine misinformation in elderly in the Netherlands: A digital randomised trial. EClinicalMedicine 2021 May;35:100881 [FREE Full text] [doi: 10.1016/j.eclinm.2021.100881] [Medline: 34124631]
27. Pennycook G, Rand DG. The psychology of fake news. Trends Cogn Sci 2021 May;25(5):388-402 [FREE Full text] [doi: 10.1016/j.tics.2021.02.007] [Medline: 33736957]
28. Bode L, Vraga EK. Correction experiences on social media during COVID-19. Social Media + Society 2021 Apr 12;7(2):205630512110088. [doi: 10.1177/20563051211008829]
29. Cook J, Lewandowsky S, Ecker UKH. Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. PLoS One 2017 May 5;12(5):e0175799 [FREE Full text] [doi: 10.1371/journal.pone.0175799] [Medline: 28475576]
30. van der Linden S, Roozenbeek J, Compton J. Inoculating against fake news about COVID-19. Front Psychol 2020 Oct 23;11:566790 [FREE Full text] [doi: 10.3389/fpsyg.2020.566790] [Medline: 33192844]
31. Jolley D, Douglas KM. Prevention is better than cure: Addressing anti-vaccine conspiracy theories. J Appl Soc Psychol 2017 Jun 28;47(8):459-469. [doi: 10.1111/jasp.12453]
32. Pennycook G, Epstein Z, Mosleh M, Arechar AA, Eckles D, Rand DG. Shifting attention to accuracy can reduce misinformation online. Nature 2021 Apr 17;592(7855):590-595. [doi: 10.1038/s41586-021-03344-2] [Medline: 33731933]
33. Pennycook G, McPhetres J, Zhang Y, Lu JG, Rand DG. Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. Psychol Sci 2020 Jul 30;31(7):770-780 [FREE Full text] [doi: 10.1177/0956797620939054] [Medline: 32603243]
34. Basol M, Roozenbeek J, Berriche M, Uenal F, McClanahan WP, Linden SVD. Towards psychological herd immunity: Cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. Big Data & Society 2021 May 11;8(1):205395172110138. [doi: 10.1177/20539517211013868]
35. Representative samples on Prolific. Prolific. 2019 Mar 05. URL: https://researcher-help.prolific.co/hc/en-gb/articles/360019236753-Representative-Samples-on-Prolific-limited-release- [accessed 2021-10-09]
36. Agley J, Xiao Y, Nolan R, Golzarri-Arroyo L. Quality control questions on Amazon's Mechanical Turk (MTurk): A randomized trial of impact on the USAUDIT, PHQ-9, and GAD-7. Behav Res Methods 2021 Aug 06:1 [FREE Full text] [doi: 10.3758/s13428-021-01665-8] [Medline: 34357539]
37. Agley J, Xiao Y, Thompson EE, Golzarri-Arroyo L. Using infographics to improve trust in science: a randomized pilot test. BMC Res Notes 2021 May 29;14(1):210 [FREE Full text] [doi: 10.1186/s13104-021-05626-4] [Medline: 34051823]
38. Dhakar L. Lightbox: the original lightbox script. URL: https://lokeshdhakar.com/projects/lightbox2/ [accessed 2021-10-09]
39. Maxmen A, Mallapaty S. The COVID lab-leak hypothesis: what scientists do and don't know. Nature 2021 Jun 08;594(7863):313-315. [doi: 10.1038/d41586-021-01529-3] [Medline: 34108722]
40. Kaufman BG, Whitaker R, Mahendraratnam N, Smith VA, McClellan MB. Comparing associations of state reopening strategies with COVID-19 burden. J Gen Intern Med 2020 Dec 06;35(12):3627-3634 [FREE Full text] [doi: 10.1007/s11606-020-06277-0] [Medline: 33021717]
41. Cheng VC, Wong S, Chuang VW, So SY, Chen JH, Sridhar S, et al. The role of community-wide wearing of face mask for control of coronavirus disease 2019 (COVID-19) epidemic due to SARS-CoV-2. J Infect 2020 Jul;81(1):107-114 [FREE Full text] [doi: 10.1016/j.jinf.2020.04.024] [Medline: 32335167]
42. Centers for Disease Control and Prevention. 2021 Aug 13. URL: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html [accessed 2021-10-09]
43. Ajzen I. Theory of planned behavior questionnaire. Measurement Instrument Database for Social Science. 2013. URL: https://www.midss.org/sites/default/files/tpb.construction.pdf [accessed 2021-10-09]

44. Agley J, Xiao Y, Thompson E, Golzarri-Arroyo L. Factors associated with reported likelihood to get vaccinated for COVID-19 in a nationally representative US survey. Public Health 2021 Jul;196:91-94 [FREE Full text] [doi: 10.1016/j.puhe.2021.05.009] [Medline: 34171616]
45. Agley J. Assessing changes in US public trust in science amid the COVID-19 pandemic. Public Health 2020 Jun;183:122-125 [FREE Full text] [doi: 10.1016/j.puhe.2020.05.004] [Medline: 32405095]
46. Bruine de Bruin W. Age differences in COVID-19 risk perceptions and mental health: Evidence from a national U.S. survey conducted in March 2020. J Gerontol B Psychol Sci Soc Sci 2021 Jan 18;76(2):e24-e29 [FREE Full text] [doi: 10.1093/geronb/gbaa074] [Medline: 32470120]
47. Yıldırım M, Güler A. COVID-19 severity, self-efficacy, knowledge, preventive behaviors, and mental health in Turkey. Death Stud 2020 Jul 16:1-8. [doi: 10.1080/07481187.2020.1793434] [Medline: 32673183]
48. Chambon M, Dalege J, Elberse JE, van Harreveld F. A psychological network approach to attitudes and preventive behaviors during pandemics: A COVID-19 study in the United Kingdom and the Netherlands. Social Psychological and Personality Science 2021 Mar 23:194855062110024 [FREE Full text] [doi: 10.1177/19485506211002420]
49. Sturgis P, Brunton-Smith I, Jackson J. Trust in science, social consensus and vaccine confidence. Nat Hum Behav 2021 May 17:1 [FREE Full text] [doi: 10.1038/s41562-021-01115-7] [Medline: 34002053]
50. Agler R, De Boeck P. On the interpretation and use of mediation: Multiple perspectives on mediation analysis. Front Psychol 2017 Nov 15;8:1984 [FREE Full text] [doi: 10.3389/fpsyg.2017.01984] [Medline: 29187828]
51. Antonakis J, Bendahan S, Jacquart P, Lalive R. Causality and endogeneity: problems and solutions. In: Day DV, editor. The Oxford Handbook of Leadership and Organizations. New York, NY: Oxford University Press; 2014:93-117.
52. Stone SI, Rose RA. Social Work Research and Endogeneity Bias. Journal of the Society for Social Work and Research 2011 Jan;2(2):54-75. [doi: 10.5243/jsswr.2011.3]
53. Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Boston, MA: Houghton Mifflin Company; 2002.
54. Byers P. Increased poison control calls due to ivermectin ingestion and potential toxicity. Mississippi State Department of Health. 2021 Aug 20. URL: https://msdh.ms.gov/msdhsite/_static/resources/15400.pdf [accessed 2021-10-09]
55. Kelly MP, Barker M. Why is changing health-related behaviour so difficult? Public Health 2016 Jul;136:109-116 [FREE Full text] [doi: 10.1016/j.puhe.2016.03.030] [Medline: 27184821]

Abbreviations

AIC: Akaike information criterion
AOR: adjusted odds ratio
BIC: Bayesian information criterion
CDC: Centers for Disease Control and Prevention
LMM: linear mixed model
LMR: Vuong-Lo-Mendel-Rubin Likelihood Ratio Test
NPB: nonpharmaceutical preventive behaviors
OR: odds ratio

Edited by G Eysenbach; submitted 27.07.21; peer-reviewed by S Boon-Itt, Z Jin, S Gordon; comments to author 18.08.21; revised version received 07.09.21; accepted 26.09.21; published 14.10.21.

Please cite as:
Agley J, Xiao Y, Thompson EE, Chen X, Golzarri-Arroyo L
Intervening on Trust in Science to Reduce Belief in COVID-19 Misinformation and Increase COVID-19 Preventive Behavioral Intentions: Randomized Controlled Trial
J Med Internet Res 2021;23(10):e32425
URL: https://www.jmir.org/2021/10/e32425
doi:10.2196/32425
PMID:34581678

Review

Ethics, Integrity, and Retributions of Digital Detection Surveillance Systems for Infectious Diseases: Systematic Literature Review

Ivy Y Zhao[1*], PhD; Ye Xuan Ma[2*], MA; Man Wai Cecilia Yu[2], MBBS; Jia Liu[3], PhD; Wei Nan Dong[2], PhD; Qin Pang[4], MSc; Xiao Qin Lu[5], MM; Alex Molassiotis[1], PhD; Eleanor Holroyd[6], PhD; Chi Wai William Wong[2,7], PhD

[1]WHO Collaborating Centre for Community Health Services, School of Nursing, The Hong Kong Polytechnic University, Hong Kong SAR, China

[2]Department of Family Medicine and Primary Care, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China

[3]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

[4]Department of Information Technology, University of Hong Kong-Shenzhen Hospital, Shenzhen, China

[5]School of General Practice and Continuing Education, Capital Medical University, Beijing, China

[6]School of Clinical Sciences, Auckland University of Technology, Auckland, New Zealand

[7]Department of Family Medicine and Primary Care, University of Hong Kong-Shenzhen Hospital, Shenzhen, China

[*]these authors contributed equally

Corresponding Author:
Chi Wai William Wong, PhD
Department of Family Medicine and Primary Care
Li Ka Shing Faculty of Medicine
The University of Hong Kong
3/F, Ap Lei Chau Clinic
161 Main Street, Ap Lei Chau
Hong Kong SAR
China
Phone: 86 2518 5657
Email: wongwcw@hku.hk

Abstract

Background: The COVID-19 pandemic has increased the importance of the deployment of digital detection surveillance systems to support early warning and monitoring of infectious diseases. These opportunities create a “double-edge sword,” as the ethical governance of such approaches often lags behind technological achievements.

Objective: The aim was to investigate ethical issues identified from utilizing artificial intelligence–augmented surveillance or early warning systems to monitor and detect common or novel infectious disease outbreaks.

Methods: In a number of databases, we searched relevant articles that addressed ethical issues of using artificial intelligence, digital surveillance systems, early warning systems, and/or big data analytics technology for detecting, monitoring, or tracing infectious diseases according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, and further identified and analyzed them with a theoretical framework.

Results: This systematic review identified 29 articles presented in 6 major themes clustered under individual, organizational, and societal levels, including awareness of implementing digital surveillance, digital integrity, trust, privacy and confidentiality, civil rights, and governance. While these measures were understandable during a pandemic, the public had concerns about receiving inadequate information; unclear governance frameworks; and lack of privacy protection, data integrity, and autonomy when utilizing infectious disease digital surveillance. The barriers to engagement could widen existing health care disparities or digital divides by underrepresenting vulnerable and at-risk populations, and patients’ highly sensitive data, such as their movements and contacts, could be exposed to outside sources, impinging significantly upon basic human and civil rights.

Conclusions: Our findings inform ethical considerations for service delivery models for medical practitioners and policymakers involved in the use of digital surveillance for infectious disease spread, and provide a basis for a global governance structure.

Trial Registration: PROSPERO CRD42021259180; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=259180

(J Med Internet Res 2021;23(10):e32328) doi:10.2196/32328

KEYWORDS

artificial intelligence; electronic medical records; ethics; infectious diseases; machine learning

Introduction

In the wake of the global COVID-19 outbreak, there is growing pressure to improve our existing practice in the prevention and ongoing monitoring of emerging infectious diseases and the adoption of targeted interventions for emerging infectious diseases. Current infectious disease surveillance systems in most countries are remarkably similar. Once a case is clinically suspected and confirmed, there are multiple levels of reporting [1]. Then, the accumulated information from local institutions is aggregated, processed, and defined at the population level before actions are subsequently disseminated through the system from a "top-down" approach. The existing process carries an inevitable time lag that can result in both reduced effectiveness for responsive public health interventions [2] and opportunities for doctors and patients to negotiate reporting, which can have catastrophic results, as had been observed in the early COVID-19 outbreak in China [3] and the Ebola outbreak in West Africa [4]. The need for timely data collection or sharing, processing, decision making, and reporting in infectious disease surveillance has been identified as one of the main drivers for introducing artificial intelligence (AI) technology.

With the establishment of electronic health records (EHRs), big data have been acquired, making it possible to build data-intensive infectious disease surveillance or early warning systems, pre-empt emergency response, and strengthen infection prevention and control. Machine learning (ML) technologies, a multiplying form of AI, has shown considerable potential in tracing the source and detecting potential outbreaks or novel infectious diseases using patients' EHRs. By utilizing real-time digital data analysis, a fully automated system could be built to transmit, through extraction, structured data and doctors' medical records in text, while new technologies, such as named entity recognition, would allow extraction of patient-related features from the unstructured text into predefined categories to support future infectious disease monitoring and surveillance [5]. Examples of enhanced timeliness resulting from this approach in COVID-19 case tracing have been reported in China and several other Asian counties [3]. When AI is being employed for infectious disease control (eg, using mobile phone apps to trace COVID-19 cases), potentially infected patients, their close contacts, and, at times, larger communities can be tracked, tested, and, if necessary, quarantined to prevent further outbreaks.

Nonetheless, the advantages of big data and ML in infectious disease control need to be weighed against the considerable ethical and legal concerns pertaining to the protections and privacy of individuals and the public in respect to access, use, and sharing of large data sets of patients' medical records. These kinds of AI interventions raise complex contemporary ethical questions regarding potential misuse of personal information and informed consent that have the potential to infringe on one's human and civil rights. Furthermore, there is the heightened risk of patients' personal information being leaked to social media when they have previously been assured of confidentiality and privacy [6]. Reidentification of named patients is a major concern when databases are hacked [7], and data custodians may sell data for financial gain to pharmaceutical, insurance, or software companies [8]. Furthermore, the routine use of big data analytics (BDA) or the ethics and widespread moral implications of ML continue to be vigorously debated around the accuracy of reporting [9,10], and the consequences of inaccuracies in the reporting of outbreaks are gaining considerable attention [11]. These developments, often in rapid response situations, have sparked issues between optimizing population health outcomes informed by epidemiology and public health, and societal or individual ethical rights and protections that inform human rights and freedom of choice.

Notwithstanding these debates, BDA is critical for managing communicable disease spread or outbreaks in today's digital world where consistency in the application of regulations or rules of privacy, confidentiality, transparency, data handling, and security safeguards for ML transmission required to protect individuals' rights is limited [12]. This study set out to examine the ethical issues of using AI and identify an intersection or balance between the protection of individuals' human rights and patients' autonomy, and common good for population-based public health outcomes specific to infectious disease control and prevention with the aim of investigating ethical issues identified from utilizing AI-augmented surveillance or early warning systems to monitor and detect common or novel infectious disease outbreaks.

Methods

Search Strategy and Selection Criteria

We searched broad search engines with no time restriction. CINAHL, PubMed, Science Direct, MEDLINE, Google Scholar, and Scopus, as well as legal and sociological databases were searched from database inception to December 8, 2020. Additionally, relevant articles from the documents' bibliography and from other articles that cited the documents were retrieved. The following search terms and combination of terms were utilized: ethic* OR "data security" OR "data privacy" OR sensitiv* OR confidential* OR anonym* OR "personally identifiable information" OR privacy or "human right*" AND "electronic health records" OR ehr OR "Clinical decision support system" OR cdss OR "Artificial Intelligence" OR ai OR "augmented surveillance" OR surveillance OR "contact tracing" OR "Machine Learning" OR ml OR "deep learning" AND "Infectious Disease" OR "Communicable Diseases." Prior to the searches, the authors reviewed all search terms. In order to identify all existing literature for this review, we considered all peer-reviewed empirical research articles, review reports, and grey reports. To enhance the rigor of the review, our approach followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Multimedia Appendix 1) [13].

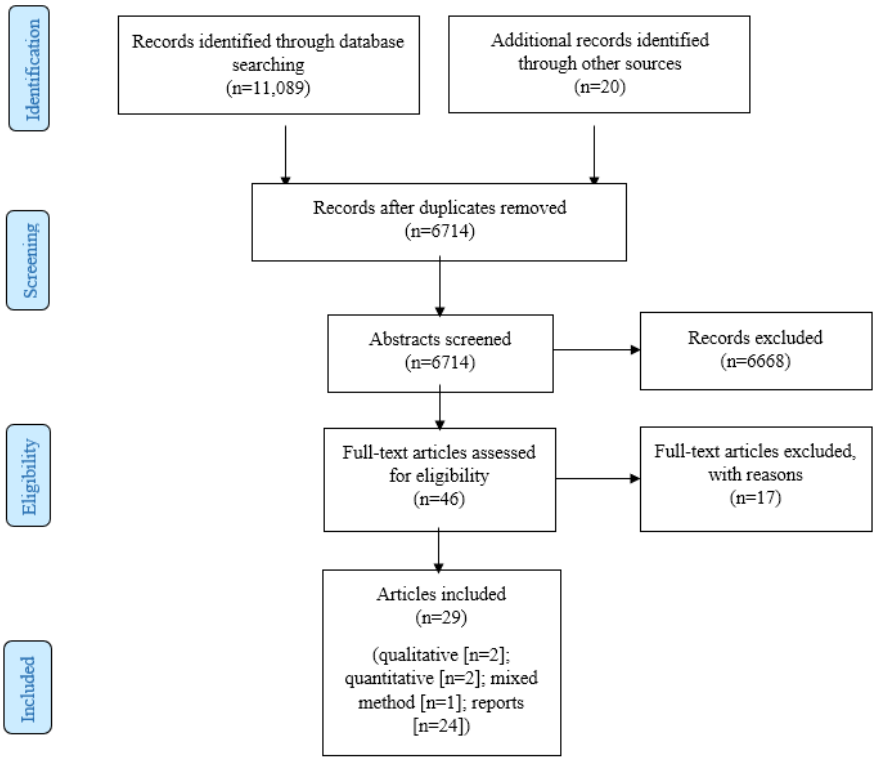
Inclusion and Exclusion Criteria

There were no language restrictions placed on the literature search. To be included for further review, the collected articles must have addressed the ethical issues of using AI, digital surveillance systems, early warning systems, and/or BDA technology for detecting, monitoring, or tracing specifically infectious diseases.

Literature Selection

We selected literature for inclusion in 2 stages to ensure rigor. In the first stage, 2 authors (IYZ and YXM) independently screened the titles and abstracts of all citations for potentially relevant articles. In the second stage, the same 2 authors independently examined the full texts of these papers against prespecified inclusion criteria. Any discrepancies were resolved with input from a third author (EH). Of the 6714 titles and abstracts reviewed, we excluded 6668 articles that did not meet our eligibility criteria, resulting in 46 articles for full-text review. Additionally, 17 articles were excluded because they (1) were not related to digital surveillance systems, (2) did not focus on infectious diseases, or (3) lacked focus on ethical concerns. In all, 29 papers satisfied our eligibility criteria and were included by consensus agreement (Figure 1).

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



Data Extraction, Quality Assessment, and Data Analysis

All authors read and reviewed the 29 articles in order to summarize the approaches, methodologies, samples, and findings. Two authors (IYZ and YXM) undertook literature quality assessment, extracted data from the literature, included them into a spreadsheet, and analyzed the data independently. Discrepancies were resolved, and data were confirmed in several rounds of discussions with other team members. We evaluated study quality and methodological rigor for 5 empirical studies by using the modified mixed-methods appraisal tool (MMAT) [14], which was not applicable for the remaining discussion papers. Lower quality scores did not result in exclusion of any articles. However, the findings of articles with lower quality scores were given less weight during data analysis.

We structured our identified ethical issues using a theoretical framework developed by Asadi et al [15], in which key concepts of prior BDA had been identified, defined, and examined using stakeholder theory and discourse ethics [16]. Key themes from our review were discussed and summarized, and the gaps in the literature and methodologies were identified.

Results

Overview and Participant Demographics

Among the 29 included articles, 2 qualitative studies [17,18], 2 quantitative studies [19,20], 1 mixed methods study [21], and 24 discussion papers were identified (Table 1). All of the selected articles were published between the years 2015 and 2021. The 5 empirical studies [17-21] adopted a combination of focus groups, a Delphi approach, a database digital ethnography, and surveys as research methods, and they were all identified as high quality. Together they included 7331 participants. Of these 5 studies, 1 was conducted in South Korea [21] and 4 in Australia [17,19-21]. Four studies [17-19,21] reported gender ratios and 3 [17,19,21] demonstrated age ranges. Research participants were the general public; policymakers; and experts in infectious diseases, epidemiology, food safety, health informatics systems, and health and technology law.

Table 1. Summary of ethical issues identified from utilizing digital surveillance systems for infectious diseases.

Article type, and authors and year	Country	Journal	1. Sampling frame 2. Sample size 3. Age 4. Gender ratio (F:M) 5. Sampling method	Study aim	Context of studies	Methodology and methods	Key and relevant findings
Journal articles (n=5)							
Degeling et al (2020) [17]	Australia	BMC Medical Ethics	1. People who had previously volunteered to take part in research and topic-blinded social media advertising on Facebook 2. n=50 3. 18-34 y (n=15); 35-54 y (n=22); and >55 y (n=11) 4. 27:21 5. Random sampling	Examine the public acceptability and ethical concerns of community juries on integration of big data analytics (BDA) into communicable disease control	BDA was perceived as intrusive and a threat to privacy.	Qualitative; deliberative group sessions; Delphi study approach	Almost all jurors supported data linkage for public health research and suggested deidentification practices. Three juries raised several conditions related to system oversight and security being met. One concern was about loss of privacy and mistrust in governments to run secure and effective systems.
Kim et al (2021) [18]	Korea	Social Science & Medicine	1. Comments from January to May 2020 made by Korean mothers on 15 internet groups called “mom cafes” 2. n=3729 3. N/A[a] 4. Female 5. Purposive sampling	Examine how Korean mothers understand morality in the context of COVID-19 contact-tracing surveillance	Korean mothers uploaded COVID-19 patient information on the boards of online groups for discussion.	Qualitative; database digital ethnography; reflexive thematic analysis	Nonmaleficence is the core morality considered by Korean mothers.
Degeling et al (2020) [19]	Australia	BMJ Open	1. The Australian general population 2. n=2008 3. Median 46 y (18-89 y) 4. 1015:993 5. Purposive sampling	Compare the value of core surveillance system attributes to the Australian public before and during the early stages of the COVID-19 pandemic	New technology raised concerns of privacy disclosure and misuse in the COVID-19 outbreak.	Quantitative; online survey	After the COVID-19 pandemic, participants demonstrated greater preference for a high data security surveillance system for public health.

Article type, and authors and year	Country	Journal	1. Sampling frame 2. Sample size 3. Age 4. Gender ratio (F:M) 5. Sampling method	Study aim	Context of studies	Methodology and methods	Key and relevant findings
Thomas et al (2020) [20]	Australia	JMIR Public Health Surveillance	1. Australians (excluding health care professionals or people who had been tested for COVID-19) 2. n=1500 3. ≥18 y 4. 1:1 5. Purposive sampling	Investigate the ethical issues of adopting the Australian government's COVID safe app	App-based contact tracing for curbing the transmission of COVID-19 needs widespread adoption.	Quantitative; online national survey	Privacy, data storage, and technical functions are ethical issues that hinder contact-tracking apps.
Degeling et al (2019) [21]	Australia	Health Research Policy and Systems	1. Australian-based policy makers and experts in infectious diseases, epidemiology, food safety, health informatics systems, and health and technology law 2. n=44 3. N/A 4. N/A 5. Purposive sampling	Identify ethical issues in the adoption and effective implementation of a digital surveillance tool	Early detection of infectious disease outbreaks involves lack of social license or ethical and legal considerations.	Mixed method; online survey; framework analysis	Infectious disease monitoring systems raise issues such as personal privacy, forensic risks, potential unintended consequences, and the weakening of public trust.
Reports (n=24)							
Sweeney (2020) [22]	United Kingdom	Nature Machine Intelligence	N/A	Balance protecting public health with safeguarding civil rights regarding contact-tracing apps	Contact-tracing apps were used in COVID-19 surveillance but were less understood by people.	Discussion paper	Contact-tracing apps debate on protecting public health with safeguarding civil rights.
Gilbert et al (2019) [6]	Australia	Asian Bioethics Review	N/A	Highlight the urgency of having an ethical framework to guide the use of new technologies in communicable disease surveillance and control	There is considerable public opposition to allowing public health authorities access to personal health data for infectious disease surveillance.	Discussion paper	Informed public discussion, greater transparency, and an ethical framework will be essential to build public trust in the use of new technology for communicable disease control.

Article type, and authors and year	Country	Journal	1. Sampling frame 2. Sample size 3. Age 4. Gender ratio (F:M) 5. Sampling method	Study aim	Context of studies	Methodology and methods	Key and relevant findings
Schwalbe et al (2020) [23]	United States	Lancet	N/A	Artificial intelligence (AI) use in low- and middle-income countries	AI-driven intervention research in global health has less addressed ethical, regulatory, or practical considerations.	Discussion paper	Addressing privacy and security in digital development involves careful consideration of which data are collected and how data are acquired, used, stored, and shared.
Garattini et al (2019) [2]	United Kingdom	Philosophy & Technology	N/A	Provide a moral foundation for the societal acceptance and responsible development of technological advancement	There are many ethical impacts when applying BDA in infectious diseases.	Discussion paper	Automation and algorithmic reliance impact freedom of choice; BDA complexity impacts informed consent; reliance on profiling impacts individual and group identities and justice/fair access; and increased surveillance and population intervention capabilities impact behavioral norms and practices.
Parker et al (2020) [3]	United Kingdom	Journal of Medical Ethics	N/A	Outline ethical considerations in the deployment of digital surveillance systems for public health response	Mobile phone contact-tracing apps have raised many ethical questions in the COVID-19 pandemic.	Discussion paper	Privacy, liberty, responsibilities, data management, public trust and confidence, equity, fairness, justice, and data consistency need to be addressed in the deployment of mobile phone apps.
Katapally (2020) [24]	Canada	Journal of Medical Internet Research	N/A	Outline an evidence-based global digital citizen science policy, which provides a theoretical and methodological basis for ethically sourcing big data from citizens to tackle pandemics such as COVID-19	A cohesive societal effort with citizens' full support is needed in pandemics.	Discussion paper	One of the biggest ethical challenges is data privacy and security. Individuals' rights to privacy and anonymity through advanced encryption and secure server storage processes, informed consent, the ability to dropout and delete their own data, and data co-ownership, should be priorities.

Article type, and authors and year	Country	Journal	1. Sampling frame 2. Sample size 3. Age 4. Gender ratio (F:M) 5. Sampling method	Study aim	Context of studies	Methodology and methods	Key and relevant findings
Mbunge (2020) [25]	Eswatini	Diabetes & Metabolic Syndrome: Clinical Research & Reviews	N/A	Analyze the potential opportunities and challenges of integrating emerging technologies, including 5G technology, AI, and big data, into COVID-19 contact tracking	Contact-tracing technologies have limitations when used in the COVID-19 pandemic.	Literature review	Ethical or legal challenges might be socioeconomic inequalities in developing counties; security risks such as data security, confidentiality, integrity, and data availability of patients and contacts in COVID-19; the privacy issues of patients, which may lead to mental health problems; consent and voluntariness; and discrimination.
Garg et al (2020) [26]	India	JMIR Public Health and Surveillance	N/A	Describe Aarogya Setu, a first-of-its-kind participatory disease surveillance initiative in India and its ethical considerations	Opt-in, data integrity, and ethical concerns need to be addressed when using the new system for the COVID-19 pandemic.	Case report	The main ethical dilemma is how to ensure data protection and proper ethics while obtaining the benefits of public health surveillance, and how to ensure the ethical use of collected data and protect individual privacy.
Denecke (2017) [27]	Switzerland	Life Sciences, Society and Policy	N/A	Highlight the ethical issues that should be considered when integrating digital epidemiology with current practice and develop an ethical assessment model for digital disease detection (DDD) technologies	Usage of digital surveillance in epidemiology has different kinds of challenges.	Discussion paper	The model developed in this study might help to make aware the ethical aspects already in the development process, and possibly address them.

Article type, and authors and year	Country	Journal	1. Sampling frame 2. Sample size 3. Age 4. Gender ratio (F:M) 5. Sampling method	Study aim	Context of studies	Methodology and methods	Key and relevant findings
Kostkova (2018) [28]	United Kingdom	Life Sciences, Society and Policy	N/A	Outline 3 major ethical and governance challenges for digital epidemiology in the 21st century	Digital surveillance has created ethical, political, and legal challenges in infectious disease control.	Commentary report	Some of the ethical challenges of sharing data across various early warning tools to support risk assessment are ownership of personal data, transparency and clarity of public health data sharing, strong transparent disclosure, data privacy and security, and the balance between data sharing, personal data protection, stakeholder needs, and public good.
Vayena et al (2015) [29]	Switzerland	PLoS Computational Biology	N/A	Identify key ethical challenges associated with DDD activities and outline a framework for addressing them	DDD has many ethical challenges in infectious disease pandemics.	Discussion paper	The ethical challenges of DDD can be divided under 3 heads: context sensitivity (privacy and contextual integrity, transparency, and global justice); nexus of ethics and methodology (risk of harm, use of resources, trust, transparency, accountability); and legitimacy requirements (shared code of practice, mechanism for quick response to inaccuracies, addressing harms caused by DDD activities, common good).

Article type, and authors and year	Country	Journal	1. Sampling frame 2. Sample size 3. Age 4. Gender ratio (F:M) 5. Sampling method	Study aim	Context of studies	Methodology and methods	Key and relevant findings
De Jong et al (2019) [30]	Belgium	Emerging Infectious Diseases	N/A	Mitigate the ethical concerns of movement mapping of potentially infected persons	Mapping the movements of potentially infected persons has ethical challenges.	Discussion paper	Ethical obstacles are privacy in relation to the principles of autonomy and nonmaleficence; and a balance between costs, risks, and benefits for participants and communities in relation to the principles of beneficence and justice, such as stigmatized community, data withholding, and whether and how to communicate information on hotspots to the general population.
Kind (2020) [31]	United Kingdom	Patterns	N/A	Examine societal, political, legal and ethical perspectives on symptom tracking, contact tracing, and immunity	The UK government asked for more information to decide the use of technology in the COVID-19 pandemic.	Rapid evidence review	The ethical issues of digital contact tracing are human rights and data protection, inequalities, data quality limitations, false reporting risks, and centralization of large amounts of personal data.
Park et al (2020) [32]	Korea	JAMA	N/A	Identify ethical concerns over privacy involving the information technology–based tracing strategy in response to COVID-19	South Korea extensively used digital tools for tracing COVID-19 patients.	Discussion paper	Privacy controversies might unveil or infer embarrassing personal details, unwanted privacy invasion, public disdain, uneven scope, and granularity of disclosures by municipal and local governments.
Fraser et al (2020) [33]	United Kingdom	University of Oxford	N/A	Minimize the invasion of privacy by using digital contact tracing	Contact tracing was used to assist people in receiving warnings about COVID-19.	Discussion paper	The ethical issues include sensitively and specifically identifying infectious individuals, user uptake and adherence, notification, integration with local health policy, and ability to evaluate effectiveness transparently.

Article type, and authors and year	Country	Journal	1. Sampling frame 2. Sample size 3. Age 4. Gender ratio (F:M) 5. Sampling method	Study aim	Context of studies	Methodology and methods	Key and relevant findings
Cho et al (2020) [34]	Singapore	ArXiv Preprint	N/A	Discuss ways of ameliorating privacy concerns without decreasing the usefulness of contact-tracing apps	The Singaporean government released a mobile phone app to assist in tracking down exposures to COVID-19 patients, but there were privacy implications.	Discussion paper	Privacy is a central feature of conversations around mobile contact-tracking apps. Some privacy trade-offs can be endured for public health.
Klenk et al (2020) [35]	Netherlands	Ethics and Information Technology	N/A	Identify factors that pose a risk for fair group composition	Digital tracing technologies for COVID-19 control were reported to have ethical risks.	Discussion paper	Digital tracking apps will introduce new psychological, social, economic, and political risks.
De Montjoye et al (2020) [36]	United Kingdom	Computational Privacy Group Blog	N/A	Propose 8 questions to assess privacy in contact-tracing apps.	A contact-tracing app was developed to assist with COVID-19 control, and record location or close contact data.	Discussion paper	Privacy protection should rely on mathematical proof, and mitigation strategies should be considered only when necessary. We should focus on privacy and ensure security.
Bernier et al (2015) [37]	Canada	University of New Brunswick Law Journal	N/A	Highlight the personal privacy in electronic public health surveillance systems	Data surveillance has become a key component of pandemic response plans.	Discussion paper	The privacy governance framework is incomplete in ensuring the effective and protective use of personal information in response to epidemics.
Ienca et al (2020) [38]	Switzerland	Nature Medicine	N/A	Identify ethical issues when using digital surveillance systems in COVID-19	The COVID-19 emergency has used much more digital tools than previous outbreaks globally.	Discussion paper	Best practices should be identified to protect privacy and public trust.
Yasaka et al (2020) [39]	United States	JMIR mHealth and uHealth	N/A	Develop an effective contact-tracing smartphone app that respects user privacy by not collecting location information or other personal data	Smartphone-based contact tracing has been used in the COVID-19 pandemic to limit disease transmission.	Discussion paper	Users may be uncomfortable with applications that track real-time locations.

Article type, and authors and year	Country	Journal	1. Sampling frame 2. Sample size 3. Age 4. Gender ratio (F:M) 5. Sampling method	Study aim	Context of studies	Methodology and methods	Key and relevant findings
Barbieri et al (2020) [40]	Italy	Istituto Affari Internazionali	N/A	Discuss the ethics of technological solutions to mitigate COVID-19	Technological solutions to mitigate the COVID-19 crisis have been implemented in China and South Korea.	Discussion paper	In a pandemic crisis, the balance between privacy and public health tends to tilt toward the latter. However, a strong legal framework should be established around any such data-driven policy, taking into account the transition to "postepidemic" life.
Chan et al (2020) [41]	United States	arXiv	N/A	Improve the privacy and anonymity standards of mobile contact tracing	The COVID-19 pandemic has been controlled by large-scale adoption of contact tracing.	Discussion paper	Ethical issues of privacy protection, transparency, and reidentification risks of anonymous information.
Peter (2020) [42]	Australia	The Guardian	N/A	Discuss the acceptability of a coronavirus tracing app by Australians and how to implement tracing technology successfully	The Australian public is seeking a way to manage the COVID-19 pandemic.	Discussion paper	The ultimate success of tracking technology will depend on confidence and mutual respect. The ultimate test of any tracking technology will be the strength of the relationship between the public and the government.

[a]N/A: not applicable.

Ethical Issues

Textbox 1 shows the key ethical issues identified in the review summarized and clustered under individual, organizational, and societal levels with key themes presented under the corresponding concepts. Six domains, namely awareness of implementing digital infectious disease surveillance, digital integrity, trust, privacy and confidentiality, civil rights, and governance were highlighted.

Textbox 1. Ethical issues in utilizing artificial intelligence–augmented infectious disease surveillance systems based on the ethical framework of Asadi et al [15].

Individual level

Data ownership

- Digital infectious disease surveillance systems challenge data ownerships rights
- Impacts of data ownership rights on public participation in digital infectious disease surveillance

Data control

- Inappropriate reidentification, sharing, or processing of personal information of infectious disease patients

Awareness

- Lack of understanding of data collection, access, processing, sharing, and storage of infectious disease surveillance systems
- Lack of and inability to give consent when enrolled into infectious disease surveillance systems

Trust

- Data governance, security, and data set bias undermines public trust
- Public mistrust in the necessity and effectiveness of contact-tracing technology for infectious diseases
- Public mistrust in governments' strategies in using digital infectious disease surveillance

Privacy

- Concerns of privacy risks and allowance to infectious disease control
- Privacy risks on contact tracing for infectious diseases and social networks
- Disclosure of infectious disease information leads to business depletion, privacy invasion, and public disdain
- Necessity of infectious disease data anonymization and risk of reidentification
- Privacy and appropriate authority oversight on artificial intelligence (AI)-augmented infectious disease surveillance systems

Self-determination

- Autonomy/personal liberty to participate in and use AI-augmented infectious disease surveillance systems

Fear

- Fears of privacy violation, institutional control/penalties, and discriminatory/stigmatized effects when using AI-augmented infectious disease surveillance systems

Organizational level

Data quality

- Data accuracy, validity, veracity, and integrity of AI-augmented infectious disease surveillance systems
- Consequences of low data quality in AI-augmented infectious disease surveillance systems

Data sourcing

- Vulnerable populations/key demographics underrepresented in data sourcing could lead to invalid infectious disease control

Data sharing/disclosure

- Data sharing/disclosure for unethical purpose when using AI-augmented infectious disease surveillance systems
- Data disclosure needs clear standards and safeguards when using AI-augmented infectious disease surveillance systems

Algorithmic decision making

- Reliability, validity, and consequences of algorithmic decision making for infectious disease outbreaks
- Vulnerability of machine learning processes for infectious disease surveillance

Presentation

- Needs for transparent and clear presentation of algorithms, data processing, and hotspots of infectious diseases

Ethical capability

- Lack of ethical training on data collection and disciplinary measures on poor data quality for infectious disease surveillance

Ethical culture

- Consideration of cultural contexts in data collection, processing, and decision making for infectious disease surveillance

Ethical governance

- Transparency of digital infectious disease surveillance systems
- Lack of common technical or ethical standards for data usage in infectious disease surveillance
- Lack of an ethical governance framework to regulate algorithms, data collection, use, and management for infectious disease surveillance

Societal level

Power

- Imbalanced power relations among decision makers, researchers, and citizens when using AI-augmented infectious disease surveillance systems
- Unequal allocation and distribution of resources and benefits when using AI-augmented infectious disease surveillance systems

Social awareness

- Social awareness in purpose, risks and benefits, and consequences when setting up infectious disease surveillance programs

Surveillance

- Ethical surveillance with full public engagement and incentives when using AI-augmented infectious disease surveillance systems
- Ensure infectious disease surveillance systems protect civil liberties and rights

Principles and guidelines

- Lack of legislation and guidelines for infectious disease outbreaks and protection of data security, individual privacy, and discrimination

Authority

- The possibility of data misuse by authorities/agencies without legal authorization

Climate

- Digital infectious disease surveillance in a social environment leads to social stigma, discrimination, rumors, and prejudice

Awareness of Implementing Digital Infectious Disease Surveillance

The informants reported insufficient understanding at every stage of data collection and distribution of digital infectious disease surveillance systems. They felt they were not informed of their rights to refuse or their ability to withdraw consent. Irresponsible decision making was identified by infectious disease patients in relation to insufficient information provided about contact tracing technology. At the societal level, digital infectious disease surveillance based on BDA was poorly accepted by the general public due to their uncertainty about its purpose and the risks posed from the potential mitigation of data sharing. Consequently, they wanted more information about the digital infectious disease surveillance systems and wanted their associated ethical concerns and consequences addressed. They also expected organizations to make public and transparent the algorithms and data processes used, and use plain language when explaining infectious disease surveillance. Simultaneously, both the general public and field experts emphasized that governments or institutions should convey the importance of infectious disease outbreak control to communities without violating ethical principles.

Data Integrity

Data integrity weaknesses were identified as common in BDA. Unreliable or invalidated data sourcing or algorithms were seen to lead to inaccurate identification of outbreaks or infected individuals, false predication of an event's trajectory or the likelihood of reoccurrence, and inaccurate notifications. The outcomes, such as inadequate data integrity, were further seen to continuously intensify economic losses to trade, tourism, and health services, causing unnecessary panic and the loss of public trust in health authorities. Some organizations and experts also worried that digital infectious disease surveillance systems would widen existing health care disparities or digital divides by underrepresenting vulnerable and at-risk populations such as older adults, children, and people in economically underresourced areas; for instance, data of COVID-19 hotspots influenced the allocation and distribution of resources [25]. It was recommended that digital infectious disease surveillance applications should consider ethical requirements and the rights of people from diverse regions and communities.

Trust

Some people questioned the necessity and effectiveness of using contact-tracing technologies in an AI-assumed situation. They mistrusted the AI-augmented systems' capacity to send correct notifications to infected individuals to instruct them to quarantine during an infectious disease outbreak in a timely manner. Emerging technology risks, such as data breaches or data set biases and government strategies of mandatory application of digital contact tracing, could further undermine public trust. Individuals and relevant experts called for an open debate or scrutiny, transparent procedures for data usage, and public consultation plans and privacy regulations.

Privacy and Confidentiality

Digital infectious disease surveillance was identified to pose considerable risks to an individual's rights to privacy and confidentiality. Contact tracing that linked a potential infectious disease with patients' movements, locations, or social networks was seen as a considerable threat to further disclosure of sensitive information. For example, personal social interactions and contact history, especially in the case of sexually transmitted infections or HIV, could be revealed. Individuals were concerned that third parties or malicious users might access large health data sets for profit and/or abuse. Disclosure of private information was seen to lead to business depletion, privacy invasion, and public discrimination and stigmatization. Data anonymity, robust encryption processes, and deidentified aggregate data were contended to be crucial to all data privacy and security procedures in compliance with data protection regulations.

Civil Rights

Data ownership rights for public health surveillance were regarded as a major ethical challenge. Some reported that authorities or institutions limited their rights to decide the adoption of digital infectious disease contact tracing or surveillance systems. Personal liberties were also impacted by movement mapping, cross-border sharing of personal health information, and frequent security checks using QR codes for filling in personal information. Countries using centralized contact-tracing apps and privacy-by-design apps were seen to have the potential to impose restrictions on civil liberties, which, in turn, impacted ethical engagement in digital health. Data co-ownership and strengthening of transparency were seen as helpful to encourage individuals to participate in data visualization, analysis, and knowledge translation, and balance the power dynamics among decision makers, researchers, and citizens.

Governance

Technical and ethical standards, as well as legislation and guidelines for infectious disease outbreaks, data security, protection of individual privacy, and avoidance of discrimination were considered poorly developed and incomplete. Technology companies were recommended to establish a mechanism to deal with inaccurate epidemic reporting and dissemination of misinformation, and were expected to develop rapid ethical assessment, training, and disciplinary measures for data collection or sharing. Governance institutions or bodies, such as national health commissions, medical councils, and company boards, were further recommended to provide appropriate oversight on the performance of algorithms and data usage. Some literature argued that an independent privacy audit was needed to secure a transparent approach for the public [33,36]. Both individuals and relevant experts also supported data sharing protocols to specify the scope and granularity of disclosure. An example provided was that personal names should not be publicly available, and personal information should only be gathered and shared within a period of time regulated by law and clearly justified based on population health needs.

Discussion

Challenges of Using AI Surveillance Systems for Infectious Diseases

The emergence of AI surveillance systems for infectious diseases promises tangible global public health benefits, but these are accompanied by significant ethical, political, and legal challenges, which span over a wide spectrum on 3 levels. Six main themes were generated from this systematic review, ranging from people's awareness and knowledge of digital infectious disease surveillance systems and personal privacy on the individual level to organizational issues of maintaining data integrity and security, and the lens was extended upward to the societal level, involving public trust, civil rights, and the need for a governance framework with ethical oversights. Some of the challenges are inherent to public health practice and only heightened by the use of digital tools, and others, such as public trust and awareness of digital surveillance, algorithmic decision making, and data security, are specific to AI approaches and largely unprecedented. It is vital to consider these challenges to enhance individuals' rights, privacy, public responsibilities, and optimal population health outcomes so that digital surveillance can tackle pandemics ethically.

Strengths and Limitations

Studies and position papers on the ethical implications of AI surveillance for infectious diseases have gained momentum since 2019 in response to the outbreak of the COVID-19 pandemic; the dearth of studies prior to this makes it challenging to reveal any time trends. Furthermore, most articles included in the review were discussion papers, limiting the ability to evaluate generalizability, transferability, and rigor or multiple stakeholders' perspectives on the use of AI-augmented infectious disease surveillance. The 5 empirical studies had restricted sampling approaches both in terms of size and stakeholder representation (health professionals and community groups), with 4 of the studies being conducted in Australia alone. Few papers involved vulnerable or diverse populations, and none involved specific cultural or socioeconomic groups, further limiting the scope of the review. Furthermore, articles acknowledged the inherent strengths and weaknesses of the public health system operating at the location of the study.

Retributions Versus Common Good for Public Health and Long-term Impacts

The tensions between human and civil right discourses and the need for rapid public health responses are exacerbated with the

use of BDA/AI in the context of a pandemic [2]. There are medicolegal and moral retributions and concerns arising from utilizing available epidemiological information through a highly effective and responsive infectious disease surveillance method that could protect local and, potentially, global communities from serious infectious disease outbreaks [3,18]. Conversely, routine collection and linkage of detailed personal information pose considerable risks for the violation of individual and civil freedoms of choice and privacy [19].

As of July 21, 2021, there have been 191 million confirmed cases and 4.11 million deaths globally from COVID-19 [43]. Given that sound public health interventions are predicated upon promoting and protecting the health of communities, timely, cost-effective, and socioculturally informed primary care interventions, advocacy, and empowerment with long-term impact evaluations are required. There are strong reasons to develop an ethical governance framework to support AI-augmented infectious disease surveillance to achieve these outcomes. The question of how to balance what is needed for the "good of public health outcomes" and human rights in pandemic crisis situations is highlighted in the literature reviewed [33,37].

Trust Building and Privacy Protection

Detection and notification of infectious disease outbreaks requires prompt accurate disease diagnosis and follow-up of infected individuals and their close contacts. An early AI-augmented warning system has exponential potential for implementing real-time, responsive, and adaptive calculations. This means that substantial personal information, such as names, ages, locations, and relevant heath data, will be accessed in a timely manner and collected by AI systems for calculation, analysis, and notification. Public opt-in and well-founded trust in the digital system, its implementation, and the governance framework are therefore essential factors regarding ethical issues at the organizational and societal levels for infectious disease surveillance, which could enable the acceptability and effectiveness of the system.

Even with optimal technical standards to maximally reduce the risks and consequences of data misuse, data safety, security, and integrity cannot be guaranteed. The role of AI-augmented infectious disease surveillance might only be supplementary for public health, while ethical issues are carefully observed. One of the most important challenges facing AI is to design and develop appropriate methods to deidentify personal information and protect privacy, yet a greater risk of false positive and false negative notifications exists with higher-level deidentified data [33].

Moreover, well-founded trust and confidence vary in different countries and between individuals. People in democratic countries tended to distrust AI-augmented surveillance systems by challenging them when personal information is reported to health authorities, often without appropriate informed consent [3]. In contrast, Chinese citizens expressed their trust in the Chinese government's response to COVID-19, which they felt had been highly successful in controlling the spread of the virus through the use of mobile phone data combined with intensive testing and restriction programs [44]. Although there is no unified standard to establish trust, the need for effective, transparent, accountable, and independent oversight is very important.

Generation of an Ethical Framework and Global Governance Structures

Digital standards and guidelines for developing and evaluating the performance of infectious disease surveillance alone are insufficient. The scope of the ethical framework in Textbox 1 needs to be expanded globally. Cross-national and national governance structures; institutional systems with regulatory, medical, ethical, and legal frameworks; and benchmarking standards have essential roles to play in the development and deployment of these new health technology systems. However, this review identified the urgent need for an ethical framework to underscore all AI-augmented infectious disease surveillance systems [18,19,30]. Considering the rapid development of global trade supply chains, mass gatherings, and international travel, World Health Organization's International Health Regulations in 2005 [45] outlined the cross-border implications of a pandemic response and provided a framework for sharing, monitoring, and evaluating information from the sources of infections [37]. In response to a pandemic emergency, cross-border sharing of personal health data is essential for tracing infectious disease patients and their contacts. At the same time, cross-border sharing further raises the heightened and unique risk for individual privacy and security breaches. In turn, a spectrum of actions regarding ethical, political, and legal implications must be framed within strict safeguards and needs to be mandated globally [37].

Implications

A systematic medical, ethical, and legal framework is necessary for governance of AI-augmented infectious disease surveillance and the protection of personal privacy and data integrity. Public health systems should maximally increase the social awareness of AI surveillance and BDA for infectious diseases, and implement new technologies for infectious disease surveillance in a more person-centered and humane manner. Future research needs to focus on the setting up and implementation of an AI-augmented infectious disease surveillance system underscored by an ethical framework based on universal human rights. Decision makers should take into account varying and diverse population needs, sociocultural status, and regulatory and legal governance in order to promote trust building between end users, including infectious disease patients, doctors, and AI system implementers. This systematic review is intended to contribute to the development of a more comprehensive and concrete ethical framework for AI-augmented infectious disease surveillance, which will enable it to ultimately maximize public health responsiveness synergized within an ethical context.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 31 KB - jmir_v23i10e32328_app1.docx]

References

1. Vlieg WL, Fanoy EB, van Asten L, Liu X, Yang J, Pilot E, et al. Comparing national infectious disease surveillance systems: China and the Netherlands. BMC Public Health 2017 May 08;17(1):415 [FREE Full text] [doi: 10.1186/s12889-017-4319-3] [Medline: 28482830]
2. Garattini C, Raffle J, Aisyah DN, Sartain F, Kozlakidis Z. Big data analytics, infectious diseases and associated ethical impacts. Philos Technol 2019;32(1):69-85 [FREE Full text] [doi: 10.1007/s13347-017-0278-y] [Medline: 31024785]
3. Parker MJ, Fraser C, Abeler-Dörner L, Bonsall D. Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. J Med Ethics 2020 Jul 04;46(7):427-431 [FREE Full text] [doi: 10.1136/medethics-2020-106314] [Medline: 32366705]
4. McCarthy M. Slow response contributed to scale of west African Ebola epidemic, CDC concludes. BMJ 2016 Jul 08;354:i3814. [doi: 10.1136/bmj.i3814] [Medline: 27401324]
5. Temko A, Marnane W, Boylan G, Lightbody G. Clinical implementation of a neonatal seizure detection algorithm. Decis Support Syst 2015 Mar;70:86-96 [FREE Full text] [doi: 10.1016/j.dss.2014.12.006] [Medline: 25892834]
6. Gilbert GL, Degeling C, Johnson J. Communicable disease surveillance ethics in the age of big data and new technology. Asian Bioeth Rev 2019 Jun 10;11(2):173-187 [FREE Full text] [doi: 10.1007/s41649-019-00087-1] [Medline: 32218872]
7. Singapore personal data hack hits 1.5m, health authority says. BBC. 2018. URL: https://www.bbc.com/news/world-asia-44900507 [accessed 2021-04-30]
8. Tanner A. How Your Medical Data Fuels a Hidden Multi-Billion Dollar Industry. Time. 2017. URL: https://time.com/4588104/medical-data-industry/ [accessed 2021-04-20]
9. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]
10. Vanessa T, Culnane C, Rubinstein B. The simple process of re-identifying patients in public health records. Pursuit. 2017. URL: https://pursuit.unimelb.edu.au/articles/the-simple-process-of-re-identifying-patients-in-public-health-records [accessed 2021-03-30]
11. Floridi L, Taddeo M. What is data ethics? Philos Trans A Math Phys Eng Sci 2016 Dec 28;374(2083):20160360 [FREE Full text] [doi: 10.1098/rsta.2016.0360] [Medline: 28336805]
12. Salerno J, Knoppers BM, Lee LM, Hlaing WM, Goodman KW. Ethics, big data and computing in epidemiology and public health. Ann Epidemiol 2017 May;27(5):297-301. [doi: 10.1016/j.annepidem.2017.05.002] [Medline: 28595734]
13. Whittemore R, Knafl K. The integrative review: updated methodology. J Adv Nurs 2005 Dec;52(5):546-553. [doi: 10.1111/j.1365-2648.2005.03621.x] [Medline: 16268861]
14. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. EFI 2018 Dec 18;34(4):285-291. [doi: 10.3233/efi-180221]
15. Asadi SI, Breidbach CF, Davern M, Shanks G. Ethical implications of big data analytics. 2016 Presented at: European Conference on Information Systems (ECIS); June 12-15, 2016; Istanbul, Turkey URL: https://aisel.aisnet.org/ecis2016_rip/24/
16. Someh I, Davern M, Breidbach CF, Shanks G. Ethical issues in big data analytics: a stakeholder perspective. CAIS 2019:718-747. [doi: 10.17705/1cais.04434]
17. Degeling C, Carter SM, van Oijen AM, McAnulty J, Sintchenko V, Braunack-Mayer A, et al. Community perspectives on the benefits and risks of technologically enhanced communicable disease surveillance systems: a report on four community juries. BMC Med Ethics 2020 Apr 25;21(1):31 [FREE Full text] [doi: 10.1186/s12910-020-00474-6] [Medline: 32334597]
18. Kim E, Chung J. Korean mothers' morality in the wake of COVID-19 contact-tracing surveillance. Soc Sci Med 2021 Feb;270:113673 [FREE Full text] [doi: 10.1016/j.socscimed.2021.113673] [Medline: 33453628]
19. Degeling C, Chen G, Gilbert GL, Brookes V, Thai T, Wilson A, et al. Changes in public preferences for technologically enhanced surveillance following the COVID-19 pandemic: a discrete choice experiment. BMJ Open 2020 Nov 18;10(11):e041592 [FREE Full text] [doi: 10.1136/bmjopen-2020-041592] [Medline: 33208337]
20. Thomas R, Michaleff ZA, Greenwood H, Abukmail E, Glasziou P. Concerns and misconceptions about the Australian government's COVIDSafe app: cross-sectional survey study. JMIR Public Health Surveill 2020 Nov 04;6(4):e23081 [FREE Full text] [doi: 10.2196/23081] [Medline: 33048826]
21. Degeling C, Johnson J, Gilbert GL. Perspectives of Australian policy-makers on the potential benefits and risks of technologically enhanced communicable disease surveillance - a modified Delphi survey. Health Res Policy Syst 2019 Apr 04;17(1):35 [FREE Full text] [doi: 10.1186/s12961-019-0440-3] [Medline: 30947721]

22. Sweeney Y. Tracking the debate on COVID-19 surveillance tools. Nat Mach Intell 2020 Jun 16;2(6):301-304. [doi: 10.1038/s42256-020-0194-1]
23. Schwalbe N, Wahl B. Artificial intelligence and the future of global health. Lancet 2020 May 16;395(10236):1579-1586 [FREE Full text] [doi: 10.1016/S0140-6736(20)30226-9] [Medline: 32416782]
24. Katapally TR. A global digital citizen science policy to tackle pandemics like COVID-19. J Med Internet Res 2020 May 26;22(5):e19357 [FREE Full text] [doi: 10.2196/19357] [Medline: 32408267]
25. Mbunge E. Integrating emerging technologies into COVID-19 contact tracing: Opportunities, challenges and pitfalls. Diabetes Metab Syndr 2020;14(6):1631-1636 [FREE Full text] [doi: 10.1016/j.dsx.2020.08.029] [Medline: 32892060]
26. Garg S, Bhatnagar N, Gangadharan N. A case for participatory disease surveillance of the COVID-19 pandemic in India. JMIR Public Health Surveill 2020 Apr 16;6(2):e18795 [FREE Full text] [doi: 10.2196/18795] [Medline: 32287038]
27. Denecke K. An ethical assessment model for digital disease detection technologies. Life Sci Soc Policy 2017 Sep 20;13(1):16 [FREE Full text] [doi: 10.1186/s40504-017-0062-x] [Medline: 28929347]
28. Kostkova P. Disease surveillance data sharing for public health: the next ethical frontiers. Life Sci Soc Policy 2018 Jul 04;14(1):16 [FREE Full text] [doi: 10.1186/s40504-018-0078-x] [Medline: 29971516]
29. Vayena E, Salathé M, Madoff LC, Brownstein JS. Ethical challenges of big data in public health. PLoS Comput Biol 2015 Feb;11(2):e1003904 [FREE Full text] [doi: 10.1371/journal.pcbi.1003904] [Medline: 25664461]
30. de Jong BC, Gaye BM, Luyten J, van Buitenen B, André E, Meehan CJ, et al. Ethical considerations for movement mapping to identify disease transmission hotspots. Emerg Infect Dis 2019 Jul;25(7) [FREE Full text] [doi: 10.3201/eid2507.181421] [Medline: 31211938]
31. Kind C. Exit through the App Store? Patterns (N Y) 2020 Jun 12;1(3):100054 [FREE Full text] [doi: 10.1016/j.patter.2020.100054] [Medline: 32548597]
32. Park S, Choi GJ, Ko H. Information technology-based tracing strategy in response to COVID-19 in South Korea-privacy controversies. JAMA 2020 Jun 02;323(21):2129-2130. [doi: 10.1001/jama.2020.6602] [Medline: 32324202]
33. Fraser C, Abeler-Dorner L, Ferretti L, Parker M, Kendall M, Bonsall D. Digital contact tracing: comparing the capabilities of centralised and decentralised data architectures to effectively suppress the COVID-19 epidemic whilst maximising freedom of movement and maintaining privacy. University of Oxford. 2020. URL: https://tinyurl.com/y4bjggke [accessed 2021-09-30]
34. Cho H, Ippolito D, Yu YW. Contact Tracing Mobile Apps for COVID-19: Privacy Considerations and Related Trade-offs. arXiv. URL: https://arxiv.org/abs/2003.11511 [accessed 2021-09-30]
35. Klenk M, Duijf H. Ethics of digital contact tracing and COVID-19: who is (not) free to go? Ethics Inf Technol 2020 Aug 24:1-9 [FREE Full text] [doi: 10.1007/s10676-020-09544-0] [Medline: 32863740]
36. de Montjoye YA, Houssiau F, Gadotti A, Guepin F. Evaluating COVID-19 contact tracing apps? Here are 8 privacy questions we think you should ask. Computational Privacy Group. 2020. URL: https://cpg.doc.ic.ac.uk/blog/evaluating-contact-tracing-apps-here-are-8-privacy-questions-we-think-you-should-ask/ [accessed 2021-02-20]
37. Bernier C, Fong L, Banks TM. Pandemics in a Connected World: Integrating Privacy with Public Health Surveillance. University of New Brunswick Law Journal. 2015. URL: https://journals.lib.unb.ca/index.php/unblj/article/view/29090 [accessed 2021-09-30]
38. Ienca M, Vayena E. On the responsible use of digital data to tackle the COVID-19 pandemic. Nat Med 2020 Apr;26(4):463-464 [FREE Full text] [doi: 10.1038/s41591-020-0832-5] [Medline: 32284619]
39. Yasaka TM, Lehrich BM, Sahyouni R. Peer-to-peer contact tracing: development of a privacy-preserving smartphone app. JMIR Mhealth Uhealth 2020 Apr 07;8(4):e18936 [FREE Full text] [doi: 10.2196/18936] [Medline: 32240973]
40. Barbieri C, Darnis JP. Technology: An Exit Strategy for COVID-19? Istituto Affari Internazionali. 2020. URL: https://www.iai.it/en/pubblicazioni/technology-exit-strategy-covid-19 [accessed 2021-09-30]
41. Chan J, Foster D, Gollakota S, Horvitz E, Jaeger J, Kakade S, et al. PACT: Privacy Sensitive Protocols and Mechanisms for Mobile Contact Tracing. arXiv. 2020. URL: https://arxiv.org/abs/2004.03544 [accessed 2021-09-30]
42. Lewis P. Trust in the government is rising – but will Australians accept the coronavirus tracing app? The Guardian. 2020. URL: https://www.theguardian.com/australia-news/commentisfree/2020/apr/20/trust-in-the-government-is-rising-but-will-australians-accept-the-coronavirus-tracing-app [accessed 2021-09-30]
43. WHO Coronavirus (COVID-19) Dashboard. World Health Organization. 2021. URL: https://covid19.who.int/ [accessed 2021-02-20]
44. COVID-19 boosts Chinese citizens' trust in their government: survey. Xinhua. 2021. URL: http://www.china.org.cn/world/2021-05/14/content_77498248.htm [accessed 2021-07-11]
45. International Health Regulations. World Health Organization. URL: https://www.who.int/health-topics/international-health-regulations#tab=tab_1 [accessed 2021-05-20]

Abbreviations

AI: artificial intelligence
BDA: big data analytics

EHR: electronic health record
ML: machine learning

Edited by C Basch; submitted 23.07.21; peer-reviewed by O Igboeli, A Benis, JA Benítez-Andrades; comments to author 03.09.21; revised version received 14.09.21; accepted 14.09.21; published 20.10.21.

Please cite as:
Zhao IY, Ma YX, Yu MWC, Liu J, Dong WN, Pang Q, Lu XQ, Molassiotis A, Holroyd E, Wong CWW
Ethics, Integrity, and Retributions of Digital Detection Surveillance Systems for Infectious Diseases: Systematic Literature Review
J Med Internet Res 2021;23(10):e32328
URL: https://www.jmir.org/2021/10/e32328
doi:10.2196/32328
PMID:34543228

Original Paper

Assessing Neonatal Intensive Care Unit Structures and Outcomes Before and During the COVID-19 Pandemic: Network Analysis Study

Hannah Mannering[1]; Chao Yan[2], MA; Yang Gong[3], MD, PhD; Mhd Wael Alrifai[4], MD; Daniel France[5], PhD; You Chen[2,6], PhD

[1]Department of Computer Science, Loyola University, Baltimore, MD, United States
[2]Department of Computer Science, Vanderbilt University, Nashville, TN, United States
[3]School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, United States
[4]Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN, United States
[5]Department of Anesthesiology, Vanderbilt University Medical Center, Nashville, TN, United States
[6]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States

Corresponding Author:
You Chen, PhD
Department of Biomedical Informatics
Vanderbilt University Medical Center
2525 West End Ave
Suite 1475
Nashville, TN, 37203
United States
Phone: 1 6153431939
Email: you.chen@vanderbilt.edu

Abstract

Background: Health care organizations (HCOs) adopt strategies (eg. physical distancing) to protect clinicians and patients in intensive care units (ICUs) during the COVID-19 pandemic. Many care activities physically performed before the COVID-19 pandemic have transitioned to virtual systems during the pandemic. These transitions can interfere with collaboration structures in the ICU, which may impact clinical outcomes. Understanding the differences can help HCOs identify challenges when transitioning physical collaboration to the virtual setting in the post–COVID-19 era.

Objective: This study aims to leverage network analysis to determine the changes in neonatal ICU (NICU) collaboration structures from the pre– to the intra–COVID-19 era.

Methods: In this retrospective study, we applied network analysis to the utilization of electronic health records (EHRs) of 712 critically ill neonates (pre–COVID-19, n=386; intra–COVID-19, n=326, excluding those with COVID-19) admitted to the NICU of Vanderbilt University Medical Center between September 1, 2019, and June 30, 2020, to assess collaboration between clinicians. We characterized pre–COVID-19 as the period of September-December 2019 and intra–COVID-19 as the period of March-June 2020. These 2 groups were compared using patients' clinical characteristics, including age, sex, race, length of stay (LOS), and discharge dispositions. We leveraged the clinicians' actions committed to the patients' EHRs to measure clinician-clinician connections. We characterized a collaboration relationship (tie) between 2 clinicians as actioning EHRs of the same patient within the same day. On defining collaboration relationship, we built pre– and intra–COVID-19 networks. We used 3 sociometric measurements, including eigenvector centrality, eccentricity, and betweenness, to quantify a clinician's leadership, collaboration difficulty, and broad skill sets in a network, respectively. We assessed the extent to which the eigenvector centrality, eccentricity, and betweenness of clinicians in the 2 networks are statistically different, using Mann-Whitney *U* tests (95% CI).

Results: Collaboration difficulty increased from the pre– to intra–COVID-19 periods (median eccentricity: 3 vs 4; *P*<.001). Nurses had reduced leadership (median eigenvector centrality: 0.183 vs 0.087; *P*<.001), and neonatologists with broader skill sets cared for more patients in the NICU structure during the pandemic (median betweenness centrality: 0.0001 vs 0.005; *P*<.001). The pre– and intra–COVID-19 patient groups shared similar distributions in sex (~0 difference), race (4% difference in White, and 3% difference in African American), LOS (interquartile range difference in 1.5 days), and discharge dispositions (~0 difference

in home, 2% difference in expired, and 2% difference in others). There were no significant differences in the patient demographics and outcomes between the 2 groups.

Conclusions: Management of NICU-admitted patients typically requires multidisciplinary care teams. Understanding collaboration structures can provide fine-grained evidence to potentially refine or optimize existing teamwork in the NICU.

(J Med Internet Res 2021;23(10):e27261) doi:10.2196/27261

KEYWORDS

neonatal intensive care unit; collaboration; health care organization structures; intensive care; length of stay; discharge dispositions; electronic health records; network analysis; COVID-19; temporal network analysis

Introduction

Health care organizations (HCOs) change intensive care unit (ICU) staffing and follow physical distancing policy during the COVID-19 pandemic to protect clinicians and patients [1,2]. For instance, many physical care activities before the COVID-19 pandemic have been transitioning to virtual systems, such as electronic health records (EHRs) or telehealth [3-5]. These changes can interfere with the structures of teamwork in the ICU, which may impact clinical outcomes. The changes in ICU structures and outcomes from pre– to intra–COVID-19 periods have not been systematically investigated. Therefore, challenges are unclear when health care delivery disruptions (eg, pandemics) or major transitions (physical to virtual collaboration) occur in the post–COVID-19 era.

One of the major challenges to analyzing ICU structures and quantifying their changes is that the ICU structures are historically developed at a coarse-grained level, which seldom considers connections among clinicians in a team owing to dynamic and complex clinical workflows, shifts, and handovers [6-9]. Understanding how clinicians connect (eg, sharing and exchanging health information) within their clinical teams when caring for patients can provide fine-grained evidence to potentially refine or optimize existing ICU structures.

In modern health care, an increasing number of clinicians utilize EHR to diagnose and treat patients by exchanging all medical statuses [10,11]. Therefore, the volume of the EHR system utilization data has been increasing exponentially in recent years, providing abundant resources to identify connections between clinicians. Recent studies applied network analysis to EHR utilization data to measure connections among clinicians [12-15]. They found EHR system utilization data to potentially be a rich resource to be leveraged to model relationships among clinicians. Recent studies have also shown that network analysis methods and data within the EHR can also be utilized to learn collaboration structures in ICUs [8,9,16]. Based on previous studies, this study leverages network analysis methods to learn structures of the neonatal ICU (NICU) in pre– and intra–COVID-19 eras in terms of collaboration among clinicians, and compares differences in the structures. Patients hospitalized in the NICU include high-risk infants who may be or are at risk for a variety of complex diseases or conditions. The management of NICU patients typically requires multidisciplinary care teams (eg, neonatal frontline providers, ancillary staff, nurses, neonatologists, residents, support staff, respiratory therapists, neonatal fellows, and highly specialized consultants) [17-19]. We investigate the connections among clinicians in a tertiary-level NICU, which has a high density of intense EHR utilization and heavy data sharing traffic per patient episode [8,9], making this environment ideal for our ICU structure study.

Methods

To describe our work systematically, we used the reporting checklist for quality improvement in health care (Multimedia Appendix 1), which is based on the SQUIRE 2.0 guidelines [20].

We extracted EHRs for all patients admitted to the NICU at Vanderbilt University Medical Center (VUMC, Nashville, Tennessee) between September 1, 2019, and June 30, 2020. We characterized pre–COVID-19 as the period of September through December 2019 and intra–COVID-19 as the period of March through June 2020. We used network analysis methods to analyze the EHRs of 712 NICU patients (pre–COVID-19 patients, n=386; intra–COVID-19 patients, n=326), excluding those with COVID-19, to assess clinician networks to describe pre–COVID-19 and intra–COVID-19 teamwork structures. These 2 groups were compared using patients' clinical characteristics, including their age, sex, race, length of stay (LOS), and discharge dispositions.

To protect patient confidentiality, analysis of EHR data was conducted at a data analysis server located at the VUMC. The EHR data used in this study were physically housed in a secure room at the VUMC's data center. All connections made to the servers were made in an encrypted manner and used Secure Shell technology from known computers. A unique login and password were set for each authorized individual. All protected health data remained on the server, and no copy of the data was provided to unauthorized parties. The Vanderbilt institutional review board reviewed and approved the study (protocol No. 200792).

Clinicians' EHR actions stemmed from different tasks, including conditions (eg, assessing a patient's condition), procedures (eg, intubation), medications (eg, prescription drugs), notes (eg, progress note writing), orders (eg, laboratory test ordering), and measurements (eg, measuring blood pressure). We leveraged the actions committed to the EHRs of patients by clinicians to measure clinician-clinician connections. Prior research shows that a 1-day window can capture meaningful collaborative relationships among clinicians [12-15]. Based on their findings, we characterized a collaboration relationship (tie) between 2 clinicians as they performed actions to EHRs of the same patient within the same day (24 hours). This definition can capture different types of interactions between clinicians. The first type

is the asynchronous interactions between clinicians; for instance, a clinician created a medication order at 9 AM, and another clinician reviewed and processed the order at 11 AM on the same day. Thus, the 2 clinicians had an asynchronous collaboration in terms of order creation and processing. The second type is the interactions between clinicians during shifts or handoffs, which are among the most critical aspects of collaboration owing to the medical errors that occur during the transition between clinicians [17-19,21]. For instance, 2 nurses (oncoming and ongoing) were responsible for a patient's handoff or shift at 7 PM. The 2 nurses may perform some actions to EHRs of the patient before, during, or after the handoff/shift, which can be captured by our definition of collaboration relationship, to build a connection between them. The third type is the interactions between clinicians built on their documentation in EHRs or messages in the basket, a communication hub where clinicians can send and receive secure messages. For instance, 2 clinicians work together during patient care without interacting with EHRs, but both made some documentation later. Based on our collaboration definition, the 2 clinicians still have a connection built between them. We quantified the weight of a relation between 2 clinicians as the number of patients they co-managed on the same day, which can be learned from EHRs. We referred to each patient on that day as a patient day. The relation's final weight is the cumulative number of patient days the 2 clinicians interacted by co-managing patients during our investigated time window (4 months). Thus, we built the pre– and intra–COVID-19 networks.

Formally, the nodes in the pre– and intra–COVID-19 networks were defined as $Z_{pre} = \{z_1, z_2, \ldots, z_p\}$ and $Z_{intra} = \{z_1, z_2, \ldots, z_q\}$, respectively. To better interpret the networks, we used the clinician's specialty (eg, respiratory therapist or NICU registered nurse) to label each node. Specialties in the pre– and intra–COVID-19 networks are referenced as $EXP_{pre} = \{exp_1, exp_2, \ldots, exp_a\}$ and $EXP_{intra} = \{exp_1, exp_2, \ldots, exp_b\}$, respectively. *Z* and *EXP* are used to describe the compositions of the pre– and intra–COVID-19 networks.

We leveraged sociometric measurements, including eigenvector centrality [22], betweenness centrality [23], and eccentricity [24], to quantify the network structures. Eigenvector centrality is the measure of the influence of a node in a network [22]. In a health care EHR setting, high eigenvector centrality implies that the clinician has a very active role and serves as a hub in information sharing and dissemination. Betweenness is a measure of centrality based on the shortest paths in a network; it is calculated as the number of times a node acts as a bridge along the shortest path between two other nodes [23]. A node with higher betweenness centrality has more control in the network because owing to its shorter paths to other nodes, more information will pass through that node. Eccentricity of a node in a network is the maximum distance from the node to any other node [24]. In a clinician setting, eccentricity is the radius from one clinician to another, which is the largest distance. A larger eccentricity implies that there many more steps to share information with another clinician. Therefore, we characterized eigenvector centrality as indicating a clinician's leadership (hub) in terms of collaboration, betweenness centrality as demonstrating a clinician cares for a wide spectrum (bridge) of patients, and eccentricity as showing the difficulty for a clinician to collaborate with others. Figure 1 shows 3 networks to illustrate each of the 3 sociometric measurements, respectively. We used Gephi, an open-source network analysis and visualization software package [25], to calculate eigenvector centrality, betweenness centrality, and eccentricity for each of the nodes in the pre– and intra–COVID-19 networks.

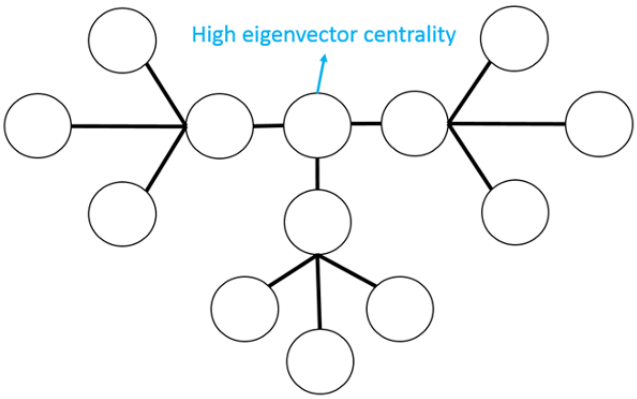
We investigated whether the differences in the clinician leadership, who care for a wide spectrum of patients, and collaboration difficulty are statistically different between pre– and intra–COVID-19 networks. In addition, we investigated changes in 2 outcome metrics, including LOS and discharge dispositions, from the pre– to the intra–COVID-19 periods. We applied Mann-Whitney *U* tests at a 95% CI to account for the non-Gaussian distribution of the sociometric measurements and outcomes. Since the pre– and intra–COVID-19 networks are composed of clinicians with different specialties, we compared the differences at both network (entire network) and specialty (each specialty) levels. We applied a Bonferroni correction to account for multiple hypothesis testing (eg, pairwise test for the specialty-level comparisons).

For each investigated specialty (eg, NICU nurse), we investigated whether clinicians affiliated with the specialty in the pre–COVID-19 network have significantly higher values of eigenvector centrality, betweenness centrality, or eccentricity than those affiliated with the same specialty in the intra–COVID-19 network. We focused on the 6 specialties that play essential roles in NICU care, which included NICU nurses, nurse practitioners, residents, respiratory therapists, cardiac ICU nurses, and neonatologists. For an investigated specialty, we created 2 arrays, one with values of a sociometric measurement (eg, eigenvector centrality) of nodes (clinicians) affiliated with the specialty in the pre–COVID-19 network and the other for the values of the same sociometric measurement of clinicians affiliated with the same specialty in the intra–COVID-19 network.

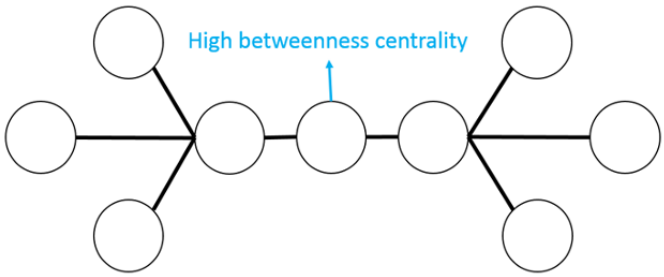
We further tested the differences in the eigenvector centrality, betweenness centrality, and eccentricity between pre– and intra–COVID-19 networks using all specialties in the 2 networks. The hypothesis test is of the following form: there are significant differences in eigenvector centrality, betweenness centrality, or eccentricity between the pre– and intra–COVID-19 networks. Within a network, we measured eigenvector centrality, betweenness centrality, or eccentricity for a specialty by calculating the mean values of eigenvector centrality, betweenness centrality, or eccentricity of all clinicians affiliated with that specialty in the network. For pre– or intra–COVID-19 networks, we developed an array of specialties, whose cell value is the value of a specialty's sociometric.

The data sets generated and analyzed during the current study are not publicly available in order to maintain the privacy of the patient information investigated in this study, but they are available from the corresponding author on reasonable request.

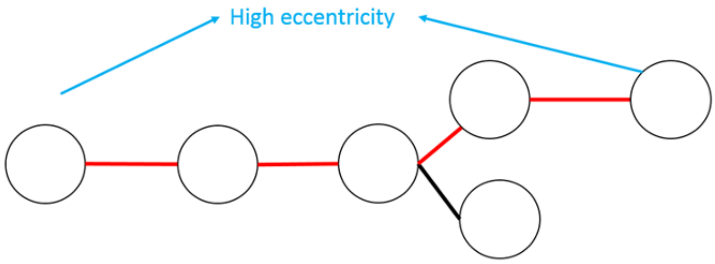
Figure 1. Example networks to illustrate (A) eigenvector centrality, (B) betweenness centrality, and (C) eccentricity.



A. A node with high eigenvector centrality serves as a hub in the network



B. A node with high betweenness centrality serves as a bottleneck in the network



C. The eccentricity is the maximum graph distance between two nodes (a red solid line is an example of high eccentricity)

Results

The pre– and intra–COVID-19 patient groups share similar distributions (Table 1) in sex (~0 difference), race (4% difference in White, and 3% difference in African American), LOS (IQR difference in 1.5 days), and discharge dispositions (~0 difference in home, 2% difference in expired, and 2% difference in others). There were no significant differences in patient demographics and outcomes between the 2 groups.

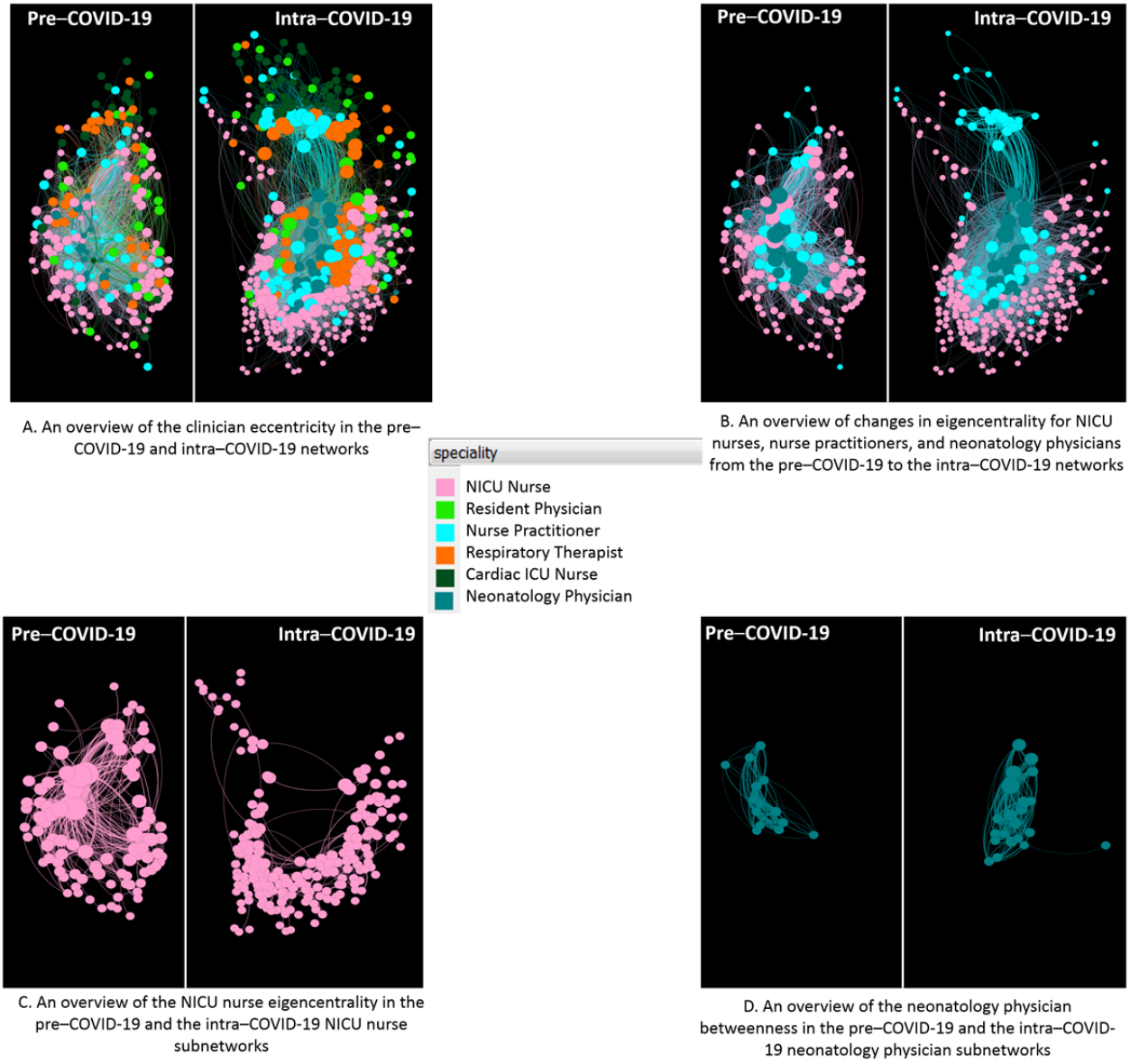
There are several notable findings in network analysis to highlight. First, the intra–COVID-19 NICU structure had a higher eccentricity (collaboration difficulty) than the pre–COVID-19 (median 3 vs 4; P<.001). Second, NICU nurses had a lower eigenvector centrality (leadership in collaboration) in the intra–COVID-19 structure than the pre–COVID-19 (median 0.183 vs 0.087; P<.001). Third, neonatology physicians had a higher betweenness centrality (care for a wider spectrum of patients) in the intra–COVID-19 than in the pre–COVID-19 structure (median 0.0001 vs 0.005; P<.005).

Table 1. Characteristics of patients admitted to the neonatal intensive care unit in the pre– and intra–COVID-19 groups (N=712).

Characteristics	Pre–COVID-19 group (n=386)	Intra–COVID-19 group (n=326)
Demographic information		
Age (days), median (IQR)	0.0 (0.0-0.0); maximum, 149.0	0.0 (0.0-0.0); max-imum, 97.0
Sex, n (%)		
Female	159 (41.2)	134 (41.1)
Male	159 (41.2)	134 (41.1)
Race, n (%)		
White	258 (66.8)	231 (70.9)
African American	64 (16.6)	43 (13.2)
Asian	16 (4.1)	6 (1.8)
Other	47 (12.2)	46 (14.1)
Ethnicity, n (%)		
Non-Hispanic	331 (85.8)	283 (86.8)
Latino	50 (13.0)	32 (9.8)
Unknown	5 (1.3)	11 (3.4)
Outcomes		
Length of stay, median (IQR)	10.0 (4.0-20.0)	9.0 (3.3-21.8)
Hospital disposition, n (%)		
Home	354 (91.7)	299 (91.7)
Expired	27 (7.0)	16 (4.9)
Hospice	2 (0.5)	2 (0.6)
Short-term hospital	3 (0.8)	8 (2.4)
Others	0 (0)	1 (0.3)

Figure 2 shows the pre– and intra–COVID-19 networks from different perspectives. Figure 2A shows that the intra–COVID-19 network is larger, implying that the 6 types of clinicians are likely to be more difficult to collaborate with others (higher eccentricity) in the intra–COVID-19 NICU care than those in the pre–COVID-19 NICU care (median 4 vs 3; P<.001). This higher eccentricity in intra–COVID-19 implies that there is more distance from one node to any other node within the network. Compared to nurse practitioners and neonatology physicians, nurses' leadership (eigenvector centrality) reduced from the pre– to intra–COVID-19 networks (median 0.183 vs 0.087; P<.001), as shown in Figure 2B. This reduced leadership implies that nurses were less active during the intra–COVID-19 period. Nurses do not occupy central positions in the intra–COVID-19 network, and they have more connections in the pre–COVID-19 period than in the intra–COVID-19 period, as shown in Figure 2C. Neonatology physicians care for a wider spectrum of patients (high betweenness centrality) in intra–COVID-19 NICU care than those in pre–COVID-19 NICU care (median 0.005 vs 0.0001; P<.001), as shown in Figure 2D. Therefore, during the COVID-19 pandemic, neonatology physicians have a higher number of shorter paths to other nodes within the network.

Figure 2. Overviews of the eccentricity of the 6 types of clinicians in (A) the pre– and intra–COVID-19 networks, (B) overviews of the eigenvector centrality of nurses, nurse practitioners, and neonatology physicians in the pre– and intra–COVID-19 networks, (C) subnetworks of NICU nurses and their eigenvector centrality in the pre– and intra–COVID-19 settings, and (D) subnetworks of neonatology physicians and their betweenness centrality in the pre– and intra–COVID-19 settings. The legend in the center shows the colors of the 6 roles, which have the largest number of clinicians affiliated. The eccentricity is directly correlated with the corresponding node size in (A), the eigenvector centrality is directly correlated with the corresponding node size in (B) and (C), and the betweenness centrality is directly correlated with the corresponding node size in (D). ICU: intensive care unit, NICU: neonatal intensive care unit.



Discussion

Principal Findings

To follow the COVID-19 physical distancing policy, the VUMC involves new EHR use practices to provide care for patients during the COVID-19 pandemic. The collaboration difficulty (increased eccentricity) can be a potential problem in the new EHR use practices. In the post–COVID-19 era, when HCOs plan to promote more collaboration in virtual platforms, they may need to develop staffing strategies to reduce the collaboration difficulty in EHRs.

Neonatologists care for a wider spectrum of patients (higher betweenness centrality) when using EHRs during the COVID-19 pandemic. HCOs may need to develop educational strategies to promote EHR collaboration between neonatologists and other clinicians to improve teamwork efficiency and NICU outcomes in the post–COVID-19 era. NICU nurses have reduced leadership (lower eigenvector centrality) in cooperation, suggesting that increased EHR use may reduce nurses' workload in the collaboration.

Our results for network analysis of collaboration structures demonstrate changes in virtual care from the pre– to the intra–COVID-19 periods. Findings in prior literature also reflect our results. Reeves et al [26] reported the increasing utilization of electronic check-in, standard ordering and documentation, secure messaging, real-time data analytics, and telemedicine during the COVID-19 pandemic, compared to those before the

pandemic. Furthermore, Wosik et al [5] examined how people, processes, and technology (EHRs) work together to support a successful virtual care transformation.

Our results show there are no significant changes in LOS and discharge dispositions, which indicates the changes in clinicians' connections to protect patients and health care professionals during the COVID-19 pandemic have few impacts on the 2 outcomes. However, the effect of the changes on the satisfaction of patients' families and clinicians has not been investigated. Although our results are dependent on an investigation of 1 health system, the network analysis methods used in our study could be used to extrapolate results in different countries with different health care systems.

Limitations

There are several limitations in this pilot study, which should be recognized. The characteristics of the NICU structures learned from this single-center analysis could provide some reference for other HCOs when they assess their NICU structures. However, VUMC NICU is a highly collaborative environment, which should be considered when interpreting the results and findings. Second, there is a lack of standard terminology for characterizing NICU specialties. Common data models for clinician types would improve the quality of our study and assist in the transition of our methodology to other institutions. Third, we assumed that 2 clinicians have a connection when they commit actions to the EHRs of patients on the same day. Although such an assumption can capture collaboration relationships among clinicians, it may also reveal many spurious relationships. This assumption also only looks at clinician connections within the EHR; failing to identify in-person (eg, discussing patient results with another clinician) or virtual interactions (eg, Zoom meetings). Furthermore, the connection between the 2 clinicians indicates potential collaboration (information sharing) rather than actual collaboration. Finally, fine-grained EHR actions are required to add contextual information to the relations between clinicians. For instance, the connection between a nurse and a consultant is created on the basis of their communications in flowsheets data.

Conclusions

The developed network methods can be effective tools to assess differences in collaboration structures in current and future disruptions in health care delivery (eg, pandemics) and major transitions (physical to virtual collaboration) adopted by HCOs. The methods and the results of our study can also be used to analyze clinician's leadership, collaboration difficulty, and broad skill sets in different health care studies. In future studies, recruiting subject matter experts (eg, clinics) to evaluate the learned connections and ICU structures will be required to validate the results. This knowledge of connections among clinicians can assist HCOs with developing more specific staffing strategies, which may improve care quality and patient outcomes.

Clinical Trial Registration

Non–clinical trial study.

Acknowledgments

This research was supported, in part, by the National Library of Medicine of the National Institutes of Health (award no. R01LM012854) and the National Science Foundation (award no. 1757644). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health and National Science Foundation.

Authors' Contributions

HM performed the data analysis, methods design and development, experiment design, evaluation and interpretation of the results, and manuscript writing. CY performed the data collection, data analysis, evaluation and interpretation of the experimental results, and the revision of the manuscript. YG performed the evaluation and interpretation of the experimental results and the revision of the manuscript. MWA performed the evaluation and interpretation of the experimental results and the revision of the manuscript. DF performed the evaluation and interpretation of the experimental results and the revision of the manuscript. YC conceived the study and performed the data collection and analysis, methods design and development, experiment design, evaluation and interpretation, and manuscript writing. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Reporting checklist for quality improvement in health care.
[DOCX File , 25 KB - jmir_v23i10e27261_app1.docx]

References

1. Black JRM, Bailey C, Przewrocka J, Dijkstra KK, Swanton C. COVID-19: the case for health-care worker screening to prevent hospital transmission. Lancet 2020 May 02;395(10234):1418-1420 [FREE Full text] [doi: 10.1016/S0140-6736(20)30917-X] [Medline: 32305073]

2. Abir M, Nelson C, Chan E, Al-Ibrahim H, Cutter C, Patel K, et al. RAND Critical Care Surge Response Tool An Excel-Based Model for Helping Hospitals Respond to the COVID-19 Crisis. RAND Corporation. 2020. URL: https://www.rand.org/pubs/tools/TLA164-1.html [accessed 2021-10-14]
3. Valdez R, Rogers C, Claypool H, Trieshmann L, Frye O, Wellbeloved-Stone C, et al. Ensuring full participation of people with disabilities in an era of telehealth. J Am Med Inform Assoc 2021 Feb 15;28(2):389-392 [FREE Full text] [doi: 10.1093/jamia/ocaa297] [Medline: 33325524]
4. Kennedy NR, Steinberg A, Arnold RM, Doshi AA, White DB, DeLair W, et al. Perspectives on Telephone and Video Communication in the Intensive Care Unit during COVID-19. Ann Am Thorac Soc 2021 May;18(5):838-847 [FREE Full text] [doi: 10.1513/AnnalsATS.202006-729OC] [Medline: 33181033]
5. Wosik J, Fudim M, Cameron B, Gellad Z, Cho A, Phinney D, et al. Telehealth transformation: COVID-19 and the rise of virtual care. J Am Med Inform Assoc 2020 Jun 01;27(6):957-962 [FREE Full text] [doi: 10.1093/jamia/ocaa067] [Medline: 32311034]
6. Shime N. ICU staffing models and patient outcomes. J Intensive Care 2019;7:24 [FREE Full text] [doi: 10.1186/s40560-019-0380-6] [Medline: 31049202]
7. Michard F, Saugel B, Vallet B. Rethinking the post-COVID-19 pandemic hospital: more ICU beds or smart monitoring on the wards? Intensive Care Med 2020 Sep;46(9):1792-1793 [FREE Full text] [doi: 10.1007/s00134-020-06163-7] [Medline: 32613430]
8. Chen Y, Lehmann CU, Hatch LD, Schremp E, Malin BA, France DJ. Modeling Care Team Structures in the Neonatal Intensive Care Unit through Network Analysis of EHR Audit Logs. Methods Inf Med 2019 Nov;58(4-05):109-123 [FREE Full text] [doi: 10.1055/s-0040-1702237] [Medline: 32170716]
9. Kim C, Lehmann C, Hatch D, Schildcrout J, France D, Chen Y. Provider Networks in the Neonatal Intensive Care Unit Associate with Length of Stay. IEEE Conf Collab Internet Comput 2019 Dec;2019:127-134 [FREE Full text] [doi: 10.1109/CIC48465.2019.00024] [Medline: 32637942]
10. Reitz R, Common K, Fifield P, Stiasny E. Collaboration in the presence of an electronic health record. Fam Syst Health 2012 Mar;30(1):72-80. [doi: 10.1037/a0027016] [Medline: 22429079]
11. Noteboom C, Motorny S, Qureshi S, Sarnikar S. Meaningful use of electronic health records for physician collaboration: A patient centered health care perspective. 2014 Presented at: 2014 47th Hawaii International Conference on System Sciences; January 6-9, 2014; Waikoloa, HI p. 656-666. [doi: 10.1109/HICSS.2014.87]
12. Chen Y, Patel M, McNaughton C, Malin B. Interaction patterns of trauma providers are associated with length of stay. J Am Med Inform Assoc 2018 Jul 01;25(7):790-799 [FREE Full text] [doi: 10.1093/jamia/ocy009] [Medline: 29481625]
13. Chen Y, Yan C, Patel MB. Network Analysis Subtleties in ICU Structures and Outcomes. Am J Respir Crit Care Med 2020 Dec 01;202(11):1606-1607 [FREE Full text] [doi: 10.1164/rccm.202008-3114LE] [Medline: 32931298]
14. Kelly Costa D, Liu H, Boltey EM, Yakusheva O. The Structure of Critical Care Nursing Teams and Patient Outcomes: A Network Analysis. Am J Respir Crit Care Med 2020 Feb 15;201(4):483-485 [FREE Full text] [doi: 10.1164/rccm.201903-0543LE] [Medline: 31626555]
15. Chen Y, Lorenzi N, Sandberg W, Wolgast K, Malin B. Identifying collaborative care teams through electronic medical record utilization patterns. J Am Med Inform Assoc 2017 Apr 01;24(e1):e111-e120 [FREE Full text] [doi: 10.1093/jamia/ocw124] [Medline: 27570217]
16. Yan C, Zhang X, Gao C, Wilfong E, Casey J, France D, et al. Collaboration Structures in COVID-19 Critical Care: Retrospective Network Analysis Study. JMIR Hum Factors 2021 Mar 08;8(1):e25724 [FREE Full text] [doi: 10.2196/25724] [Medline: 33621187]
17. Vandenberg KA. Individualized developmental care for high risk newborns in the NICU: a practice guideline. Early Hum Dev 2007 Jul;83(7):433-442. [doi: 10.1016/j.earlhumdev.2007.03.008] [Medline: 17467932]
18. Purdy I, Melwak M, Smith J, Kenner C, Chuffo-Siewert R, Ryan D, et al. Neonatal Nurses NICU Quality Improvement: Embracing EBP Recommendations to Provide Parent Psychosocial Support. Adv Neonatal Care 2017 Feb;17(1):33-44. [doi: 10.1097/ANC.0000000000000352] [Medline: 27763909]
19. Subedi D, DeBoer MD, Scharf RJ. Developmental trajectories in children with prolonged NICU stays. Arch Dis Child 2017 Jan;102(1):29-34. [doi: 10.1136/archdischild-2016-310777] [Medline: 27637907]
20. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. Squire 2.0 (Standards for Quality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. Am J Crit Care 2015 Nov;24(6):466-473. [doi: 10.4037/ajcc2015455] [Medline: 26523003]
21. Greenwald J, Denham C, Jack B. The hospital discharge: a review of a high risk care transition with highlights of a reengineered discharge process. J Patient Saf 2007;3(2):97-106. [doi: 10.1097/01.jps.0000236916.94696.12]
22. Laporta L, Afonso J, Mesquita I. The need for weighting indirect connections between game variables: Social Network Analysis and eigenvector centrality applied to high-level men's volleyball. International Journal of Performance Analysis in Sport 2018 Dec 03;18(6):1067-1077. [doi: 10.1080/24748668.2018.1553094]
23. Meghanathan N. A computationally lightweight and localized centrality metric in lieu of betweenness centrality for complex network analysis. Vietnam J Comput Sci 2016 Jun 17;4(1):23-38. [doi: 10.1007/s40595-016-0073-1]

24. Lü L, Chen D, Ren X, Zhang Q, Zhang Y, Zhou T. Vital nodes identification in complex networks. Phys Rep 2016 Sep;650:1-63. [doi: 10.1016/j.physrep.2016.06.007]
25. Bastian M, Heymann S, Jacomy M. Gephi : An Open Source Software for Exploring and Manipulating Networks. 2009. URL: https://gephi.org/publications/gephi-bastian-feb09.pdf [accessed 2021-10-14]
26. Reeves J, Hollandsworth H, Torriani F, Taplitz R, Abeles S, Tai-Seale M, et al. Rapid response to COVID-19: health informatics support for outbreak management in an academic health system. J Am Med Inform Assoc 2020 Jun 01;27(6):853-859 [FREE Full text] [doi: 10.1093/jamia/ocaa037] [Medline: 32208481]

Abbreviations

EHR: electronic health record
HCO: health care organization
ICU: intensive care unit
LOS: length of stay
NICU: neonatal intensive care unit
VUMC: Vanderbilt University Medical Center

Edited by R Kukafka; submitted 18.01.21; peer-reviewed by YL Weideman, M Lotto; comments to author 10.05.21; revised version received 18.05.21; accepted 08.10.21; published 20.10.21.

Please cite as:
Mannering H, Yan C, Gong Y, Alrifai MW, France D, Chen Y
Assessing Neonatal Intensive Care Unit Structures and Outcomes Before and During the COVID-19 Pandemic: Network Analysis Study
J Med Internet Res 2021;23(10):e27261
URL: https://www.jmir.org/2021/10/e27261
doi:10.2196/27261
PMID:34637393

Original Paper

Topics and Sentiments of Public Concerns Regarding COVID-19 Vaccines: Social Media Trend Analysis

Michal Monselise[1], MSc; Chia-Hsuan Chang[2], MSc; Gustavo Ferreira[1], MSc; Rita Yang[3], PharmD; Christopher C Yang[1], PhD

[1]College of Computing and Informatics, Drexel University, Philadelphia, PA, United States

[2]Department of Information Management, National Sun Yat-sen University, Kaohsiung, Taiwan

[3]Virtua Voorhees Hospital, Voorhees Township, NJ, United States

Corresponding Author:
Christopher C Yang, PhD
College of Computing and Informatics
Drexel University
3675 Market St
10th Floor
Philadelphia, PA, 19104
United States
Phone: 1 215 895 1631
Email: chris.yang@drexel.edu

Abstract

Background: As a number of vaccines for COVID-19 are given emergency use authorization by local health agencies and are being administered in multiple countries, it is crucial to gain public trust in these vaccines to ensure herd immunity through vaccination. One way to gauge public sentiment regarding vaccines for the goal of increasing vaccination rates is by analyzing social media such as Twitter.

Objective: The goal of this research was to understand public sentiment toward COVID-19 vaccines by analyzing discussions about the vaccines on social media for a period of 60 days when the vaccines were started in the United States. Using the combination of topic detection and sentiment analysis, we identified different types of concerns regarding vaccines that were expressed by different groups of the public on social media.

Methods: To better understand public sentiment, we collected tweets for exactly 60 days starting from December 16, 2020 that contained hashtags or keywords related to COVID-19 vaccines. We detected and analyzed different topics of discussion of these tweets as well as their emotional content. Vaccine topics were identified by nonnegative matrix factorization, and emotional content was identified using the Valence Aware Dictionary and sEntiment Reasoner sentiment analysis library as well as by using sentence bidirectional encoder representations from transformer embeddings and comparing the embedding to different emotions using cosine similarity.

Results: After removing all duplicates and retweets, 7,948,886 tweets were collected during the 60-day time period. Topic modeling resulted in 50 topics; of those, we selected 12 topics with the highest volume of tweets for analysis. Administration and access to vaccines were some of the major concerns of the public. Additionally, we classified the tweets in each topic into 1 of the 5 emotions and found fear to be the leading emotion in the tweets, followed by joy.

Conclusions: This research focused not only on negative emotions that may have led to vaccine hesitancy but also on positive emotions toward the vaccine. By identifying both positive and negative emotions, we were able to identify the public's response to the vaccines overall and to news events related to the vaccines. These results are useful for developing plans for disseminating authoritative health information and for better communication to build understanding and trust.

(J Med Internet Res 2021;23(10):e30765) doi:10.2196/30765

KEYWORDS

health care informatics; topic detection; unsupervised sentiment analysis; COVID-19; vaccine hesitancy; sentiment; concern; vaccine; social media; trend; trust; health information; Twitter; discussion; communication; hesitancy; emotion; fear

Introduction

Background

In late 2020, the COVID-19 pandemic had approached the year mark when a number of pharmaceutical companies began to release their vaccine clinical trial results. A global sense of relief was felt when the results of the clinical trials looked promising. The first vaccine developed by Pfizer and BioNTech was given for emergency use authorization in December 2020 by the US Food and Drug Administration [1]. While this timeline seemed too fast for some, most vaccines for COVID-19 relied on many years of previous scientific work. For example, mRNA-based vaccines had been in development for over a decade at that point [2-4]. Despite efforts of the scientific community to assure the public that these vaccines are safe and effective, public sentiment has been mixed. There has been a significant amount of public hesitancy toward vaccination against COVID-19 [5]. At the same time, many have expressed excitement over the prospect of returning to a prepandemic world. Given this mixed reaction, it is essential to investigate the actual public sentiment regarding COVID-19 vaccines. Particularly, we were interested in learning about public sentiment for a period of 60 days when the vaccines were started in the United States. Social media provides a great data source for listening to the public on what they are thinking and what concerns and questions they have. We used Twitter as a proxy for public sentiment and were able to find the most important discussion topics that pertained to COVID-19 vaccines in the early days of the vaccine rollout. Additionally, we were able to classify public sentiment as it pertained to the vaccines and how this sentiment changed over time overall and in each topic as well. The goal of this research was to examine the discussion topics and public sentiment toward COVID-19 vaccines. By studying the topic and sentiment of the discussion on COVID-19 vaccines on Twitter, we may understand public concerns as they happen and learn more accurately about the source of vaccine hesitancy. By learning what drives vaccine hesitancy, we can better address it and formulate tailored and targeted communication. Conversely, we may also learn about the excitement toward the vaccine and study what is going well and what resonates well with the public on social media. This research will use the results uncovered by the topic and sentiment analysis of the Twitter data and suggest actionable insights for practitioners to address COVID-19 vaccine hesitancy. This research will also address how to utilize positive sentiment toward the vaccine.

Previous Works

Public Sentiment on COVID-19 Vaccine

A number of studies about vaccine hesitancy on social media have been published during the pandemic. Before any vaccine was approved, research showed hesitancy on social media. Harrison and Wu [6] examined vaccine hesitancy at the start of the pandemic and discussed methods to reduce vaccine hesitancy in preparation for the vaccine that would eventually come. This paper critiques current approaches for combating vaccine hesitancy with the goal of improving on these approaches when the COVID-19 vaccines are authorized for emergency use. A study by Chou and Budenz [7] discusses both methods for reducing hesitancy as well as for fostering positive emotions toward the vaccine. They propose acknowledging fear, anger, and other negative emotions and addressing them to convince the public to get vaccinated. A study by Wilson and Wiysonge [8] showed the existence of organized disinformation campaigns against the vaccines for COVID-19. However, this study focused on exposing negative sentiment against the vaccine and did not measure the positive sentiment toward the vaccine on social media. While the abovementioned studies discuss public sentiment, they do not measure both positive and negative sentiment, and some just make recommendations rather than looking at empirical evidence.

Topic Detection in COVID-19–Related Tweet Sentiment Analysis

Owing to the pandemic and quarantine policy, social media platforms such as Twitter became the main channel for people to share thoughts and to express their opinions about any impacts caused by the COVID-19. The hidden topics underneath such massive textual contents on social media help governments and health care units to understand the demand of the general public so as to make better decision and quick response. Cinelli et al [9] extracted topics using Partitioning Around Medoids algorithm on word vector representations and proposed a custom epidemic model for characterizing misinformation spreading speed in different social platforms. Since the temporal trends of the hidden topics reflect concerns of the general public through time, Chang et al [10] proposed 2 temporal models based on nonnegative matrix factorization (NMF), which help to identify the trends of several important themes such as government policy, economic crisis, COVID-19 case updates, COVID-19 urgent events, prevention, vaccines and treatments, and COVID-19 testing.

Sentiment Analysis

Sentiment analysis is a research area that involves the classification of text, images, or audio into a set of one or more sentiments [11]. In the context of this research, we will be classifying the sentiment of short snippets of text. When classifying text, we can classify at the word, sentence, or document level. There are different classification methods, including rule-based [12-14], support vector machine [15,16], random forest [17], Naive Bayes [18,19], embedding-based [20,21], as well as sentiment analysis using neural networks [22-25]. Additionally, we may classify sentiment by using unsupervised methods such as methods using rule-based unsupervised sentiment analysis [26], embeddings such as Word2Vec and Doc2Vec [27], and lexical resources for sentiment analysis [28].

Sentiment Analysis in Twitter

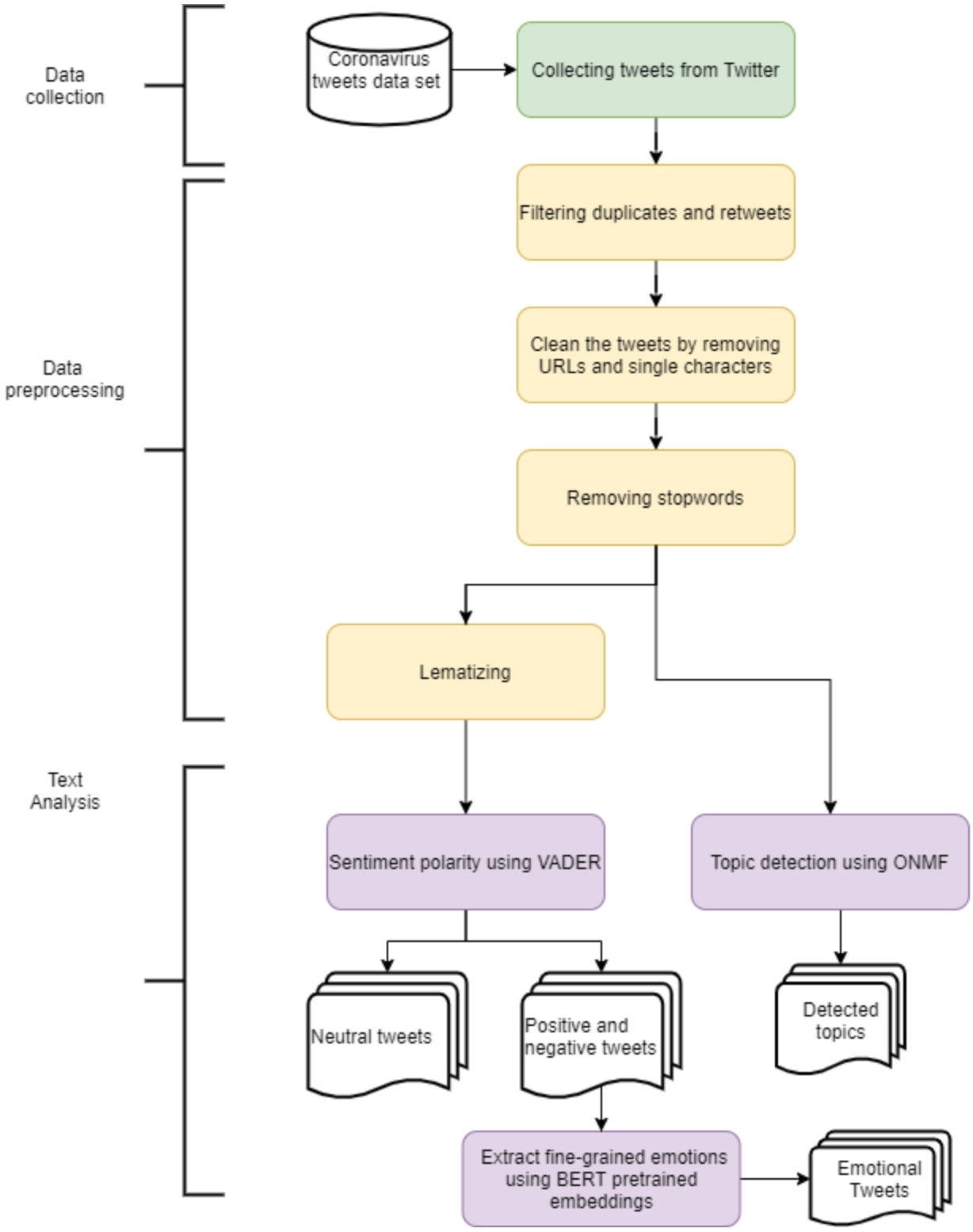
Sentiment analysis is an established research field in the area of natural language processing. However, performing sentiment analysis on tweets is a slightly different task. Zimbra et al [29] reviewed a number of techniques for classifying sentiment in tweets. They found that due to factors such as the brevity of tweets, Twitter-specific language [30], and a class imbalance [31], classification algorithms achieved an accuracy of around 70%. However, Adwan et al [32] also reviewed a large number

of techniques and they found a mix of accuracy scores, with some papers passing 80% accuracy while others still perform below 80% even with new algorithms [33]. Among those who have improved their accuracy, some only focus on specific politics-related data sets [34], some propose methods that require a large number of steps [35], while others address the issues with tweets, such as Twitter-specific language [36].

Methods

Our entire pipeline is described in Figure 1. We first introduce the data collection and preprocessing. We then detail our topic detection algorithm and procedure of sentiment and emotion classification.

Figure 1. Pipeline of our text analysis. BERT: bidirectional encoder representations from transformers; ONMF: online nonnegative matrix factorization; VADER: Valence Aware Dictionary and sEntiment Reasoner.



Data Collection

We adopted the coronavirus tweets data set [37] as our data source, which uses over 90 keywords and hashtags [38] to monitor the real-time coronavirus-related tweets from February 05, 2020 till present. Since the US Food and Drug Administration authorized Pfizer-BioNTech COVID-19 vaccine and Moderna vaccine for emergency use in mid-December, we only kept tweets that were created during a 60-day period between December 16, 2020 and February 13, 2021 for extracting discussion topics and their sentiment from the general public about COVID-19 vaccines. Owing to the data sharing policy of Twitter, the coronavirus tweets data set only shares the IDs of the collected tweets. Therefore, we employed Twitter's tweet lookup application programming interface [39] to retrieve the content and metainformation of each retained tweet. In order to downsize the corpus and retain vaccine-related tweets, we only selected tweets that contained at least 1 keyword in our predefined keyword list: "vaccine," "vaccines," "#vaccine," "#vaccines," "corona vaccine," "corona vaccines," "#coronavaccine," "#coronavaccines," "pfizer," "biontech," "moderna," "Pfizer-BioNTech," "Pfizer/BioNTech," "Pfizer BioNTech," "#PfizerBioNTech," "COVAX," "COVAX," "Sinopharm," "Sinovac," "AstraZeneca," "Sputnik V," and "Gamaleya." The list of keywords was generated by the authors with the intention of collecting data on COVID-19 vaccines in general as well as the specific vaccines that were available to the public at the start of the data collection period. We also filtered out duplicated content, for example, retweets and non-English contents for providing more consistent data. Thus, we had 7,948,886 tweets for further text analysis.

Topic Detection

There are 2 types of models for topic detection: latent Dirichlet allocation [40] and NMF [41]. In this study, we chose NMF because its superiority has been proved in extracting topics from tweets [42]. NMF is a matrix factorization algorithm that learns and maps high-dimensional data into low-dimension representations. In this study, our tweet corpus $V \in$  is represented as a matrix with rows (words) and columns (tweets). After the preprocessing process detailed in Figure 1, we constructed the corpus using tf-idf weighting scheme:

Where $n_{i,j}$ is the count of word $i \in F$ appearing in tweet $j \in N$, and $N^{(i)}$ is the number of tweets containing word i. With such weighting scheme, the word has more weights, as it is an important word for a tweet. After encoding the corpus, we apply NMF for extracting topics, whose objective of factorization is as follows:

We can exploit the topic word distribution using $W \in$  because each column represents a hidden topic, where the representative words will be encoded more weights. $H \in$  can be served as document topic distribution since each column indicated a topic weight distribution of each tweet. For coping with the large-scale tweets and the subsequent memory issue, we adopted online NMF (ONMF) [10,43] to solve both and in an online learning fashion. Specially, the whole tweet corpus will be divided into a set of small batches () and be sequentially used for updating W^q and H^q of each batch. The step for updating the coefficient of the current batch H^q is to fix the word dictionary of previous batch W^{q-1} and find a H that recovers W^q with least error (see line 6 in Algorithm 1). Similarly, to update the dictionary of the current batch W^q,H^q is then fixed, and the best W is solved using line 8 in Algorithm 1. The mathematical details of the 2 updating forms can be seen in Zhao and Tan [43]. As a result, Algorithm 1 is the whole procedure for topic detection.

Note that we will use the final topic word dictionary $W^{q=Q}$ to infer topic weights of each tweet (ie, H). The representative topic of a tweet is determined by selecting the topic with maximum weight: $\text{argmax}_{k \in K} H_{k,j}$ and we recorded the representative topics of all tweets as $H^{rep} \in$  .

Sentiment Analysis

To detect the sentiment conveyed in the tweets, we utilized a two-step approach. In the first step, we computed the polarity score of our tweets, and based on this score, we classified the tweets as either positive, neutral, or negative. In the second step, we classified the emotional content of the tweet into 1 of the 5 emotions: anger, fear, joy, hopefulness, and sadness.

Polarity Classification

The first classification step was performed using the VADER (Valence Aware Dictionary and sEntiment Reasoner) Python library [14]. The VADER library is a rule-based model for general sentiment analysis. VADER is constructed using existing well-established sentiment lexicons such as Linguistic Inquiry and Word Count and supplemented using lexical features commonly used to express sentiment in social media. After expanding using social media lexical terms, VADER was then human validated and is currently considered a gold standard in social media lexicons [44]. VADER evaluates the sentiment of each tweet by returning a compound sentiment score between –1 and 1. Based on the classification thresholds determined by the developers of the library, we assigned a negative sentiment to a compound score less than or equal to –0.05, a positive sentiment to all compound scores greater than or equal to 0.05, and a neutral sentiment to a compound score between –0.05 and 0.05 [14]. Since VADER is more sensitive to expressions of sentiment in the social media context, it performs better than other rule-based classification algorithms in this context [45]. It has been found that VADER outperforms individual human raters [14] in the F1 score.

Emotion Classification

In the second step, we separated our data into positive, negative, and neutral and detected 1 of the 2 emotions for positive polarity, that is, joy and hopefulness, and 1 of the 3 emotions for negative polarity, that is, anger, fear, and sadness. Since

VADER only includes positive, negative, and neutral sentiment, to detect more fine-grained emotions, we used zero-shot classification, an unsupervised method for discovering the applicable emotion for each tweet. Zero-shot classification is used in machine learning to classify things such as images and text [46,47]. We detected the emotion by finding the BERT (bidirectional encoder representations from transformers) [48] embeddings of the tweets and of the emotion words (fear, joy, hopefulness, anger, and sadness) and then computing the cosine similarity of the emotion words and each tweet and selecting the emotion with the highest cosine similarity as the emotion associated with the tweet.

BERT [48] is a word representation model that uses unannotated text to perform various natural language processing tasks such as classification and question answering. By considering the context of a word using the words both before and after the word, we were able to produce embeddings for words that are more context aware. Our research used the pretrained sentence BERT [49] model to generate the embedding vectors for our emotion classification task.

Given our tweet corpus $V \in$ $^{F \times N}$, we represented our emotion results as $X \in$ $^{C \times N}$, where C is number of emotion categories and N is the number of tweets. For each emotion, we computed an embedding vector Ec_i where $i = 1,\ldots,C-1$ and for each tweet, we computed an embedding vector Ev_j where $j = 1,\ldots N$ using the pretrained sentence BERT model. To populate our emotion matrix, we first computed the VADER sentiment score and assigned the score to the neutral category in our matrix. We then computed the cosine similarity between each of the remaining *C-1* categories and each tweet using the following equation:

Where $i = 1,\ldots,C-1$ and $j = 1,\ldots N$. We assigned a representative emotion to each tweet by finding $\text{argmax}_{c \in C} X_{c,i}$ for $i = 1,\ldots N$, resulting into $X^{rep} \in$ $^{1 \times N}$, which records representative emotions of all tweets.

Combining Topic and Sentiment

We merged the detected topics $H^{rep} \in$ $^{1 \times N}$ and identified emotions $X^{rep} \in$ $^{1 \times N}$ using the unique IDs of tweets, resulting into a matrix $O \in$ $^{2 \times N}$. By referring to the timestamp of each tweet, we were able to track the changes in the sentiment and topics over time to see how the public responded to the different vaccines as time passed.

Results

Tracking Topic Over Time

We started by generating 50 topics (K=50) using the ONMF algorithm with 2000 as the batch size (s=2000). In order to only retain the representative topics about vaccines, we calculated the ratio of each topic k using the following equation:

With the topic ratio, we could estimate how many tweets belonged to topic k and filtered out 38 insignificant topics whose topic ratios were below the average, that is, 2%. As listed in Table 1, the remaining 12 topics were then labeled by reviewing the most contributed keywords in each topic.

Table 1. The most significant 12 vaccine-related topics and the percentage of tweets in each topic (N=7,948,886).

Topic ID	Topic label	Topic totals, n (%)
1	Vaccination of frontline workers	690,357 (8.7)
2	Access to vaccines, signing up online	658,115 (8.3)
3	South African variant	593,425 (6.8)
4	Biden stimulus plan	540,065 (3.7)
5	mRNA vaccines	292,217 (3.2)
6	Complaints about pharmaceutical company profits	250,337 (3.1)
7	Vaccine conspiracy theories online	243,934 (2.9)
8	Trials in non-mRNA vaccines	232,780 (2.5)
9	Vaccine distribution in Canada	202,164 (2.5)
10	Supply and herd immunity	198,967 (2.5)
11	Genetic concerns about vaccines and kids	194,578 (2.2)
12	Low distribution of AstraZeneca vaccine	189,468 (2.1)

Figure 2 shows the trends for the 6 most important topics whose topic ratios were greater than 3%. The most important topic discussed the vaccination of frontline workers (topic 1), wherein the ratio stayed above 7% from mid-December to mid-February. Such a high attention of topic 1 indicated that people were concerned about the eligibility of vaccination and relevant plans from governments, especially in the early roll-out phases (ie, phase 1a and phase 1b). A discussion peak was observed on December 20, 2020, and December 21, 2020, as shown in Figure 3 because some congress members got vaccinated before frontline workers, which triggered heated debates. The

representative tweets of topic 1 during that period were as follows:

> …*Speakers: Finding eligible #candidates for #COVID19 vaccine have to be ensured* [December 20, 2020]
>
> …*What makes Blumenthal and Murphy eligible for the vaccine. Are they frontline workers?* [December 20, 2020]
>
> …*They are depriving frontline workers of a vaccine. They are literally scum.* [December 20, 2020]

The above tweets reported that the priority of accepting COVID-19 vaccines and justice were also critical concerns of the people. The second largest topic was about access to vaccines—signing up online (topic 2). After the early distribution of the vaccines, we observed that people started to be concerned about the access to the vaccines, resulting in a growth starting from the last week of 2020. The following relevant tweets of topic 2 indicated that governments and health care facilities [50] began implementing online appointments for vaccination.

> …*Heads up Ottawa County-you can sign up for vaccine notifications online* [May 1, 2021]
>
> …*@drharshvardhan: Please implement Aadhaar-based online appointment for Covid vaccine as applicable in case of appointment for passport and driving license* [May 1, 2021]
>
> …*A step-by-step guide for the online vaccine appointment process* [wenatcheeworld]

Figure 2. The topic trends for the most significant 6 topics that had a topic ratio above 3%.

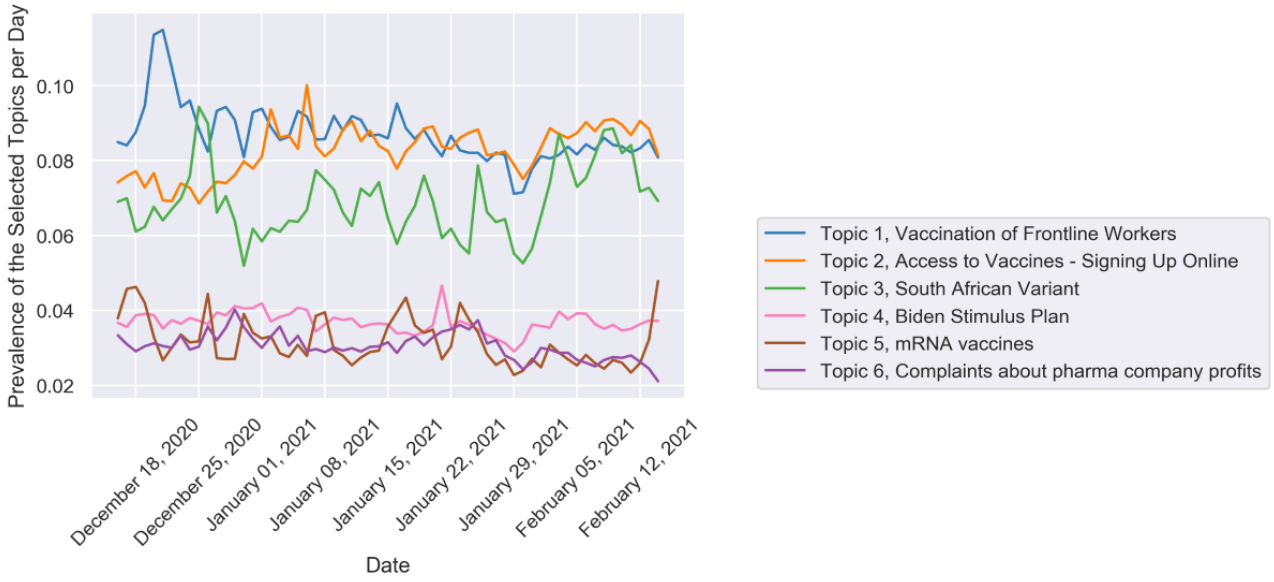
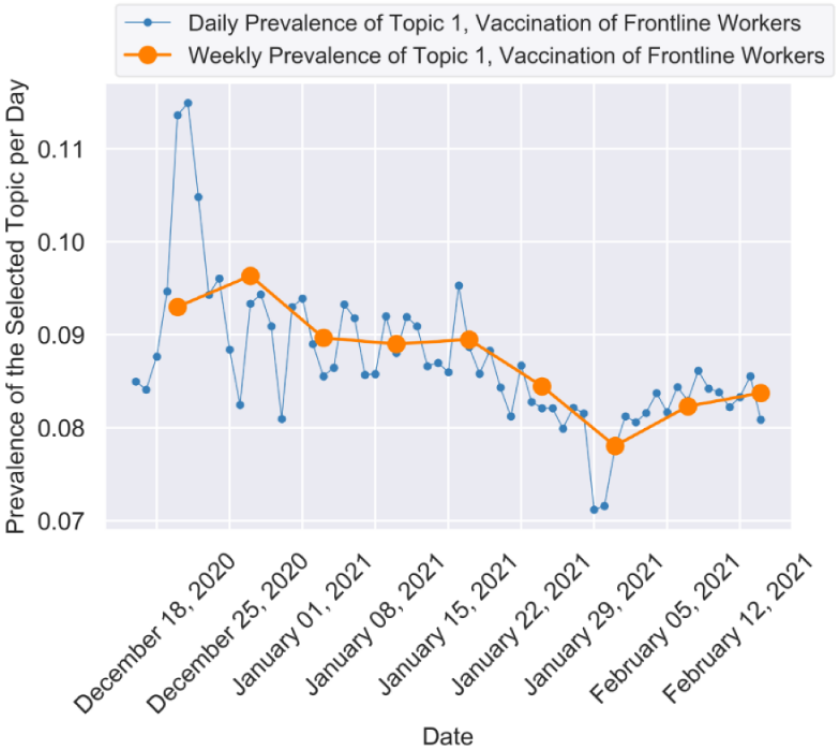


Figure 3. The daily and weekly trends of topic 1.



The third largest topic was about the South African variant (topic 3), which peaked in late December and was relevant to the announcement of the South African variant from the South African health officials [51] and the first variant case detected in the United States [52], resulting in a rising trend from the late January of 2021. The high ratio of topic 3 indicated that the effectiveness of the released vaccines was of great concern, and people were skeptical and conservative. Finally, comparing the top 3 significant topics, topics 4-6 (ie, Biden Stimulus Plan, mRNA Vaccines, and Complaints about pharma company profits) showed relatively steady discussion trends.

Figure 4 presents the remaining 6 important topics. Topic 8 (Trials in non-mRNA vaccines) and topic 12 (Low distribution of AstraZeneca vaccine) had apparent spikes on January 29, 2021. For the peak of topic 8 (see Figure 5), we found that the emerging event "the positive trial results of Johnson & Johnson's single-shot vaccine" caught the public's eye and stimulated discussion. The relevant contents were tweeted frequently at that moment, and most of them cited news sources [52-54]. The sample tweets were as follows:

> …*Single-shot Johnson & Johnson vaccine 66 percent effective against moderate and severe illness* [cited from Washington post, January 29, 2021]
>
> …*Johnson & Johnson says its single-shot vaccine is 66% effective overall at preventing moderate to severe illness* [cited from Fox8live, January 29, 2021]
>
> …*Johnson & Johnson's one-shot #COVID19 vaccine is effective against severe disease* [cited from Science News, January 29, 2021]

The spike on Topic 12 (see Figure 6) can be related to the dispute between the European Union and AstraZeneca in the third week of January [55]. The citizens in the European Union expressed their depression about the delay and inefficiency of vaccine ordering, and the representative tweets were as follows:

> …*EU vaccine delays prompt press frustration* [January 28, 2021]
>
> …*AstraZeneca is supplying European Union vaccine at cost with zero profit. European Union has a cheek to talk about suing AZ* [January 29, 2021]
>
> …*The actions of the European Union to cover their abject failure to obtain vaccine* [January 29, 2021]

Figure 4. The topic trends for the rest of the topics that had a topic ratio below 3%.

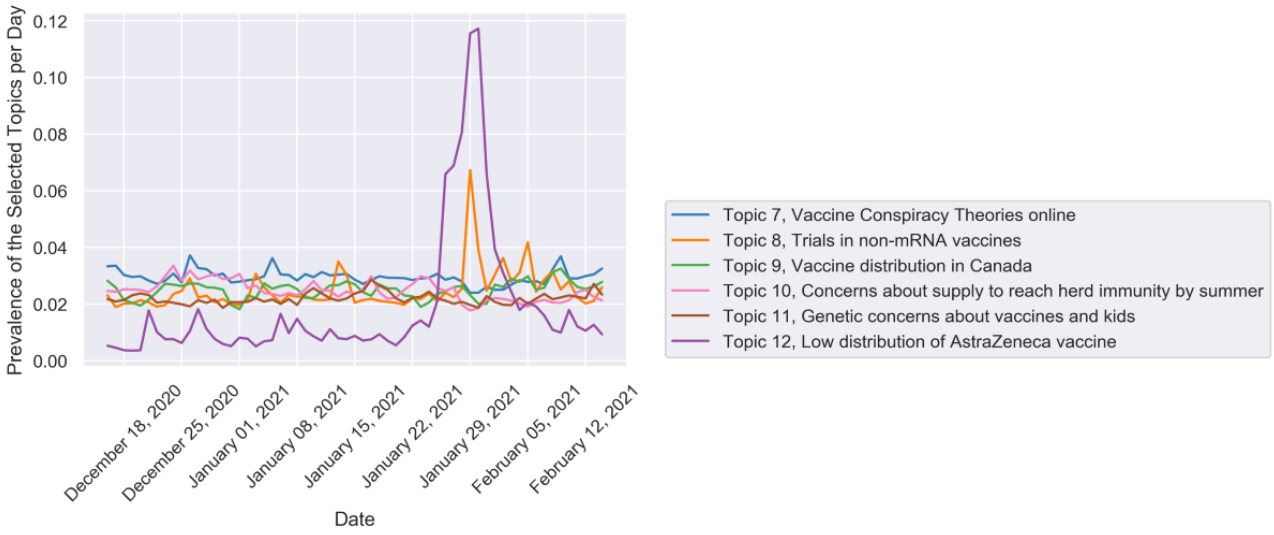


Figure 5. The daily and weekly trends of topic 8.

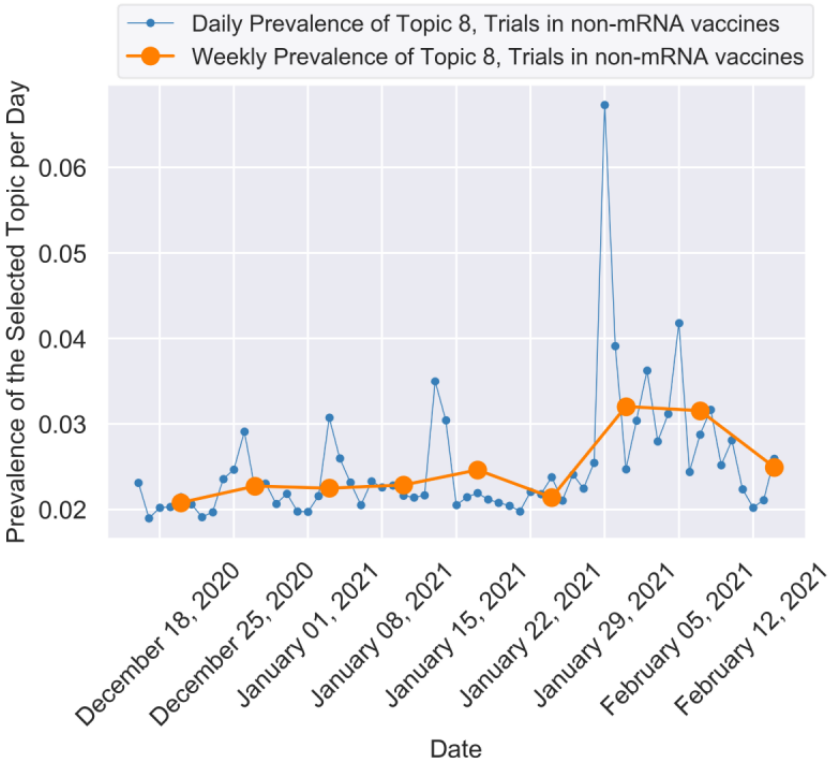
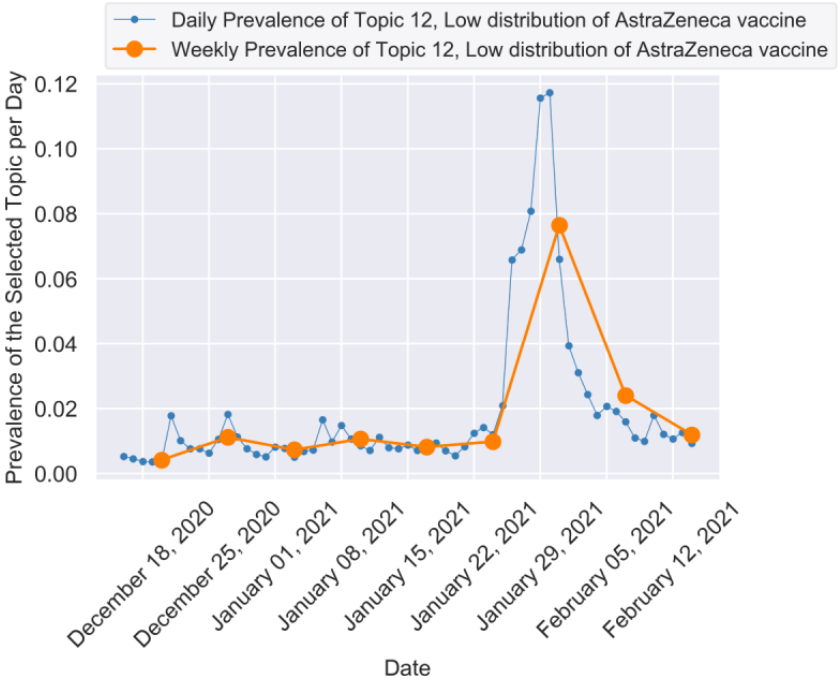


Figure 6. The daily and weekly trends of topic 12.



Tracking Sentiment Over Time

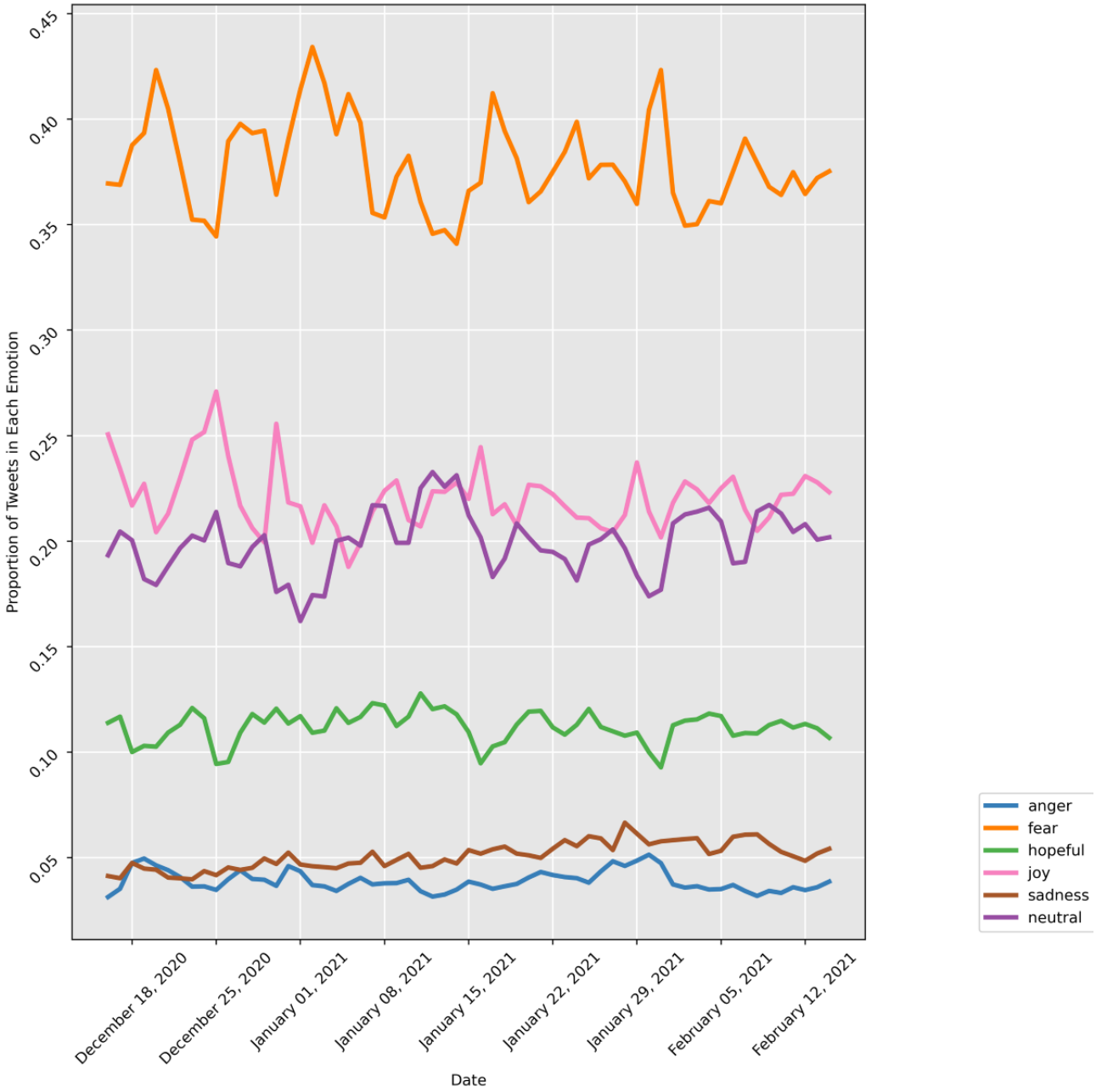
When summarizing the sentiment in all 7,948,886 million tweets throughout the entire period, we observed that the top emotion that appeared in our tweets was fear followed by joy. The percentage of tweets containing each of the emotions from the tweets collected during the entire period is described in Table 2.

Figure 7 presents the trends of the 5 emotions during the 60-day period starting from December 16, 2020. It shows that fear was consistently the most frequently detected emotion. Joy was the second most common emotion followed by neutral sentiment. Hopefulness, sadness, and anger were reflected in a lower proportion of tweets. The Augmented Dickey-Fuller test showed that all emotions, except for sadness, were stationary throughout the entire period, while sadness increased throughout the period.

Table 2. Proportion of tweets by emotion (N=7,948,886).

Sentiment	Tweet totals, n (%)
Negative emotion	
Fear	3,002,467 (37.8)
Sadness	406,095 (5.1)
Anger	312,398 (3.9)
Neutral emotion	1,582,221 (19.9)
Positive emotion	
Joy	1,751,729 (21.9)
Hopefulness	406,095 (11.2)

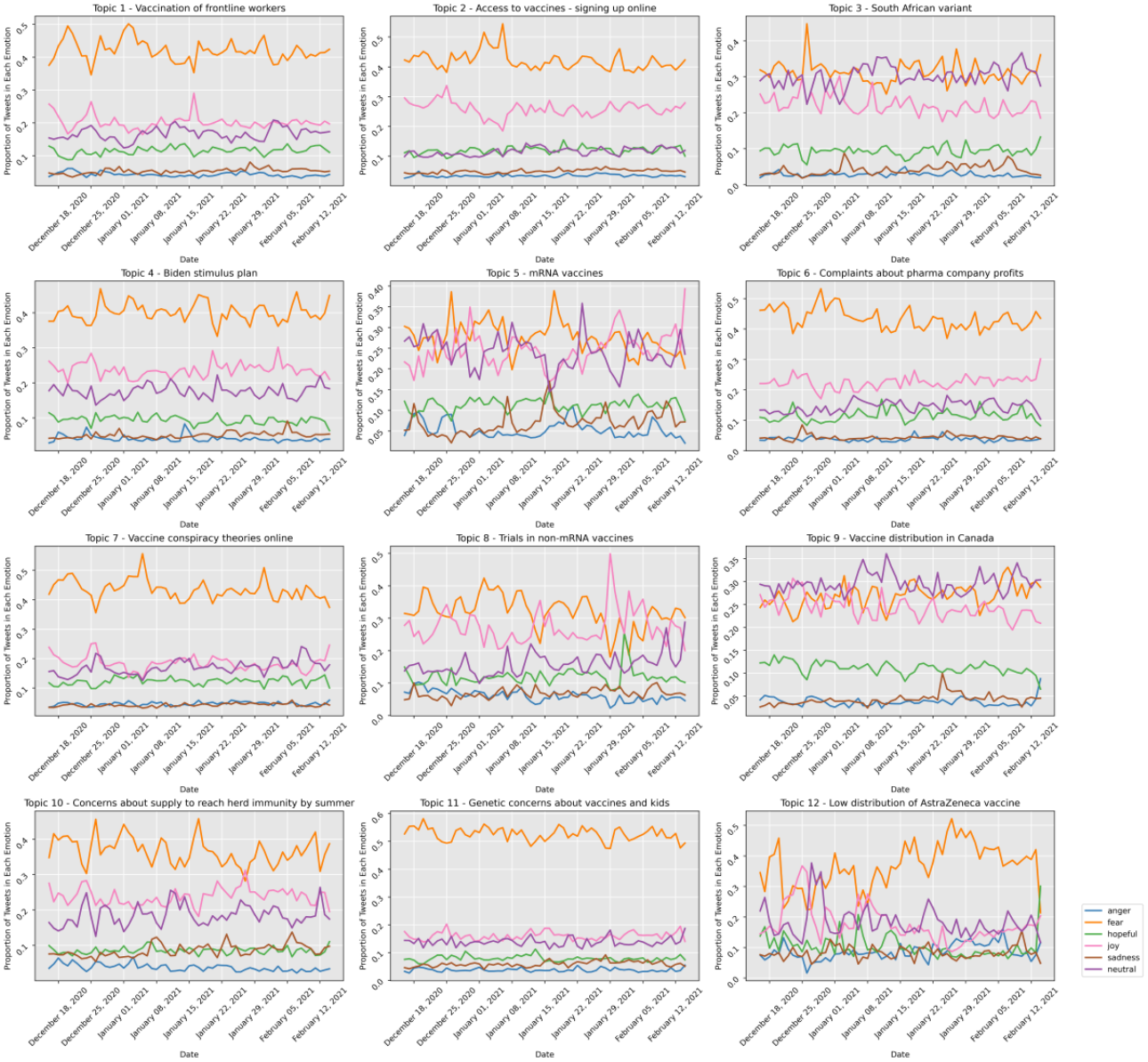
Figure 7. Emotional trends over time.



Sentiment Trends in the 12 Detected Topics

To analyze the sentiment in each of the top 12 topics, we plotted the proportion of each sentiment for each topic and observed how the percentages changed over time. The percentage of tweets in each sentiment is described in Figure 8.

Figure 8. Emotional trends over time regarding topics 1-12.



Negative Sentiment

Negative sentiment was the leading sentiment in our tweets, with fear as the leading emotion.

Fear

Our graphs show that for the majority of topics, fear was the most observed emotion. In topics 1, 2, 4, 6, 7, 10, and 11, fear was the most observed emotion throughout the majority of the time period. Topic 1 discussed the vaccination of the frontline workers. Some representative tweets from this topic that contained fear were as follows:

> *…@POTUS Mr. President, I'm really worried about my state (GA) and the rollout with vaccines. There doesn't seem to be a plan and we are being pushed to have school and teachers are not vaccinated and barely hospital workers and senior citizens have.* [February 13, 2021]

> *…@CTVNews Hi, I am an Ontario resident and my wife works at X-ray & Ultrasound clinic in Newmarket. I am worried about her and her Associates not getting the vaccine along with hospital workers, she sees patients every day and I think they must be vaccinated ASAP. Thanks, Charlie* [January 12, 2021]

The main theme in these tweets was fear that frontline workers would not be vaccinated soon enough and that they would not receive the highest priority in the vaccine rollout.

Topic 2 discussed access to vaccines and signing up online. The most prominent emotion in this topic throughout the period was fear. Below are some example tweets from this topic:

> *…I got vaccinated. I'm Latino. Making my appt was confusing and my 2nd appt kept getting cancelled even though I work in a hospital. Also lots of fear, distrust and misinformation, people saying the vaccine gives you the 666 sign of the devil, etc. Many people are scared of it. https://t.co/98pguyfuiJ* [January 31, 2021]

> *…I'm very concerned my 82-year-old mother must go online to a website; register for the vaccine in Nevada that is still not available until February 28? How do we solve this for our older generation with no computer knowledge to help them get vaccines quicker?* [January 06, 2021]

We could identify with the struggle to obtain an appointment for vaccinations in many states. There were also technical difficulties with multiple websites that caused concern among many Twitter users.

Topic 4 discussed Biden's stimulus plan. The plan contained funding for COVID-19 vaccine distribution [56]. In January, the gap between fear and joy widened; however, after Biden took office in January, joy increased and the gap between fear and joy became smaller.

Examples of tweets from topic 4 that conveyed fear are as follows:

> *…@GovInslee I'm a fan Jay, but I'm worried Washington is going to screw up the vaccine distribution.* [January 13, 2021]

> *…@JoeBiden Please save Texas from @GovAbbott's ignorance and massive logistical failures with respect to distribution of the vaccine* [January 17, 2021]

Many of the tweets in this topic conveyed fear with respect to not executing Biden's plan rather than fear of the plan itself.

Anger

While fear was the most prominent emotion followed by joy, some topics contained spikes of anger-related tweets. Topic 5 contained a few spikes of anger. Here are some examples of angry tweets from topic 5:

> *…Coronavirus: European Union anger over reduced Pfizer vaccine deliveries. Why to rely on profiteering Pfizer ? There are other vaccine! https://t.co/E27tWB71IJ* [January 15, 2021]

> *…@latimes Is that why it's killing old people? 20+ dead in Norway alone. Global scientists calling for immediate stoppage of Pfizer drug. Btw it's not a vaccine by definition. Its mRNA therapy. A vaccine uses a dead virus that's incubated and cultured.* [January 16, 2021]

There was anger due to lack of trust of the vaccine manufacturers as well as anger over rumors of deaths and injuries due to the vaccines.

Sadness

Sadness was one of the least prominent emotions in our data. It was the highest in topic 10, which discussed concerns about vaccine supply that would enable reaching herd immunity by summer 2021. Here are some representative tweets from this topic containing sadness:

> *…My dad was so close to getting his vaccine. But he didn't make it. Meredith pays tribute to her father who died 4 days ago with COVID-19. He was a Cumbrian farmer. She describes him as grumpy but in a charming way. https://t.co/OR5NsNVuZG* [January 13, 2021]

> *…@SHCGreen @NicolaSturgeon @jasonleitch @edinburghpaper @lothianlmc @NHS_Lothian @DrGregorSmith Glad to see some people getting the vaccine. Sadly my aunt didn't get to have hers. Died early hours from COVID. Will miss her very much.*

Many of the tweets in this topic containing sad emotion discussed deaths due to COVID-19 that could have been prevented by a quicker vaccine rollout.

Additionally, we saw the following tweets from topic 1 showing sadness:

> *…Some of these are so painful. 65-year-old local pharmacist, kept working, hence couldn't social-distance like, well, a writer. Dead as a consequence. Why frontline workers should be further up in the vaccine queue than even 78-year-olds like me. https://t.co/EYZ6uUr5K8* [January 21, 2021]

> *…An extended family member was a carer in a home, no vaccine, was in a coma for 2 weeks and passed last week. I didn't personally know her but her niece is heartbroken. Thought all care home staff had the vaccine according to the Government.* [February 03, 2021]

These tweets showed sadness and concern that frontline workers would not be vaccinated soon enough and might contract COVID-19.

Neutral Sentiment

Many neutral tweets contained information from news websites or from official sources. As a result, we observed that many of these tweets contained links or media. Neutral sentiment was not the leading emotion in any of the topics; however, we still detected many neutral tweets in all topics. Below are the tweets from different topics containing neutral sentiment from the top 6 topics:

Topic 1:

> *…Westminster residents ages 65 and older are now eligible to receive the COVID-19 vaccine. Read the full press release below for instructions. #westminsterca #covidvaccine #orangecounty https://t.co/7cgiOLQLl5 (January 13 2021)*

> *After 40 hours of work, the volunteers of Broadbent Arena, in Louisville, KY, are eligible for their own vaccines. Every day, the oldest volunteers with 40 hours under their belts get the leftover doses. https://t.co/tB3NY2ECSE (February 04 2021)*

Topic 2:

> *…#Health care workers, anyone 70 years and older, and state/local government employees and contractors who perform #COVID_19 vaccinations and testing in SC can make appointments to get a #vaccine. https://t.co/65iyk1qJWi* [January 15, 2021]

> *…The fastest way to register into this system will be online, WV rolling out new vaccine registration system https://t.co/gTzl9s54vq* [January 22, 2021]

Topic 3:

> *Virus Updates: S. Africa Halts AstraZeneca Shot; COVID Reinfections May Be Overlooked https://t.co/VRvgEd0DDV (February 08 2021)*
>
> *Moderna says it's working on COVID booster shot for variant in South Africa, says current vaccine provides some protection https://t.co/UQLInvRVoO* [January 25, 2021]

Topic 4:

> *…COVID-19 vaccine distribution ramps up for 20 million to be immunized by the start of the new year https://t.co/zWUVzjxTNw* [December 21,2020]
>
> *…The $900 billion stimulus package includes unemployment support of up to $300 per week. The bill also includes $45 billion in support for transportation, $82 billion for schools, $20 billion for coronavirus vaccine distribution and $25 billion in emergency assistance to renters.* [December 20, 2020]

Topic 5:

> *…Sir Ian McKellen says he feels 'euphoric' after receiving the Pfizer/BioNTech vaccine; https://t.co/Jr4XvRUDlH* [December 17, 2020]
>
> *…The @nytimes reported Pfizer announced that they will ship fewer vials of their coronavirus #vaccine to the US, in response to the FDA approving a change to the label saying the vials contain six doses rather than five: https://t.co/w8pmbwWBoB* [January 25, 2021]

Topic 6

> *…Column: Pfizer, Moderna expect billions in profits from COVID vaccines. That's a scandal https://t.co/LIhZT0uTlB* [January 04, 2021]
>
> *…The pharmaceutical company expects around $15 billion of revenue from sales of its COVID-19 vaccine this year, while Wall Street had anticipated $12.7 billion. https://t.co/KkjT4vur1d* [February 02, 2021]

In all topics, there were a multitude of articles and opinion pieces from different media outlets. The articles typically followed the theme in the topic to which they were classified.

Positive Sentiment

Positive sentiment was the second most common in our data and contained 2 emotions: joy and hopefulness.

Joy

In topics 3, 5, 8, and 9, the leading emotion fluctuated throughout the time period. While joy was not the leading topic throughout the entire period, in these few topics, the expression of joy exceeded fear for at least some days during the period.

Topic 5 discussed mRNA vaccines. The vaccines discussed in this topic were only the Pfizer and Moderna vaccines since they were given emergency use authorization for use at the time of data collection.

Examples of tweets from topic 5 that contain joy are as follows:

> *…Congratulations! Still wear your mask and wash those hands, keep yourself safe! I get my second one tomorrow. Moderna or Pfizer? I got the Pfizer, people I know who have gotten their second dose are having a rough couple days. Molly must be so happy!* [February 06, 2021]
>
> *…Pfizer and Moderna seem to be the clear vaccine winners* [January 29, 2021]
>
> *…Wow vaccine is looking awesome. I'm super impressed with Moderna and Pfizer-- and in record time:)* [February 13, 2021]

Topic 8 discussed trials of non-mRNA vaccines. While there were many days where fear was the top emotion in this topic, joy was a prominent emotion in the tweets discussing this topic since it was the leading emotion in some days during the time period. Below are examples of tweets containing joy from topic 8:

> *…Waking up to great news on the COVID vaccines front: Novavax 89% efficacy, Johnson&Johnson single dose, and 100% protected from death 28 days after single shot, AstraZeneca fully approved in EU. #VaccinesSaveLives* [January 30, 2021]
>
> *…I participated in the Janssen/Johnson & Johnson #ENSEMBLE2 COVID-19 vaccine trial Only time will tell whether I received vaccine or placebo. But so happy to be taking part. Thanks to all the amazing staff at St. Thomas' Hospital London @GSTTnhs #janssen # COVID-19 https://t.co/brHCDOJC6u* [January 13, 2021]

The possibility of having a variety of vaccines that were approved was a cause for joy for many Twitter users.

Hopefulness

Topic 12 contained a spike of hopefulness in late December. This topic discussed the concerns of low distribution of the AstraZeneca vaccine. Below are examples of hopefulness in topic 12:

> *…Hopefully the Oxford vaccine can help out those countries, not just in EU, who don't have enough vaccines. https://t.co/BrC3dJ71tN* [December 21, 2020]
>
> *…@ChristinaSNP What a smashing day. Sun is shining, a British vaccine for COVID is approved. The European Union approved #brexit deal is being flown in at the moment. When signed the @theSNP can surely let us know their plans for our future, not merely criticize others like #NoDealNicola #BetterTogether* [December 30, 2020]

We can see that there was some hopefulness regarding the distribution of the AstraZeneca vaccine. However, hopefulness was not the leading emotion during that time period. Additionally, by the end of the time period, fear was by far the most prominent emotion.

Discussion

Principal Results

Our study aimed to detect the topics and sentiments of public concerns of COVID-19 vaccines by performing a trend analysis on tweets collected for a period of 60 days when the vaccines were started in the United States and to make practical suggestions to address the concerns of different groups in the public as expressed on social media. Approximately 8 million tweets related to COVID-19 vaccines were collected and 12 important topics were selected for analysis. The 3 most important topics with the highest topic ratio were "Vaccination of Frontline Workers," "Access of Vaccines–Signing Up Online," and "South African Variant." The other topics were mostly related to the concerns about the vaccines as well as their supply and distribution. There were also topics related to the stimulus plan, profits of pharmaceutical companies, and conspiracy theories. Through the trend analysis, it was found that the peaks of the topics were impacted by the events reported in the news and spread through social media. The sentiment analysis showed that 46.9% (3,720,960/7,948,886) of the tweets were negative with emotions of mostly fear, followed by sadness and anger, 33.2% (2,645,705/7,948,886) of tweets were positive with emotions of joy and hopefulness, and 19.9% (1,582,221/7,948,886 tweets) of tweets were neutral. Fear and joy were the most detected emotions. Our analysis examined the 6 different sentiments detected in the tweets and their change over time. We observed that the keywords in each topic did not change much over time; therefore, we were able to track our tweets using the same topics throughout the entire period. In some topics, sentiment was stationary throughout the period, while in others, there were significant trends. For example, in topic 3 "South African variant," we saw an increase in fear and neutral sentiment over the period and a decrease in joy at the same time. Similarly, we saw an increase in fear and a decrease in joy in topic 12 "Low Distribution of the AstraZeneca Vaccine." Overall, fear was the top emotion followed by joy. Sadness and hopefulness remained low in most topics throughout the entire period.

Identifying Specific Concerns in Each Topic by Using Emotional Content

The most notable conclusion from the data is that the main reaction to the COVID-19 vaccines on social media was fear. However, we could identify every one of the emotions in each topic. In each topic, we could find tweets related to the topic containing each of the emotions. By looking at the representative tweets for each topic and each emotion, we were able to learn what specific concerns people may have that may lead to vaccine hesitancy. For example, from topic 1, we found that there was fear surrounding the vaccination of government officials prior to frontline workers. By addressing this publicly and assuring the public that the frontline workers would receive their vaccines as soon as possible, this would help to build public confidence in the vaccine rollout. We could also identify tweets that contained sadness to identify further concerns about the rollout to frontline workers and see Twitter users expressing sadness regarding frontline workers possibly dying due to lack of vaccines. This could be addressed by being more transparent about vaccination timelines or by advocating for more vaccine supply. By being aware of specific concerns as they happen (eg, the vaccination of frontline workers), we will be better able to address the source of concern and reduce vaccine hesitancy.

Vaccine Administration

The very first dose of the mRNA COVID-19 vaccine by Pfizer and BioNTech was given to a health care worker on December 14, 2020. This may explain why the most significant topic at the start of the study was vaccination of frontline workers (topic 1). As more vaccines were administered, reports of anaphylaxis began to surface, especially with the Moderna vaccine [57]. In the United States alone, 10 cases of anaphylaxis were reported after 4,041,496 (0.002%) vaccines were given between December 21, 2020, and January 10, 2021. This created fear as indicated in the trend, and fear dominated all other emotions throughout the course of the study period. It will be interesting to find out how many of these tweets are from health care personnel versus that from the general public. According to the Centers for Disease Control recommendation, both health care personnel and residents of long-term care facilities were the first to be offered the COVID vaccine [58]. Health care personnel include both clinical and nonclinical staffs such as those who work in food, environmental, and administrative services. It can be assumed that clinical staff have adequate knowledge of vaccines and need not to be afraid to take it. Therefore, public health authorities and health care systems can focus on educating the adverse effects of the vaccine to the nonclinical staff and the general public. For example, anaphylactic reactions occur mostly in people who have a similar reaction to other food and drugs, and it usually occurs within minutes after injection. Better understanding of the adverse effects will minimize fear of the vaccine and thus reduce vaccine hesitancy.

Access to Vaccines

Signing up online (topic 2), vaccine distribution in Canada (topic 9), and low distribution of AstraZeneca vaccine (topic 12) can all be categorized as accessibility of vaccines. A good amount of positive emotion all through the study period in topic 2 indicated that there was a sense of hope in the midst of the daily rising COVID cases. There is still a large amount of fear regarding COVID-19. It may be the fear of the inability to obtain an appointment for the vaccine. Unlike the United States, Canada does not have her own domestic manufacturers to produce vaccines. As a result, Canada relies on international vaccine manufacturers. The advance purchase contract was signed but there was no specific date for delivery except for "first quarter of 2021." There was a shortage of supply of vaccines in Canada because of which the Canadian government prioritized giving the first dose to the population first and the second dose 16 weeks later [59] as opposed to after 3 or 4 weeks. The European Union was furious when in early January, AstraZeneca announced that there would be 60% fewer doses of vaccines than it had promised to deliver in the first quarter of 2021. The spikes of fear and anger emotions during this period in topic 12 were the direct reflection of this news. Being able to have access to the vaccines is important once COVID-19

vaccines are authorized for emergency use. Therefore, public health authorities must have plans to work with vaccine manufacturers to manufacture and deliver the vaccines in a timely manner. The transparency of the access information from social media and public health officials is helpful to reduce the fear and anger in the public.

Practical Implications

In December 2020, the World Health Organization released a safety surveillance manual for COVID-19 vaccines. This manual addressed a number of topics with regards to vaccine administration, including how to communicate information regarding the vaccine on social media [60]. Among other points, the report offers proposals to listen proactively and craft tailored messages to different audiences and address specific concerns of different groups. Using this research, we can take the World Health Organization's recommendations to provide more specific advice to clinicians and policy makers. To address specific concerns, we divided the 12 topics into 3 groups: favoring vaccines, vaccine hesitant, and vaccine opposed.

Favoring Vaccines

The topics that leaned toward those who favor vaccines were topic 1 (vaccination of frontline workers), topic 2 (access to vaccines–signing up online), topic 9 (vaccine distribution in Canada), topic 10 (concerns about supply to reach herd immunity by summer), and topic 12 (low distribution of AstraZeneca vaccine). While these topics also produced negative feelings of fear, anger, and sadness, these negative feelings were regarding concern about not having enough vaccines or not having access to vaccines fast enough. It is crucial to monitor topics that contain tweets from individuals who do want to get vaccinated and keep them informed. Here are some examples of tweets that conveyed fear or concern by individuals who wanted to get vaccinated:

> *…Anybody know what's going on with BAT 24-hour appts? Are they fully back up and running again after being shut down for lack of vaccine? My second shot is at 2:45 a.m. next week, and I'm wary of getting up in the middle of the night to go down there to find them closed.*
>
> *…To be honest, I'd rather risk my life / keep myself in lockdown, for younger key workers to have the vaccine. They are the ones keeping the country going after all.*
>
> *…Blocking access to a vaccine that could save my life is, oh I don't know, attempted murder? So is exhaling their COVID breath around me, but the former is active and so much more egregious. Ain't nobody got time for that mess.*

Identifying the topics that vaccine-favoring individuals discussed was crucial to reducing their concern. In accordance with the World Health Organization document, communication on vaccine availability should be active and frequent. An example of using the analysis from this study to inform the public is looking at the visualizations in real time to produce the right messaging on social media. We observed a spike in the volume of topic 1 in the week of December 18. Figure 8 shows that the leading emotion for that week and topic was fear; further, there was a spike in fear during that week for topic 1. Therefore, it was crucial to post messages on social media that week that address the public fear that health care workers would not have adequate access to vaccines. Another key component in keeping the public informed was updating official websites with vaccine information very frequently. During the early days of vaccination, there was a lack of information in many states about the timeline of vaccination for each risk group. Providing more information on the rollout schedule would help ease the concern of individuals in this group. It is crucial to look at the tweets that convey fear and anger in these topics to create the right messaging and address points that concern this group of the public.

Vaccine Hesitant

This group of individuals was the most crucial to reach since they can be persuaded to get vaccinated. Topics that discussed vaccine hesitancy were topic 3 (South African variant), topic 5 (mRNA vaccines), topic 8 (Trials in non-mRNA vaccines), and topic 11 (Genetic concerns about vaccines and kids). Below are examples of tweets of the vaccine hesitant from these topics:

> *…Just keep in mind that some small percentage of those who received the vaccine did not develop immunity, during the clinical trials. And its effectiveness against variant strains is still not fully known.*
>
> *…The fact that 3 vaccines all appeared to show lowered effectiveness against the variant from South Africa is not encouraging, and the results Novavax announced Thurs were the 1st to occur outside of a lab, testing how well a vaccine worked in people infected with a new variant.*
>
> *…There were obviously several people in the United Kingdom who had had a severe allergic reaction to this vaccine and had a history of severe allergic reaction, said Offit Several people!!!!! #vaccine*

Like the vaccine favorable group, we should also target this group with facts and do so often. However, with this group, we should focus on messages that can be detected in these topics such as those related to side effects of the vaccine, the efficacy of the different types of the vaccine for the original strain of COVID-19 as well as for variants, and why you can still contract COVID-19 even after being vaccinated. We can craft helpful messaging for this group by looking at the topic and emotion data for these topics. For example, we saw an increase in the volume of topic 3 (South African variant) toward the end of January. The most prominent emotion for that topic during that time was fear. Therefore, we can craft messaging on social media regarding the variant that will help with this fear. As the World Health Organization recommends, we should mainly focus on facts and provide up-to-date information to the public through social media regarding the variant.

Vaccine Opposed

This group was the least likely to be persuaded by messaging on the vaccine but should not be ignored. This is because they produce messaging on social media that may convince others.

Therefore, we should attempt to counter their messaging with up-to-date and correct information. Topics that contained a large number of tweets from individuals that were vaccine opposed were topic 6 (Complaints about pharma company profits), but we can find a small number of tweets from this group in all topics, particularly in tweets that were labeled angry or fearful. Examples of tweets from this group were as follows:

> *…We have been here before with the Nazis and Thalidomide yet the whole world rushes to take an untested vaccine. People are dying after having the vaccine yet no enquiries into what happened just a rapid cremation and silence. We should all be very worried.*
>
> *…I bind you up Satan in the name of Jesus, no weapon formed against us shall prosper, and I mean this vaccine is Satan here. "Mark of the beast" read your bibles people.*
>
> *…He didn't take the vaccine! He's a Eugenics partner with Bill Gates they don't take their own vaccines! How about some proof! He's just trying to coverup the ill side effects and deaths that are already happening!*

Those who were opposed to vaccines were hard to persuade, but we must spread truthful messages to counteract the messages that they spread. Many of the tweets by these individuals did not even discuss concerns that could be addressed but were more about vaccine refusal and the freedom to refuse vaccines. It is important to amplify stories of those who suffered severe consequences by refusing to take the vaccine. This is mostly for the sake of the vaccine hesitant rather than the vaccine opposed. An example of messaging can be obtained by looking at the patterns for topic 6. This topic was stable over time and did not experience any spikes. Therefore, we should stay consistent with our messaging over time and counteract any information on this topic with facts on a consistent basis as recommended by the World Health Organization report.

Limitations of This Research

Limitations of Twitter

Twitter is a large social network with 353 million monthly active users [61]. While this is a significant number of users, there is no guarantee that Twitter users are representative of the global or the US population as a whole. Mislove et al [62] have investigated the ability of Twitter data to represent the US population and have found that areas that are more densely populated tend to be overrepresented in Twitter. Additionally, Gore et al [63] and Padilla et al [64] found geographical bias in their analysis of Twitter data. Both studies found an overrepresentation of urban areas in the demographic data of Twitter users included in their studies. Given this prior research, we must assume that users from urban areas are overrepresented in this data set as well.

Keyword Selection

The keywords that were chosen to generate this data set were selected by the authors. The list of keywords described in the data collection section contains keywords that name the colloquial names for the available vaccines at the time of the study. The list also contains terms such as "vaccine" and "coronavaccine" that were included in order to capture a more general discussion regarding COVID-19 vaccines. The list is not meant to be exhaustive and represents the vaccines publicly available at the start of data collection in December 2020.

Duplicated Tweets

Bots posting on Twitter are a well-documented phenomenon [65-67]. One of the issues our study faced was the duplication of content due to bot activity on the topic of vaccines. Other research has documented bot activity on COVID-19 and COVID-19 vaccine misinformation as well [65,68,69]. The main issue this may cause in our analysis is that bot activity may overinflate the importance of certain topics. To combat this, we deduped the Twitter data as part of our analysis and reduced the number of tweets from approximately 20 million to approximately 8 million tweets.

Conclusion

We used topic detection and sentiment analysis as social media trend analysis to better understand the discourse on COVID-19 vaccines tweets. Using this methodology, we could identify the trending topics that reflected the public concerns on COVID-19 vaccines and their responses to the topics indicated by the polarity and emotions on the sentiments. We found that the administration and access to vaccine were some of the major concerns. While most of the information was received from the internet, they were not directly obtained from the health organization. Misinformation may cause negative emotions. In some cases, conspiracy spreading in social media may cause substantial amount of fear. The findings in social media trend analysis are helpful for the health organizations to develop strategies for better communication to the target groups and assist them in coping with their concerns that cause negative emotions or vaccine hesitancy. Disseminating accurate information of COVID-19 vaccines will reduce the negative emotion caused by misinformation or rumors. A report on COVID-19 vaccines by the World Health Organization suggested careful examination of social media to detect specific concerns regarding the vaccines [60]. By understanding what drives different emotions regarding the vaccines, tailored and targeted communication can be developed to provide authoritative health information, which will be helpful to achieve herd immunity and end the pandemic.

Acknowledgments

This work was supported in part by the National Science Foundation under grant NSF-1741306 and grant IIS-1650531. Any opinions and conclusions or recommendations expressed in this study are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Conflicts of Interest

None declared.

References

1. Oliver SE, Gargano JW, Marin M, Wallace M, Curran KG, Chamberland M, et al. The Advisory Committee on Immunization Practices' Interim Recommendation for Use of Pfizer-BioNTech COVID-19 Vaccine - United States, December 2020. MMWR Morb Mortal Wkly Rep 2020 Dec 18;69(50):1922-1924 [FREE Full text] [doi: 10.15585/mmwr.mm6950e2] [Medline: 33332292]
2. Midoux P, Pichon C. Lipid-based mRNA vaccine delivery systems. Expert Rev Vaccines 2015 Feb;14(2):221-234. [doi: 10.1586/14760584.2015.986104] [Medline: 25540984]
3. Schlake T, Thess A, Fotin-Mleczek M, Kallen K. Developing mRNA-vaccine technologies. RNA Biol 2012 Nov;9(11):1319-1330 [FREE Full text] [doi: 10.4161/rna.22269] [Medline: 23064118]
4. Pardi N, Hogan MJ, Weissman D. Recent advances in mRNA vaccine technology. Curr Opin Immunol 2020 Aug;65:14-20. [doi: 10.1016/j.coi.2020.01.008] [Medline: 32244193]
5. Vergara R, Sarmiento P, Lagman J. Building public trust: a response to COVID-19 vaccine hesitancy predicament. J Public Health (Oxf) 2021 Jun 07;43(2):e291-e292 [FREE Full text] [doi: 10.1093/pubmed/fdaa282] [Medline: 33454769]
6. Harrison EA, Wu JW. Vaccine confidence in the time of COVID-19. Eur J Epidemiol 2020 Apr;35(4):325-330 [FREE Full text] [doi: 10.1007/s10654-020-00634-3] [Medline: 32318915]
7. Chou WS, Budenz A. Considering Emotion in COVID-19 Vaccine Communication: Addressing Vaccine Hesitancy and Fostering Vaccine Confidence. Health Commun 2020 Dec;35(14):1718-1722. [doi: 10.1080/10410236.2020.1838096] [Medline: 33124475]
8. Wilson SL, Wiysonge C. Social media and vaccine hesitancy. BMJ Glob Health 2020 Oct;5(10):e004222. [doi: 10.1136/bmjgh-2020-004206] [Medline: 33097547]
9. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, et al. The COVID-19 social media infodemic. Sci Rep 2020 Oct 06;10(1):16598 [FREE Full text] [doi: 10.1038/s41598-020-73510-5] [Medline: 33024152]
10. Chang C, Monselise M, Yang CC. What Are People Concerned About During the Pandemic? Detecting Evolving Topics about COVID-19 from Twitter. J Healthc Inform Res 2021 Jan 17:1-28 [FREE Full text] [doi: 10.1007/s41666-020-00083-3] [Medline: 33490856]
11. Prabowo R, Thelwall M. Sentiment analysis: A combined approach. Journal of Informetrics 2009 Apr;3(2):143-157. [doi: 10.1016/j.joi.2009.01.003]
12. Asghar MZ, Khan A, Ahmad S, Qasim M, Khan IA. Lexicon-enhanced sentiment analysis framework using rule-based classification scheme. PLoS One 2017;12(2):e0171649 [FREE Full text] [doi: 10.1371/journal.pone.0171649] [Medline: 28231286]
13. Tan L, Phang W, Chin K, Patricia A. Rule-based sentiment analysis for financial news. 2015 Oct 09 Presented at: IEEE International Conference on Systems, Man, and Cybernetics; 2015; Hong Kong, China p. 1601. [doi: 10.1109/smc.2015.283]
14. Hutto C, Gilbert E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. 2014 May 16 Presented at: Proceedings of the International AAAI Conference on Web and Social Media; June 1-4; Ann Arbor, MI, USA p. 216-225 URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14550
15. Ahmad M, Aftab S, Ali I. Sentiment Analysis of Tweets using SVM. IJCA 2017 Nov 15;177(5):25-29. [doi: 10.5120/ijca2017915758]
16. Ahmad M, Aftab S, Salman M, Hameed N, Ali I, Nawaz Z. SVM Optimization for Sentiment Analysis. International Journal of Advanced Computer Science and Applications 2018;9(4):393-398. [doi: 10.14569/ijacsa.2018.090455]
17. Fauzi MA. Random Forest Approach for Sentiment Analysis in Indonesian Language. IJEECS 2018 Oct 01;12(1):46. [doi: 10.11591/ijeecs.v12.i1.pp46-50]
18. Tan S, Cheng X, Wang Y, Xu H. Adapting naive bayes to domain adaptation for sentiment analysis. In: Advances in Information Retrieval. 2009 Apr 06 Presented at: 31th European Conference on IR Research; April 6-9; Toulouse, France p. 337-349. [doi: 10.1007/978-3-642-00958-7_31]
19. Dey L, Chakraborty S, Biswas A, Bose B, Tiwari S. Sentiment Analysis of Review Datasets Using Naïve Bayes‘ and K-NN Classifier. IJIEEB 2016 Jul 08;8(4):54-62. [doi: 10.5815/ijieeb.2016.04.07]
20. Tang D, Wei F, Qin B, Yang N, Liu T, Zhou M. Sentiment Embeddings with Applications to Sentiment Analysis. IEEE Trans. Knowl. Data Eng 2016 Feb 1;28(2):496-509. [doi: 10.1109/tkde.2015.2489653]
21. Li Q, Shah S, Fang R, Nourbakhsh A, Liu X. Tweet sentiment analysis by incorporating sentiment-specific word embedding and weighted text features. 2016 Oct 13 Presented at: IEEE/WIC/ACM International Conference on Web Intelligence; 13-16 October; Omaha, NE, USA p. 568-571. [doi: 10.1109/wi.2016.0097]
22. Arras L, Montavon G, Müller K, Samek W. Explaining recurrent neural network predictions in sentiment analysis. In: The 8th Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media Analysis. 2017 Sep Presented at: The 2017 Conference on Empirical Methods in Natural Language Processing; September 2017; Copenhagen, Denmark p. A. [doi: 10.18653/v1/w17-5221]

23. dos Santos C, Gatti M. Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts. 2014 Presented at: COLING 2014, the 25th International Conference on Computational Linguistics; 2014; Dublin, Ireland URL: https://www.aclweb.org/anthology/C14-1008
24. Wang X, Jiang W, Luo Z. Combination of convolutional and recurrent neural network for sentiment analysis of short texts. 2016 Presented at: COLING 2016, the 26th International Conference on Computational Linguistics; December 20; Osaka, Japan URL: https://www.aclweb.org/anthology/C16-1229
25. Ouyang X, Zhou P, Li CH, Liu L. Sentiment Analysis Using Convolutional Neural Network. 2015 Oct 26 Presented at: 2015 IEEE international conference on computer and information technology; ubiquitous computing and communications; dependable, autonomic and secure computing; pervasive intelligence and computing; 26-28 Oct 2015; Liverpool, UK p. 2359-2364 URL: https://ieeexplore-ieee-org.ezproxy2.library.drexel.edu/document/7363395
26. Vashishtha S, Susan S. Fuzzy rule based unsupervised sentiment analysis from social media posts. Expert Systems with Applications 2019 Dec;138:112834. [doi: 10.1016/j.eswa.2019.112834]
27. Chen Q, Sokolova M. Unsupervised sentiment analysis of objective texts. In: Advances in Artificial Intelligence. Cham: Springer International Publishing; 2019:460-465.
28. Pandarachalil R, Sendhilkumar S, Mahalakshmi GS. Twitter Sentiment Analysis for Large-Scale Data: An Unsupervised Approach. Cogn Comput 2014 Nov 7;7(2):254-262. [doi: 10.1007/s12559-014-9310-z]
29. Zimbra D, Abbasi A, Zeng D, Chen H. The State-of-the-Art in Twitter Sentiment Analysis. ACM Trans. Manage. Inf. Syst 2018 Sep 05;9(2):1-29. [doi: 10.1145/3185045]
30. Bermingham A, Smeaton A. Classifying sentiment in microblogs: is brevity an advantage? 2010 Presented at: Proceedings of the 19th ACM international conference on Information knowledge management; October; New York, USA. [doi: 10.1145/1871437.1871741]
31. Hagen M, Potthast M, Büchner M, Stein B. Webis: An ensemble for Twitter sentiment detection. 2015 Presented at: CIKM '10: Proceedings of the 19th ACM international conference on information and knowledge management; October 26-30, 2010; Toronto, ON, Canada. [doi: 10.18653/v1/s15-2097]
32. Adwan OY, Al-Tawil M, Huneiti A, Shahin R, Abu Zayed A, Al-Dibsi R. Twitter Sentiment Analysis Approaches: A Survey. Int. J. Emerg. Technol. Learn 2020 Aug 14;15(15):79. [doi: 10.3991/ijet.v15i15.14467]
33. Abbas A, Salih A, Hussein H, Hussein Q, Abdulwahhab S. Twitter Sentiment Analysis Using an Ensemble Majority Vote Classifier. Journal of Southwest Jiaotong University 2020;55(1):9. [doi: 10.35741/issn.0258-2724.55.1.9]
34. Wang M, Hu G. A Novel Method for Twitter Sentiment Analysis Based on Attentional-Graph Neural Network. Information 2020 Feb 08;11(2):92. [doi: 10.3390/info11020092]
35. Zarisfi Kermani F, Sadeghi F, Eslami E. Solving the twitter sentiment analysis problem based on a machine learning-based approach. Evol. Intel 2019 Oct 22;13(3):381-398. [doi: 10.1007/s12065-019-00301-x]
36. Chen Y, Yuan J, You Q, Luo J. Twitter Sentiment Analysis via Bi-sense Emoji Embedding and Attention-based LSTM. 2018 Oct Presented at: Proceedings of the 26th ACM international conference on Multimedia; October 22-26, 2018; Seoul, Republic of Korea. [doi: 10.1145/3240508.3240533]
37. Lamsal R. Design and analysis of a large-scale COVID-19 tweets dataset. Appl Intell 2020 Nov 06;51(5):2790-2804. [doi: 10.1007/s10489-020-02029-z]
38. Lamsal R. Coronavirus (COVID-19) tweets data set. IEEE Data Port. URL: https://ieee-dataport.org/open-access/coronavirus-covid-19-tweets-dataset [accessed 2021-08-31]
39. Tweet lookup. Twitter Developers. URL: https://developer.twitter.com/en/docs/twitter-api/tweets/lookup/introduction [accessed 2021-08-31]
40. Blei D, Ng A, Jordan M. Latent dirichlet allocation. The Journal of Machine Learning Research. 2003 Mar 01. URL: https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf [accessed 2021-10-13]
41. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature 1999 Oct 21;401(6755):788-791. [doi: 10.1038/44565] [Medline: 10548103]
42. Chen Y, Zhang H, Liu R, Ye Z, Lin J. Experimental explorations on short text topic mining between LDA and NMF based Schemes. Knowledge-Based Systems 2019 Jan;163:1-13. [doi: 10.1016/j.knosys.2018.08.011]
43. Zhao R, Tan VYF. Online Nonnegative Matrix Factorization With Outliers. IEEE Trans. Signal Process 2017 Feb 1;65(3):555-570. [doi: 10.1109/tsp.2016.2620967]
44. Bonta V, Kumaresh N, Janardhan N. A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis. AJCST 2019 Mar 05;8(S2):1-6. [doi: 10.51983/ajcst-2019.8.s2.2037]
45. Elbagir S, Yang J. Twitter Sentiment Analysis Using Natural Language Toolkit and VADER Sentiment. 2019 Mar 13 Presented at: Proceedings of the International MultiConference of Engineers and Computer Scientists; 13-15 March, 2019; Hong Kong, China p. 5. [doi: 10.1142/9789811215094_0005]
46. Rios A, Kavuluru R. Few-shot and zero-shot multi-label learning for structured label spaces. 2018 Presented at: Conference on Empirical Methods in Natural Language Processing; October-November 2018; Brussels, Belgium p. A. [doi: 10.18653/v1/d18-1352]
47. Biswas S, Annadani Y. Preserving semantic relations for zero-shot learning. 2018 Presented at: The IEEE Conference on Computer Vision and Pattern Recognition; Jun 18-22; Seattle, WA, USA. [doi: 10.1109/cvpr.2018.00793]

48. Devlin J, Chang M, Lee K, Toutanova K. BERT: pretraining of deep bidirectional transformers for language understanding. 2019 Presented at: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics; June 2019; Minneapolis, Minnesota. [doi: 10.18653/v1/n18-2]
49. Reimers N, Gurevych I. Sentence-BERT: Sentence Embedding using Siamese BERT-Networks. 2019 Presented at: The 2019 Conference on Empirical Methods in Natural Language Processing; November; Hong Kong, China URL: http://arxiv.org/abs/1908.10084 [doi: 10.18653/v1/d19-1410]
50. Happold M. A step-by-step guide for the online vaccine appointment process. The Wenatchee World. URL: https://www.wenatcheeworld.com/news/coronavirus/a-step-by-step-guide-for-the-online-vaccine-appointment-process/article_25dcefa8-6725-11eb-bbf2-df282a071772.html [accessed 2021-08-31]
51. South Africa announces a new coronavirus variant. The New York Times. URL: https://www.nytimes.com/2020/12/19/world/south-africa-announces-a-new-coronavirus-variant.html [accessed 2021-08-31]
52. Single-shot Johnson & Johnson vaccine prevents illness but shows the threat of variants. The Washington Post. URL: https://www.washingtonpost.com/health/2021/01/29/covid-vaccine-johnson-and-johnson/ [accessed 2021-08-31]
53. J&J; 1-dose shot prevents COVID-19, but less than some others. Fox8live. URL: https://web.archive.org/web/20210129132041/https://www.fox8live.com/2021/01/29/jj-dose-shot-prevents-covid-less-than-some-others/ [accessed 2021-08-31]
54. One-shot COVID-19 vaccine is effective against severe disease. Science News. URL: https://www.sciencenews.org/article/covid-19-coronavirus-vaccine-johnson-and-johnson-results-variants [accessed 2021-10-13]
55. COVID: EU-AstraZeneca disputed vaccine contract made public. BBC News. URL: https://www.bbc.com/news/world-europe-55852698 [accessed 2021-08-31]
56. Higgins T. COVID relief bill includes about $8 billion for vaccine distribution. CNBC. URL: https://www.cnbc.com/2020/12/21/covid-relief-bill-includes-about-8-billion-for-vaccine-distribution.html [accessed 2021-08-31]
57. CDC COVID-19 Response Team, FoodDrug Administration. Allergic Reactions Including Anaphylaxis After Receipt of the First Dose of Pfizer-BioNTech COVID-19 Vaccine - United States, December 14-23, 2020. MMWR Morb Mortal Wkly Rep 2021 Jan 15;70(2):46-51 [FREE Full text] [doi: 10.15585/mmwr.mm7002e1] [Medline: 33444297]
58. Dooling K, McClung N, Chamberland M, Marin M, Wallace M, Bell BP, et al. The Advisory Committee on Immunization Practices' Interim Recommendation for Allocating Initial Supplies of COVID-19 Vaccine - United States, 2020. MMWR Morb Mortal Wkly Rep 2020 Dec 11;69(49):1857-1859 [FREE Full text] [doi: 10.15585/mmwr.mm6949e1] [Medline: 33301429]
59. COVID-19 vaccination coverage in Canada. Government of Canada. URL: https://health-infobase.canada.ca/covid-19/vaccination-coverage/ [accessed 2021-05-21]
60. COVID-19 vaccine: safety surveillance manual. World Health Organization. URL: https://www.who.int/vaccine_safety/committee/Module_Communication.pdf?ua=1 [accessed 2021-08-30]
61. The Global State of Digital 2021. Hootsuite. URL: https://www.hootsuite.com/pages/digital-trends-2021 [accessed 2021-08-30]
62. Mislove A, Lehmann S, Ahn Y, Onnela J, Rosenquist J. Understanding the Demographics of Twitter Users. 2011 Jul 17 Presented at: Proceedings of the International AAAI Conference on Web and Social Media; July 17-21; Barcelona, Spain p. 554-557.
63. Gore RJ, Diallo S, Padilla J. You Are What You Tweet: Connecting the Geographic Variation in America's Obesity Rate to Twitter Content. PLoS One 2015;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]
64. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]
65. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. Am J Public Health 2018 Oct;108(10):1378-1384. [doi: 10.2105/AJPH.2018.304567] [Medline: 30138075]
66. Lee K, Eoff B, Caverlee J. Seven months with the devils: A long-term study of content polluters on Twitter. 2011 Jul 17 Presented at: Proceedings of the International AAAI Conference on Web and Social Media; July 17-21; Barcelona, Spain p. 185-192.
67. Gilani Z, Farahbakhsh R, Tyson G, Wang L, Crowcroft J. Of bots and humans (on Twitter). 2017 Presented at: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining; July 31; New York, USA. [doi: 10.1145/3110025.3110090]
68. Scannell D, Desens L, Guadagno M, Tra Y, Acker E, Sheridan K, et al. COVID-19 Vaccine Discourse on Twitter: A Content Analysis of Persuasion Techniques, Sentiment and Mis/Disinformation. J Health Commun 2021 Jul 03;26(7):443-459. [doi: 10.1080/10810730.2021.1955050] [Medline: 34346288]
69. Egli A, Rosati P, Lynn T, Sinclair G. Bad Robot: A Preliminary Exploration of the Prevalence of Automated Software Programmes and Social Bots in the COVID-19 #antivaxx Discourse on Twitter. 2021 Presented at: The International Conference on Digital Society 2021; July 18-22; Nice, France.

Abbreviations

BERT: bidirectional encoder representations from transformers
NMF: nonnegative matrix factorization
VADER: Valence Aware Dictionary and sEntiment Reasoner

Edited by C Basch; submitted 27.05.21; peer-reviewed by S Wilson, A Ramazan, R Poluru, R Gore; comments to author 30.08.21; revised version received 17.09.21; accepted 17.09.21; published 21.10.21.

Please cite as:
Monselise M, Chang CH, Ferreira G, Yang R, Yang CC
Topics and Sentiments of Public Concerns Regarding COVID-19 Vaccines: Social Media Trend Analysis
J Med Internet Res 2021;23(10):e30765
URL: https://www.jmir.org/2021/10/e30765
doi:10.2196/30765
PMID:34581682

Original Paper

The State of Mind of Health Care Professionals in Light of the COVID-19 Pandemic: Text Analysis Study of Twitter Discourses

Aviad Elyashar[1,2], MSc; Ilia Plochotnikov[1,2]; Idan-Chaim Cohen[3], BEMS; Rami Puzis[1,2*], PhD; Odeya Cohen[4*], RNC, PhD

[1]Software and Information Systems Engineering, Ben-Gurion University of the Negev, Beer Sheva, Israel

[2]Cyber@BGU, Ben-Gurion University of the Negev, Beer Sheva, Israel

[3]School of Public Health, Ben-Gurion University of the Negev, Beer Sheva, Israel

[4]Department of Nursing, Ben-Gurion University of the Negev, Beer Sheva, Israel

[*]these authors contributed equally

Corresponding Author:
Rami Puzis, PhD
Software and Information Systems Engineering
Ben-Gurion University of the Negev
Rager st. POB 653
Beer Sheva, 8410501
Israel
Phone: 972 544764010
Email: puzis@bgu.ac.il

Abstract

Background: The COVID-19 pandemic has affected populations worldwide, with extreme health, economic, social, and political implications. Health care professionals (HCPs) are at the core of pandemic response and are among the most crucial factors in maintaining coping capacities. Yet, they are also vulnerable to mental health effects caused by managing a long-lasting emergency with a lack of resources and under complicated personal concerns. However, there are a lack of longitudinal studies that investigate the HCP population.

Objective: The aim of this study was to analyze the state of mind of HCPs as expressed in online discussions published on Twitter in light of the COVID-19 pandemic, from the onset of the pandemic until the end of 2020.

Methods: The population for this study was selected from followers of a few hundred Twitter accounts of health care organizations and common HCP points of interest. We used active learning, a process that iteratively uses machine learning and manual data labeling, to select the large-scale population of Twitter accounts maintained by English-speaking HCPs, focusing on individuals rather than official organizations. We analyzed the topics and emotions in their discourses during 2020. The topic distributions were obtained using the latent Dirichlet allocation algorithm. We defined a measure of topic cohesion and described the most cohesive topics. The emotions expressed in tweets during 2020 were compared to those in 2019. Finally, the emotion intensities were cross-correlated with the pandemic waves to explore possible associations between the pandemic development and emotional response.

Results: We analyzed the timelines of 53,063 Twitter profiles, 90% of which were maintained by individual HCPs. Professional topics accounted for 44.5% of tweets by HCPs from January 1, 2019, to December 6, 2020. Events such as the pandemic waves, US elections, or the George Floyd case affected the HCPs' discourse. The levels of joy and sadness exceeded their minimal and maximal values from 2019, respectively, 80% of the time (P=.001). Most interestingly, fear preceded the pandemic waves, in terms of the differences in confirmed cases, by 2 weeks with a Spearman correlation coefficient of ρ(47 pairs)=0.340 (P=.03).

Conclusions: Analyses of longitudinal data over the year 2020 revealed that a large fraction of HCP discourse is directly related to professional content, including the increase in the volume of discussions following the pandemic waves. The changes in emotional patterns (ie, decrease in joy and increase in sadness, fear, and disgust) during the year 2020 may indicate the utmost importance in providing emotional support for HCPs to prevent fatigue, burnout, and mental health disorders during the postpandemic period. The increase in fear 2 weeks in advance of pandemic waves indicates that HCPs are in a position, and with adequate qualifications, to anticipate pandemic development, and could serve as a bottom-up pathway for expressing morbidity and clinical situations to health agencies.

(J Med Internet Res 2021;23(10):e30217) doi:10.2196/30217

KEYWORDS

health care professionals; Twitter; COVID-19; topic analysis; emotion analysis; sentiment analysis; social media; machine learning; active learning

Introduction

The COVID-19 pandemic has affected nations and societies worldwide. The pandemic became a significant health crisis, with extreme health, economic, social, and political implications. COVID-19 created a unique situation, one which requires that people from different countries, cultures, and life circumstances stand against the same emergency situation [1].

Online social networks may provide insights on the state of mind and the experience of people during COVID-19, from emotional effect [2] to adherence to restrictions [3].

Studies found that negative sentiments were dominant in tweets posted by people during the pandemic [2,4], and Twitter could advance social stigmas in those situations [5]. According to Park et al [6], the spread of information related to COVID-19 on Twitter was faster than in other content networks. They showed that the spillover effect of information that included medical knowledge about COVID-19 was more significant than news with nonmedical content [6]. Many publications relating to Twitter analyses during the pandemic highlighted health care agencies' and professionals' critical roles in providing reliable information disseminated via online social networks during intense situations [4,6]. Contrary to studies that analyzed the general population, others focused on specific populations, such as policy makers [7], students [8], and health care professionals (HCPs) [9].

HCPs are an essential resource for public health. The World Health Organization recognized the health workforce as one of the six foundations for improving health outcomes [10] and perceived the development of HCPs as an opportunity for the sustainable development of society as a whole [11]. During an emergency, HCPs are one of the most crucial factors in developing surge capacity to satisfy health care demands [12]. Yet, they are also vulnerable to personal concerns, fear, and anxiety [13] caused by managing a long-lasting emergency situation with a lack of resources and under accelerated conditions. The situation of HCPs struggling to balance day-to-day self-management while having to intensify their work is exacerbated by their circumstances, such as having elderly family members, children at home, or family members with special needs [14].

Most existing studies of the HCP experience during COVID-19 are (1) cross-sectional: demonstrating associations with no causality between study variables, (2) hypothesis driven: unlikely to produce new findings that are not grounded in existing theory, and (3) small scale: study populations range from a few hundred to a few thousand. There are a lack of data-driven longitudinal studies based on large-scale analysis of the HCP experience during COVID-19.

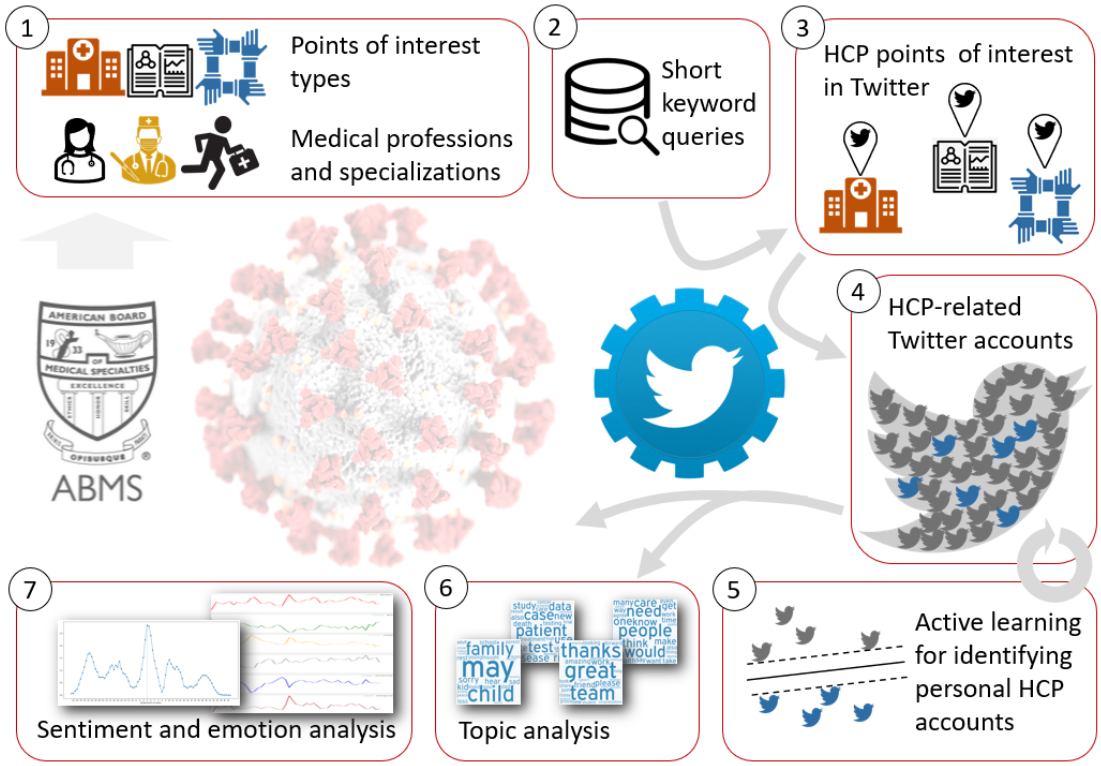
This study aims to analyze the state of mind of HCPs as expressed in online discussions published on Twitter in light of COVID-19 from the onset of the pandemic until the end of 2020. The analyzed data include 16.6 million English tweets from 53,063 HCP accounts identified using a tailor-made machine learning classifier. We present the major topics concerning the study population and the dynamics of emotions during the pandemic.

Methods

Identifying the Study Population

This study's first and significant challenge in this study was to collect tweets published by the heterogeneous HCP population, while excluding formal communication by health care facilities and organizations. We tackled this challenge through a multistep process depicted in Figure 1, steps 1 to 5.

Figure 1. Analysis of the health care professional (HCP) discourse in light of COVID-19. Steps 1 to 5 identify the study population. Steps 6 and 7 analyze the HCP discourse. In step 1, we identify search terms. In step 2, we construct short keyword queries. In step 3, we pinpoint and manually distill HCP points of interest. In step 4, we collect a pool of health care–related Twitter accounts. In step 5, we conduct active learning for filtering out organizational and non-HCP accounts. In step 6, we collect the tweets and analyze the major topics. In step 7, we analyze sentiment and emotions.



Twitter Search Queries

First, two health care experts defined a list of general medical professions and specializations recognized by the American Board of Medical Specialties (Figure 1, step 1, and Table S1, list 1, in Multimedia Appendix 1). Second, a list of points of interest (POIs) types, such as unions, conferences, and journals, were defined (Figure 1, step 1, and Table S1, list 2, in Multimedia Appendix 1). In addition, we added countries and regions related to English speakers. We created short keyword queries by taking one keyword from each list, for example, “immunology journal” or “family medicine England” (Figure 1, step 2).

HCP Points of Interest

Overview

We used the Twitter search engine to collect accounts matching the short keyword queries (Figure 1, step 3). The results were manually inspected to identify accounts managed by facilities, organizations, and venues related to health care. The resulting 563 accounts were divided into four groups of medical professions and a fifth group of general health care interest (Table S2 in Multimedia Appendix 1). We removed from this list 41 POIs that are likely to be followed by non-HCPs.

HCP Twitter Accounts

Using the Twitter application programming interface, we collected 434,825 followers of the remaining 522 HCP POIs prioritizing those that followed multiple POIs. Twitter accounts collected in this manner included private accounts and accounts managed by organizations, for example, official accounts of medical centers in the United States. There are also many non-HCP accounts, for example, patients, reporters, or academic scholars.

The definition of HCP varies considerably. Broad definitions of HCP consider any professional that contributes to people’s well-being as an HCP. In this study, we define an HCP as an individual working in the health care system or being a student of any medical profession. Students were also included since they typically go through hands-on training that incorporates interaction with patients. Effectively, the medical professionals that were considered HCPs were mainly medical doctors, nurses, medical management and administration staff, pharmacists, medical students, psychologists, and others. We excluded therapists that work in fields that are mostly considered complementary or alternative medicine (eg, naturopaths and shamans) and art therapy (eg, drama and music therapists).

Training Machine Learning Classifiers Using Active Learning

To filter out organizational and non-HCP Twitter accounts, we trained two respective machine learning classifiers through an iterative process known as active learning [15]. As done by Lo et al [16], in this research, we used the support vector machine (SVM) classification model. Term frequency–inverse document frequency features were extracted from the accounts’ descriptions and full names. The latter occasionally included a profession or a relevant title, such as “MD” (medical doctor) or “RN” (registered nurse).

Supervised machine learning requires a labeled training set to build accurate statistical models. Active learning allows

reduction of the manual annotation effort by focusing on accounts that would contribute the most to improving the classification model. Therefore, we used the uncertainty-sampling active learning strategy [17], where we manually annotated accounts that the classifier is least certain about.

The descriptions and timelines of 90 accounts were inspected by two human annotators in every active learning iteration. In case of a conflict, a third annotator determined the label after an open discussion. In case of uncertainty, the LinkedIn profile associated with the Twitter account was inspected as well. If the panel of three annotators could not agree about the label of an account, it was excluded from the training set and replaced with a random unlabeled account on which the panel had agreed. Every iteration ended with training two SVM classifiers that were used for sampling accounts for the next iteration. Initially, the accuracy of the classifiers increased by 1% to 3% with each iteration. Between the 15th and 20th iteration, the marginal increase in accuracy was as low as 0.5%. Thus, we decided to halt the manual labeling process after 20 iterations (Figure S1 in Multimedia Appendix 1).

Overall, the final training data set contained 1392 profiles on which we performed 10-fold cross-validation for evaluation. The results of this training process are presented in the Results section. Using the trained classifiers, we selected accounts classified as individual HCPs with a confidence of 70% or higher and validated the quality of this selection by manual inspection of 100 random accounts.

Social Discourse: Topic Detection

We collected all public tweets published by the individual HCP accounts from January 1, 2019, to December 6, 2020; see Table 1 for details about the data set of collected tweets.

We used Python (version 3.6.8; Python Software Foundation) and several libraries for our analysis. We applied standard text preprocessing techniques, such as removing line breaks, emojis, nonalphabetic words, stop words, short words with less than three characters, user mentions (@), and hyperlinks; we applied lower-casing and lemmatization using Natural Language Toolkit (NLTK) (version 3.5; NLTK Team) [18] and WordNet (version 3.0; Princeton University) [19]. Hashtag (#) terms were not removed because they carry a significant informative load on Twitter. The hashtag symbol itself was removed.

Most of the social discourse during 2020 revolved around COVID-19. This theme overshadowed other topics discussed both with and without a relationship to the pandemic. Therefore, to have a precise dissection of topics discussed during 2020, we removed the terms signifying the disease or the virus. The full list of COVID-19 terms can be found in Table S3 in Multimedia Appendix 1.

Table 1. The study data set.

Statistic	Value
Accounts, n	53,063
Tweets, n	16,616,970
Tweets, mean (SD)	313 (1386.6)
Friends, mean (SD)	511 (1767.1)
Followers, mean (SD)	475 (4466.8)
Total tweets published in 2019, n	7,168,088
Total tweets published in 2020 (up to December 6, 2020), n	9,448,882

Topic models were obtained using the latent Dirichlet allocation (LDA) algorithm [20] implemented in the gensim library (version 3.8.3; Python Software Foundation) [21]. The number of iterations was set to 150 with a chunk size of 130,000 tweets. All other parameters were set to their default values. We searched for the optimal number of topics between 5 and 55. To quantify the quality of the topic distributions, we used one of several coherence scores that were proposed by Roder et al [22]. In their work, human judgments of the interpretability of topics extracted from several benchmark data sets were recorded. The authors then examined the correlation of each measure against the human interpretability scores. We chose the coherence score, which is the most correlative to human topic rankings. In addition to coherence scores, we also examined the Jaccard similarity index between the sets of top 100 words of topics originating in different distributions using the multiple correspondence analysis algorithm (configured with px=py=0.95) from the Jaccard package (version 0.1.0) in the R language (version 4.1.0; The R Foundation). To determine the topics that persist in different distributions, we studied the connections between distributions to eventually decide on the best one regarding the number of similar topics that it shares with other distributions.

Provided the topic distributions, volumes of the topics, and their persistence over different distributions, we chose to analyze the topic distribution containing 20 topics. A manual subjective inspection was conducted to make sense of the automatically generated topics. The manual inspection included an assessment of the topics' subjective cohesion level and topic naming. Each topic received a subjective cohesion score based on our ability to name the topic. The naming relies on the top 50 words within the topic and the contents of a few hundred tweets with the highest probability of belonging to the topic. There are three cohesion levels: high, medium, and low. A high cohesion score was given to topics where most of the 50 top words could easily be associated with a single well-defined theme, and most of the inspected tweets matched this theme. A medium cohesion score was given to a topic when some of the 50 top words could easily

be associated with a specific theme, yet a relatively high ratio of words and tweets could not be associated with it. In cases where no single name could be identified that described a significant number of the topic's 50 top words and top tweets, the topic received a low subjective cohesion score and no name.

We chose the main topics for further analysis based on their volumes, coherence, and cohesion scores. Weekly volumes of the chosen topics were tracked throughout the year 2020. We identified the major changes in topic volumes and associated them with the most significant events presumably corresponding to the changes in topic volumes. The topic analysis results are discussed in the Topic Detection subsection within the Results section.

Analysis of Sentiment and Emotions

To estimate the sentiments expressed by HCPs, we used the Valence Aware Dictionary and Sentiment Reasoner (VADER), which is a lexicon and rule-based sentiment analysis tool [23]. Given a tweet, VADER provides a sentiment score, which ranges from –1 (the highest negative score) to 1 (the highest positive score). To track the changes in sentiment, we calculated the average sentiment of each topic every week, as well as the Bayesian credible interval of 95% to ensure that the average sentiment was a good representative of the sample obtained.

Sentiment is a coarse-grained measure that does not allow for understanding the different emotional tones expressed in text. Therefore, we used a pretrained recurrent neural network model developed by Colnerič and Demšar [24] for quantifying the probabilities of Ekman's six basic emotions [25] expressed in the text. We inferred the distribution of emotions in a topic by aggregating the emotion probabilities across all tweets associated with the topic. We included retweets and tweets containing quotes when aggregating emotions and sentiment, since by retweeting or quoting an emotionally loaded tweet, a person shows their support for it. Differences of emotions between the years 2019 and 2020 were calculated based on the emotion distribution: the Welch *t* test for normally distributed emotions (ie, anger, sadness, and joy) and the Mann-Whitney *U* test for nonnormally distributed emotions (ie, fear, surprise, and disgust).

Next, we analyzed the time course of each emotion and quantified their correlation during the study periods (47 weeks) with the following: number of new COVID-19 cases (*Confirmed*), number of deaths caused by COVID-19 (*Deaths*), their weekly change (Δ*Confirmed* and Δ*Deaths*, respectively), and the estimated reproduction rate of SARS-CoV-2 (R_t). A Shapiro-Wilk test was conducted to examine the distribution of the variables. Cross-correlation analysis was performed to account for possible lags in a range of 1 to 8 weeks between pandemic development and emotional response. This study was approved by the Institutional Review Board of Ben-Gurion University of the Negev (1879-1).

Results

HCP Accounts on Twitter

During the active learning process, we manually labeled 1800 Twitter accounts. Out of these accounts, 1192 (66.2%) were labeled as profiles of individuals and 299 (16.6%) were labeled as organizational. The best classifier differentiated between individuals and organizations with an accuracy score of 0.88, an F1 score of 0.88, a precision score of 0.884, and a recall score of 0.88. The best classifier, trained on accounts labeled as individuals to take HCP and non-HCP accounts apart, obtained an accuracy score of 0.786, an F1 score of 0.785, a precision score of 0.795, and a recall score of 0.787. Performance scores are reported for balanced test sets with subsampling of the majority class.

Out of the 434,825 HCP POI followers, 53,063 profiles were classified as individual HCPs with a confidence of 70% or higher. Random manual validation of 100 accounts confirmed that 90 of these accounts belonged to health care individuals.

Topic Detection

The LDA algorithm's topic detection resulted in the highest average coherence of 0.433 for the distribution of 30 topics. The distributions of 25 and 20 topics exhibited average coherence scores of 0.427 and 0.402, respectively. The coherence values for each topic distribution were normally distributed. We linked topics in different distributions based on the Jaccard coefficient of the sets of tweets associated with the topics. As can be seen in Figure 2, there are no topics linked to more than one topic in a different distribution with a Jaccard coefficient higher than 0.3. The alignment of topic distributions also shows that four topics persisted throughout distributions of 15, 20, 25, and 30 topics. These topics were "public health and social values" (topic 0), "day-to-day life" (topic 1), "food" (topic 2), and "medical studies and COVID-19 information" (topic 8). Topics 0, 1, and 2 contained the largest number of tweets during 2020: 25.9%, 27.7%, and 13.5%, respectively.

Topic 8 ("medical studies and COVID-19 information") also contained a significant fraction of tweets (7.6%) and had the highest coherence score in the distribution of 20 topics (Figure 3). The distribution of 20 topics contained the highest number of persistent topics (n=14).

Topic volumes are provided in thousands or millions of tweets over the entire period. The percentage of the total volume is provided in parentheses. Topics are sorted and color coded according to the average sentiment in the main chart and the legend. The 95% CIs of the sentiments are provided in parentheses alongside the topic names.

Further manual inspection of the distribution of 20 topics revealed that the most cohesive topics were also the most persistent, except topic 10 ("account promotion"), which received a high cohesion score but appeared only in this distribution. This topic accounted for 1.8% of the tweets. Overall, topics 0, 1, 3, 6, 8, 9, 10, and 16 received high cohesion scores and are listed in Table 2.

Figure 2. Alignment of topic distributions. Topics, represented with word clouds, were obtained from four different distributions of 15, 20, 25, and 30 topics. Every pair of aligned topics is connected using a weighted link. The weight and the width of each link represent the Jaccard coefficient of the sets of words associated with the two aligned topics; coefficients are reported in the top rows of each set of values. Links having a Jaccard coefficient lower than 0.3 are not shown. Topics that have no link are also not shown (except topic 10).

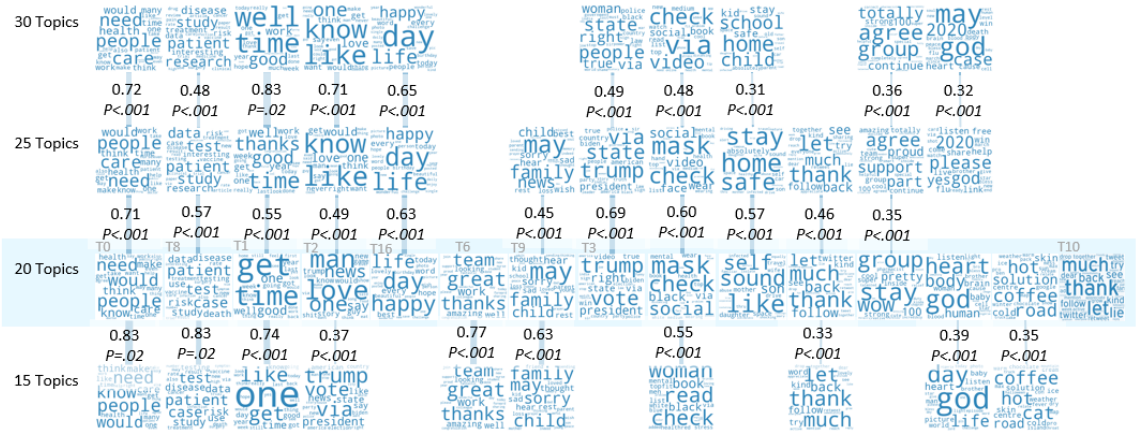


Figure 3. Topic (T) volumes over time in 2020 sorted by sentiment and their coherence. The 95% CIs of the sentiments are provided in parentheses alongside the topic names in the legend.

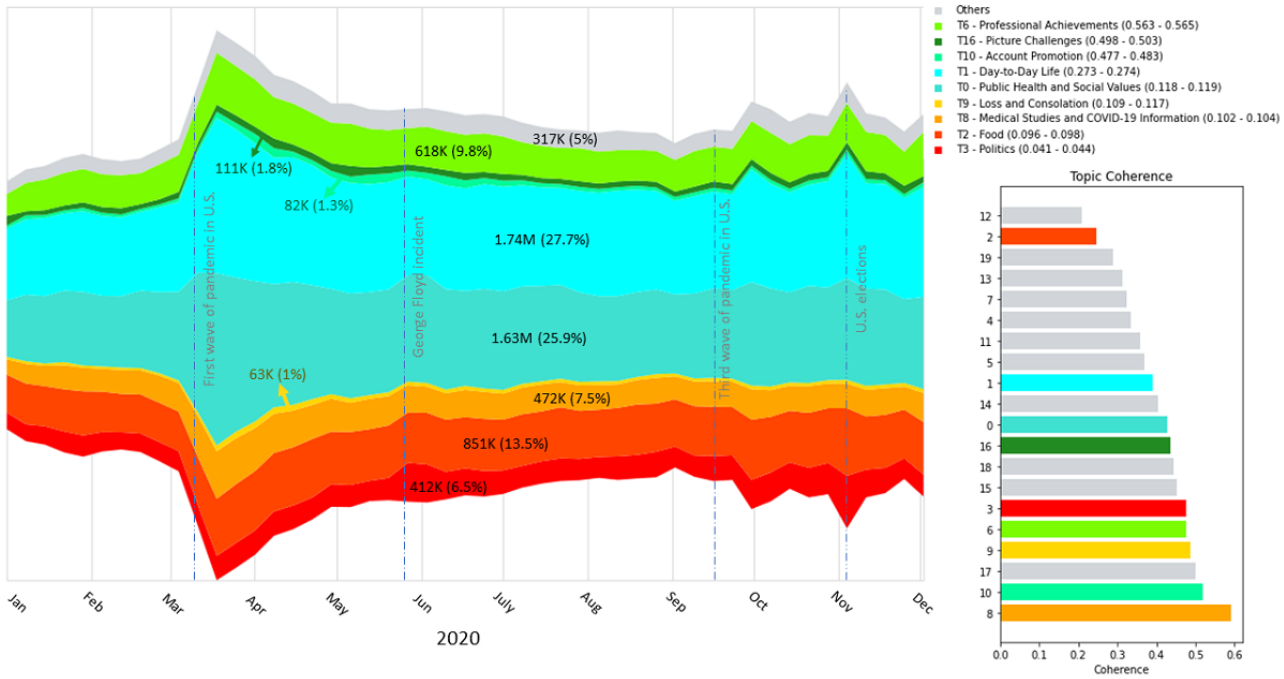


Table 2. The most significant topics discussed by health care professionals during the COVID-19 pandemic.

Title	Example tweet	Explanation	Top 25 words[a]
Public health and social values (topic 0)	"Medicare being defunded for primary care services. There's a huge opportunity to reform healthcare to invest where the best evidence sits. Primary care & preventive activities. (Public health too cos investment in public housing education & early life is essential part of this)"	Tweets on this topic discuss issues of public health policy and its applications as well as social values that are mostly related to health aspects.	People, need, would, know, think, care, make, health, many, one, get, time, take, want, work, even, way, agree, thing, also, help, say, right, public, doctor
Day-to-day life (topic 1)	"10 years ago I was sitting on the sofa waiting to watch Jools Holland's Hootenanny. Tonight, I am sitting on the sofa waiting to watch Jools Holland's Hootenanny. In ten years time, I hope to be sitting on the sofa waiting to watch Jools Holland's Hootenanny. #Hootenanny"	This topic contains tweets that mainly describe typical everyday situations.	Get, time, good, one, well, year, got, going, know, week, last, thing, back, really, still, done, work, would, home, see, think, yes, look, first feel
Food (topic 2)	"This is so delicious and easy. I have been eating and making it for 35 years..... 4oz (125g) blanched almonds, toasted 4oz cherries 2oz mixed candied peel 2oz raisins 2oz sultanas 2oz currants 1 des sp mixed spice 1 tsp cinnamon 1 tsp nutmeg"	This topic consists of tweets about restaurants, impressions from dishes, recipes, and other food-related matters.	Love, man, one, news, say, trump, story, guy, know, new, would, shit, game, look, show, think, eat, ever, via, song, want, make, food, said, house
Politics (topic 3)	"I want my President to respect Americans. Republican Americans, Democrat Americans, members of our American military, our democracy, women, law and order, the Constitution, our country..."	The tweets on this topic are generally related to politics and government, particularly to the United States' contemporary affairs.	Trump, vote, president, right, state, via, video, true, win, biden, party, sign, country, police, election, american, america, sir, law, watch, india, funny, stop, joe, republican
Professional achievements (topic 6)	"Proud of all the amazing accomplishments of our fellowship graduates Deputy Medical Director of [hospital], will be presenting at EMSWorld (EMS - EMT - Paramedic Emergency Medical Services) Expo."	This topic contains tweets that commend teams and individuals for professional accomplishments.	Great, thanks, team, please, work, amazing, friend, looking, congratulation, forward, proud, student, support, share, today, colleague, sharing, fantastic, join, new, welcome, awesome, look, help, brilliant
Medical studies and COVID-19 information (topic 8)	"Nice article on Remdesivir & renal dysfxn. Interesting a 117% (over 2x) increase in end of therapy AKI occurred in pts w/ renal dysfxn was not statistically significant (5% vs 2.3%). Stats are so... interesting... sometimes similar numbers (sample size?) result in FDA approval."	The tweets on this topic are primarily about science communication of medical studies and epidemiological information related to COVID-19.	Patient, case, test, study, data, risk, disease, use, death, testing, treatment, vaccine, new, rate, surgery, cancer, result, care, also, interesting, number, via, symptom, trial, infection
Loss and consolation (topic 9)	"Sad to hear May Allah rest him in peace and give you and family Sabar! Ameen"	Tweets on this topic convey sorrow and are meant to console individuals and families for the loss of loved ones.	May, family, child, sorry, hear, thought, kid, sad, rest, parent, school, young, loss, loved, sending, thinking, peace, memory, adult, indeed, dad, prayer, anxiety, mom, soul
Account promotion (topic 10)	"Active? Drop your username and retweet lets follow each other Grow together ..FOLLOW BACK.. follow back immediately..No Lie.. Try me"	This topic essentially contains tweets meant to promote accounts and tweets with information about accounts' activities on Twitter.	Thank, much, let, back, follow, twitter, see, kind, together, tweet, lie, try, dear, drop, retweet, appreciate, active, mention, reach, immediately, grow, following, fan, week, reply
Picture challenges (topic 16)	"Day 5/7, I've been challenged by [user mention] to produce a picture a day for 7 days to illustrate my current life (no captions, no people). Nominate a person a day, asking to copy these words, add photo and repeat the challenge. Today I nominate [URL]"	This topic's tweets are mainly a response to an online social media challenge that prompted Twitter users to post photographs that represent their lives.	Day, happy, life, every, word, today, best, hope, lovely, beautiful, birthday, wish, photo, person, picture, challenge, year, new, save, people, one, enjoy, posted, single, wonderful

[a]Words are listed in order of their prevalence.

Analysis of the Discussion Topics

Content of Discourse

We identified 9 out of 20 topics (45%) that constituted 95% of the total discourse. Table 2 presents the final 9 topics that describe the HCPs' discussions during the year 2020. For each topic, the top 25 words are presented alongside a short explanation and a representative tweet. The titles of the topics were chosen to match the majority of manually inspected tweets. The tweets in the selected topics discussed both professional (45.5%) and personal (54.5%) issues. All professional topics presented as high coherence levels as the topics of "professional achievements" and "medical studies and COVID-19 information."

Out of the 9 topics, 8 (89%) received a high subjective cohesion score, while 1 topic (11%; topic 2: "food") received low

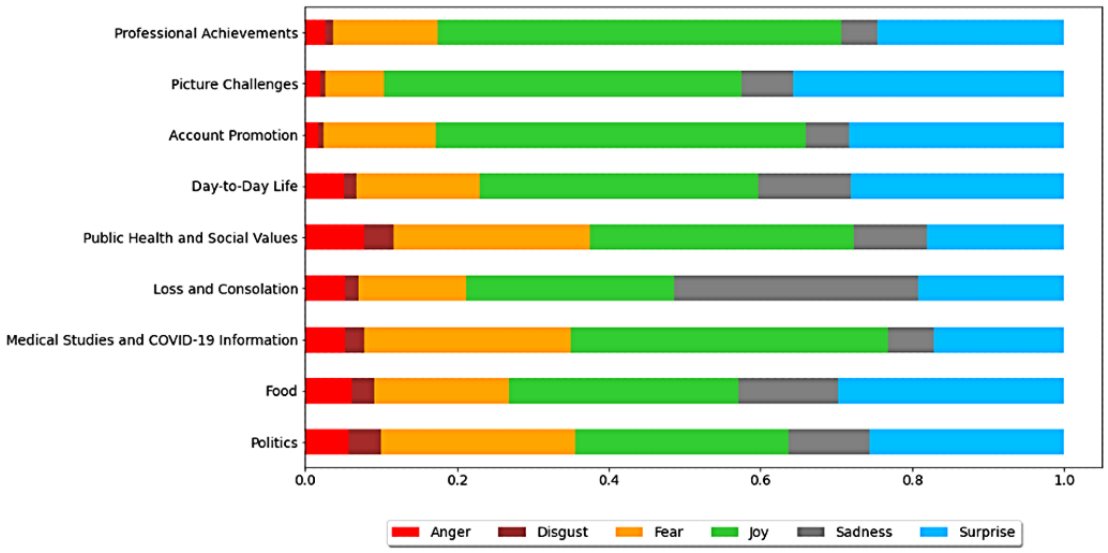
coherence and medium cohesion scores. The cohesion of this topic was set to medium because the top 50 words were loosely associated with a common theme. Still, when examining the topic's tweets, we found that the vast majority of them were related to food. Note that types of food were not discernible in the topic's top 50 words due to the high diversity of the food types used in the tweets (eg, salad, chicken, and BBQ). The volume of the food topic remained consistently high (13.5%) throughout the year 2020.

The distribution of tweets across HCPs (Figure S2 in Multimedia Appendix 1) was a long-tailed (power-law) distribution. However, the user that tweeted the most accounted for only 0.57% of the data, and the top 60 users accounted for 10%. To check the influence of the top users on the social discourse, we compared their topic distribution to that of the rest of the users (Table S4 in Multimedia Appendix 1). The topic distributions were highly correlated (Spearman ρ[9 pairs]=0.9, $P<.001$), showing no significant influence of the top users on the social discourse.

Emotion Analysis of the Detected Topics

We computed the average levels of six emotions in each topic: anger, disgust, fear, joy, sadness, and surprise. Topics with the highest levels of joy (about 50%) were "professional achievements" (53.1%), "account promotion" (48.8%), and "picture challenges" (47.0%). As expected, the "loss and consolation" topic presented the highest average sadness (32%) in comparison to the average of 8.6% in the other topics. Similarly, the topic of "politics" presented the lowest ratio of joy and was also the topic with the lowest sentiment score (Figure 4).

Figure 4. Distribution of emotions per topic.



Trends of Discourse Throughout the Year 2020

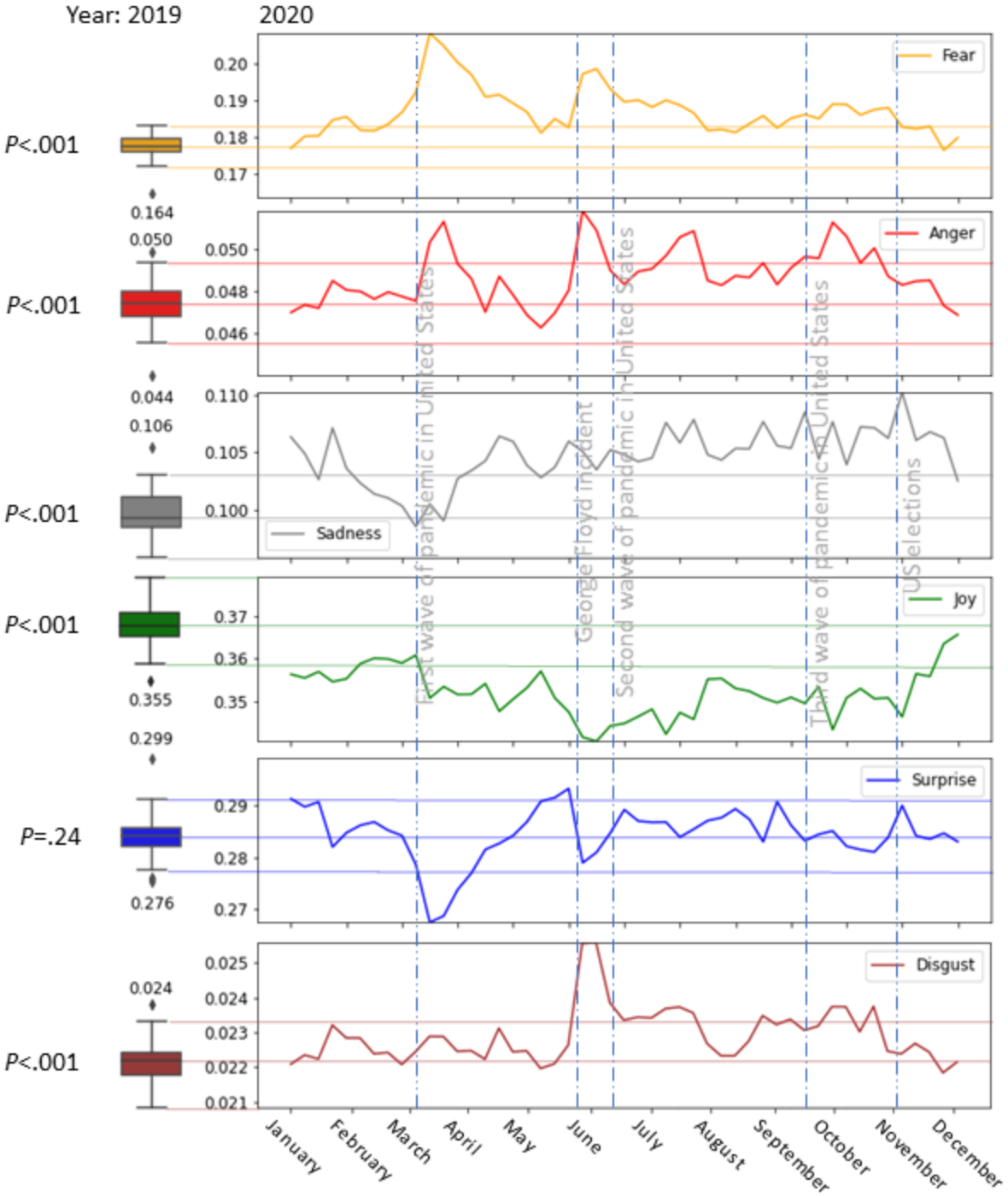
Figure 3 describes the topic volumes over time in 2020 sorted by average sentiment and their coherence. It includes the Bayesian credible interval of 95% for each topic's average sentiment. In addition, the distribution of the sentiment scores for each topic can be found in Figure S3 in Multimedia Appendix 1. Among the identified topics, the "professional achievements" topic presented the highest positive sentiment score of 0.56, while the "politics" topic obtained the lowest sentiment score of 0.04. Trends of discourse during the year 2020 revealed that HCPs express in their discussions special events that have occurred, most of them in the United States. The increase in tweets responded to the global crisis, ahead of the first wave of the COVID-19 pandemic in the United States. Analyzing modifications from the content point of view, the topic of "public health and social values" (topic 0) exhibited the highest increase. "Medical studies and COVID-19 information" (topic 8), together with "day-to-day life" (topic 1), achieved relatively moderate growth. In a milder tendency, these topics reacted similarly to the third wave of the pandemic. The case of George Floyd sparked political (topic 3) discussions at the expense of day-to-day life (topic 1), with a minor impact on the context of public health. Around the US elections, discussions on political issues rose sharply with the reverberation of the elections in the public health topic.

Emotion Dynamics of the HCPs

Overview

We analyzed the average weekly emotion values in the HCP Twitter discourse during 2020. Figure 5 presents the dynamics of Ekman's six basic emotions: anger, joy, fear, sadness, surprise, and disgust. We analyzed the sentiment dynamics as well. Since sentiment closely follows the values of joy, we did not present it in Figure 5.

Figure 5. Emotions over time. On the left of the time series, the box plots show the statistics of the year 2019 for each emotion. The boxes represent the IQRs, and the whiskers represent the upper and lower quartiles. The dashed gray vertical lines indicate the important events during 2020. The time series detail the average weekly emotions during 2020. The y-axes represent the intensity of six emotions and are the same for the time series and the box plots. *P* values on the left side refer to the significance of the difference between the emotion levels in 2019 and 2020.

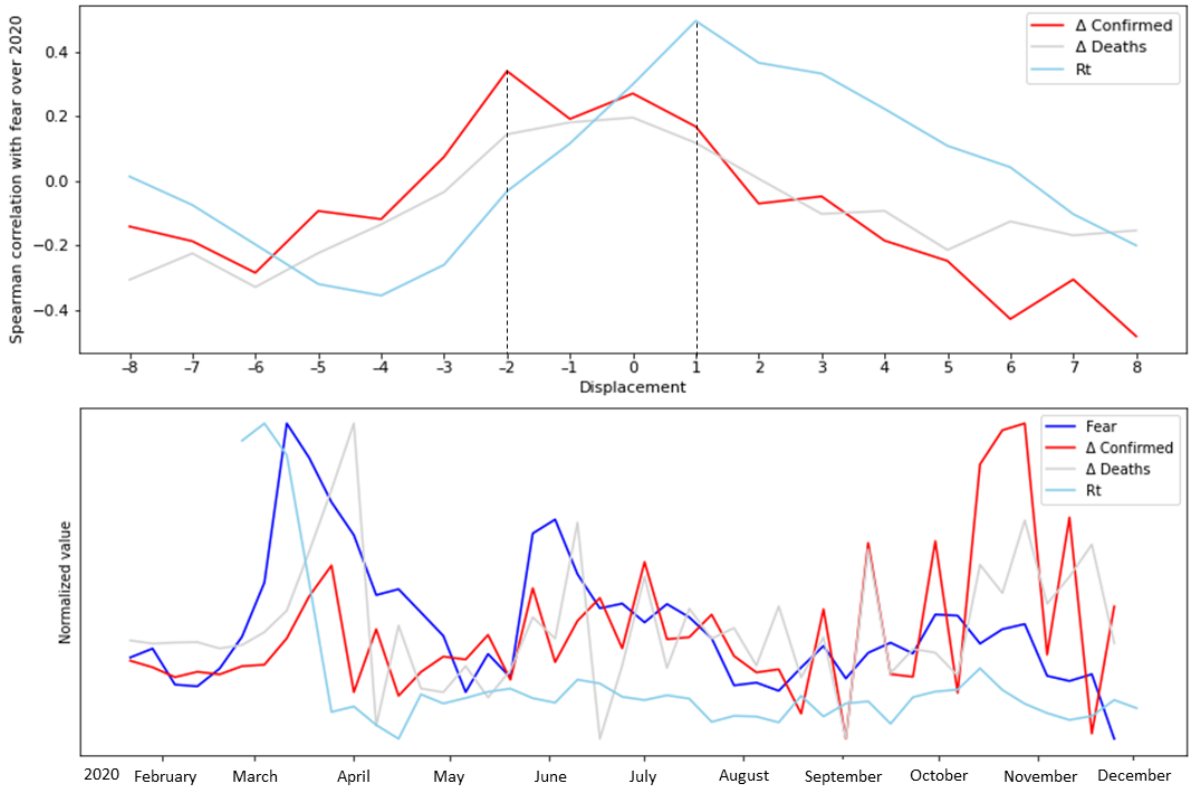


Acute Effects

Tracking the emotions over time showed that anger, fear, and disgust exhibited fluctuations that corresponded to the three COVID-19 pandemic waves worldwide (Figure 5). Fear exhibited the three clear waves, with the first wave being the largest and the second and third magnitudes gradually decreasing. Cross-correlation analysis revealed that fear was found to correlate most highly with the virus reproduction rate R_t with a delay of 1 week (ρ[47 pairs]=0.486, *P*=.001). A lesser but significant correlation (ρ[47 pairs]=0.340, *P*=.03) was found between fear and Δ*Confirmed*, where fear preceded Δ*Confirmed* by 2 weeks. The normalized values of fear, R_t, Δ*Deaths*, and Δ*Confirmed* are shown in Figure 6. The average weekly levels of anger were correlated (ρ[47 pairs]=0.386, *P*=.009) with the change in the average number of death cases (Δ*Deaths*), with Δ*Confirmed* (ρ[47 pairs]=0.308, *P*=.04), and with the virus reproduction rate R_t (ρ[47 pairs]=0.316, *P*=.04) after displacement of 2 weeks (anger after R_t). A positive correlation

was found between sadness and Δ*Confirmed*, where sadness preceded Δ*Confirmed* by 3 weeks (ρ[47 pairs]=0.423, *P*=.005). A negative correlation arose between joy and Δ*Confirmed*, where joy decreased after 1 week. Two periods of increased disgust roughly corresponded to the second and the third pandemic wave. Interestingly, disgust remained at an average level in the HCP discourse during the first wave, in contrast to fear and anger. Disgust and anger spiked after the George Floyd case and the second wave of the pandemic.

Figure 6. Normalized weekly fear levels (dark blue) during 2020 vs the SARS-CoV-2 reproduction rate (R_t; light blue), the difference in the average number of daily death cases (Δ*Deaths*; gray), and the difference in the average number of daily new confirmed cases (Δ*Confirmed*; red). The top panel shows the cross-correlation results with displacement ranging from –8 (before fear) to +8 (after fear) in terms of Spearman correlation coefficients.



Long-term Effects

We can see in Figure 5 that joy was low throughout 2020. Average weekly joy during 2020 was lower than the minimal weekly joy during 2019. We also saw a rise in joy by the end of 2020, bringing it closer to the values observed during 2019. Sadness increased during 2020, with the minimal sadness between April and December 2020 being higher than the maximal sadness before December 2019. The intensities of all emotions, except surprise, were significantly different during 2020 than in 2019 (*P*<.001).

Discussion

Overview

HCPs are at the core of pandemic response, and the impact of emergencies, especially a pandemic, on the frontline workforce was found to be significant during former events [26]. This study was designed to explore the state of mind of HCPs in light of the COVID-19 pandemic, as expressed in Twitter discussions during the year 2020. The study was conducted by a multidisciplinary group of researchers from data and health sciences to reflect professional health aspects and data-driven approaches throughout the study. Beyond the results presented in this paper, the study tries to serve as documentation of the broad reality of HCPs during the pandemic. Such documentation was found to be missing during former pandemics, such as the influenza pandemic of 1918-1919 [27].

Principal Findings

This study explored three principal results:

1. A rigorous multistep process of selecting the study population, which involved active machine learning and manual curation, resulted in a high-quality data set of English tweets from 53,063 HCPs.
2. Analysis of the topics discussed by the HCPs during the year 2020 revealed that Twitter serves as a platform for expressing both day-to-day and professional content during the pandemic. The analyses revealed that discussions could be attributed to events that occurred. As such, the volume of tweets increased appropriately with the progression of the COVID-19 outbreak, and the analyzed data articulated events with influential aspects, such as the pandemic waves and the George Floyd case.
3. Analysis of emotions expressed in Twitter explored the significant emotional impacts of the COVID-19 pandemic on HCPs that were sharpened compared to emotion levels found in 2019. The results portrayed significant long-term and acute responses to the pandemic.

Comparison With Prior Work

Methodological Aspects

Recently, many studies examined different aspects of the COVID-19 pandemic using public information published within social media. Some of the studies attempted to predict the number of reported cases associated with the COVID-19 pandemic based on posts published on Sina Weibo [28,29] or Twitter [30]. Multiple articles analyzed online discourses, sentiments, and general dynamics during the pandemic [4,31-35]. These studies made their conclusions based on posts, which were published in the early stages of the pandemic (ie, over a few months), as opposed to this study, which analyzed online discourses during a single year of the pandemic, including a comparison of the emotional dimension during 2020 with the emotional dimension a year before the pandemic emerged. In addition, most of the studies collected posts using predefined COVID-19–related keywords [31,33] or hashtags [34]. However, it is important to understand the overall state of mind of HCPs during the crisis, including aspects not directly related to the COVID-19 pandemic. Therefore, in this study, we analyzed posts of HCPs across a variety of topics. Although many studies investigated COVID-19 health aspects, only a few of them focused on HCPs.

Similar to this study, a few articles focused on the HCP population [9,32]. Ojo et al [9] addressed a specific research question comparing the COVID-19–related social media reaction of HCPs to gun violence. Wahbeh et al [32] studied a small-scale population of HCPs. The limited scope or population of related work could be explained by the difficulties in identifying HCPs as a study population in social media.

Social Media as a Community in Times of Crisis

Studies found that virtual communities provide their members a sense of community, especially in emergencies [36]. According to Neubaum et al [37], in a time of crisis, social media platforms serve as a space for social sharing of emotions and pursuing empathetic concerns. Twitter is currently the most popular form of social media used for health care communication [38]. Moorhead et al [39] recognized social media as a dimension of health care and a powerful tool that offers collaboration between users and a social interaction mechanism for a range of individuals.

Current studies call for HCPs to join the social media space, take part in relevant conversations, and increase their involvement in providing professional information [4,40]. As far as we know, their involvement was not measured. Ojo et al [9] found that tweets from HCPs were more positive and action oriented in the context of the COVID-19 than tweets that dealt with gun violence, but their study addressed only two issues in a short-term period.

Analyzing the topics discussed on Twitter by HCPs during 2020 revealed two interesting findings:

1. About 44.5% of the topics addressed by HCPs during the COVID-19 pandemic were related directly to their professional context (“public health and social values” [26%], “professional achievements” [10%], “medical studies and COVID-19 information” [7.5%], and “loss and consolation” [1%]).
2. Most of the volume of discussions that increased following the pandemic’s waves was related to professional aspects. Although we did not measure the involvement of HCPs in social media during the COVID-19 pandemic, our findings could indicate the role of HCPs in the social media expanse.

HCPs’ discussions on Twitter during the study period showed that HCPs maneuver between their day-to-day reality and their professional aspects in the virtual space. The emotion analysis of the topics (Figure 4) fit the topic themes, validating their appropriateness. For example, “professional achievements” exhibited the highest joy, while “loss and consolation” exhibited the highest sadness. Based on the topic analysis (eg, Figure 2), four topics constituted a solid, persistent part of the HCP discourse unaffected by sampling and the stochastic nature of the topic detection algorithms: “public health and social values” (topic 0), “day-to-day life” (topic 1), “food” (topic 2), and “medical studies and COVID-19 information” (topic 8).

Various studies investigated social media concerning the rise of the COVID-19 pandemic [2]. Several studies analyzed the discussions of public users within social media. Saleh et al [41] attempted to understand public perception of COVID-19 social distancing on Twitter, and Xue et al [42] analyzed users’ discourse and psychological reactions to the pandemic on Twitter. Other studies focused on specific populations, for example, US governors and presidential cabinet members [7], the social media activity and mental health of students in Switzerland [8], detection of users who were found to suffer from depression using transformer-based deep learning models on the Twitter platform [43], among others. Recently, Ojo et al [9] examined the behavior of health care workers on social media concerning two specific public health crises—the COVID-19 pandemic and gun violence—using analysis of two online discussions derived from two selected hashtags.

Sentiment and Emotional Effects Among HCPs During the COVID-19 Pandemic

Many studies published during the COVID-19 pandemic assessed the emotional effects of the pandemic on frontline HCPs. Most of these studies were cross-sectional, used questionnaires, and were disseminated through social media platforms. In general, anxiety, stress, and posttraumatic stress disorder were identified among HCPs working in different countries, such as Singapore and India [44], Spain [45], and Italy [46]. Some of the studies revealed that mental health symptoms, such as depression, anxiety, and stress, are associated with the presence of physical symptoms [44].

This study is different, since it did not begin with a known theory framework but was designed as a data-driven exploration of the emotional status of HCPs during the COVID-19 pandemic. Figure 5 shows trends of emotions over time, with comparison to patterns explored in 2019. The apparent difference in all emotions, except surprise, expressed during 2020 as compared to 2019 confirmed the results of previous studies, this time at a large scale in the social media expanse.

We recognized two pathways: (1) acute emotional response to COVID-19 progression and (2) long-term effects (ie, structures of emotions developed over time beyond the direct association with the pandemic waves). Acute responses (ie, anger, fear, and disgust) were associated with the pandemic waves, portraying a singular trend for each emotion. It is interesting to see the differences between the emotions: fear exhibited a decreasing trend over time, although the overall pandemic impact was increasing. This trend and the positive correlation between fear and the development of the reproduction number (R_t) was also found among the general population [47]. Still, the fact that the increase in fear expressed by HCPs preceded the change in confirmed cases may indicate that HCPs express their feelings according to the population behavior they observe. These results emphasize the fact that emotions (eg, fear) among HCPs could be an indicator of the current situation and the near-anticipated future. The results of this study imply that beyond the traditional role of HCPs in providing reliable information to the population, they may also serve as a bottom-up pathway for expressing morbidity and clinical situations to health agencies. Therefore, we suggest that decision makers invest additional resources into listening to the HCP community in the broadest sense, expanding beyond epidemiology professionals. Brief surveys, 1- to 2-minute interviews at workplaces, and online social media analyses may be good sources of such elicitation.

Although this study did not measure the effect of COVID-19 on HCPs' mental health, other studies explored the correlation between fear and mental health impacts, such as anxiety, stress, and depressive symptoms [48,49]. Based on the study of Braquehais et al [50], the high prevalence of anxiety and depressive symptoms among HCPs were developed due to the exposure to COVID-19 aspects, material and mental resources, and personal factors. Following former studies regarding the effects of the pandemic on the risk behavior of HCPs in the postpandemic period, our results describe the accumulated sadness and the decreasing joy over the year 2020. These findings should be an additional warning sign to health organizations regarding the immense importance of providing available and accessible mental health support to HCPs, assisting them in coping with the pandemic's consequences.

Limitations

Our findings should be considered while bearing several limitations. The analyses did not account for the voice of HCPs who do not use the Twitter platform. In addition, we did not compare the discussions of HCPs to the general population's discussions to explore similarities and differences. This study described the emotional status of HCPs as expressed in their Twitter discussions, unconfirmed by questionnaires or interviews. The correlations presented in this paper do not imply causality. Yet, the correlation between fear and epidemic measurements was confirmed by another empirical study [47] and theory [51].

Recent works showed that geographical differences [52,53], seasonality, and mass media [54] can influence social media discourses. However, in this study, we assumed that HCPs worldwide were exposed to similar conditions during the pandemic and had similar professional backgrounds and training. Hence, we considered the HCPs as a single study population without addressing cultural and geographical differences. Further studies should focus on local circumstances and cultural aspects of each location, seasonality, and the effects of mass media.

Conclusions

HCPs are at the core of pandemic response, and the impacts of the pandemic were found to have severe mental health and risk behavior implications during former events. A rigorous multistep process of selecting the study population, which involved active machine learning and manual curation, resulted in a high-quality data set of English tweets from 53,063 HCPs.

Analyses of longitudinal data over the year 2020 revealed that about 44.5% of Twitter discussions from HCPs were directly related to professional content. The rise in discussions following the pandemic waves were mostly focused on professional content. Exploring emotional trends expressed in Twitter discussions showed that the emotional realm of HCPs was affected during the COVID-19 pandemic. Thus, it may indicate the utmost importance in providing emotional support for HCPs to prevent fatigue, burnout, and mental health disorders in the postpandemic period.

In addition, the results clearly showed that fear and other emotions in the HCP discourse carried the signal reflecting the current situation and the near-anticipated future. Therefore, decision makers should invest resources into listening to the HCP community in the broadest sense expanding beyond epidemiology professionals. Brief surveys, 1- to 2-minute interviews at workplaces, and online social media analyses may be good sources of such elicitation. Also, the increase in fear 2 weeks before the pandemic waves ($\Delta Confirmed$) indicated that HCPs were in a position, and with adequate qualifications, to anticipate the pandemic development. Future research directions could include identifying and examining the major factors leading to fatigue and burnout among HCPs using machine learning techniques. Also, recommendations for preventing these adverse effects could be helpful in improving HCPs' experience in the face of long-term emergencies like a pandemic.

Script Availability

The scripts used for the analysis, as well as those used to create the figures for the paper, are available in Multimedia Appendix 2 and on GitHub [55].

Acknowledgments

This research was funded by an internal grant from the Ben-Gurion University of the Negev, Coronavirus Taskforce, and an Israeli Ministry of Science and Technology Coronavirus research grant. Sponsors were not involved in the study's design nor its interpretation.

Authors' Contributions

RP and OC conceptualized, designed, and led the study. AE, IP, and ICC identified and validated the study population's data set and analyzed and interpreted the data. All authors drafted, reviewed, and approved the final version of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary figures and tables.
[DOCX File , 91 KB - jmir_v23i10e30217_app1.docx]

Multimedia Appendix 2
Scripts that were used for analysis and to create the figures for the paper.
[ZIP File (Zip Archive), 2344 KB - jmir_v23i10e30217_app2.zip]

References

1. Hiscott J, Alexandridi M, Muscolini M, Tassone E, Palermo E, Soultsioti M, et al. The global impact of the coronavirus pandemic. Cytokine Growth Factor Rev 2020 Jun;53:1-9 [FREE Full text] [doi: 10.1016/j.cytogfr.2020.05.010] [Medline: 32487439]
2. Singh P, Singh S, Sohal M, Dwivedi YK, Kahlon KS, Sawhney RS. Psychological fear and anxiety caused by COVID-19: Insights from Twitter analytics. Asian J Psychiatr 2020 Dec;54:102280 [FREE Full text] [doi: 10.1016/j.ajp.2020.102280] [Medline: 32688277]
3. Xu P, Dredze M, Broniatowski DA. The Twitter social mobility index: Measuring social distancing practices with geolocated tweets. J Med Internet Res 2020 Dec 03;22(12):e21499 [FREE Full text] [doi: 10.2196/21499] [Medline: 33048823]
4. Naseem U, Razzak I, Khushi M, Eklund PW, Kim J. COVIDSenti: A large-scale benchmark Twitter data set for COVID-19 sentiment analysis. IEEE Trans Comput Soc Syst 2021 Aug;8(4):1003-1015. [doi: 10.1109/tcss.2021.3051189]
5. Budhwani H, Sun R. Creating COVID-19 stigma by referencing the novel coronavirus as the "Chinese virus" on Twitter: Quantitative analysis of social media data. J Med Internet Res 2020 May 06;22(5):e19301 [FREE Full text] [doi: 10.2196/19301] [Medline: 32343669]
6. Park HW, Park S, Chong M. Conversations and medical news frames on Twitter: Infodemiological study on COVID-19 in South Korea. J Med Internet Res 2020 May 05;22(5):e18897 [FREE Full text] [doi: 10.2196/18897] [Medline: 32325426]
7. Sha H, Al Hasan M, Mohler G, Brantingham PJ. Dynamic topic modeling of the COVID-19 Twitter narrative among US governors and cabinet executives. In: Proceedings of the 5th International Workshop on Social Sensing (SocialSens 2020). 2020 Presented at: 5th International Workshop on Social Sensing (SocialSens 2020); June 8, 2020; Virtual URL: https://arxiv.org/pdf/2004.11692v1.pdf
8. Elmer T, Mepham K, Stadtfeld C. Students under lockdown: Comparisons of students' social networks and mental health before and during the COVID-19 crisis in Switzerland. PLoS One 2020;15(7):e0236337 [FREE Full text] [doi: 10.1371/journal.pone.0236337] [Medline: 32702065]
9. Ojo A, Guntuku SC, Zheng M, Beidas RS, Ranney ML. How health care workers wield influence through Twitter hashtags: Retrospective cross-sectional study of the gun violence and COVID-19 public health crises. JMIR Public Health Surveill 2021 Jan 06;7(1):e24562 [FREE Full text] [doi: 10.2196/24562] [Medline: 33315578]
10. Everybody's Business: Strengthening Health Systems to Improve Health Outcomes. WHO'S Framework for Action. Geneva, Switzerland: World Health Organization; 2007. URL: https://apps.who.int/iris/bitstream/handle/10665/43918/9789241596077_eng.pdf [accessed 2021-10-15]
11. Working for Health and Growth: Investing in the Health Workforce. Report of the High-Level Commission on Health Employment and Economic Growth. Geneva, Switzerland: World Health Organization; 2016. URL: https://apps.who.int/iris/bitstream/handle/10665/250047/97892415?sequence=1 [accessed 2021-10-15]
12. Veenema TG, Griffin A, Gable AR, MacIntyre L, Simons RN, Couig MP, et al. Nurses as leaders in disaster preparedness and response--A call to action. J Nurs Scholarsh 2016 Mar;48(2):187-200. [doi: 10.1111/jnu.12198] [Medline: 26869230]
13. Shanafelt T, Ripp J, Trockel M. Understanding and addressing sources of anxiety among health care professionals during the COVID-19 pandemic. JAMA 2020 Apr 07:2133-2134. [doi: 10.1001/jama.2020.5893] [Medline: 32259193]
14. Adams JG, Walls RM. Supporting the health care workforce during the COVID-19 global epidemic. JAMA 2020 Mar 12:1439-1440. [doi: 10.1001/jama.2020.3972] [Medline: 32163102]
15. Olsson F. A Literature Survey of Active Machine Learning in the Context of Natural Language Processing. SICS Technical Report T2009:06. Kista, Sweden: Swedish Institute of Computer Science; 2009 Apr 17. URL: http://ri.diva-portal.org/smash/get/diva2:1042586/FULLTEXT01.pdf [accessed 2021-10-15]
16. Lo SL, Chiong R, Cornforth D. Using support vector machine ensembles for target audience classification on Twitter. PLoS One 2015;10(4):e0122855 [FREE Full text] [doi: 10.1371/journal.pone.0122855] [Medline: 25874768]

17. Lewis DD, Gale WA. A sequential algorithm for training text classifiers. In: Proceedings of the 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval.: Springer; 1994 Presented at: 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval; July 3-6, 1994; Dublin, Ireland p. 3-12. [doi: 10.1007/978-1-4471-2099-5_1]
18. Loper E, Bird S. NLTK: The Natural Language Toolkit. In: Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics. 2002 Presented at: ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics; July 7, 2002; Philadelphia, PA p. 63-70 URL: https://dl.acm.org/doi/pdf/10.3115/1118108.1118117 [doi: 10.3115/1118108.1118117]
19. Fellbaum C, editor. WordNet: An Electronic Lexical Database. Cambridge, MA: The MIT Press; 1998.
20. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003;3:993-1022 [FREE Full text]
21. Řehůřek R, Sojka P. Gensim—Statistical semantics in Python . In: Proceedings of the 4th European Meeting on Python in Science (EuroScipy 2011). 2011 Presented at: 4th European Meeting on Python in Science (EuroScipy 2011); August 25-28, 2011; Paris, France URL: https://www.fi.muni.cz/usr/sojka/posters/rehurek-sojka-scipy2011.pdf
22. Röder M, Both A, Hinneburg A. Exploring the space of topic coherence measures. In: Proceedings of the 8th ACM International Conference on Web Search and Data Mining. New York, NY: Association for Computing Machinery; 2015 Presented at: 8th ACM International Conference on Web Search and Data Mining; February 2-6, 2015; Shanghai, China p. 399-408. [doi: 10.1145/2684822.2685324]
23. Hutto C, Gilbert E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media. 2014 Presented at: 8th International AAAI Conference on Weblogs and Social Media; June 1-4, 2014; Ann Arbor, MI p. 216-225 URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14550/14399
24. Colneric N, Demsar J. Emotion recognition on Twitter: Comparative study and training a unison model. IEEE Trans Affect Comput 2020 Jul 1;11(3):433-446. [doi: 10.1109/taffc.2018.2807817]
25. Ekman P. Basic emotions. In: Dalgleish T, Power MJ, editors. Handbook of Cognition and Emotion. West Sussex, UK: John Wiley & Sons; 1999:45-60.
26. Jun J, Tucker S, Melnyk BM. Clinician mental health and well-being during global healthcare crises: Evidence learned from prior epidemics for COVID-19 pandemic. Worldviews Evid Based Nurs 2020 Jun;17(3):182-184. [doi: 10.1111/wvn.12439] [Medline: 32246793]
27. Beiner G. Out in the cold and back: New-found interest in the great flu. Cult Soc Hist 2006;3(4):496-505. [doi: 10.1191/1478003806cs070ra]
28. Li J, Xu Q, Cuomo R, Purushothaman V, Mackey T. Data mining and content analysis of the Chinese social media platform Weibo during the early COVID-19 outbreak: Retrospective observational infoveillance study. JMIR Public Health Surveill 2020 Apr 21;6(2):e18700 [FREE Full text] [doi: 10.2196/18700] [Medline: 32293582]
29. Shen C, Chen A, Luo C, Zhang J, Feng B, Liao W. Using reports of symptoms and diagnoses on social media to predict COVID-19 case counts in mainland China: Observational infoveillance study. J Med Internet Res 2020 May 28;22(5):e19421 [FREE Full text] [doi: 10.2196/19421] [Medline: 32452804]
30. Sarker A, Lakamana S, Hogg-Bremer W, Xie A, Al-Garadi MA, Yang Y. Self-reported COVID-19 symptoms on Twitter: An analysis and a research resource. J Am Med Inform Assoc 2020 Aug 01;27(8):1310-1315 [FREE Full text] [doi: 10.1093/jamia/ocaa116] [Medline: 32620975]
31. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: A surveillance study. J Med Internet Res 2020 Apr 09;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
32. Wahbeh A, Nasralah T, Al-Ramahi M, El-Gayar O. Mining physicians' opinions on social media to obtain insights into COVID-19: Mixed methods analysis. JMIR Public Health Surveill 2020 Jun 18;6(2):e19276 [FREE Full text] [doi: 10.2196/19276] [Medline: 32421686]
33. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, et al. Global sentiments surrounding the COVID-19 pandemic on Twitter: Analysis of Twitter trends. JMIR Public Health Surveill 2020 May 22;6(2):e19447 [FREE Full text] [doi: 10.2196/19447] [Medline: 32412418]
34. Pobiruchin M, Zowalla R, Wiesner M. Temporal and location variations, and link categories for the dissemination of COVID-19-related information on Twitter during the SARS-CoV-2 outbreak in Europe: Infoveillance study. J Med Internet Res 2020 Aug 28;22(8):e19629 [FREE Full text] [doi: 10.2196/19629] [Medline: 32790641]
35. Chandrasekaran R, Mehta V, Valkunde T, Moustakas E. Topics, trends, and sentiments of tweets about the COVID-19 pandemic: Temporal infoveillance study. J Med Internet Res 2020 Oct 23;22(10):e22624 [FREE Full text] [doi: 10.2196/22624] [Medline: 33006937]
36. Mahmudul H. Exploring sense of community in self-governed virtual communities in times of crisis. In: Proceedings of the Information Systems for Crisis Response And Management (ISCRAM) Asia Pacific Conference. 2018 Presented at: Information Systems for Crisis Response And Management (ISCRAM) Asia Pacific Conference; November 5-7, 2018;

Wellington, New Zealand URL: https://www.confer.co.nz/iscramasiapacific2018/wp-content/uploads/2019/01/Submission_0043_-_PosterFile_ID-24895-80-156-Hasan-Mahmudul.pdf
37. Neubaum G, Rösner L, Rosenthal-von der Pütten AM, Krämer NC. Psychosocial functions of social media usage in a disaster situation: A multi-methodological approach. Comput Human Behav 2014 May;34:28-38. [doi: 10.1016/j.chb.2014.01.021]
38. Pershad Y, Hangge PT, Albadawi H, Oklu R. Social medicine: Twitter in healthcare. J Clin Med 2018 May 28;7(6):121 [FREE Full text] [doi: 10.3390/jcm7060121] [Medline: 29843360]
39. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: Systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
40. Stukus DR. The role of medical professionals in social media. In: Social Media for Medical Professionals. Cham, Switzerland: Springer; 2019:65-82.
41. Saleh SN, Lehmann CU, McDonald SA, Basit MA, Medford RJ. Understanding public perception of coronavirus disease 2019 (COVID-19) social distancing on Twitter. Infect Control Hosp Epidemiol 2021 Feb;42(2):131-138 [FREE Full text] [doi: 10.1017/ice.2020.406] [Medline: 32758315]
42. Xue J, Chen J, Chen C, Zheng C, Li S, Zhu T. Public discourse and sentiment during the COVID 19 pandemic: Using latent Dirichlet allocation for topic modeling on Twitter. PLoS One 2020;15(9):e0239441 [FREE Full text] [doi: 10.1371/journal.pone.0239441] [Medline: 32976519]
43. Zhang Y, Lyu H, Liu Y, Zhang X, Wang Y, Luo J. Monitoring depression trends on Twitter during the COVID-19 pandemic: Observational study. JMIR Infodemiology 2021;1(1):e26769 [FREE Full text] [doi: 10.2196/26769] [Medline: 34458682]
44. Chew NWS, Lee GKH, Tan BYQ, Jing M, Goh Y, Ngiam NJH, et al. A multinational, multicentre study on the psychological outcomes and associated physical symptoms amongst healthcare workers during COVID-19 outbreak. Brain Behav Immun 2020 Apr 21;88:559-565 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.049] [Medline: 32330593]
45. Ruiz-Fernández MD, Ramos-Pichardo JD, Ibáñez-Masero O, Cabrera-Troya J, Carmona-Rega MI, Ortega-Galán Á. Compassion fatigue, burnout, compassion satisfaction and perceived stress in healthcare professionals during the COVID-19 health crisis in Spain. J Clin Nurs 2020 Nov;29(21-22):4321-4330. [doi: 10.1111/jocn.15469] [Medline: 32860287]
46. Di Tella M, Romeo A, Benfante A, Castelli L. Mental health of healthcare workers during the COVID-19 pandemic in Italy. J Eval Clin Pract 2020 Dec;26(6):1583-1587. [doi: 10.1111/jep.13444] [Medline: 32710481]
47. Trnka R, Lorencova R. Fear, anger, and media-induced trauma during the outbreak of COVID-19 in the Czech Republic. Psychol Trauma 2020 Jul;12(5):546-549. [doi: 10.1037/tra0000675] [Medline: 32538654]
48. Fitzpatrick KM, Harris C, Drawve G. Fear of COVID-19 and the mental health consequences in America. Psychol Trauma 2020 Aug;12(S1):S17-S21. [doi: 10.1037/tra0000924] [Medline: 32496100]
49. Yıldırım M, Arslan G, Özaslan A. Perceived risk and mental health problems among healthcare professionals during COVID-19 pandemic: Exploring the mediating effects of resilience and coronavirus fear. Int J Ment Health Addict 2020 Nov 16:1-11 [FREE Full text] [doi: 10.1007/s11469-020-00424-8] [Medline: 33223977]
50. Braquehais MD, Vargas-Cáceres S, Gómez-Durán E, Nieva G, Valero S, Casas M, et al. The impact of the COVID-19 pandemic on the mental health of healthcare professionals. QJM 2020 Jun 22;113(9):613-617 [FREE Full text] [doi: 10.1093/qjmed/hcaa207] [Medline: 32569374]
51. Epstein JM, Parker J, Cummings D, Hammond RA. Coupled contagion dynamics of fear and disease: Mathematical and computational explorations. PLoS One 2008;3(12):e3955 [FREE Full text] [doi: 10.1371/journal.pone.0003955] [Medline: 19079607]
52. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]
53. Gore RJ, Diallo S, Padilla J. You are what you tweet: Connecting the geographic variation in America's obesity rate to Twitter content. PLoS One 2015;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]
54. Sousa-Pinto B, Anto A, Czarlewski W, Anto JM, Fonseca JA, Bousquet J. Assessment of the impact of media coverage on COVID-19-related Google Trends data: Infodemiology study. J Med Internet Res 2020 Aug 10;22(8):e19611 [FREE Full text] [doi: 10.2196/19611] [Medline: 32530816]
55. HCP_State_of_Mind_covid_19. GitHub. URL: https://github.com/aviade5/HCP_State_of_Mind_covid_19 [accessed 2021-10-19]

Abbreviations

HCP: health care professional
LDA: latent Dirichlet allocation
MD: medical doctor
NLTK: Natural Language Toolkit
POI: point of interest

RN: registered nurse
SVM: support vector machine
VADER: Valence Aware Dictionary and Sentiment Reasoner

Edited by C Basch; submitted 05.05.21; peer-reviewed by A Rovetta, R Gore; comments to author 17.06.21; revised version received 08.07.21; accepted 23.07.21; published 22.10.21.

Please cite as:
Elyashar A, Plochotnikov I, Cohen IC, Puzis R, Cohen O
The State of Mind of Health Care Professionals in Light of the COVID-19 Pandemic: Text Analysis Study of Twitter Discourses
J Med Internet Res 2021;23(10):e30217
URL: https://www.jmir.org/2021/10/e30217
doi:10.2196/30217
PMID:34550899

Original Paper

Using Twitter Comments to Understand People's Experiences of UK Health Care During the COVID-19 Pandemic: Thematic and Sentiment Analysis

Esther Ainley[1], BSc, PhD; Cara Witwicki[1], BSc; Amy Tallett[1], BSc, PhD; Chris Graham[1], MA (Oxon)

Picker Institute Europe, Oxford, United Kingdom

Corresponding Author:
Esther Ainley, BSc, PhD
Picker Institute Europe
Buxton Court
3 West Way
Oxford, OX2 0JB
United Kingdom
Phone: 44 01865208168
Email: esther.ainley@pickereurope.ac.uk

Abstract

Background: The COVID-19 pandemic has led to changes in health service utilization patterns and a rapid rise in care being delivered remotely. However, there has been little published research examining patients' experiences of accessing remote consultations since COVID-19. Such research is important as remote methods for delivering some care may be maintained in the future.

Objective: The aim of this study was to use content from Twitter to understand discourse around health and care delivery in the United Kingdom as a result of COVID-19, focusing on Twitter users' views on and attitudes toward care being delivered remotely.

Methods: Tweets posted from the United Kingdom between January 2018 and October 2020 were extracted using the Twitter application programming interface. A total of 1408 tweets across three search terms were extracted into Excel; 161 tweets were removed following deduplication and 610 were identified as irrelevant to the research question. The remaining relevant tweets (N=637) were coded into categories using NVivo software, and assigned a positive, neutral, or negative sentiment. To examine views of remote care over time, the coded data were imported back into Excel so that each tweet was associated with both a theme and sentiment.

Results: The volume of tweets on remote care delivery increased markedly following the COVID-19 outbreak. Five main themes were identified in the tweets: access to remote care (n=267), quality of remote care (n=130), anticipation of remote care (n=39), online booking and asynchronous communication (n=85), and publicizing changes to services or care delivery (n=160). Mixed public attitudes and experiences to the changes in service delivery were found. The proportion of positive tweets regarding access to, and quality of, remote care was higher in the immediate period following the COVID-19 outbreak (March-May 2020) when compared to the time before COVID-19 onset and the time when restrictions from the first lockdown eased (June-October 2020).

Conclusions: Using Twitter data to address our research questions proved beneficial for providing rapid access to Twitter users' attitudes to remote care delivery at a time when it would have been difficult to conduct primary research due to COVID-19. This approach allowed us to examine the discourse on remote care over a relatively long period and to explore shifting attitudes of Twitter users at a time of rapid changes in care delivery. The mixed attitudes toward remote care highlight the importance for patients to have a choice over the type of consultation that best suits their needs, and to ensure that the increased use of technology for delivering care does not become a barrier for some. The finding that overall sentiment about remote care was more positive in the early stages of the pandemic but has since declined emphasizes the need for a continued examination of people's preference, particularly if remote appointments are likely to remain central to health care delivery.

(J Med Internet Res 2021;23(10):e31101) doi:10.2196/31101

KEYWORDS

patient experience; COVID-19; remote health care; phone consultation; video consultation; Twitter; sentiment analysis; social media; digital health; public health; public opinion

Introduction

The COVID-19 pandemic has presented many challenges to health and care services. New methods of care delivery have been rapidly introduced to create capacity in hospitals, enable health care professionals to work remotely, and to reduce the risk of transmitting the virus in care settings [1]. Care was adapted at speed, and people had to rapidly learn new ways of navigating the health and social care system, such as accessing care remotely. Since the onset of COVID-19 in spring 2020, general practitioner (GP) practices have provided a much higher proportion of consultations by phone, although not by video consultation. In 2020, the proportion of phone consultations increased from 14% in February to 28% in March and then stabilized at 48% between April and June [2]. In contrast, the proportion of online/video appointments (including nonvideo-based online consultations such as live chat or internet telephony [VoIP]) remained at less than 1% over the same period [2]. However, the quality of these data is likely to be impacted by variations in the approach to appointment management between practices, and it is suggested that many video consultations start as a telephone appointment and then switch to video, which therefore may be undercounted [2]. The uptake of video consultation (Near Me) in general practice within Scotland increased markedly from 38 consultations in February 2020 to a peak of 14,602 in May 2020, although it then decreased by 20% in June-August 2020 when lockdown restrictions eased [3].

Although research prior to the COVID-19 outbreak has examined patients' experiences of receiving care remotely, it is argued that the findings may not be applicable to the current climate where services are being impacted by COVID-19 [4]. Several studies have examined the impact of COVID-19 on service delivery changes, but there has been little published research examining patients' experiences of accessing remote consultations since the COVID-19 pandemic. A patient survey conducted by Oxleas National Health Service (NHS) Foundation Trust showed that the "convenience" of video consultations was the main theme that arose [5]. It is vital that such research is undertaken to understand and learn directly from people's experiences, particularly as remote methods for delivering care are likely to be maintained [1].

Social media sites such as Twitter provide opportunities for research to understand how people are experiencing care. Twitter data can be useful for exploring people's opinions on health issues or treatment [6-9], insights into previous pandemics [10,11], and public reactions to the COVID-19 outbreak [12-14]. There are both strengths and limitations of using Twitter data as a source for research. One advantage is that it allows quick and relatively easy access to people's views on particular topics, and the data can be used without obtaining explicit informed consent since it is part of the public domain [15,16]. Moreover, Twitter data are useful to be able to explore people's views when it may be inappropriate and difficult to conduct primary research. Research using Twitter may also allow the voices of people who may be more critical of services to be heard, which may be missed if only face-to-face methods are used [17]. However, using Twitter as a source of data is limited to those with access to the internet and who engage in this particular social media platform. Furthermore, there is evidence that British Twitter users are not representative of the general population; they are generally younger, wealthier, and better educated [18,19]. This means that caution should be taken when extrapolating the findings from this research to the wider population.

The aim of the study was to use content from Twitter to understand discourse around health and care delivery in the United Kingdom as a result of COVID-19, focusing on the views and attitudes related to care being delivered remotely (including through video consultations and telephone calls, as well as other innovative methods).

Methods

Identification and Collection of Tweets

Three search terms were used to collect relevant tweets to address the research objective (Table 1). For each search term, the following criteria were specified: date range, January 1, 2018 to October 10, 2020 (date of extraction); location, restrict to the United Kingdom; and language, English.

Table 1. Number of tweets extracted from Twitter for each search term.

Search term	Number of tweets extracted
[Video|Virtual|Remote|*phone|Telehealth|Telecare|Online AND Consultation|Appointment AND GP|Doctor|Dr]	764
["Video|Virtual|Remote|*phone|Telehealth|Telecare|Online Consultation|Appointment" AND Care|NHS|Nurse|Physiotherapist|"Occupational therapist"|Chiropodist|Podiatrist|"Health visitor"|Dietician]	494
[Video|Virtual|Remote|*phone|Telehealth|Telecare|Online AND Consultation|Appointment AND "chronic|ongoing condition"|Hypertension|"High Blood Pressure"|Depression|Diabetes|Asthma| "Kidney disease"|Heart|Cardiovascular|Cancer|COPD|"Chronic obstructive pulmonary disease"|Stroke|"mental health"]	150
Total	1408

The third search term sought to extract tweets posted by or referring to people with long-term conditions to understand their experiences of remote care. The names of specific long-term conditions were included in the search term rather

than more general terms such as "long-term condition" or "chronic condition" that are less likely to be used in tweets. These were based on the most prevalent conditions in England reported in the Quality and Outcome Framework [20].

Twitter data acquisition was achieved using a scraper written in Python 3, which interfaced with the official Twitter search application programming interface (API). Search terms and specifications were converted into Twitter API query language. The scraper made requests to the API for data fitting a particular set of criteria as outlined in the research brief, and then would scroll through that data, writing it to files for delivery and processing. To restrict the search to the United Kingdom, "place" information (a form of geographic tagging) was used to restrict to UK countries. This was a more favorable approach than using longitude and latitude data, which might include tweets posted outside the United Kingdom (such as parts of France or the Republic of Ireland) or exclude areas that should be included (such as the Isle of Wight).

A total of 1408 comments across the three search terms (detailed in Table 1) were extracted. In addition to the tweet text, the following metadata were acquired: date the tweet was posted; username and ID; tweet ID; numbers of Likes, retweets, replies; user bio information (eg, user description, user follower count, geographical location ["place ID" and "place name"]).

Data Cleaning and Analysis

Tweets and the associated metadata extracted from the scrape were imported into Excel. Of the 1408 tweets extracted, 161 duplicate tweets across the search terms were identified via the unique tweet identification number. After removal of duplicates, 1247 comments remained. The tweets and dates posted were imported into NVivo software for manual coding. To develop the coding frame, two researchers analyzed a sample of 300 comments each and coded them thematically, using an inductive and deductive approach to coding. An initial codebook was discussed and agreed upon. This was revised following further coding and additional nodes were added to the codebook when new topics were identified. After final development of the codebook, Cohen κ scores were calculated both for the primary theme coding and sublevel coding between the two researchers for 200 jointly coded tweets. This showed a very high level of agreement for the primary theme coding (κ=0.93) and a good level of agreement for the sublevel coding (κ=0.76).

Seven percent of the tweets were coded to more than one theme. Many of the codes had a positive, neutral, and negative subcategory to aid comparison across different types of remote care delivery and to understand sentiment. The tweets assigned to a neutral sentiment referred to remote care without any opinions expressed (such as people stating that they had accessed/attended a telephone appointment).

Following the coding process, both researchers examined the tweets assigned to each of the codes and grouped comments into key themes. These themes were then analyzed to identify topics and patterns in the data.

Identifying Irrelevant Comments

During the manual coding, 610/1247 (48.92%) tweets were identified as irrelevant to the research question and were coded as "unusable." These comments were varied in nature and covered a range of topics (see Textbox 1).

A total of 637 tweets were included in the analysis following the removal of duplicate and irrelevant tweets.

Textbox 1. Types of tweets identified as irrelevant to the research objective.

- Tweets about accessing general practitioner appointments that did not refer to remote care (most of these were posted before the pandemic and appeared to refer to face-to-face consultations)
- Tweets about COVID-19 that were not directly related to the research question, such as people tweeting about their symptoms, the National Health Service test and trace service, or the virus in general
- Health-related tweets but not about people's views or experiences of care and/or how these have been impacted as a result of COVID-19
- Nonhealth-related tweets such as tweets referring to virtual appointments for British Gas, hair salons, etc
- Tweets that included words such as "online" or "video," but were not relevant to the research question, such as those referring to people watching health-related videos or health care providers reminding patients to book flu vaccines online
- Tweets that could not be understood out of context, such as replies to tweets that made little sense on their own

Results

Overall Frequency and Sentiment of Tweets

Table 2 displays the frequency of tweets by month that referred to remote care, showing that the COVID-19 outbreak increased the discourse on remote care delivery. Based on the three search terms, there was an average of 10 monthly tweets between January 2018 and February 2020, compared with a monthly average of 50 tweets between March and September 2020 (October was not included in this calculation as the Twitter data were only extracted up until October 10). There was a sharp rise in the number of tweets in March 2020 when the United Kingdom first went into a country-wide lockdown at the onset of the pandemic.

Table 2. Number and percentage of tweets referencing remote care over time (N=637).

Year	Tweets, n (%)
2018	
January	8 (1.3)
February	11 (1.7)
March	7 (1.1)
April	8 (1.3)
May	7 (1.1)
June	16 (2.5)
July	6 (0.9)
August	8 (1.3)
September	13 (2.0)
October	13 (2.0)
November	9 (1.4)
December	3 (0.5)
2019	
January	18 (2.8)
February	7 (1.1)
March	6 (0.9)
April	6 (0.9)
May	8 (1.3)
June	4 (0.6)
July	6 (0.9)
August	15 (2.4)
September	13 (2.0)
October	10 (1.6)
November	9 (1.4)
December	8 (1.3)
2020	
January	16 (2.5)
February	17 (2.7)
March	90 (14.1)
April	36 (5.7)
May	48 (7.5)
June	27 (4.2)
July	42 (6.6)
August	55 (8.6)
September	54 (8.5)
October	33 (5.2)

Thematic Analysis

Overall Themes

There were five main themes identified in the tweets extracted, which are summarized in Textbox 2.

The largest number of tweets were related to accessing remote care. As some tweets were coded under more than one theme, the total number of tweets used in the thematic analysis (N=681) is greater than the overall number of tweets in the dataset (N=637).

Textbox 2. Main themes identified.

1. Access to remote care appointments (267/681 tweets, 39.2%)

Views on accessing phone or video appointments, including the ease/difficulty of getting an appointment

2. Quality of remote care delivery (130/681 tweets, 19.1%)

Views/experiences on the standard of care provided and the nature of the interaction with health care professionals

3. Anticipation of remote care (39/681 tweets, 5.7%)

Views and attitudes toward remote care appointments ahead of receiving such care

4. Online booking and asynchronous communication (85/681 tweets, 12.5%)

Attitudes/experiences of using online appointment booking systems or asynchronous approaches to communicating with health care professionals (eg, messaging systems)

5. Publicizing changes to services or care delivery (160/681 tweets, 23.5%)

Tweets publicizing remote ways of delivering care or informing people of changes to care as a result of COVID-19

The "online booking/asynchronous communication" and "publicizing changes to care" themes are not discussed in this paper as they were considered to be less relevant to this research, which is focused on the views to care being delivered remotely. Each of the other themes is discussed in more detail below.

Access to Remote Care

Proportion of Tweets Related to Theme

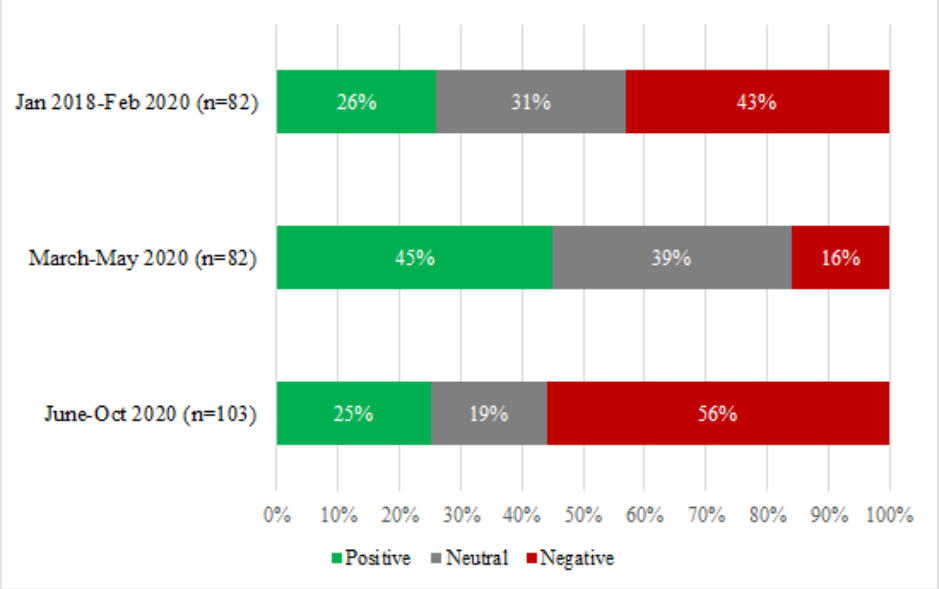
Of the 267 tweets that were related to accessing care remotely, those referring to phone consultations accounted for 81.3% (n=217), with video/online consultations accounting for 18.4% (n=49) of tweets. One tweet (0.4%) referred to accessing both a phone and video consultation. Comments were posted about accessing phone or video appointments both before and since COVID-19, although the number of tweets increased markedly since March 2020 in a similar pattern to that observed with all coded comments (Table 2). Based on our search terms, there was an average of 3 monthly tweets on accessing remote care appointments between January 2018 and February 2020, compared with an average of 24 tweets a month between March and September 2020.

Sentiment

The sentiment of tweets coded in this theme were mixed, with a similar proportion of positive (84/267, 31.5%), neutral (77/267, 28.8%), and negative (106/267, 39.7%) comments. To identify any changes over time, the sentiment of tweets was compared across three time periods: before COVID-19 (January 2018-February 2020), in the early stages of the pandemic (March-May 2020), and in the following stages of the pandemic when some UK restrictions had been eased (June-October 2020).

The overall sentiment of these tweets changed at different time periods (Figure 1). During the initial stages of the pandemic, there was almost double the proportion of comments with a positive tone (37/82, 45%) when compared with those posted in the time periods both before COVID-19 (21/82, 26%) and subsequently when some of the restrictions had been lifted (26/103, 25.2%).

Figure 1. The proportion of positive, neutral and negative tweets on access to remote care at different time periods (n=267).



Some tweets posted in the later period (June-October 2020) highlighted people's frustrations that only remote care appointments were still being offered, despite restrictions having eased across the United Kingdom: "Can still only get a phone

appointment with the dr yet people in service industries been back on the front line since July" [September 2020].

In terms of the positive tweets posted in the initial period following COVID-19 (March-May 2020), there were some tweets that expressed people's gratitude to receiving an appointment, which were not as evident in the later time period: "Very grateful to have just had my respiratory consultant appointment by telephone this afternoon" [April 2020].

Possible explanations for the overall change in sentiment shown in Figure 1 are considered further in the Discussion. Within the overarching theme on "access to remote care," two key subthemes were identified: ease/difficulty of getting a telephone/video appointment (including the use of remote appointments as a preliminary to a face-to-face consultation) and the lack of specific phone appointment timings.

Ease/Difficulty of Accessing a Remote Consultation

Tweets that were *positive* about accessing phone or video appointments centered on the efficiency of the remote care service, such as the speed of booking and "attending" an appointment, and the convenience of not needing to travel to a GP practice or wait for the appointment in the surgery.

> *Had to speak to my GP about a minor thing this morning and v impressed - called to arrange a phone appointment which was set up within about an hour - then could do a video call from a browser on my phone to do an examination. Hope this is something they'll continue to offer* [May 2020]

Some tweets also reflected positively on the safety of accessing care remotely during the pandemic:

> *I've just had a very interesting video consultation with Dr [name]. (My GP) We were using AccRX. This is a brilliant use of technology which means I don't have to go down to the practice. Especially useful during the current Coronovirus situation. Brilliant!* [March 2020]

In contrast, tweets that were *negative* in tone within this subtheme highlighted people's frustrations with needing to wait a long time for a phone appointment, both prior to and since COVID-19.

> *Once answered our surgery then tells us a Dr will phone us back within 5 days, and then if you're lucky the appointment may be a month away. Shocking!* [January 2019]

> *Phoned GP for appointment. Not doing appointments at this time. I can have a phone consult on 13th July!!! [2 weeks later].I'm in pain now or I wouldn't have called #NHS* [July 2020]

The difficulty in actually being able to book a phone consultation was also mentioned in several tweets due to lengthy waiting times to get through to the GP practice initially (ie, phone queues), the practice of GP receptionists triaging patients first, and phone consultations being carried out as a preliminary to a face-to-face appointment: "Patients, who telephone a GP Surgery, may have to submit to a 'telephone interrogation' by...a receptionist...before any appointment is arranged" [July 2020].

Lack of Specific Appointment Times

Some of the negative tweets about accessing remote appointments (particularly phone) were around the lack of specific appointment times or appointments running late. Such tweets were posted both prior to and since the COVID-19 pandemic.

> *I dunno why GP surgeries are using COVID as an excuse to be even more useless. I've just spent 18 minutes of my 30 minute lunch break making 44 phone calls to get an appointment. Have to ring at this exact time. In return they've given me a 2.5 hour slot they might call back in.* [September 2020]

Some people expressed their frustration that not having a specific time for a phone consultation impacted their ability to carry out daily activities, with a particular reference to confidentiality of discussions:

> *Waited weeks for a phone appointment with the doctor to discuss PCOS diagnosis. Of course I just got the Spanish inquisition about my ovaries in the queue for Morrisons. Soz queue.* [June 2020]

Quality of Remote Care

Proportion of Tweets Related to Theme

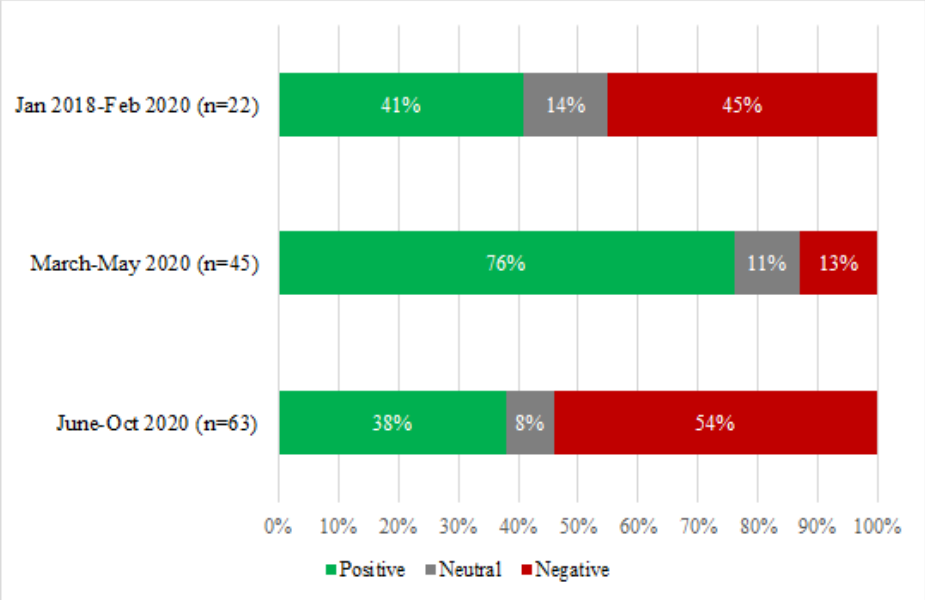
Tweets were coded in this theme if the quality of care/service received via phone or video consultations was mentioned, including the patients' interaction with health care professionals. Just over half of the tweets coded in this theme were positive in sentiment (67/130, 51.5%). Although remote consultations were being carried out before COVID-19, only 16.9% (22/130) of the tweets extracted from our search terms discussed the quality of the care delivered via phone/video before March 2020.

Sentiment

The sentiment of tweets relating to care quality changed at different time periods in a similar pattern to that noted previously for tweets about accessing remote care. During the initial stages of the pandemic (March-May 2020), there was double the proportion of comments with a positive tone when compared with the time period when some restrictions had been lifted (June-Oct 2020), and almost double the proportion of positive tweets in March-May 2020 (34/45, 76%) when compared with the prepandemic time period (Figure 2). Possible reasons for the change in sentiment noted at these two themes are considered in the Discussion.

The key subcategories identified in this theme were: efficacy of prescribing (positive comments only), standard of the care provided, and the nature of the interaction with health care professionals.

Figure 2. The proportion of positive, neutral and negative tweets on the quality of remote care at different time periods (n=130).



Efficient Prescribing

The most common positive subcategory, accounting for over one third of the positive tweets in the quality of remote care theme, was on prescribing. Tweets referred to the ease and efficiency at which prescriptions had been issued and/or received following their remote appointment. Almost all of these tweets were posted in the early stages of the COVID-19 pandemic (March-May 2020). Some tweets specifically noted that prescribing medications via remote consultations was an improvement to traditional face-to-face consultations and that this way of delivering care should be retained following the pandemic.

> *I had a speedy consultation. Sent photos in advance. A quick phone call & prescription sent to my local pharmacy. Saved me and the GP loads of time. This so should be the new normal.* [April 2020]

Standard of Care

There were several *positive* tweets about the quality of the care received via remote consultations. Some tweets were quite general and simply referred to a good service or a positive experience. Other tweets specified how the quality of care was good such as obtaining a quick referral to secondary care and receiving follow-up information.

> *Big shout out to the Trauma Physios at the [hospital] today. I had my physio appointment on the phone and an email sent with a list of exercises. I was also given another appointment in 6 weeks which could be audio visual. Fantastic service. The whole NHS are amazing.* [March 2020]

The advantage of remote consultations for people with long term conditions was also noted in a small number of tweets:

> *Had my first hospital consultation online with a Dr at [name] such an incredible experience during these difficult times, think it is going to be a way forward for people with #chronicillness Ok, now I need to attend hospital for tests, scans etc but it was less stressful.* [May 2020]

Being able to link to other patient data through technology during a remote consultation was also noted as an advantage in a few tweets:

> *It does depend on the type. My video appt with my diabetes consultant was great as I can download my insulin pump info plus daily blood tests so we could look at it together and do what was necessary. Fab appointment & no hanging about!* [August 2020]

Tweets with a *negative sentiment* in this theme were mainly in relation to phone rather than video appointments. Some people felt that not seeing a health care professional face-to-face provided a less thorough consultation; a lack of visuals and/or not being physically examined were noted as issues by some:

> *…All I could get was a telephone consultation which was alright to a point, but he can't see the area where the trouble is. When I asked would he make physical appointment he said they were emergency only, but I could send pictures by email* [July 2020]

It was implied in some tweets that the level of care or treatment received via a phone consultation was inadequate or had not fully resolved the health problem. Other tweets expressed concern and frustration that phone consultations had resulted in the condition being incorrectly treated or diagnosed by GPs:

> *My 94yo mum finally got an appointment with consultant after being in agony for 13 weeks. GP would only do phone consult and wouldn't refer her, just gave out morphine. Turns out her leg is broken and displaced and her hip is fractured. Her operation is on Wednesday* [June 2020]

Some tweets were more general in nature with people expressing their reservations about the care provided via phone appointments when compared with face-to-face interactions

with health care professionals. For instance, one person considered the change from a face-to-face appointment to a phone appointment with a hospital consultant a "downgrade":

> *four weeks to give me the anti clot injection. I have been troubled with high potassium levels but this was dealt with arms length by my GP. I was supposed to see my surgeon next week but this has been downgraded to a telephone consultation. I am beginning feel so alone* [June 2020]

A small number of tweets (both before and since COVID-19) also questioned the quality of care that could be provided via video consultation, with concerns that it could not offer the same standard of care as a face-to-face consultation. Interestingly, a mixed experience of remote care delivery was noted by one person who reflected that it may not be appropriate for all health issues:

> *When I scratched my arm in the garden, I had video consultation. GP sent prescription for antibacterial cream to pharmacy near me. That worked well. Practice nurse did asthma review over phone which seemed odd -not sure if that's really the best way. Smear test next week...* [August 2020]

Interaction With Health Care Professionals

Tweets from people that were *positive* about the interaction with health care professionals during remote consultations noted good interpersonal skills of the health professional (listening, helpful, and reassuring) in addition to their professionalism.

> *My virtual appointment today was actually the best appointment I've had in years. The dietician is going to contact me as well as the nurse to help me combat the hypos and my gastroparesis. It was great having a Dr listen to me #type1diabetes #gbdoc* [May 2020]

Negative tweets about the interaction with health care professionals referred to phone appointments being rushed and/or a lack of interest shown by the health care professional.

> *Seeing a doctor? Fat chance. A brief rushed phone consultation. No examination. And for the flu jab, guess what? It's being done as a drive-through in the car park. I might just say thanks, but no thanks. But I'm nearly 76.* [August 2020]

There were also a small number of tweets that noted some functional issues associated with video consultations that impacted the quality of care provided:

> *My son had a follow up appointment via a video chat from the hospital where the doctor said she couldn't see as the picture was out of focus.* [August 2020]

Although some of the tweets in this theme implied that the poor experiences were due to the care being delivered remotely, others appeared to be more related to the doctor's knowledge and/or interpersonal skills, which could be the same when delivering care face-to-face.

Anticipation of Remote Care

This theme captured people's views of remote care appointments before they had actually received them. This was not one of the main themes to emerge from the data, with only 39 comments coded under it and 79% of these tweets were only posted since March 2020. Many tweets under this theme were written in either a curious or sarcastic manner, possibly due to a lack of explanation or understanding of how a telephone/video appointment could work: "Receptionist at the GP surgery booked me in for a phone appointment. For my blood test." [June 2020]

Reservation was expressed by some people of how a remote care appointment could work effectively when they felt a physical examination was needed for their particular condition/health problem:

> *tried making an appointment at the dr for [name's] rash and it's a telephone appointment now I'm no expert but surely the dr needs to actually see the rash* [June 2020]
>
> *6 month cancer check soon, letter from hospital today saying it will be a telephone appointment! How's that going to work? Also referred to hospital by my GP for a throat problem, they're giving me a telephone appointment for that! Went to the dentist last week and had a filling* [July 2020]

Some people suggested in their tweets that remote appointments were potentially a waste of time, when they knew they would need a physical examination or procedure anyway, such as vaccinations and blood tests. There was a view expressed in a small number of tweets that moving to remote care appointments would put people's health at risk:

> *Sorry but how can you do a examination / consultation over the phone, appointment cancelled in August due to COVID, you are putting cancer patients at risk, and leaving them to fend for themselves #FeelingLetDown #COVID19 #AtRisk #CancerPatients2nd #melanoma* [July 2020]

Concern was also expressed in some tweets about anticipating bad news about their, or their family members', health or condition over the phone rather than in a face-to-face appointment:

> *I had a few blood tests taken 2 days ago and now the doctor surgery have phoned to say I have to arrange a telephone consultation about the results. I'm left here thinking like "what if there's something bad wrong with me?!"* [May 2018]

Although most tweets in this theme were negative in tone, there were tweets where people were supportive of the need to carry out their appointment remotely due to COVID-19:

> *So my cancer follow up appointment on Monday to check bloods and neck is to be done over the phone. must be some new technology I'm not aware of but to be fair good decision as it's non urgent so good on you [name] hospital. #NHSheroes* [March 2020]
>
> *My Acute asthma appointment is now a telephone appointment rather than F2F #CommonSense wins #cornoravirusuk* [March 2020]

There were also some tweets where people were positively anticipating their remote appointment and were appreciative that this form of care delivery was now a possibility. One person mentioned how they had been apprehensive about remote appointments, but after having had their first, was really reassured:

> *Super excited to be waiting for my video appointment with the [hospital name]! I remember back in the day when this was all a future reality*  [July 2020]

Discussion

Principal Findings

This study explored tweets about remote care delivery between January 2018 and October 2020 in the United Kingdom, and showed that the volume of related tweets increased markedly following the COVID-19 outbreak. The key themes identified in the tweets were access to remote care, quality of remote care, and anticipation of remote care. Mixed public attitudes and experiences to the changes in service delivery were found. The proportion of positive tweets regarding access to, and quality of, remote care was higher in the immediate period following the COVID-19 outbreak (March-May 2020) when compared to the time before COVID-19 onset, and the time when restrictions from the first lockdown eased (June-October 2020).

This research showed that discourse about remote health care delivery increased markedly on Twitter since the onset of the COVID-19 pandemic in March 2020. This finding is perhaps unsurprising given the changes to how services have been delivered. The pandemic has resulted in the rapid adoption of digital technology and has revolutionized the use of remote care [1,21,22]. Our research allowed us to use online data to explore how Twitter users have communicated about changes in care delivery during a time when it would have been very difficult to conduct primary data collection with patients.

Although the search terms were not restricted to primary care, and included the search terms “doctor” and “dr” in addition to “GP,” the majority of the tweets extracted were about GP consultations. This is perhaps unsurprising given that primary care is the first point of contact in the health care system for people seeking advice or treatment. There was also a much higher proportion of tweets about phone consultations rather than other types of remote care delivery such as video consultations or online messaging systems. This reflects remote primary care delivery patterns since the onset of COVID-19 in March 2020, with GP practices providing a much higher proportion of phone rather than video consultations [2]. It has been suggested that the low proportion of video consultations in general practice may, in part, be due to the limited usefulness of video consultations over telephone or face-to-face in most cases [23]. Although the uptake of primary care video consultation in Scotland increased significantly in response to COVID-19, an evaluation of the program noted variation in uptake by GP practices both within and between NHS boards, and there was also a fall in usage between June and August 2020 [3]. The reasons suggested for the limited use, and fall in usage, of video consultations across GP services in Scotland include the case mix (where telephone was sufficient for dealing with straightforward problems in patients who are well known to the clinician), the logistical challenges of using video due to high variability in appointments, problems accessing video call technology in a busy and complex work environment, and internet connectivity and local information technology helpdesk support difficulties [3].

Tweets about accessing remote consultations were more common than those referring to care quality. This could suggest that access is more pertinent to people or that there are more issues with accessing remote care compared to the quality of such care. However, this could simply reflect the limited number of characters for tweets, making it difficult to express views about care quality. Previous research that examined the content of tweets about hospitals showed that comments describing care were in the minority with various other topics being discussed [24].

Examining the sentiment of tweets about remote consultations over time revealed an interesting pattern in the data. The proportion of positive tweets regarding access to, and quality of, remote care was higher in the immediate period following the COVID-19 outbreak (March-May 2020) when compared to the time before the COVID-19 onset, and the time when restrictions from the first lockdown eased (June-October 2020). This is perhaps surprising as it might be expected that people would be less positive immediately following the lockdown when some services were in a state of flux and people would be unfamiliar with navigating new approaches to care delivery. Analysis of the tweets at different time periods provides insufficient detail to draw strong conclusions on the reason for this finding. One explanation might be that people were more understanding of the changes to care delivery initially when services were perceived to be under pressure and/or when changes were regarded to be temporary. There appeared to be several tweets posted in the later period since the pandemic (June-October 2020) that highlighted people’s frustrations that face-to-face consultations were still not being widely conducted despite the easing of restrictions.

Another explanation could relate to the fall in the number of people that sought health care during the initial period following lockdown (March-May 2020). With less people seeking health care during this period, there may have been greater availability of remote primary care appointments, resulting in a more positive experience for those people that did seek health care. This suggestion appears to be supported, in part, by experimental statistics on the length of time between booking and appointment dates. Over 60% of consultations took place on the same day as requested in April and May 2020, whereas there was a monthly downward trend subsequently, falling to 41% in October 2020 [2]. Furthermore, some patients may have decided to wait until they could have a face-to-face consultation, but then resorted to having a remote appointment in the later period (June-October) when it became clear that the virus was still having an impact on care delivery. These patients may have been less positive about the care being delivered remotely if their preference had been for face-to-face interaction.

Tweets that were negative about accessing care remotely centered on the difficulties of booking an appointment, lengthy waiting times for an appointment, and a lack of specific appointment times for phone consultations. These were noted in tweets posted both before and since COVID-19, although the volume of such tweets increased after March 2020. Results from the GP Patient Survey (prior to COVID-19) have shown a downward trend since 2012 in the proportion of patients reporting that it was "easy" to get through to their GP practice on the phone [25]. During March 2020, GP practices were advised by NHS England to triage patients before an appointment was made and to provide care remotely as much as possible [22]. Although practices moved toward more appointments being delivered remotely, our research shows that the issues surrounding the booking of appointments remained unchanged from those associated with a face-to-face appointment. Difficulties in getting through to the GP practice on the phone and issues with online appointment booking systems were frequently cited. The approach of triaging patients was tweeted about with mixed views; while some supported the need to triage patients, others were frustrated that decisions about clinical need appeared to be taken by GP practice receptionists.

The benefits of accessing care remotely were highlighted in the data. The efficiency of getting a remote appointment, and the convenience and safety of not needing to go to the GP practice were noted in tweets, with some people calling for remote consultations to be maintained in the long-term. Other studies have shown similar findings with patients valuing the convenience and time saved by video consultations when compared to face-to-face consultations [5,26,27]. Reduced travel time/expenses and convenience were also highlighted as benefits of telephone consultations by patients experiencing hospital-based telemedicine [28].

The contrast in views and experiences of accessing remote consultations may in part reflect differences between practices in the approach to appointment management and how well set up they were to deliver remote care prior to the pandemic. The Care Quality Commission found that some providers, especially larger ones, were able to move to remote consultations more easily due to already having the right technology in place [29].

There was a mix of views on the quality of care provided in remote consultations, although a higher proportion of tweets had a positive sentiment. A common positive theme, particularly immediately after the onset of COVID-19 (March-May 2020), was the ease and efficiency in which prescriptions had been issued. There is some evidence of a rapid increase in the prescribing of new medications for remote GP appointments, and it has been suggested that this may be the result of GPs being more cautious and prescribing medication "just in case," or due to a shift in the case mix where more patients with new symptoms accessed remote appointments compared to face-to-face consultations [21]. Other positive aspects about the quality of care delivered remotely were quick referrals to secondary care, receiving follow-up information, and the interpersonal skills of health care professionals. A study conducted by Oxleas NHS Foundation Trust between March and July 2020 showed that patients reported receiving the same level of care and treatment during their remote appointment as they had received previously in face-to-face appointments, although it was noted that there was a preference for being seen face-to-face [5]. Further research is required to determine people's willingness to receive remote care instead of face-to-face appointments in the longer term.

Tweets that were less positive about the quality of remote care implied that the standard of care was not as high as in face-to-face consultations, with a lack of visuals and physical examination being highlighted. Concerns were expressed by a small number of Twitter users that their condition had not be diagnosed or treated correctly via a telephone appointment. These findings support research carried out before COVID-19, which showed that, compared to face-to-face appointments, patients were less positive about the care received via remote consultations [25,30,31]. For instance, the 2020 GP Patient Survey (fieldwork January-March 2020, before the pandemic) showed that compared to face-to-face appointments, patients who had received a telephone appointment were 2% less likely to have their needs met, 4% less likely to say they were given enough time, and 4% less likely to feel that any mental health needs were recognized or understood [21].

The mixed attitudes toward remote care evident from our Twitter data support the view that although care delivered remotely can offer efficiency and convenience for patients and allows easier access for some groups of people, face-to-face consultations are more appropriate for others, or for certain conditions or situations [1,5,32-34]. It is important that patients can choose the type of consultation that best suits their needs and that the increased use of technology for delivering care does not become a barrier for some.

Policy Implications

There has been a rapid shift to delivering health care remotely since the COVID-19 pandemic, and it is important to learn from those who have been at the forefront of experiencing such changes. Being able to deliver care remotely may have potential for improving efficiencies in health and social care systems, but further research is needed. Our analysis of Twitter comments has shown mixed attitudes and experiences to these changes. The finding that overall sentiment about remote care was more positive in the early stages of the pandemic but has since declined is important. Although the reasons for this can only be speculated, it does emphasize the need for a continued examination and understanding of people's experiences as remote services continue to evolve.

There have been calls by policymakers for the increased use of remote care delivery to continue after the pandemic [35], although some GPs have found the high levels of remote care delivery a strain, have missed face-to-face contact with patients, and have been concerned about the clinical risk associated with delivering care in this way [23]. It has also been argued that sufficient funding and technical infrastructure are required to ensure that the increase in remote care provision can be successfully embedded [1,26]. A report by the Care Quality Commission noted that information technology systems were sometimes a barrier for patients and providers, with a lack of equipment in some sectors and some patients finding it difficult

to adapt to using the new digital systems [29]. Although our research has highlighted some Twitter users with positive views about remote care and the benefits that it can offer, this study also shows where improvements are needed. Further research is needed to explore the challenges and barriers associated with remote care delivery to inform the future planning and delivery of remote care. Despite the shift to more consultations now being delivered remotely, the difficulties surrounding booking and getting an appointment remain an issue for some. Health care providers should offer specific appointment times for telephone consultations, not only to improve patient's experiences but also to minimize appointments being missed and to protect patient confidentiality.

Our research also found that some people were negatively anticipating care being delivered remotely, including confusion as to why appointments requiring tests or physical procedures had been booked as a telephone consultation. Some of this confusion likely reflects initial difficulties experienced by providers due to the speed at which remote care was implemented. However, it does suggest that providing people with more information about how remote consultation works may improve public attitudes and acceptance, in addition to giving patients a better experience and avoiding the anxieties leading up to a remote care appointment.

Limitations

There were some limitations of using Twitter comments to understand views on care being delivered remotely. The data are limited to internet users who engage with this social media platform and who tweeted between January 2018 and October 2020, and therefore do not represent the views of all health care users. Although the number of internet nonusers (ie, adults who have either never used the internet or have not used it in the last 3 months) has been declining over time, 10% of the adult UK population in 2018 were internet nonusers [36]. There is a "digital divide" as internet use and digital skills vary for different groups of the population; for example, internet nonusers are disproportionately disabled, women, those aged over 75 years, and those who are not in employment [36,37]. There are therefore concerns that the delivery of care remotely may negatively impact some groups of people more than others, such as those with limited digital literacy and/or lack of access to technology [38,39]. This is difficult to examine in an analysis of Twitter comments, as such groups are less likely to use social media platforms such as Twitter to share their views.

Since tweets are unprompted, despite carefully planned search terms, the extracted data are not always relevant. Over 1000 tweets were extracted using three search terms, but more than one third of these comments were irrelevant to the research question. Several tweets were also coded into themes that did not directly relate to understanding people's views and attitudes of remote care delivery. The relatively low number of relevant comments is important to consider when comparing the proportion of tweets with a negative/neutral/positive sentiment within themes. Although there was a sufficient volume for identifying overall themes and trends, there was an insufficient level of detail to elucidate the reasons for some findings, highlighting the need for further research.

Another limitation of the study was the inability to examine any variations in views of remote care delivery by geographical region or by demographic factors such as age or gender. This was either due to the relatively low volume of relevant tweets extracted from our search terms (for geographic comparisons) or because such information was not available in the metadata (gender or age). Similarly, although one of the search terms sought to explore any changes in care delivery for people living with long-term conditions, there was an insufficient number of tweets referring to particular long-term conditions to allow for analysis by health condition. Long-term conditions are more prevalent in older and in more deprived groups [40], which are the same groups experiencing greater digital exclusion. As mentioned previously, the "economically inactive" are the most likely to be internet nonusers, particularly adults on long-term sick leave (due to health conditions that last 12 months or more) or who are disabled [36]. It is therefore likely that those with long-term conditions were underrepresented in the tweets analyzed in this study.

As previously mentioned, most of the tweets extracted were about GP consultations, and therefore our research does not provide as much insight into Twitter users' views of remote secondary care. There were also very few tweets that referenced alternative methods of remote care delivery other than telephone or video consultations. This may reflect that other approaches such as online messaging systems or live chat are not yet being widely used to deliver remote care. It may, however, point toward a limitation of the search terms used, which did not include words such as "live chat" or "message/messaging," although the word "online" was used.

Conclusion

Using Twitter data to address our research questions was beneficial for providing rapid access to Twitter users' attitudes about remote care delivery at a time when it was difficult to conduct primary research due to the COVID-19 pandemic. This approach allowed us to examine Twitter users' views and experiences of remote care, and to explore shifting attitudes at a time of rapid changes in care delivery. However, we recognize that our findings do not represent the views of all health care users, and further research using alternative methodologies such as in-depth interviews with patients could complement our findings to provide further insight into people's experiences of receiving remote care.

Acknowledgments

This research is funded by the National Institute for Health Research (NIHR) Policy Research Programme, conducted through the Quality, Safety and Outcomes Policy Research Unit, PR-PRU-1217-20702. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. The authors would like to thank the wider members

of the Quality, Safety and Outcomes Policy Research Unit for their valuable input and advice. The authors are also grateful to Hertzian Ltd for supporting access to Twitter data using specified search terms.

Authors' Contributions

Study design: CG and AT. Analysis and interpretation of data: EA, CW, AT, and CG. Drafting of manuscript: EA, CW, and AT.

Conflicts of Interest

None declared.

References

1. Hutchings R. The impact of Covid-19 on the use of digital technology in the NHS. Nuffield Trust. 2020 Aug. URL: https://www.nuffieldtrust.org.uk/files/2020-08/the-impact-of-covid-19-on-the-use-of-digital-technology-in-the-nhs-web-2.pdf [accessed 2021-01-14]
2. Appointments in General Practice October 2020. Official statistics, Experimental statistics. NHS Digital. 2020 Oct. URL: https://digital.nhs.uk/data-and-information/publications/statistical/appointments-in-general-practice/october-2020 [accessed 2021-02-10]
3. Wherton J, Greenhalgh T. Coronavirus (COVID-19) - Near Me video consulting service: evaluation 2020 - main report. Scottish Government. 2021 Mar. URL: https://www.gov.scot/publications/evaluation-near-video-consulting-service-scotland-during-covid-19-2020-main-report/ [accessed 2021-05-13]
4. Greenhalgh T. Using virtual consultations in the fight against COVID-19: Interview with Professor Trish Greenhalgh. The Health Foundation. 2020 Mar 30. URL: http://tracker.health.org.uk/news-and-comment/newsletter-features/using-virtual-consultations-in-the-fight-against-covid-19 [accessed 2021-04-06]
5. Oxleas NHS. Patient Experience of Remote Consultations during the COVID-19 Pandemic. Health Innovation Network South London. 2020 Sep. URL: https://healthinnovationnetwork.com/wp-content/uploads/2020/12/MH-Remote-Consultations-Patient-Experiences.pdf [accessed 2021-02-10]
6. Beguerisse-Díaz M, McLennan AK, Garduño-Hernández G, Barahona M, Ulijaszek SJ. The 'who' and 'what' of #diabetes on Twitter. Digit Health 2017 Jan;3:2055207616688841 [FREE Full text] [doi: 10.1177/2055207616688841] [Medline: 29942579]
7. Cheng TY, Liu L, Woo BK. Analyzing Twitter as a platform for Alzheimer-related dementia awareness: thematic analyses of tweets. JMIR Aging 2018 Dec 10;1(2):e11542 [FREE Full text] [doi: 10.2196/11542] [Medline: 31518232]
8. Gohil S, Vuik S, Darzi A. Sentiment analysis of health care tweets: review of the methods used. JMIR Public Health Surveill 2018 Apr 23;4(2):e43 [FREE Full text] [doi: 10.2196/publichealth.5789] [Medline: 29685871]
9. Conway M, Hu M, Chapman WW. Recent advances in using natural language processing to address public health research questions using social media and consumer-generated data. Yearb Med Inform 2019 Aug;28(1):208-217 [FREE Full text] [doi: 10.1055/s-0039-1677918] [Medline: 31419834]
10. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
11. Mamidi R, Miller M, Banerjee T, Romine W, Sheth A. Identifying key topics bearing negative sentiment on Twitter: insights concerning the 2015-2016 Zika epidemic. JMIR Public Health Surveill 2019 Jun 04;5(2):e11036 [FREE Full text] [doi: 10.2196/11036] [Medline: 31165711]
12. Saleh SN, Lehmann CU, McDonald SA, Basit MA, Medford RJ. Understanding public perception of coronavirus disease 2019 (COVID-19) social distancing on Twitter. Infect Control Hosp Epidemiol 2021 Feb 06;42(2):131-138 [FREE Full text] [doi: 10.1017/ice.2020.406] [Medline: 32758315]
13. Mustafa A, Ansari I, Mohanta SK, Balla S. Public reaction to COVID-19 on Twitter: a thematic analysis. EPRA Int J Multidisc Res 2020 May 25;19:349-358. [doi: 10.36713/epra4518]
14. Xue J, Chen J, Chen C, Zheng C, Li S, Zhu T. Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. PLoS One 2020 Sep 25;15(9):e0239441 [FREE Full text] [doi: 10.1371/journal.pone.0239441] [Medline: 32976519]
15. Van Hoof W, Provoost V, Pennings G. Reflections of Dutch patients on IVF treatment in Belgium: a qualitative analysis of internet forums. Hum Reprod 2013 Apr;28(4):1013-1022. [doi: 10.1093/humrep/des461] [Medline: 23335612]
16. Taylor J, Pagliari C. #Deathbedlive: the end-of-life trajectory, reflected in a cancer patient's tweets. BMC Palliat Care 2018 Jan 22;17(1):17 [FREE Full text] [doi: 10.1186/s12904-018-0273-9] [Medline: 29357865]
17. Chilman N, Morant N, Lloyd-Evans B, Wackett J, Johnson S. Twitter users' views on mental health crisis resolution team care compared with stakeholder interviews and focus groups: qualitative analysis. JMIR Ment Health 2021 Jun 29;8(6):e25742 [FREE Full text] [doi: 10.2196/25742] [Medline: 34185017]
18. Blank G. The digital divide among Twitter users and its implications for social research. Soc Sci Comput Rev 2016 Oct 12;35(6):679-697. [doi: 10.1177/0894439316671698]

19. Mellon J, Prosser C. Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. Res Polit 2017 Jul 13;4(3):205316801772000. [doi: 10.1177/2053168017720008]
20. Quality and Outcomes Framework, 2019-20. NHS Digital. 2020 Aug 20. URL: https://digital.nhs.uk/data-and-information/publications/statistical/quality-and-outcomes-framework-achievement-prevalence-and-exceptions-data/2019-20/main-findings [accessed 2020-10-07]
21. Morris J. The remote care revolution in the NHS: understanding impacts and attitudes. Nuffield Trust. 2020 Dec 16. URL: https://www.nuffieldtrust.org.uk/resource/the-remote-care-revolution-in-the-nhs-understanding-impacts-and-attitudes [accessed 2021-01-14]
22. Clarke G, Pariza P, Wolters A. How has COVID-19 affected service delivery in GP practices that offered remote consultations before the pandemic? The Health Foundation. 2020 Jul 16. URL: https://www.health.org.uk/news-and-comment/charts-and-infographics/how-has-covid-19-affected-service-delivery-in-gp-practices [accessed 2021-01-14]
23. Murphy M, Scott LJ, Salisbury C, Turner A, Scott A, Denholm R, et al. Implementation of remote consulting in UK primary care following the COVID-19 pandemic: a mixed-methods longitudinal study. Br J Gen Pract 2021;71(704):e166-e177 [FREE Full text] [doi: 10.3399/BJGP.2020.0948] [Medline: 33558332]
24. Greaves F, Laverty AA, Cano DR, Moilanen K, Pulman S, Darzi A, et al. Tweets about hospital quality: a mixed methods study. BMJ Qual Saf 2014 Oct;23(10):838-846 [FREE Full text] [doi: 10.1136/bmjqs-2014-002875] [Medline: 24748372]
25. GP Patient Survey National Report. Ipsos MORI. 2020. URL: https://www.gp-patient.co.uk/SurveysAndReports [accessed 2021-02-10]
26. Donaghy E, Atherton H, Hammersley V, McNeilly H, Bikker A, Robbins L, et al. Acceptability, benefits, and challenges of video consulting: a qualitative study in primary care. Br J Gen Pract 2019 Sep;69(686):e586-e594 [FREE Full text] [doi: 10.3399/bjgp19X704141] [Medline: 31160368]
27. Wherton J, Greenhalgh T. Evaluation of the Attend Anywhere/Near Me video consulting service in Scotland, 2019-20. Scottish Government. 2020 Jul 23. URL: https://www.gov.scot/publications/evaluation-attend-anywhere-near-video-consulting-service-scotland-2019-20-main-report/ [accessed 2021-02-18]
28. Smrke A, Younger E, Wilson R, Husson O, Farag S, Merry E, et al. Telemedicine during the COVID-19 pandemic: impact on care for rare cancers. JCO Global Oncol 2020 Nov(6):1046-1051. [doi: 10.1200/go.20.00220]
29. The state of health care and adult social care in England 2019/20. Care Quality Commission. 2020 Oct 15. URL: https://www.cqc.org.uk/sites/default/files/20201016_stateofcare1920_fullreport.pdf [accessed 2021-02-18]
30. Hammersley V, Donaghy E, Parker R, McNeilly H, Atherton H, Bikker A, et al. Comparing the content and quality of video, telephone, and face-to-face consultations: a non-randomised, quasi-experimental, exploratory study in UK primary care. Br J Gen Pract 2019 Sep;69(686):e595-e604 [FREE Full text] [doi: 10.3399/bjgp19X704573] [Medline: 31262846]
31. McKinstry B, Hammersley V, Burton C, Pinnock H, Elton R, Dowell J, et al. The quality, safety and content of telephone and face-to-face consultations: a comparative study. Qual Saf Health Care 2010 Aug 29;19(4):298-303. [doi: 10.1136/qshc.2008.027763] [Medline: 20430933]
32. The Doctor Will Zoom You Now: getting the most out of the virtual health and care experience. Healthwatch. 2020 Aug 18. URL: https://www.healthwatch.co.uk/report/2020-08-18/doctor-will-zoom-you-now-getting-most-out-virtual-health-and-care-experience [accessed 2021-02-10]
33. RCGP survey provides snapshot of how GP care is accessed in latest stages of pandemic. Royal College of General Practitioners. 2020 Jul 30. URL: https://www.rcgp.org.uk/about-us/news/2020/july/rcgp-survey-provides-snapshot-of-how-gp-care-is-accessed-in-latest-stages-of-pandemic.aspx [accessed 2021-02-10]
34. Rapson J. Digital appointments could be storing up problems for the future. HSJ. 2020 Apr 16. URL: https://www.hsj.co.uk/technology-and-innovation/digital-appointments-could-be-storing-up-problems-for-the-future/7027421.article [accessed 2021-02-10]
35. Bridge-Wilkinson A. Health secretary calls for non-urgent GP consultations to be conducted remotely. The Parliamentary Review News. 2020 Jul 31. URL: https://www.theparliamentaryreview.co.uk/news/health-secretary-calls-for-non-urgent-gp-consultations-to-be-conducted-remotely [accessed 2021-02-10]
36. Exploring the UK's digital divide. Office for National Statistics. 2019 Mar 04. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/articles/exploringtheuksdigitaldivide/2019-03-04 [accessed 2021-07-15]
37. Access and Inclusion in 2018. OFCOM. 2019 Jan 14. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0018/132912/Access-and-Inclusion-report-2018.pdf [accessed 2021-07-15]
38. Scobie S. Will a digital NHS reap the rewards policy-makers are aiming for? Nuffield Trust. 2019 Aug 20. URL: https://www.nuffieldtrust.org.uk/news-item/will-a-digital-nhs-reap-the-rewards-policy-makers-are-aiming-for [accessed 2021-02-10]
39. Majeed A, Maile E, Coronini-Cronberg S. Covid-19 is magnifying the digital divide. The BMJ Opinion. 2020 Sep 01. URL: https://blogs.bmj.com/bmj/2020/09/01/covid-19-is-magnifying-the-digital-divide/ [accessed 2021-02-10]
40. Department of Health and Social Care. Long Term Conditions Compendium of Information: Third Edition. UK Government (GOV.UK). 2012 May 30. URL: https://www.gov.uk/government/publications/long-term-conditions-compendium-of-information-third-edition [accessed 2021-01-14]

Abbreviations

API: application programming interface
GP: general practitioner
NHS: National Health Service
NIHR: National Institute for Health Research

Edited by C Basch; submitted 09.06.21; peer-reviewed by E Donaghy, M Hu; comments to author 23.06.21; revised version received 12.08.21; accepted 30.08.21; published 25.10.21.

Please cite as:
Ainley E, Witwicki C, Tallett A, Graham C
Using Twitter Comments to Understand People's Experiences of UK Health Care During the COVID-19 Pandemic: Thematic and Sentiment Analysis
J Med Internet Res 2021;23(10):e31101
URL: https://www.jmir.org/2021/10/e31101
doi:10.2196/31101
PMID:34469327

Original Paper

Emotional Tone, Analytical Thinking, and Somatosensory Processes of a Sample of Italian Tweets During the First Phases of the COVID-19 Pandemic: Observational Study

Dario Monzani[1,2*], PhD; Laura Vergani[1,2*], MSc; Silvia Francesca Maria Pizzoli[1,2], PhD; Giulia Marton[1,2], MSc; Gabriella Pravettoni[1,2], PhD

[1]Department of Oncology and Hemato-Oncology, University of Milan, Milan, Italy

[2]Applied Research Division for Cognitive and Psychological Science, IEO, European Institute of Oncology IRCCS, Milan, Italy

[*]these authors contributed equally

Corresponding Author:
Laura Vergani, MSc
Department of Oncology and Hemato-Oncology
University of Milan
Via Festa del Perdono 7
Milan, 20122
Italy
Phone: 39 029 4372099
Email: laura.vergani@ieo.it

Abstract

Background: The COVID-19 pandemic is a traumatic individual and collective chronic experience, with tremendous consequences on mental and psychological health that can also be reflected in people's use of words. Psycholinguistic analysis of tweets from Twitter allows obtaining information about people's emotional expression, analytical thinking, and somatosensory processes, which are particularly important in traumatic events contexts.

Objective: We aimed to analyze the influence of official Italian COVID-19 daily data (new cases, deaths, and hospital discharges) and the phase of managing the pandemic on how people expressed emotions and their analytical thinking and somatosensory processes in Italian tweets written during the first phases of the COVID-19 pandemic in Italy.

Methods: We retrieved 1,697,490 Italian COVID-19–related tweets written from February 24, 2020 to June 14, 2020 and analyzed them using LIWC2015 to calculate 3 summary psycholinguistic variables: emotional tone, analytical thinking, and somatosensory processes. Official daily data about new COVID-19 cases, deaths, and hospital discharges were retrieved from the Italian Prime Minister's Office and Civil Protection Department GitHub page. We considered 3 phases of managing the COVID-19 pandemic in Italy. We performed 3 general models, 1 for each summary variable as the dependent variable and with daily data and phase of managing the pandemic as independent variables.

Results: General linear models to assess differences in daily scores of emotional tone, analytical thinking, and somatosensory processes were significant ($F_{6,104}$=21.53, P<.001, R^2= .55; $F_{5,105}$=9.20, P<.001, R^2= .30; $F_{6,104}$=6.15, P<.001, R^2=.26, respectively).

Conclusions: The COVID-19 pandemic affects how people express emotions, analytical thinking, and somatosensory processes in tweets. Our study contributes to the investigation of pandemic psychological consequences through psycholinguistic analysis of social media textual data.

(J Med Internet Res 2021;23(10):e29820) doi:10.2196/29820

KEYWORDS

internet; mHealth; infodemiology; infoveillance; pandemic; public health; COVID-19; Twitter; psycholinguistic analysis; trauma

Introduction

Background

As a way to express information, news, opinions, and even private emotions and to connect people worldwide, Twitter, established in 2006, is a microblogging service that is the 13th most-used social media platform, with 340 million users [1]. In the first quarter of 2020, it registered 166 million average monetizable daily active users, with a 24% growth from 2019 [2]. And Twitter itself attributes part of this exceptional growth to a "global conversation related to the COVID-19 pandemic" [3]. While the coronavirus disease pandemic is affecting the world, regional and national lockdowns are restraining the possibility to travel and physically meet other people: Social networks, including Twitter, now represent a way to keep in touch, exchange information, solve problems, and conversate together and globally. And there is even something more.

Infodemiology is a new research field using online data and defined as "the science of distribution and determinants of information in an electronic medium, specifically the internet, or in a population, with the ultimate aim to inform public health and public policy" [4]. Among the infodemiology indicators, "metrics on the 'chatter' in discussion groups, blogs, and microblogs (eg, Twitter)" [4] are considered. Along this line, various researchers have successfully used this kind of data in the health context. Specifically, Twitter represents a unique opportunity for scholars to recruit participants, deliver interventions, or directly obtain data [5]. In particular, as a data source, it can provide population-level, real-time, high-volume, easily, publicly accessible data [5]: These are beneficial features, especially in the psychosocial field that normally relies on self-report, time-lagged questionnaires, with limited numbers of participants.

Today, Twitter-based health research represents a rapidly developing field, combining different methodologies and applying to various contexts, such as public health, infectious diseases including Ebola and influenza, neurology, and psychiatry [5]. Some studies have also been conducted in the COVID-19 pandemic context, demonstrating the feasibility of using Twitter as a means to collect valuable data to obtain deep insights in this emergency situation. Lwin and colleagues [6] collected more than 20 million tweets written worldwide during the first phases of the pandemic and studied the emotional responses to COVID-19 by using sentiment analysis. Xue and colleagues [7] used sentiment analysis alongside unsupervised machine learning and qualitative methods to identify main COVID-19–related themes discussed on Twitter, such as news, cases, and deaths, accompanied by a sentiment of fear.

Other studies, instead, relied on psycholinguistic analysis of Twitter data. Su and colleagues [8] used psycholinguistic analysis on Weibo and Twitter posts to investigate the psychological impact of lockdown measures in China and Italy: After lockdown, people used more cognitive processes and home words.

Indeed, as demonstrated by a vast amount of literature [9], the words we use in our daily lives have various links to different psychosocial variables, including mental health, psychological status, and "ongoing emotional and cognitive coping processes, and idiosyncratic reactions to crisis" [10]. In fact, the pandemic could be considered as "[…] the cause of individual and collective traumas" [11], that is, also having tremendous consequences on mental and psychological health [12,13].

Overall, psycholinguistic analysis of textual data coming from Twitter allows some advantages. Usually, assessing psychological variables requires the recruitment of a sample of participants, relying on their availability to individually administer questionnaires and instruments. This process is expensive and time-consuming, resulting in a limited amount of data, often biased by the issues associated with self-report instruments, such as a time lag between the event of interest experienced by people and the moment of data collection. Psycholinguistic analysis of Twitter data requires downloading, in a quite fast and automatic way, a massive amount of population-level data in near real time—as tweets written immediately after the event of interest—in a discreet and unobtrusive way, resulting in a faster and less expensive process.

Among the psychological variables, psycholinguistic analysis of textual Twitter data could provide information about emotional expressions, analytical thinking, and somatosensory processes, which are particularly important in traumatic event contexts.

Specifically, emotional tone is a psycholinguistic variable that summarizes the presence of positive and negative emotions in written text as the difference between positive-emotion words and negative-emotion words [10]. Individuals' expressions of emotions in language are connected to the way they experience the world and also react to, and cope with, traumatic events [9]. In particular, experiencing positive emotions after a challenging event is important for resilience [14], while some studies highlighted how, after a traumatic experience such as the September 11 attacks, the emotional tone in journal entries by people in the United States was low, in other words characterized by a negative tone, which slowly rebuilt after some time [10].

Analytical thinking is a psycholinguistic variable that reflects "the degree to which people use words that suggest formal, logical, and hierarchical thinking patterns" [15]. A lower level of analytical thinking reflects a more narrative and personal thinking pattern. The value of cognitive words in trauma narratives remains controversial: These types of words are linked with positive or negative effects on people's well-being [16].

In trauma narratives, somatosensory words, such as words related to body, sensory, and perceptual processes, assume great relevance, with a stronger presence than in other neutral or positive-tone narratives [17]. The use of this type of words is associated with the symptoms of posttraumatic stress disorder (PTSD) and depression [17-19].

Our aim was to analyze the influence of the pandemic—such as official Italian COVID-19 daily data (new cases, deaths, and hospital discharges) and the phase of managing the pandemic—on psycholinguistic variables in Italian tweets

written during the first phases of the COVID-19 pandemic in Italy.

Objective

The pandemic is characterized by daily information about new cases and deaths and by governments' decisions and restrictions that impact everyone's lives: It could be considered a collective and individual traumatic experience [11]. This traumatic experience can have profound psychological consequences on mental health and the well-being of citizens that can also be reflected, as discussed earlier, in people's use of words, specifically the emotional tone, analytical thinking, and somatosensory processes variables.

Our aim was to analyze the way people express emotions, their analytical thinking, and somatosensory processes in a sample of Italian tweets during the first phases of the COVID-19 pandemic in Italy. Specifically, we were interested in assessing the influence of official Italian COVID-19 daily data (eg, new cases, new deaths, and hospital discharges) as well as the phase of managing the outbreak on tweets occurring during the following 24 hours, specifically on the emotional tone, analytical thinking, and somatosensory processes in tweets.

Methods

Dataset

The dataset used in this study came from a large-scale COVID-19 Twitter chatter project that actively collected COVID-19 tweets from January 1, 2020 (for a brief overview, see [20]). Specifically, this dataset, which has been made freely available by Banda and colleagues [20] through Zenodo, includes tweets collected from the publicly available Twitter Stream API with a collection process that gathered any available tweets with keywords related to COVID-19 (eg, "coronavirus," "2019ncov," "COVID19," "COVID-19"). See [20] for further information on the full list of keywords and the rationale for their selection and inclusion. As of September 20, 2020, this project had collected almost 166 million unique tweets. The project only released the Tweet IDs of the collected tweets; thus, the software DocNow Hydrator was used to extract tweets. This user-friendly software has been proven effective by previous research [21,22]. We only selected tweets in the Italian language created between 6:00 pm on February 24, 2020 and 11:59 pm on June 14, 2020. Both the language and timestamp of tweets are provided directly by Twitter through its API, a tool to contribute to, engage with, and analyze the conversation happening on Twitter. We chose to focus on this period because official data about the COVID-19 outbreak were available since 6:00 pm on February 24, 2020 (ie, 3 days after Italian Patient One was tested positive), and "Phase 3" started on June 15, 2020, characterized by a sharp loosening of previous public health measures and restrictions.

In addition, official data about daily new cases, new deaths, and new discharges from hospital were also retrieved from the GitHub page of the Italian Prime Minister's Office and Civil Protection Department. From February 24, 2020 to April 17, 2020, data on the COVID-19 outbreak in Italy were communicated in a press conference held daily at 6:00 pm by the head of the Civil Protection Department. After April 17, 2020, the daily press conference was no longer held, but official information about the pandemic continued to be released at 6:00 pm through a daily bulletin.

We considered 3 different phases of managing COVID-19, characterized by distinct restrictions and measures to counteract virus spreading. The first was the outbreak, from February 24, 2020 (ie, the day on which the official Civil Protection Department 6:00 pm press conference began) to March 8, 2020: Along with the first confirmed indigenous cases, regional and national governments began to take action, including school and university closures, postponing or canceling some public events, and strict lockdown for 11 municipalities in northern Italy. The second was Phase 1, from March 9, 2020 to May 3, 2020: A "I stay home" national decree imposed lockdown in all Italian regions, and citizens were allowed to leave their homes only for documented work, health, or emergency reasons, while nonessential commercial activities were closed. The third was Phase 2, from May 4, 2020 to June 14, 2020: A gradual relaxing of lockdown restrictions began, with reopening of some services and activities, such as parks, museums, restaurants, and bars for take-away service; practicing social distance remained mandatory.

Data use complied with ethical guidelines for internet research [23]. The European Union General Data Protection Regulation 2016/679 allows for the use of anonymous data for research purposes under certain conditions. Since all analyses have been performed on public and anonymized meta-data, no institutional review board approval was required for the use of this database or the completion of this study.

Statistical Analysis

Text mining and text analysis were performed with R version 3.4.3 and Linguistic Inquiry and Word Count (LIWC) 2015. We were interested in understanding whether daily data on the COVID-19 outbreak would affect how people express emotion, cognition, and somatosensory processes in their tweets during the following 24 hours. Thus, before analysis, all tweets were preprocessed: Daily tweets from 6:00 pm to 5:59 pm the following day were merged into a single text file. For instance, the overall corpus for March 1, 2020 included aggregated text coming from 11,707 tweets from 6:00 pm on March 1 to 5:59 pm on March 2. There was a total of 1,692,181 tweets from 6:00 pm on February 24, 2020 to 23:59 pm on June 14, 2020; the number of tweets per day ranged from 6977 (on June 14, 2020) to 33,356 (on May 25, 2020) with a daily average of 15,108.76 (SD 3895.29) tweets.

Then, each daily text was analyzed with the Italian LIWC2007 Dictionary [24] and the Italian Function Words Dictionary 2015 of LIWC2015 [25]. LIWC calculates the percentage of total words in each text that falls into predefined linguistic and psycholinguistic categories. We then computed separate indexes for emotional tone, analytical thinking, and somatosensory processes. Based on previous research, each of these 3 summary variables are constructed from different LIWC categories. First, to calculate the emotional tone score, we employed the procedure described by Cohn et al [10]. Specifically, tone was computed as (positive emotion) – (negative emotion): thus, the

higher the score, the more positive the emotional tone of daily tweets. Second, analytical thinking is a factor-analytically derived dimension based on 8 function word dimensions. This dimension "captures the degree to which people use words that suggest formal, logical, and hierarchical thinking patterns" [15]. It was computed as (articles) + (prepositions) - (total pronouns) - (auxiliary) - (negations) - (conjunctions) - (adverbs) [26]: the higher the score, the higher the analytical thinking of the daily tweets. Third, as somatosensory details, in particular words related to body and perception, have been found to be common and important in different studies examining trauma narratives [16], we decided to calculate a somatosensory index, namely somatosensory processes: This index was computed as (perceptual processes) + (body). These 2 categories captured the use of words related to perceptual experiences (such as "observing, heard, feeling, rumors, touch") and body parts, processes, or diseases (such as "cheek, hands, spit, cough, flesh, brain, hearth, pain, contagious, headache, sick"), tapping into perceptual and sensory features that are meant to be common in this type of narrative. Higher scores in this index imply higher somatosensory experiences expressed in daily tweets. Since emotional tone, analytical thinking, and somatosensory processes were computed for each day by considering all the text coming from daily tweets, in all subsequent analyses, the total sample was the number (ie, 122) of days from February 24, 2020 to June 14, 2020 (with days as the unit of analysis).

We performed 3 general linear models using Jamovi 1.1 [27,28], 1 for each of the 3 LIWC summary variables, namely emotional tone, analytical thinking, and somatosensory processes. In each model, the LIWC summary variable was entered as the dependent variable; daily official data about new cases of COVID-19, new deaths, and new discharges were entered as continuous independent variables, while the phase of managing the COVID-19 outbreak was entered as a categorical independent variable (coded as 1=COVID-19 spreading; 2=Phase 1; 3=Phase 2). Specifically, these general linear models assessed whether daily new cases, new deaths, and new hospital discharges, alongside the phases of managing the COVID-19 pandemic, influenced the 3 daily summary variables constructed through LIWC. Besides the main effects, we included second- and third-order interaction terms for the continuous independent variable. We adopted a stepwise backward regression analysis approach. Thus, starting from the full model, nonsignificant, higher-order terms were eliminated one at a time, in order to obtain a final, more parsimonious model. If not one of the interaction terms was significant, the final model included only the main effects of all the predictors. For significant interactions, simple slope analysis was performed to test the effect of a specific predictor at different levels (ie, 1 standard deviation above and below the mean) of another predictor. All continuous independent variables were mean centered. The magnitude of each effect was interpreted by considering its associated partial eta squared (ie, η_p^2). Specifically, effects were considered weak ($.01 < \eta_p^2 \leq .06$), moderate ($.06 < \eta_p^2 \leq .14$), or strong ($\eta_p^2 > .14$). The final dataset and the scripts to perform data analysis are available in Multimedia Appendix 1 and Multimedia Appendix 2, respectively.

Results

Figure 1 displays the trends over time for emotional tone, analytical thinking, and somatosensory processes (as z scores) as expressed in daily tweets from February 24, 2020 to June 14, 2020.

Figure 1. Trends over time for emotional tone, analytical thinking, and somatosensory processes.

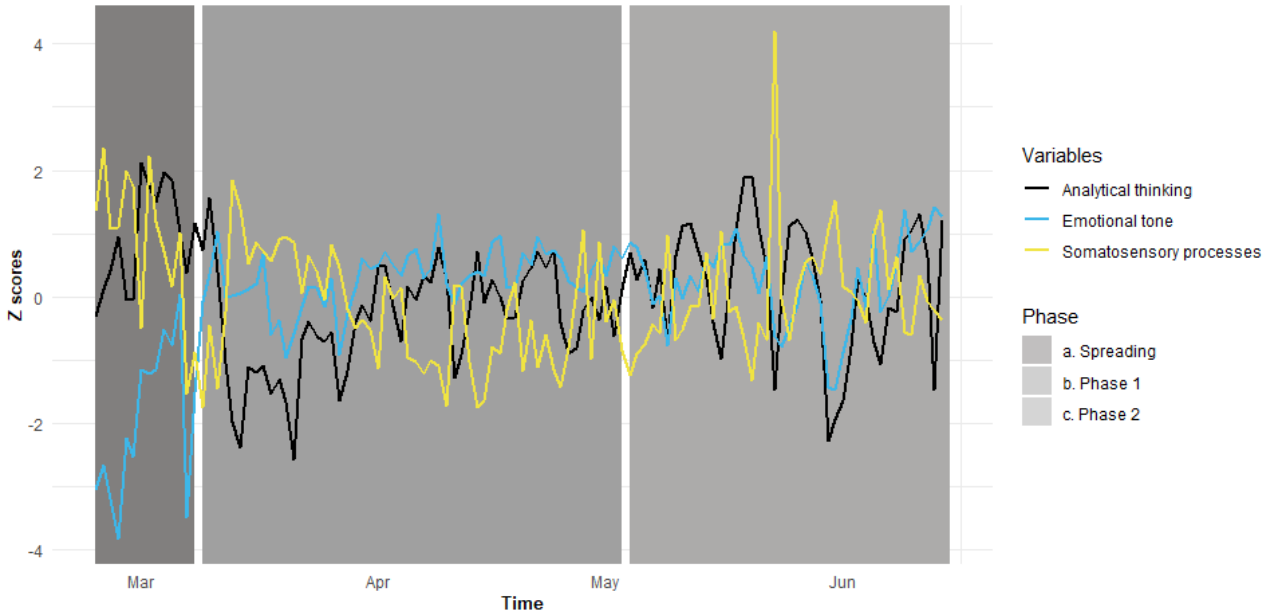
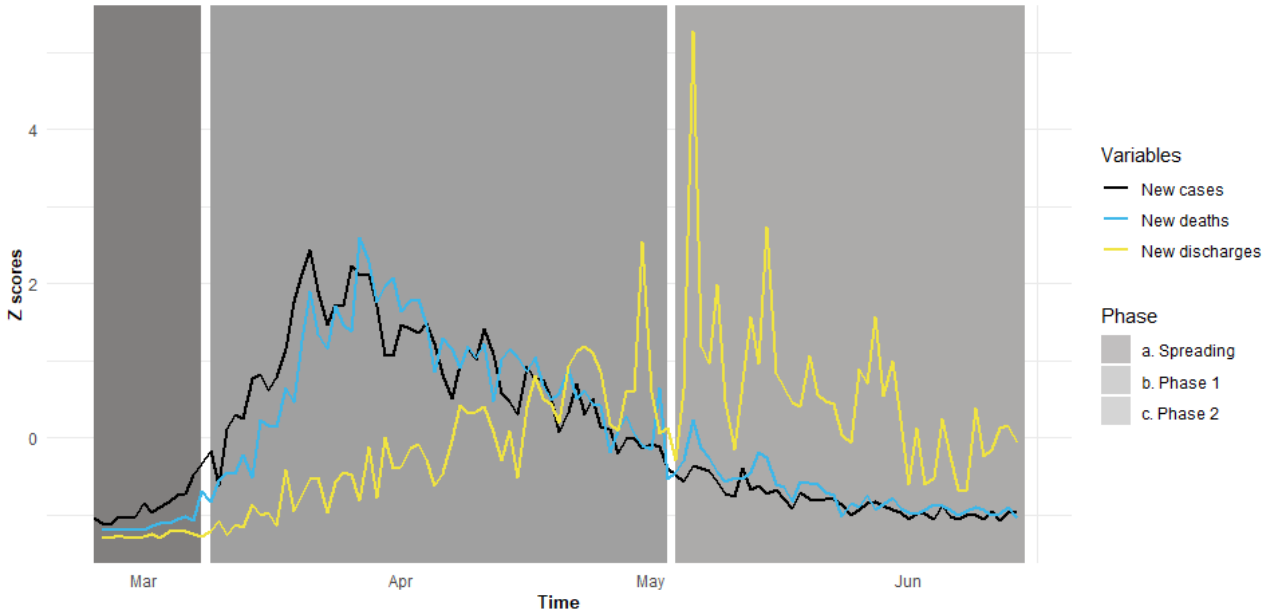


Figure 2 displays trends over time for daily new cases, new deaths, and new hospital discharges (as z scores) from February 24, 2020 to June 14, 2020.

Figure 2. Trends over time for daily new cases, new deaths, and new hospital discharges.



Results of the 3 general linear models assessing influences on each of the 3 LIWC summary variables, namely emotional tone, analytical thinking, and somatosensory processes, are reported in Table 1. By considering emotional tone, the final general linear model was significant ($F_{6,104}$=21.53, *P*<.001) and explained more than 55% of the dependent variable. Specifically, we found a significant interaction between daily new cases and new deaths for COVID-19 in explaining emotional tone ($F_{1,104}$=4.10, β=–.24, *P*=.045, η_p^2=.04). The simple slope analysis showed that, when the number of deaths was low (b=–0.00, SE=0.00, t_{104}=–0.13, *P*=.900) or average (b=–0.00, SE=0.00, t_{104}=–1.27, *P*=.207), daily new cases of COVID-19 were not related to tone. On the other hand, when the number of deaths was high, the higher the number of daily new cases, the lower the estimated emotional tone was (b=–0.00, SE=0.00, t_{104}=–2.42, *P*=.017). Other interactions were not significant and, thus, were excluded one at a time by adopting a stepwise backward regression analysis approach. The main effect of daily new cases was not significant ($F_{1,104}$=1.62, β=–.27, *P*=.207, η_p^2=.02), while the effect of daily new deaths was significant but weak ($F_{1,104}$=3.63, β=.41, *P*=.048, η_p^2=.03). Moreover, emotional tone was not related with daily number of new hospital discharges ($F_{1,104}$=0.12, β=–.03, *P*=.729, η_p^2=.00). Finally, phases of managing the COVID-19 outbreak were responsible for strong differences in emotional tone ($F_{2,104}$=30.27, *P*<.001, η_p^2=.37). Estimated marginal means of daily scores of emotional tone were –1.08 (SE=0.04) during the outbreak, –0.83 (SE=0.02) during Phase 1, and –0.83 (SE=0.03) during Phase 2. As highlighted by post hoc analyses with a Bonferroni correction (*P*<.05), daily scores of emotional tone during the first outbreak were lower than the ones reported in both Phase 1 and Phase 2. The 2 latter phases did not differ in daily scores of tone.

Table 1. Summary of the results of the 3 general linear models.

Variables	Emotional tone[a]			Analytical thinking[b]			Somatosensory processes[c]		
	β	*P* value	η_p^2	β	*P* value	η_p^2	β	*P* value	η_p^2
Phase	_[d]	<.001	0.37	_[d]	.001	0.12	_[d]	.55	0.01
New cases	–.27	.21	0.02	–.84	.001	0.10	.51	.07	0.03
New deaths	.41	.06	0.03	.55	.02	0.05	–1.01	<.001	0.12
New discharges	–.03	.73	0.00	.10	.33	0.01	.11	.38	0.01
New case*New deaths	–.24	.045	0.04	_[d]	_[d]	_[d]	.44	.004	0.08

[a]$F_{6,104}$=21.53, *P*<.001, R^2=.55.

[b]$F_{5,105}$=9.20, *P*<.001, R^2=.30.

[c]$F_{6,104}$=6.15, *P*<.001, R^2=.26.

[d]Not applicable.

The final general linear model performed to assess differences in daily scores of analytical thinking was significant ($F_{5,105}$=9.20, *P*<.001) and explained 30% of the dependent variable. No significant second- and third-order interactions were observed; thus, all interaction terms were excluded one at a time by adopting a stepwise backward regression analysis approach. Analytical thinking was not related to daily new discharges from hospital ($F_{1,105}$= 0.95, β=.10, *P*=.332, η_p^2=.01), while it was negatively and moderately related to daily new cases of COVID-19 ($F_{1,105}$=11.14, β=–.84, *P*=.001, η_p^2=.10) and positively but weakly linked to new deaths related to COVID-19 ($F_{1,105}$=5.48, β=.55, *P*=.021, η_p^2=.05). Daily scores of analytical thinking differed moderately among different phases of managing the COVID-19 outbreak ($F_{1,105}$=7.27, *P*=.001, η_p^2=.12). Estimated marginal means for daily scores of analytical thinking were –2.14 (SE=0.51) during the outbreak, –4.01 (SE=0.29) during Phase 1, and –4.22 (SE=0.32) during Phase 2. As highlighted by post hoc analyses with a Bonferroni correction (*P*<.05), daily scores of analytical thinking during the first outbreak were lower than the scores reported in both Phase 1 and Phase 2. The 2 latter phases did not differ in daily scores of analytical thinking.

By considering somatosensory processes, the final general linear model was significant ($F_{6,104}$=6.15, *P*<.001) and explained more than 26% of the dependent variable. Specifically, we found a significant interaction between daily new cases and new deaths related to COVID-19 in explaining somatosensory processes ($F_{1,104}$=8.79, β=.44, *P*=.004, η_p^2=.08). The simple slope analysis showed that, when the number of deaths was low (b=0.00, SE=0.00, t_{104}=0.19, *P*=.851) or average (b=0.00, SE=0.00, t_{104}=1.87, *P*=.065), daily new cases of COVID-19 were not related to somatosensory processes. On the other hand, when the number of deaths was high, the higher the number of daily new cases, the higher the estimated score of somatosensory processes was (b=0.00, SE=0.00, t_{104}=3.55, *P*<.001). Other interactions were not significant and, thus, were excluded one at a time by adopting a stepwise backward regression analysis approach. The main effect of daily new cases was not significant ($F_{1,104}$=3.48, β=.51, *P*=.065, η_p^2=.03), while the main effect of daily new deaths was significant and moderate ($F_{1,104}$=13.69, β=–1.01, *P*<.001, η_p^2=.12). Moreover, daily number of new hospital discharges was not related with somatosensory processes ($F_{1,104}$=0.77, β=.11, *P*=.383, η_p^2=.01). Finally, phases of managing the COVID-19 outbreak were not responsible for differences in somatosensory processes ($F_{2,104}$=0.60, *P*=.551, η_p^2=.01).

Discussion

General Considerations

All 3 general linear models to assess differences in daily scores of analytical thinking, emotional tone, and somatosensory processes were significant, with specific and different patterns. As already pointed out, we might discuss our results considering this pandemic as "the cause of individual and collective traumas" [11]. In fact, different people dealing with the same stressful event could develop various reactions: Some individuals could develop a nonpathological response, with emotional, cognitive, and physical symptoms resolving spontaneously after some days or weeks, the successful implementation of resilience and coping strategies, and a return to a previous baseline without long-lasting consequences. For these individuals, the stressful event remains only "potentially" traumatic. Other individuals, instead, develop more pathological reactions, ranging from adjustment disorders to PTSD, with trauma lived as "a complex emotional response to a stressful event, that overwhelms the individual's capacity to cope" [11].

Various studies have analyzed individuals' language use after a traumatic event (eg, Cohn et al [10]), but, to the best of our knowledge, this is the first study using these summary variables in a sample of Italian tweets during the first phases of the pandemic. First, in all our general models, we did not find any significant effect of the daily number of new hospital discharges on our variables of interest. Daily hospital discharges, compared to daily new cases and new deaths, was the only "positive" data considered. The absence of any effect could be due to the negativity bias, which is the human tendency to give more importance and attention to negative data—or entities in general—[29] such as COVID-19 deaths and new cases, while ignoring positive data, such as hospital discharges. The negativity bias has been demonstrated to be related to life stressors and PTSD [30,31], as individuals affected by PTSD tend to focus their attention on potential threats [32]. This could also explain the fact that, in each model, summary variables were related to the negative data. So, when experiencing a stressful event, such as the pandemic period, individuals may experience negativity bias, focusing more on negative data. Experiencing these data could be considered a stressful event. In particular, we found that increases in daily deaths and daily new cases, in other words the worst situation possible, increased negative emotional tone. This seems intuitive: Negative emotion words are habitually used when writing about a negative event, such as the situation described before, and have been linked with suicide and depression [9]. Moreover, negative alterations in mood experience, negative affect, and difficulty in experiencing positive emotions are typical reactions experienced after a stressful event and, in some cases, could be symptoms of PTSD [33].

The same interaction was found to have an effect on increased use of somatosensory words. This result also seems intuitive, as the use of sensory, body, and perceptual words in narratives related to traumatic events are common and often linked to PTSD and its symptoms, even in studies of more individual traumatic events such as traffic accidents [16,18].

Regarding analytical thinking, we found 2 opposite effects: Daily new cases were negatively linked with this linguistic marker, while new deaths were positively linked with it. High scores in the analytical thinking variable are related to a formal and logical thinking pattern, while a low score is related with a more narrative style, focused on the here and now [15]. Various studies have considered the use of cognitive words after

traumatic events: More cognitive words are often present in trauma- or distress-related narratives [34]. Using cognitive words is linked to an individual's effort to elaborate and integrate the event in their own memories [16], reflecting "an active search for meaning and understanding of the stressful event" [34]. In fact, using cognitive words, in particular causal and insight ones, when writing about a past event is linked to "the active process of reappraisal" [9]. So, using more cognitive words is associated with better physical health [35,36], fewer PTSD symptoms [37], and adaptive coping strategies [38]. On the contrary, some studies have shown a link between cognitive words and PTSD symptoms [39]. In fact, as some authors pointed out [39,40] using LIWC, it is difficult to understand how these words are used, for example referring to "organized or disorganized thoughts" or linked to "ruminative processes and fruitless attempts to assimilate what happened" [16]. After a traumatic event, indeed, individuals' thoughts could be affected in different ways. For instance, PTSD symptoms include intrusive and upsetting memories or negative thoughts about themselves and the world or avoidance of thoughts related to trauma [33]. These different reactions and discordance about the meaning of cognitive words after traumatic events may account for these opposite effects. Even if more data and research are needed, we may cautiously think that, when confronted with new deaths data—the worst news—individuals may try to react using a formal and logical way of writing, trying to make sense of this negative information. Considering new cases data, so slightly less negative, people may try to avoid the data or react with a more narrative tone, feeling less the need to elaborate them.

The last interesting result we retrieved is the effect of the phase of managing the pandemic on emotional tone and analytical thinking variables. In particular, both emotional tone and analytical thinking were lower during the outbreak, then increased in the first and second phases. As explained, the initial phase of the pandemic in Italy was characterized by different restrictions and measures taken by the government in order to counteract the spread of the virus. These measures differed, in particular, between the outbreak and the first and second phases. As the first indigenous cases were confirmed at the end of February, but maybe the gravity of the situation was still not clear, different day-to-day actions and initiatives were taken in each part of Italy: Universities and schools were closed first only in northern regions and initially only for some days; 11 municipalities in Lombardy and Veneto were in strict lockdown; some major public events, such as the Carnival of Venice, were postponed or cancelled; in other regions, considered at minor risk, schools remained open with some events confirmed, such as Series A soccer matches with the presence of fans in southern Italy. However, contradictory messages hit the population: Fake news stating the closure of all Italian schools circulated at the end of February, while some ads and initiatives reassured people, even in the northern areas, to continue to live their normal lives; all of this contributed to creating a climate of uncertainty. The first and second phases, instead, were characterized by national-level and long-term measures, with a strict lockdown and suspension of nonnecessary activities in all Italian regions, which gradually loosened at the beginning of May. These 2 phases marked a tragic and dramatic situation but were more stable and predictable in their restrictions. These differences between the very first and the other phases could account for the differences retrieved in our summary variables. Uncertainty about future events, as people may experience during the outbreak phase about future restrictions and development of the emergency, is common in threat contexts and could elicit negative emotions, such as anxiety and fear [41]. After the situation became more stable in the subsequent phases, with less uncertainty, emotional tone may increase. This emotional tone pattern confirms other results retrieved in the COVID-19 pandemic and in other trauma contexts: Sadiković and colleagues [42] found decreased worry, fear, and boredom over 5 weeks after the first COVID-19–confirmed case in Serbia. Cohn and colleagues [10] found that, immediately after the September 11 attacks, emotional tone measured in a sample of online journals was low, returning slowly to baseline after 1 week. Experiencing negative emotions is a typical reaction after an emotional upheaval and uncertain and threatening situation, even representing a specific criterion for PTSD disorder [33]; experiencing positive emotions after a crisis acts as a buffer against depression in resilient individuals [14], and positive emotions in trauma narratives are linked to better adaptation or less severe PTSD symptoms [37,43]. So, after the initial, negative reaction, the situation changed, becoming more predictable and less uncertain, and people enact their resilience and coping strategies, using more positive emotions to overcome the emotional upheaval and resulting again in a more positive way of expressing themselves.

The uncertainty of the outbreak situation—with different restrictions and even contradictory circulating messages—may also have had an impact on people's analytical thinking and use of words: Reasoning and trying to make sense of events are difficult in such contexts [41]. People might have reacted with a more logical thinking style, trying to find meaning from the situation only during the first and second phases when things were more stable, the gravity and seriousness of the emergency became clearer, and a consistent view was reached. This result seems in contrast with the one obtained by Cohn and colleagues [10], who highlighted a rapid increase in cognitive word use immediately after the attacks; their level returned to baseline after some days and then decreased again. We have to point out that our study and the study by Cohn et al [10] used different writing samples (tweets vs journal entries) and also different words in the analysis: Even if theoretically tapping the same construct, such as a sort of thinking style, analytical thinking is based on function words while the cognitive processing index used by Cohn and colleagues [10] reflects words such as because, think, and question. However, we think that these differences in results could be due to the reasons already explained: September 11 was a punctual, intense, and disruptive outbreak, leading to a rapid need to make sense of what was happening. This pandemic outbreak phase, instead, was very different, with a slower unravelling and uncertainty that persisted for weeks and weeks.

As COVID-19 could be considered "the cause of individual and collective traumas" [11], we discussed our results considering previous studies both concerning individual (for example, traffic accidents or relationship breakups [18,34]) and collective trauma

(for example, the September 11 attacks [10]). With heterogeneous yet similar consequences for individuals, more research is needed to highlight pandemic-specific psycholinguistic trauma at both individual and collective levels.

Limitations

Our study is not exempt from certain limitations. Our data consist of publicly available Italian tweets, so our results could not be generalized to other Twitter users with private accounts nor to the general Italian population. Even if it is used by a considerable amount of people—3.7 million users as of January 2020—Twitter is now only the sixth most used social media platform in Italy.

Moreover, we did not collect any information about users actually writing the analyzed tweets: Some demographic and other characteristics (eg, gender, age, working status, coping strategies) could account for differences in reactions to official COVID-19 data and for different use of words in their tweets. Specifically, some studies showed that even the area from which people tweet could account for some differences in their tweets: Gore et al [44], for example, showed that geotagged tweets in US areas with lower obesity rates have, among other results, a higher level of happiness. Another study [45] found that weather, days, and type of activities done during the day impact on emotions expressed in tourists' tweets.

So, specifically regarding our context, we might think that urban areas and their characteristics, days, and seasonal weather could have influenced the emotional tone and, globally, the words people use in their tweets.

Implications and Future Work

We think that our study could have relevant implications for actionable policies in the health care context and for future related works expanding our research questions.

These results prove the feasibility and importance of infodemiological indicators and psycholinguistic analysis to monitor mental health–related variables in a fast and cost-effective way. While traditional psychology instruments and measures (such as self-reported questionnaires and surveys) provide a one-time measure of the variable of interest in a limited sample, this method could provide longitudinal and population-level data. Considering all the limitations and influences, this method could be used as active surveillance of the impact of a pandemic and the related daily sharing of information on people's mental health, providing dynamic knowledge to inform relevant health policies. Knowing in advance or in real time which type of information—as new daily cases, new daily deaths, or the phase of the pandemic—could have an impact and how it impacts emotions, analytical thinking, and the mental health of a population could allow the implementation of ad hoc and concrete responses. As a pandemic is constantly and heavily affecting our daily lives and mental health [11,12,13], we think that monitoring psychological health and intervening to prevent costly consequences or improve well-being with tailored psychological interventions are essential.

Future studies are needed to approach this active surveillance approach as a useful and concrete instrument for institutions and health policy.

Moreover, as our study contributes to the growing field of infodemiology in the pandemic context, further research could expand our research questions, analyzing and controlling for other factors that could influence word use in tweets in this pandemic period, such as geotagging, days, and seasonal weather [44,45], as well as age, gender, working status, and other sociodemographic and spatial-temporal characteristics.

Conclusions

An increasing amount of literature has demonstrated the vast effects this pandemic is having on mental health, emotions, and cognition of the global and Italian populations. However, to the best of our knowledge, this is the first study analyzing psycholinguistic summary variables and their relationships with official COVID-19 Italian data and phases of managing the pandemic in a sample of Italian tweets during the first phases of the pandemic.

Our results show a powerful picture of the effects of COVID-19–related data and phases on emotions, analytical thinking, and somatosensory processes of Italian Twitter users: Specifically, when there was an increase in daily deaths and daily new cases, negative emotions and somatosensory words, often linked to traumatic events and PTSD symptoms, increased too. Moreover, emotional tone and analytic thinking were lower in the first phase of the pandemic, which was characterized by uncertainty, and increased during the first and second phases. As new instruments are implemented to monitor patients' psychological status [46], having information on how the pandemic may affect the use of words with its relationships with psychosocial variables could be useful for institutions and health policies to develop specific interventions in order to mitigate the effects of this or future situations on the population's mental health. Even if more studies are necessary, our results showed the feasibility and importance of infodemiological indicators and psycholinguistic analysis to monitor mental health–related variables in these unprecedented situations.

Acknowledgments

This work was partially supported by the Italian Ministry of Health with Ricerca Corrente and 5x1000 funds for IEO European Institute of Oncology IRCCS.

LV and GM are PhD students within the European School of Molecular Medicine (SEMM).

Authors' Contributions

DM and LV planned the study, constructed the dataset, performed statistical analysis, and drafted the manuscript. GM and SFMP drafted the manuscript. GP supervised all the processes, provided critical guidance, and revised the manuscript. All authors contributed to the article and approved the submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Final dataset.
[XLSX File (Microsoft Excel File), 15 KB - jmir_v23i10e29820_app1.xlsx]

Multimedia Appendix 2
Jamovi script.
[DOCX File , 13 KB - jmir_v23i10e29820_app2.docx]

References

1. Sehl K. Top Twitter Demographics That Matter to Social Media Marketers. Hootsuite. 2020 May 28. URL: https://blog.hootsuite.com/twitter-demographics/ [accessed 2021-09-14]
2. Twitter Announces First Quarter 2020 Results. Twitter. 2020 Apr 30. URL: https://s22.q4cdn.com/826641620/files/doc_financials/2020/q1/Q1-2020-Earnings-Press-Release.pdf [accessed 2021-09-14]
3. @TwitterIR. Q1 2020 Letter to Shareholders. URL: http://q4live.s22.clientfiles.s3-website-us-east-1.amazonaws.com/826641620/files/doc_financials/2020/q1/Q1-2020-Shareholder-Letter.pdf [accessed 2021-10-10]
4. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
5. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a Tool for Health Research: A Systematic Review. Am J Public Health 2017 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
6. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, et al. Global Sentiments Surrounding the COVID-19 Pandemic on Twitter: Analysis of Twitter Trends. JMIR Public Health Surveill 2020 May 22;6(2):e19447 [FREE Full text] [doi: 10.2196/19447] [Medline: 32412418]
7. Xue J, Chen J, Hu R, Chen C, Zheng C, Su Y, et al. Twitter Discussions and Emotions About the COVID-19 Pandemic: Machine Learning Approach. J Med Internet Res 2020 Nov 25;22(11):e20550 [FREE Full text] [doi: 10.2196/20550] [Medline: 33119535]
8. Su Y, Xue J, Liu X, Wu P, Chen J, Chen C, et al. Examining the Impact of COVID-19 Lockdown in Wuhan and Lombardy: A Psycholinguistic Analysis on Weibo and Twitter. Int J Environ Res Public Health 2020 Jun 24;17(12):1 [FREE Full text] [doi: 10.3390/ijerph17124552] [Medline: 32599811]
9. Tausczik YR, Pennebaker JW. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. Journal of Language and Social Psychology 2009 Dec 08;29(1):24-54. [doi: 10.1177/0261927X09351676]
10. Cohn MA, Mehl MR, Pennebaker JW. Linguistic markers of psychological change surrounding September 11, 2001. Psychol Sci 2004 Oct 01;15(10):687-693. [doi: 10.1111/j.0956-7976.2004.00741.x] [Medline: 15447640]
11. Masiero M, Mazzocco K, Harnois C, Cropley M, Pravettoni G. From Individual To Social Trauma: Sources Of Everyday Trauma In Italy, The US And UK During The Covid-19 Pandemic. J Trauma Dissociation 2020 Jul 12;21(5):513-519. [doi: 10.1080/15299732.2020.1787296] [Medline: 32654633]
12. Marton G, Vergani L, Mazzocco K, Garassino M, Pravettoni G. 2020s Heroes Are Not Fearless: The Impact of the COVID-19 Pandemic on Wellbeing and Emotions of Italian Health Care Workers During Italy Phase 1. Front Psychol 2020;11:588762 [FREE Full text] [doi: 10.3389/fpsyg.2020.588762] [Medline: 33178088]
13. Monzani D, Gorini A, Mazzoni D, Pravettoni G. Brief report - "Every little thing gonna be all right" (at least for me): Dispositional optimists display higher optimistic bias for infection during the Italian COVID-19 outbreak. Pers Individ Dif 2021 Jan 01;168:110388 [FREE Full text] [doi: 10.1016/j.paid.2020.110388] [Medline: 32921860]
14. Fredrickson BL, Tugade MM, Waugh CE, Larkin GR. What good are positive emotions in crisis? A prospective study of resilience and emotions following the terrorist attacks on the United States on September 11th, 2001. Journal of Personality and Social Psychology 2003;84(2):365-376. [doi: 10.1037/0022-3514.84.2.365]
15. Interpreting LIWC Output. Pennebaker Conglomerates. URL: https://liwc.wpengine.com/interpreting-liwc-output/ [accessed 2021-09-14]
16. Crespo M, Fernández-Lansac V. Memory and narrative of traumatic events: A literature review. Psychol Trauma 2016 Mar;8(2):149-156. [doi: 10.1037/tra0000041] [Medline: 25915647]

17. Beaudreau SA. Are trauma narratives unique and do they predict psychological adjustment? J Trauma Stress 2007 Jun;20(3):353-357. [doi: 10.1002/jts.20206] [Medline: 17598170]
18. Jones C, Harvey AG, Brewin CR. The organisation and content of trauma memories in survivors of road traffic accidents. Behav Res Ther 2007 Jan;45(1):151-162. [doi: 10.1016/j.brat.2006.02.004] [Medline: 16563341]
19. Buck N, Kindt M, van den Hout M, Steens L, Linders C. Perceptual Memory Representations and Memory Fragmentation as Predictors of Post-Trauma Symptoms. Behav. Cogn. Psychother 2006 Nov 15;35(3):259-272. [doi: 10.1017/s1352465806003468]
20. Banda J, Tekumalla R, Wang G, Yu J, Liu T, Ding Y, et al. A large-scale COVID-19 Twitter chatter dataset for open scientific research -- an international collaboration. Cornell University. URL: http://arxiv.org/abs/2004.03688 [accessed 2021-09-14]
21. Arafat M. A Review of Models for Hydrating Large-scale Twitter Data of COVID-19-related Tweets for Transportation Research. Advance: a SAGE preprints community. 2020 Apr 27. URL: https://advance.sagepub.com/articles/preprint/A_Review_of_Models_for_Hydrating_Large-scale_Twitter_Data_of_COVID-19-related_Tweets_for_Transportation_Research/12192693 [accessed 2021-09-14]
22. Hadi M, Xiao Y, Iqbal MS, Wang T, Arafat M, Hoque F, Florida International University Lehman Center for Transportation Research, Florida Department of Transportation Office of Research and Development. Estimation of System Performance and Technology Impacts to Support Future Year Planning. United States Department of Transportation. URL: https://rosap.ntl.bts.gov/view/dot/53966 [accessed 2021-09-14]
23. Ess C, Jones S. Ethical decision-making and internet research: Recommendations from the AoIR ethics working committee. In: Buchanan E, editor. Readings in Virtual Research Ethics: Issues and Controversies. Hershey, PA: IGI Global; 2004:27-44.
24. Agosti A, Rellini A. The Italian LIWC Dictionary. LIWC. 2007. URL: http://liwc.wpengine.com/ [accessed 2021-09-14]
25. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The Development and Psychometric Properties of LIWC2015. University of Texas at Austin. 2015. URL: https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf [accessed 2021-09-14]
26. Pennebaker JW, Chung CK, Frazee J, Lavergne GM, Beaver DI. When small words foretell academic success: the case of college admissions essays. PLoS One 2014;9(12):e115844 [FREE Full text] [doi: 10.1371/journal.pone.0115844] [Medline: 25551217]
27. GAMLj: General Analyses for the Linear Model in Jamovi. GAMLj. URL: https://gamlj.github.io/ [accessed 2021-09-14]
28. Jamovi. URL: https://www.jamovi.org/ [accessed 2021-09-14]
29. Rozin P, Royzman EB. Negativity Bias, Negativity Dominance, and Contagion. Pers Soc Psychol Rev 2016 Dec 21;5(4):296-320. [doi: 10.1207/S15327957PSPR0504_2]
30. Williams LM, Gatt JM, Schofield PR, Olivieri G, Peduto A, Gordon E. 'Negativity bias' in risk for depression and anxiety: brain-body fear circuitry correlates, 5-HTT-LPR and early life stress. Neuroimage 2009 Sep;47(3):804-814. [doi: 10.1016/j.neuroimage.2009.05.009] [Medline: 19446647]
31. Kimble M, Batterink L, Marks E, Ross C, Fleming K. Negative expectancies in posttraumatic stress disorder: neurophysiological (N400) and behavioral evidence. J Psychiatr Res 2012 Jul;46(7):849-855 [FREE Full text] [doi: 10.1016/j.jpsychires.2012.03.023] [Medline: 22595869]
32. Hayes JP, Vanelzakker MB, Shin LM. Emotion and cognition interactions in PTSD: a review of neurocognitive and neuroimaging studies. Front Integr Neurosci 2012;6:89 [FREE Full text] [doi: 10.3389/fnint.2012.00089] [Medline: 23087624]
33. Diagnostic and Statistical Manual of Mental Disorders, 5th ed. Washington, DC: American Psychiatric Association; 2013.
34. Boals A, Klein K. Word Use in Emotional Narratives about Failed Romantic Relationships and Subsequent Mental Health. Journal of Language and Social Psychology 2016 Jul 26;24(3):252-268. [doi: 10.1177/0261927x05278386]
35. Pennebaker JW, Francis ME. Cognitive, Emotional, and Language Processes in Disclosure. Cognition and Emotion 2010 Sep 10;10(6):601-626. [doi: 10.1080/026999396380079]
36. Pennebaker JW, Mayne TJ, Francis ME. Linguistic predictors of adaptive bereavement. Journal of Personality and Social Psychology 1997;72(4):863-871. [doi: 10.1037/0022-3514.72.4.863]
37. Follmer Greenhoot A, Sun S, Bunnell SL, Lindboe K. Making sense of traumatic memories: memory qualities and psychological symptoms in emerging adults with and without abuse histories. Memory 2013 Jan;21(1):125-142. [doi: 10.1080/09658211.2012.712975] [Medline: 22943468]
38. Goldberg DP, Gater R, Sartorius N, Ustun TB, Piccinelli M, Gureje O, et al. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. Psychol Med 1997 Jan 01;27(1):191-197. [doi: 10.1017/s0033291796004242] [Medline: 9122299]
39. D'Andrea W, Chiu PH, Casas BR, Deldin P. Linguistic Predictors of Post-Traumatic Stress Disorder Symptoms Following 11 September 2001. Appl. Cognit. Psychol 2011 Oct 24;26(2):316-323. [doi: 10.1002/acp.1830]
40. Jelinek L, Randjbar S, Seifert D, Kellner M, Moritz S. The organization of autobiographical and nonautobiographical memory in posttraumatic stress disorder (PTSD). J Abnorm Psychol 2009 May;118(2):288-298. [doi: 10.1037/a0015633] [Medline: 19413404]

41. Aitken C, Mavridis D. Reasoning under uncertainty. Evid Based Ment Health 2019 Feb 24;22(1):44-48. [doi: 10.1136/ebmental-2018-300074] [Medline: 30679196]
42. Sadiković S, Branovački B, Oljača M, Mitrović D, Pajić D, Smederevac S. Daily Monitoring of Emotional Responses to the Coronavirus Pandemic in Serbia: A Citizen Science Approach. Front Psychol 2020 Aug 19;11:2133 [FREE Full text] [doi: 10.3389/fpsyg.2020.02133] [Medline: 32973636]
43. Sheard J, Johnsen BH, Saus ER. Trauma narratives and emotional processing. Scand J Psychol 2005 Dec;46(6):503-510. [doi: 10.1111/j.1467-9450.2005.00482.x] [Medline: 16277651]
44. Gore RJ, Diallo S, Padilla J. You Are What You Tweet: Connecting the Geographic Variation in America's Obesity Rate to Twitter Content. PLoS One 2015;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]
45. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018 Jun 14;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]
46. Kondylakis H, Bucur A, Dong F, Renzi C, Manfrinati A, Graf N, et al. iManageCancer: Developing a platform for empowering patients and strengthening self-management in cancer diseases. 2017 Presented at: 30th IEEE Int Symp on Computer-Based Medical Systems (CBMS); June 22-24, 2017; Thessaloniki, Greece p. 2017. [doi: 10.1109/cbms.2017.62]

Abbreviations

LIWC: Linguistic Inquiry and Word Count
PTSD: posttraumatic stress disorder

Edited by C Basch; submitted 21.04.21; peer-reviewed by R Gore, J Chen, B Green; comments to author 11.05.21; revised version received 30.06.21; accepted 16.07.21; published 27.10.21.

Please cite as:
Monzani D, Vergani L, Pizzoli SFM, Marton G, Pravettoni G
Emotional Tone, Analytical Thinking, and Somatosensory Processes of a Sample of Italian Tweets During the First Phases of the COVID-19 Pandemic: Observational Study
J Med Internet Res 2021;23(10):e29820
URL: https://www.jmir.org/2021/10/e29820
doi:10.2196/29820
PMID:34516386

Original Paper

ClinicalTrials.gov as a Source of Information About Expanded Access Programs: Cohort Study

Jan Borysowski[1,2], MD, PhD; Andrzej Górski[3], MD, PhD

[1]Department of Clinical Immunology, Medical University of Warsaw, Warsaw, Poland

[2]Centre for Studies on Research Integrity, Institute of Law Studies, Polish Academy of Sciences, Warsaw, Poland

[3]Laboratory of Bacteriophages, Ludwik Hirszfeld Institute of Immunology and Experimental Therapy, Polish Academy of Sciences, Wrocław, Poland

Corresponding Author:
Jan Borysowski, MD, PhD
Department of Clinical Immunology
Medical University of Warsaw
Nowogrodzka Str 59
Warsaw, 02-006
Poland
Phone: 48 22 502 10 58
Email: jborysowski@interia.pl

Abstract

Background: ClinicalTrials.gov (CT.gov) is the most comprehensive internet-based register of different types of clinical studies. Expanded access is the use of unapproved drugs, biologics, or medical devices outside of clinical trials. One of the key problems in expanded access is the availability to both health care providers and patients of information about unapproved treatments.

Objective: We aimed to evaluate CT.gov as a potential source of information about expanded access programs.

Methods: We assessed the completeness of information in the records of 228 expanded access programs registered with CT.gov from February 2017 through May 2020. Moreover, we examined what percentage of published expanded access studies has been registered with CT.gov. Logistic regression (univariate and multivariate) and mediation analyses were used to identify the predictors of the absence of some information and a study's nonregistration.

Results: We found that some important data were missing from the records of many programs. Information that was missing most often included a detailed study description, facility information, central contact person, and eligibility criteria (55.3%, 54.0%, 41.7%, and 17.5% of the programs, respectively). Multivariate analysis showed that information about central contact person was more likely to be missing from records of studies registered in 2017 (adjusted OR 21.93; 95% CI 4.42-172.29; $P<.001$). This finding was confirmed by mediation analysis ($P=.02$). Furthermore, 14% of the programs were registered retrospectively. We also showed that only 33 of 77 (42.9%) expanded access studies performed in the United States and published from 2014 through 2019 were registered with CT.gov. However, multivariate logistic regression analysis showed no significant association between any of the variables related to the studies and the odds of study nonregistration ($P>.01$).

Conclusions: Currently, CT.gov is a quite fragmentary source of data on expanded access programs. This problem is important because CT.gov is the only publicly available primary source of information about specific programs. We suggest the actions that should be taken by different stakeholders to fully exploit this register as a source of information about expanded access.

(J Med Internet Res 2021;23(10):e26890) doi:10.2196/26890

KEYWORDS

ClinicalTrials.gov; expanded access; expanded access program; compassionate use; unapproved drug; investigational drug

Introduction

Expanded access, also termed compassionate use, is the use of unapproved drugs, biologics, or medical devices outside of clinical trials [1,2]. Regulations to permit expanded access to unapproved treatments have been introduced in many countries worldwide including the United States, many Member States of the European Union (EU), Canada, Australia, Japan, and Brazil [3-5]. In short, expanded access is a regulatory pathway that enables doctors to use unapproved treatments for patients with serious or life-threatening diseases who have run out of

approved treatments and are not eligible for enrollment in a clinical trial [6].

ClinicalTrials.gov (CT.gov) is a comprehensive web-based repository of information about clinical studies performed both in the United States and other countries [7]. Originally, it was established to provide potential study participants with information about interventional clinical trials [8]. However, with time, other types of studies started to be registered with CT.gov, especially observational studies and expanded access programs. As of May 29, 2021, 758 expanded access programs have been registered with CT.gov [7]. According to the terminology adopted by CT.gov, those programs are termed expanded access studies; the latter term will be used throughout the article.

Importantly, not all types of studies have to be registered with CT.gov. Current regulations oblige the responsible parties to register, in particular, so-called applicable clinical trials, that is, trials meeting certain criteria. In addition, if the drug being investigated in an applicable clinical trial is available through expanded access and the responsible party in the trial is also the drug's manufacturer, then the corresponding expanded access study has to be registered by statute as well [9]. However, there are no regulations to mandate the registration of observational studies.

Since expanded access involves the use of unapproved treatments, one of the key problems that can be encountered by both health care providers and patients is the availability of information about those treatments [10,11]. Access to data about unapproved treatments will certainly be more limited compared with standard approved therapies that are commonly known and used by many doctors. The objective of this study was to evaluate CT.gov as a source of information about expanded access studies. We focused on 3 main problems: (1) completeness of information required when registering an expanded access study with CT.gov, (2) mode of registration (prospective vs retrospective), (3) the percentage of expanded access studies that has been registered with CT.gov. These 3 problems are key in the evaluation of CT.gov because, to be helpful for patients and health care providers, CT.gov has to present complete data that are posted prospectively.

Methods

Selection of Expanded Access Studies Registered With ClinicalTrials.gov

Eligible studies were searched for in CT.gov [7] on June 8, 2020. Using the "Advanced Search" function, we selected expanded access studies involving the use of a drug, biologic, or medical device, registered with CT.gov from February 2017 through May 2020. We did not include studies registered by February 2017 because, in January 2017, some changes were introduced to the range of data that are required when registering expanded access studies with CT.gov. In particular, expanded access type and facility information started to be required in accordance with the Food and Drug Administration (FDA) Amendments Act (FDAAA) 801 final rule (42 CFR Part 11) [12].

Extraction of Data From Records of ClinicalTrials.gov-Registered Expanded Access Studies

The record of each expanded access study registered with CT.gov includes a range of data classified either as required, conditionally required, or optional [12]. The range of data that we evaluated included unique protocol identification number; brief title; expanded access type; record verification date; expanded access status; responsible party, by official title; name of the sponsor; brief summary; detailed description; condition or focus of the study; intervention type; intervention name; eligibility including patient sex, age limits, and eligibility criteria; central contact person; facility information; citations to publications related to the expanded access; and links to web sites directly relevant to the expanded access.

Moreover, from the "History of Changes" field, we extracted the first recruitment status ("Available," "No longer available," "Temporarily not available," or "Approved for marketing").

Publications on Expanded Access

We searched for publications on expanded access in Medline through Pubmed using the following search string: "Compassionate Use Trials" [MeSH] OR "expanded access" OR "compassionate use" OR "early access" OR "managed access" OR "named patient" OR "humanitarian device exemption." The following inclusion criteria were used: (1) publication year 2014-2019; (2) a study involving the use of a drug, biologic, or medical device, with an explicit statement in the published article that the treatment was performed in an expanded accesss program or on a compassionate use basis; (3) at least 1 center located in the United States. We included expanded access studies regardless of the number of participants and study design (case studies, case series, and cohort studies; prospective and retrospective). The search for eligible publications was performed in May 2020.

For each of the included publications, we searched for a corresponding entry in CT.gov. This search was performed in 2 stages. First, we searched for an identifier typical of CT.gov using an automated search function (Ctrl-F). Since each CT.gov identifier starts with the prefix "NCT," this prefix was used as the search term. In the second stage, for each publication that did not contain a CT.gov identifier, we searched CT.gov using keywords from the publication, especially intervention name and disease. This search was limited to expanded access studies performed in the United States (field "Country"). When assessing whether an article matched a study registered with CT.gov, we considered the type of study (expanded access), the intervention that was used, and center location.

Statistical Analysis

Summary statistics were used to show absolute numbers and frequencies of main study characteristics. Different statistical tests were employed to evaluate whether the variables related to the expanded access studies are interrelated. These included a chi-square test (a discrete variable vs a discrete variable), Mann-Whitney test (a continuous variable vs a binary variable), analysis of variance (a continuous variable vs a discrete variable with more than 2 levels). For these tests, $P<.05$ was considered significant. Multicollinearity of the variables was assessed by

determining the generalized variance inflation factor (GVIF). We considered that a GVIF value >5 may be indicative of multicollinearity.

Logistic regression (univariate and multivariate) analyses were performed to check whether posting of some data and registration of published expanded access studies depended on different variables. All variables with $P<.05$ in univariate analysis were entered into the multivariate model. For each regression analysis, the level of statistical significance was set using Bonferroni correction. The results of the logistic regression analyses were verified by mediation analysis. All computations were performed in R v. 3.6.1.

Results

Selection and Characteristics of Expanded Access Studies

Eligible expanded access studies were identified in CT.gov [7]. The flow diagram showing the selection process is presented in Figure 1. We selected 228 studies registered with CT.gov from February 2017 through May 2020. The detailed characteristics of those studies are presented in Table 1. Overall, 195 different interventions were used in those studies. Most studies concerned oncology (96; 42.1%), neurology (29; 12.7%), or infectious diseases (28; 12.3%).

Figure 1. Flow diagram showing the selection of ClinicalTrials.gov-registered expanded access studies.

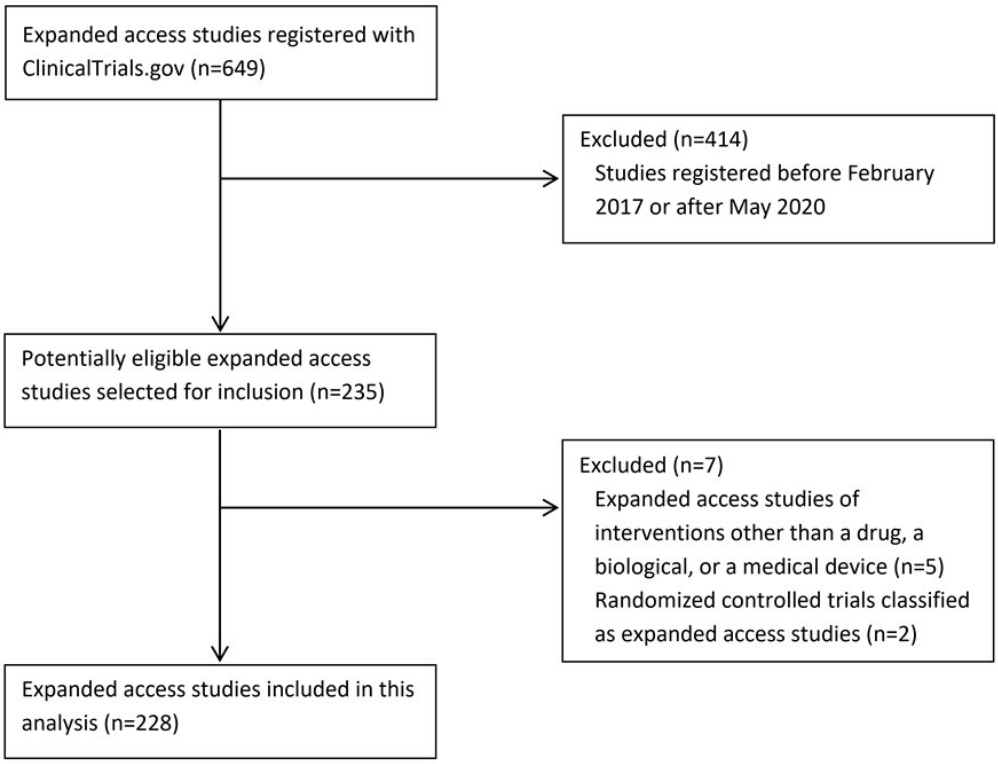


Table 1. Characteristics of the 228 ClinicalTrials.gov-registered, expanded access studies.

Characteristics	Number of studies, n (%)
Medical specialty	
Oncology	96 (42.1)
Neurology	29 (12.7)
Infectious diseases	28 (12.3)
Hematology	16 (7.0)
Metabolic diseases	15 (6.6)
Gastroenterology	8 (3.5)
Cardiovascular diseases	7 (3.1)
Pulmonology	5 (2.2)
Other	17 (7.5)
Unknown	7 (3.1)
Intervention type	
Drug	162 (71.1)
Biologic	61 (26.8)
Medical device	5 (2.2)
Sponsor	
Industry	182 (79.8)
Nonindustry	32 (14.0)
Mixed	14 (6.1)
Center location	
United States	75 (32.9)
International	16 (7.0)
Other	14 (6.1)
Unknown	123 (54.0)
Expanded access type	
IP[a]	103 (45.2)
IS[b]	44 (19.3)
Tr. IND/Pr.[c]	46 (20.2)
Mixed[d]	24 (10.5)
Unknown	11 (4.8)
Multicenter studies	
Yes	50 (21.9)
No	55 (24.1)
Unknown	123 (54.0)
Multinational studies	
Yes	16 (7.0)
No	91 (39.9)
Unknown	121 (53.1)

[a]IP: individual patients.

[b]IS: intermediate-size population.

[c]Tr. IND/Pr.: treatment investigational new drug (IND)/protocol.

[d]Any combination of individual patients, intermediate-size population, and treatment IND/protocol.

Assessment of Data Contained in Records of Expanded Access Studies

We evaluated the completeness of information that is required or optional when registering an expanded access study with CT.gov. Remarkably, detailed description, facility information, central contact person, and eligibility criteria were missing from the records of 126 (126/228, 55.3%), 123 (123/228, 54.0%), 95 (95/228, 41.7%), and 40 (40/228, 17.5%) studies, respectively (Table 2). Except for detailed description, all those data are required by CT.gov.

Table 2. ClinicalTrials.gov-registered expanded access studies (n=228) with different missing data.

Data missing from the record	Number of studies, n (%)
Detailed description[a]	126 (55.3)
Facility information[b]	123 (54.0)
Central contact person[b]	95 (41.7)
Eligibility criteria[c]	40 (17.5)
Sex of patients[c]	31 (13.6)
Expanded access type[b]	11 (4.8)
Conditions or focus of study[c]	7 (3.1)
Expanded access status[b]	0 (0)
Responsible party[b]	0 (0)
Name of the sponsor[b]	0 (0)
Brief summary[b]	0 (0)
Intervention type[b]	0 (0)
Intervention name[b]	0 (0)
Patient age limits[c]	0 (0)
Unique protocol identification number[b]	0 (0)
Brief title[b]	0 (0)
Record verification date[b]	0 (0)
Responsible party[b]	0 (0)

[a]Optional data.

[b]Required data.

[c]Data conditionally required.

In this analysis, as a comparator for expanded access studies, we used a random sample of 220 clinical trials registered with CT.gov in the same period of time. We found that detailed description, facility information, and central contact person were missing from the records of 88 (88/220, 40.0%), 6 (6/220, 2.7%), and 82 (82/220, 37.3%) trials, respectively. However, eligibility criteria, as well as all other types of required data were contained in each clinical trial record that we examined.

Using logistic regression analysis, we also determined the predictors of the absence of some data in the records of expanded access studies. This analysis was performed for detailed description, central contact person, and facility location (ie, the data that were missing from the records most often). The following variables were entered into the analysis: registration year, funding source, US involvement, whether the study was multicenter and multinational, and type of expanded access. Registration year is the year in which a study was registered with CT.gov (we examined studies registered between 2017 and 2020). Funding sources were divided into industry and nonindustry based on information contained in the "Sponsor" field of an expanded access study record (the category "Industry" included all for-profit organizations, especially pharmaceutical companies, while nonindustry sources included all non-for-profit organizations such as universities, academic reasearch centers, and federal agencies). The variable "US involvement" refers to the presence of at least 1 center located in the United States. A multicenter study was defined as a study performed at 2 or more centers. Studies performed in at least 2 countries were considered multinational. The type of expanded access was divided into 3 categories: (1) individual patients, (2) intermediate-size population, (3) treatment investigational new drug (IND)/protocol (ie, use of an unapproved treatment in a large population of patients). This division is in line with the FDA regulations [1] and was also adopted by CT.gov [12].

In this analysis, we adjusted for multiple comparisons using Bonferroni correction. The adjusted level of statistical significance was set at *P*=.002. In addition, prior to performing logistic regression analysis, we evaluated the multicollinearity of the variables. However, the GVIF value for each variable was between 1.0 and 1.7, indicative of a lack of multicollinearity (detailed data not shown).

Detailed results of the logistic regression analysis are shown in Table 3, Table 4, and Table 5. Univariate analysis showed that facility location was less likely to be missing from records of treatment IND/protocol studies related to individual patient expanded access studies (odds ratio [OR] 0.76; 95% CI 0.64-0.9; *P*=.001). Information about the central contact person was more likely to be missing in studies registered in 2017 (OR 1.53; 95% CI 1.26-1.86; *P*<.001) compared with those registered in 2020. Multivariate analysis showed that information about central contact person was more likely to be missing from records of studies registered in 2017 (adjusted OR [aOR] 21.93; 95% CI 4.42-172.29; *P*<.001).

Table 3. Predictors of the absence of study detailed description in the records of expanded access studies registered with ClinicalTrials.gov.

Variables	Univariate analysis		Multivariate analysis	
	Odds ratio (95% CI)[a]	*P* value	Adjusted odds ratio (95% CI)[a]	*P* value
First posted date				
2020	referent	N/A[b]	referent	N/A
2019	1.02 (0.84-1.23)	.84	N/A	N/A
2018	1.12 (0.93-1.36)	.22	N/A	N/A
2017	1.0 (0.81-1.23)	.98	N/A	N/A
Sponsor				
Industry	referent	N/A	referent	N/A
Nonindustry	0.76 (0.63-0.91)	.003	0.46 (0.19-1.08)	.07
Multicenter study				
No	referent	N/A	referent	N/A
Yes	1.08 (0.63-0.91)	.41	N/A	N/A
Multinational study				
No	referent	N/A	referent	N/A
Yes	0.93 (0.89-1.31)	.57	N/A	N/A
US involvement				
No	referent	N/A	referent	N/A
Yes	0.95 (0.71-1.21)	.66	N/A	N/A
Type of expanded access				
IP[c]	referent	N/A	referent	N/A
IS[d]	0.92 (0.75-1.2)	.33	0.75 (0.36-1.55)	.42
Tr. IND/Pr.[e]	0.82 (0.77-1.09)	.02	0.47 (0.23-0.96)	.03

[a]For each variable (first posted date; sponsor; multicenter study; multinational study; US involvement; type of expanded access), odds ratios and 95% CI are shown related to the referent value.

[b]N/A: not applicable.

[c]IP: individual patients.

[d]IS: intermediate-size population.

[e]Tr. IND/Pr.: Treatment IND/protocol.

Table 4. Predictors of the absence of information about the central contact person in the records of expanded access studies registered with ClinicalTrials.gov.

Variables	Univariate analysis		Multivariate analysis	
	Odds ratio (95% CI)[a]	*P* value	Adjusted odds ratio (95% CI)[a]	*P* value
First posted date				
2020	referent	N/A[b]	referent	N/A
2019	1.17 (0.97-1.4)	.10	3.62 (0.8-26.08)	.13
2018	1.26 (1.05-1.51)	.01	11.9 (2.5-89.64)	.004
2017	1.53 (1.26-1.86)	<.001	21.93 (4.42-172.29)	<.001
Sponsor				
Industry	referent	N/A	referent	N/A
Nonindustry	0.89 (0.74-1.07)	.19	N/A	N/A
Multicenter study				
No	referent	N/A	referent	N/A
Yes	0.79 (0.65-0.95)	.01	0.33 (0.13-0.79)	.01
Multinational study				
No	referent	N/A	referent	N/A
Yes	0.83 (0.64-1.09)	.18	N/A	N/A
US involvement				
No	referent	N/A	referent	N/A
Yes	1.04 (0.83-1.31)	.73	N/A	N/A
Type of expanded access				
IP[c]	referent	N/A	referent	N/A
IS[d]	0.84 (0.71-1.0)	.05	N/A	N/A
Tr. IND/Pr.[e]	0.95 (0.8-1.13)	.54	N/A	N/A

[a]For each variable (first posted date; sponsor; multicenter study; multinational study; US involvement; type of expanded access), odds ratios and 95% CIs are shown related to the referent value.

[b]N/A: not applicable.

[c]IP: individual patients.

[d]IS: intermediate-size population.

[e]Tr. IND/Pr.: Treatment IND/protocol.

Table 5. Predictors of the absence of information about the facility location in the records of expanded access studies registered with ClinicalTrials.gov.

Variables	Univariate analysis		Multivariate analysis	
	Odds ratio (95% CI)[a]	*P* value	Adjusted odds ratio (95% CI)[a]	*P* value
First posted date				
2020	referent	N/A[b]	referent	N/A
2019	0.83 (0.68-1.0)	.05	1.0 (0.9-1.1)	.94
2018	1.0 (0.83-1.21)	.99	1.09 (0.98-1.22)	.10
2017	0.85 (0.7-1.04)	.12	1.0 (0.9-1.11)	.98
Sponsor				
Industry	referent	N/A	referent	N/A
Nonindustry	0.89 (0.74-1.07)	.21	0.98 (0.9-1.07)	.65
Multinational study				
No	referent	N/A	referent	N/A
Yes	0.98 (0.91-1.05)	.55	N/A	N/A
Type of expanded access				
IP[c]	referent	N/A	referent	N/A
IS[d]	0.84 (0.71-1.0)	.05	0.99 (0.91-1.08)	.85
Tr. IND/Pr.[e]	0.76 (0.64-0.9)	.001	1.04 (0.96-1.13)	.30

[a]For each variable (first posted date; sponsor; multicenter study; multinational study; US involvement; type of expanded access), odds ratios and 95% CIs are shown related to the referent value.

[b]N/A: not applicable.

[c]IP: individual patients.

[d]IS: intermediate-size population.

[e]Tr. IND/Pr.: Treatment IND/protocol.

In order to verify the results of the logistic regression analysis, we checked whether the variable "Registration date" was related with the other variables. Indeed, we found that this variable was related to the variable "Type of expanded access" (*P*=.002). By contrast, the relationships between "Registration date" and all the remaining variables were not statistically significant (*P*>.05). Therefore, we performed mediation analysis to verify whether the effect of registration date on the risk of the absence of information about the central contact person was significant per se or rather is a result of its relationship with the variable "Type of expanded access" (potential mediator). This analysis showed that the average causal mediation effect, that is the effect dependent on the mediator, was not statistically significant (*P*>.05; Table 6). By contrast, the average direct effect, that is the effect of the variable "Registration date" itself, was significant (*P*=.02; Table 6). Thus, the effect of the registration date on the risk of the absence of information about the central contact person is in itself statistically significant and is not a result of its relationship with the variable "Type of expanded access."

We also found that the records of very few studies contained citations to publications (17/228, 7.5%) and links (31/228, 13.6%). Most of the links provided access to sponsors' general web sites (12/31, 39%) or sponsors' general policies of expanded access (8/31, 26%). As few as 3 links provided access to further information about a given expanded access program. The remaining links provided access to other related information.

Moreover, we examined the mode of registration of expanded access studies. Most of the studies (193/228, 84.7%) had the status "Available" in the first entry. This means that expanded access for a given intervention is available to patients. However, there were also several studies that, in the first entry, had the status "Temporarily not available" (16/228, 7.0%), "No longer available" (15/228, 6.6%), or "Approved for marketing" (1/228, 0.4%). According to the definitions of those statuses [11], each of them means that expanded access for a given intervention was available in the past. Thus, 32 studies (32/228, 14.0%) were *de facto* registered retrospectively. Remarkably, 21 studies (21/228, 9.2%) were not available to patients at any point. Three studies (3/228, 1.3%) had the status "Not yet recruiting" in the first entry; this is an unexpected finding because this status is typical of interventional clinical trials and not of expanded access studies [13].

Table 6. Mediation analysis to evaluate whether the effect of registration date on the risk of the absence of information about the central contact person is mediated by the type of expanded access.

Variable	Estimate	95% CI[a]	*P* value[b]
ACME[c]	0.05	–0.009-0.200	.18
ADE[d]	0.48	0.08-0.76	.02
TE[e]	0.53	0.19-0.77	<.001
PM[f]	0.05	–0.01-0.59	.18

[a]Quasi-Bayesian 95% CI.

[b]*P* value as determined by mediation analysis.

[c]ACME: average causal mediation effect.

[d]ADE: average direct effect.

[e]TE: total effect.

[f]PM: proportion mediated.

Registration With ClinicalTrials.gov of the Published Expanded Access Studies

Another important problem that we addressed is the percentage of expanded access studies that has been registered with CT.gov. To that end, we examined papers reporting on results of expanded access treatment, with at least 1 center located in the United States, and published from 2014 through 2019. We excluded from this analysis studies that did not involve at least 1 US center because most expanded access studies registered with CT.gov were available to patients in the United States. Our inclusion criteria were met by 77 papers, and 70 different interventions were used in the published studies. In most of the studies, a drug was used (38/77, 49%), followed by a biologic (28/77, 36%) and medical device (11/77, 14%). The studies mostly concerned oncology (30/77, 39%), infectious diseases (10/77, 13%), cardiovascular diseases (10/77, 13%), and other medical specialties (27/77, 35%); 24 (24/77, 31%) studies were multinational, and 51 (51/77, 66%) were multicenter; 39 (39/77, 51%) studies were funded from industry sources, 9 (9/77, 12%) studies were funded from nonindustry sources, and the remaining 29 studies (29/77, 38%) did not receive any funding. The median number of patients was 23 (interquartile range, 3-149).

Only 21 papers (21/77, 27%) included an identifier typical of CT.gov. However, we assumed that some studies may have been registered without providing a relevant identifier in the corresponding publication. Therefore, using keywords from the publications, we examined whether CT.gov contains records matching the remaining 56 studies. We found that 12 of these (12/77, 16%) were actually registered. Thus, overall, 33 of the 77 published studies (43%) were registered with CT.gov.

To put these findings into broader context, we also evaluated the registration of clinical trials. In the first step, for each expanded access study, we tried to identify a corresponding clinical trial in PubMed of the same therapeutic intervention that involved at least 1 center located in the United States and was published in the same period of time. We identified 71 such trials (for the remaining 6 expanded access studies, there was no trial evaluating the same intervention). We found that 68 of 71 (96%) trials were registered with CT.gov. As few as 3 (3/71, 4%) trials were unregistered.

In addition, using logistic regression analysis, we identified the predictors of an expanded access study not being registered (Table 7). The following variables were entered into the analysis: funding source, number of patients, and whether the study was multicenter and multinational. Number of patients (sample) was a continuous variable. Funding sources (sponsors) were divided into 3 categories: (1) industry (when a study was at least in part funded by the pharmaceutical industry), (2) nonindustry (when a study was funded solely from nonindustry sources), and (3) none (when a study received no funding). A multicenter study was defined as a study performed at 2 or more centers. Studies performed in at least 2 countries were considered multinational. Logistic regression analysis was preceded by the evaluation of the multicollinearity of the variables. However, the GVIF value for each variable was between 1.1 and 1.4, indicative of a lack of multicollinearity (detailed data not shown).

In this analysis, we adjusted for multiple comparisons using Bonferroni correction. The adjusted level of statistical significance was set at P=.01. In the univariate analysis, studies with a lower number of participants and studies funded from nonindustry sources were more likely to be unregistered (P=.009 and P=.008, respectively; Table 7). We also found that single-center studies were less likely to be unregistered (P=.002; Table 7). However, none of the analyzed variables was a predictor of study nonregistration in the multivariate analysis (P>.01; Table 7).

While it was a lower number of patients that was a predictor of study nonregistration in the univariate analysis, a substantial percentage of studies involving a higher number of participants has not been registered either. In particular, 14 of 40 studies (35%) involving at least 20 participants have not been registered. Among the studies involving at least 100 participants, 5 of 22 (23%) have not been registered.

Table 7. Predictors of nonregistration of the published expanded access studies.

Variables	Univariate analysis		Multivariate analysis	
	Odds ratio (95% CI)[a]	*P* value	Adjused odds ratio (95% CI)[a]	*P* value
Sample	0.99 (0.99-0.99)	.009	0.99 (0.99-1.0)	.48
Sponsor				
Industry	referent	N/A[b]	N/A	
Nonindustry	1.6 (1.14-2.23)	.008	1.55 (1.08-2.24)	.02
None	1.63 (1.06-2.51)	.02	1.6 (1.0-2.55)	.05
Multicenter study				
Yes	referent	N/A	N/A	N/A
No	0.7 (0.56-0.87)	.002	1.02 (0.73-1.43)	.87
Multinational study				
Yes	referent	N/A	N/A	N/A
No	0.9 (0.7-0.1.14)	.40	N/A	N/A

[a]Except for sample (which is a continuous variable), odds ratios and 95% CIs are shown related to the referent value for each of the remaining variables (sponsor, multicenter study, multinational study).

[b]N/A: not applicable.

Discussion

Overall, our results show that the information about expanded access studies posted on CT.gov is quite fragmentary. Remarkably, less than one-half of expanded access studies performed in the United States have been registered with CT.gov. In some cases, this may result from noncompliance with the statutory requirement to register some expanded access studies with CT.gov. The FDAAA of 2007 obliges the responsible party of each applicable clinical trial being registered with CT.gov to specify whether the drug (or other intervention) evaluated in the trial is also available through expanded access [14]. If expanded access is available and the responsible party is the drug's manufacturer, then the sponsor has to register the corresponding expanded access with CT.gov [8]. However, as shown by our results, some sponsors failed to meet this requirement. Other studies, especially those not linked with applicable clinical trials may remain unregistered because there are simply no regulations to mandate their posting on CT.gov.

A specific form of expanded access studies that are very likely to remain unregistered is small studies performed outside of expanded access programs. In general, expanded access has 2 different forms [10]. The first is expanded access programs (some of which are linked with applicable clinical trials). Those programs are launched by manufacturers of investigational treatments and open to patients who meet specific eligibility criteria. However, even in the absence of a specific program providing access to an investigational drug, a doctor can submit to the manufacturer a request for that drug for a limited number of patients under his or her care. Thus, some published expanded access studies (especially those involving a small number of participants) can be a result of not expanded access programs but treatment of single patients outside of a formal program. It is rather unlikely that studies performed outside of a formal expanded access program have been registered with CT.gov. Indeed, the univariate logistic regression analysis showed that studies with a lower number of participants are significantly less likely to be registered. While multivariate analysis did not confirm this finding, a subgroup analysis for the published studies involving a higher number of patients revealed that the proportion of unregistered studies was also considerable (eg, almost one-quarter for studies involving at least 100 participants). Thus, study nonregistration is an important problem even in the case of large expanded access programs.

For each of the 228 CT.gov-registered expanded access studies, we also evaluated the completeness of the posted information. We found that the records of many studies were incomplete, and the information that was missing most often included the detailed description of a study, facility information, central contact person, and eligibility criteria. The absence of some of those data depended on specific variables, especially the registration year. Specifically, study registration in 2017 significantly increased the odds of the absence of information about the central contact person. This may be associated with the FDAAA 801 Final Rule. It expanded requirements for the submission of clinical trial registration and results information to CT.gov [8]. While this rule was issued in September 2016, it became effective in January 2017, and the responsible parties were expected to be in compliance by April 18, 2017 [8]. We believe that at that time, many sponsors devoted most of their resources to ensure compliance with these regulations. This may have resulted in a situation where fewer resources could be devoted to expanded access (from a point of view of drug development, expanded access is certainly not as important as clinical trials). Therefore, for some responsible parties, it might take some time to improve standards of posting of information about expanded access studies on CT.gov.

However, we were unable to identify a single factor associated with an increased risk of nonposting of all types of data. In particular, in the multivariate analysis, whether the data were

missing from records was not significantly associated with the funding source. Lack of some of those data, especially facility information, can result from the nature of expanded access. While some studies can be open to patients at specific medical centers, investigational drugs can also used by doctors residing in different locations [11]. In such cases, information about facility location would be irrelevant. However, we cannot see any justification for a lack of data such as central contact person or eligibility criteria (at least the main criteria should be listed). These are data that are particularly important for doctors and patients who seek access to unapproved tretments.

We also asked whether the absence of some of the required data is specific to expanded access studies or rather is a broader problem that pertains to clinical trials as well. We found that certain data were indeed missing from the records of some clinical trials. However, in the case of clinical trials, the scale of this problem was much smaller. Furthermore, we found that the percentage of unregistered clinical trials was very low compared with expanded access studies. Overall, CT.gov is a more complete source of information about clinical trials than expanded access studies, at least for studies performed in the United States.

An important question is whether, apart from CT.gov, there are other publicly available comprehensive sources of information about expanded access studies. Generally, information about investigational treatments can be obtained from a few different sources. The first of these are patient advocacy organizations' web sites. However, it was shown that, while most of these present data on clinical trials, very few post any information about expanded access studies [15]. Another potential source of information is web-based expanded access navigators. In the United States, the primary navigator of that kind was developed by the Reagan-Udall Foundation for the FDA [16]. However, it posts data about single-patient expanded access only. Furthermore, data on specific expanded access studies posted by this resource are actually pulled from CT.gov [16]. Thus, if the information posted on CT.gov is incomplete, this navigator will not present complete data about single-patient expanded access studies. Furthermore, unlike CT.gov, the navigator posts no information about programs dedicated to intermediate-sized groups of patients and treatment IND/protocols. Overall, the navigator is a much less complete source of information about expanded access than CT.gov.

Some information about expanded access can be also found on pharmaceutical companies' web sites [17]. In particular, in the United States, the 21st Century Cures Act of 2016 required manufacturers of investigational drugs to post key information about their general policies on evaluating and responding to expanded access requests [18]. However, this act does not explicitly require the manufacturers to post full listings of available expanded access studies. Rather, it obliges them to post a reference (eg, a hyperlink) to pertinent information on CT.gov.

Thus, currently CT.gov is the only primary source of information about specific expanded access studies, at least in the United States. In our opinion, CT.gov could be a very useful resource for patients and health care providers because it enables searching studies based on several criteria that are important in practice, including disease, intervention name, facility location, sponsor, or any combination thereof. However, some actions have to be undertaken to fully exploit the potential of CT.gov as a source of information about expanded access. First, CT.gov should implement some measures to promote submission of all data that are required when registering expanded access studies. Moreover, further research should be performed to evaluate what percentage of expanded access studies linked with applicable clinical trials have been registered with CT.gov. Such research could be performed on the data collected by the FDA. All expanded access studies in the United States have to be approved by the FDA [1]. Thus, by comparing the FDA's data with CT.gov records, one could determine the extent of noncompliance with the statutory requirement to register expanded access studies linked with applicable clinical trials on CT.gov. If a substantial number of such studies have not been registered, then some measures should be introduced to enforce higher compliance. It is noteworthy that FDA-affiliated authors have already published some reports on expanded access based on the data contained in expanded access requests submitted to the FDA [19-21].

Moreover, we postulate registration with CT.gov of all expanded access studies available to patients in the United States (regardless of whether these are linked with applicable clinical trials). Since the registration of the latter category is not mandatory, a decision about possible registration rests with the manufacturer of the unapproved drug, biologic, or medical device. However, to ensure fair access of patients and health care providers to information about unapproved treatments, all planned expanded access studies should be registered. We hope that this article will raise awareness of this problem among sponsors of expanded access studies.

Our study also enabled us to make some interesting observations about the included expanded access studies. First, we found that most studies involved drugs used in oncology, neurology, and infectious diseases. In the case of oncology, this is likely caused by generally poor prognosis of patients with different kinds of cancer who have run out of approved treatments [22]. The high number of expanded access studies in neurology may be associated with the fact that some relatively frequent chronic neurological diseases (eg, Alzheimer's disease) do not have effective treatments [23]. The high demand for unapproved treatments in infectious diseases can be a result of the development of resistance to approved drugs [24,25]. We also showed that only a small subset of studies in our cohort was funded by nonindustry sources. Therefore, in our view, in the future, noncommercial sponsors might consider more involvement in expanded access studies.

A limitation to our study is that we relied on data posted by CT.gov only and did not verify their validity. For instance, most of the included expanded access studies had the status "Available" in the first entry in the "History of Changes" field. However, theoretically, it is possible that in some of those studies, the actual start date was earlier and preceded the date of registration with CT.gov of a given study.

Another limitation is that we were unable to get access to some data about the factors that may impact the availability of information about expanded access. These include, in particular, the factors related to the sponsors of individual studies. For instance, the number of the staff members to prepare and enter relevant data into CT.gov records will likely affect the efficiency of the posting of information. Another important factor is whether the sponsor has any previous experience with posting on CT.gov of expanded access studies. Unfortunately, those data are generally unavailable, so we were unable to include them in our statistical analyses. We also could not analyze the exact number of patients that were treated in individual studies because CT.gov records do not contain relevant data. We hope that the publication of this article will encourage some sponsors to share their experiences with posting on CT.gov of expanded access studies.

In conclusion, CT.gov is the only primary, publicly available source of information about expanded access studies, especially those available in the United States. However, the data posted by this register are fragmentary, which is an important factor restricting access to information about investigational treatments. Introduction of the measures postulated in this paper will be beneficial to patients and health care providers who seek information about possibilities to use unapproved drugs, biologics, and medical devices.

Acknowledgments

This work was done at the Center of Bioethics, Harvard Medical School during the Kosciuszko Foundation Fellowship awarded to the first author (JB). The sponsor had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. We are very grateful to Magdalena Zielenkiewicz, PhD (Institute of Mathemathics, University of Warsaw) for expert statistical analysis of the results.

Conflicts of Interest

None declared.

References

1. Expanded Access. Food and Drug Administration. URL: https://www.fda.gov/news-events/public-health-focus/expanded-access [accessed 2021-09-29]
2. Van Norman GA. Expanded Patient Access to Investigational New Devices: Review of Emergency and Nonemergency Expanded Use, Custom, and 3D-Printed Devices. JACC Basic Transl Sci 2018 Aug;3(4):533-544 [FREE Full text] [doi: 10.1016/j.jacbts.2018.06.006] [Medline: 30175277]
3. Borysowski J, Górski A. Compassionate use of unauthorized drugs: Legal regulations and ethical challenges. Eur J Intern Med 2019 Jul;65:12-16. [doi: 10.1016/j.ejim.2019.04.008] [Medline: 31036436]
4. Nakada H, Takashima K. Where Can Patients Obtain Information on the Preapproval Access Pathway to Investigational Treatment in Japan? A Survey of Patient Advocacy Organizations' Websites. Clin Pharmacol Drug Dev 2019 Nov 08;8(8):978-983 [FREE Full text] [doi: 10.1002/cpdd.745] [Medline: 31592580]
5. da Silva RE, Lima E, Novaes M, Osorio-de-Castro C. The High "Cost" of Experimental Drugs Obtained Through Health Litigation in Brazil. Front Pharmacol 2020;11:752 [FREE Full text] [doi: 10.3389/fphar.2020.00752] [Medline: 32508660]
6. Fountzilas E, Said R, Tsimberidou AM. Expanded access to investigational drugs: balancing patient safety with potential therapeutic benefits. Expert Opin Investig Drugs 2018 Feb 22;27(2):155-162 [FREE Full text] [doi: 10.1080/13543784.2018.1430137] [Medline: 29353505]
7. ClinicalTrials.gov. URL: https://www.clinicaltrials.gov/ [accessed 2021-09-29]
8. Clinical Trials Registration and Results Information Submission. Federal Register: the Daily Journal of the United States Government. 2016 Sep 21. URL: https://www.federalregister.gov/documents/2016/09/21/2016-22129/clinical-trials-registration-and-results-information-submission [accessed 2021-09-29]
9. Tse T, Fain KM, Zarin DA. How to avoid common problems when using ClinicalTrials.gov in research: 10 issues to consider. BMJ 2018 May 25;361:k1452 [FREE Full text] [doi: 10.1136/bmj.k1452] [Medline: 29802130]
10. Caplan AL, Bateman-House A. Should patients in need be given access to experimental drugs? Expert Opin Pharmacother 2015 Jun 22;16(9):1275-1279. [doi: 10.1517/14656566.2015.1046837] [Medline: 26001178]
11. Caplan AL, Teagarden JR, Kearns L, Bateman-House AS, Mitchell E, Arawi T, et al. Fair, just and compassionate: A pilot for making allocation decisions for patients requesting experimental drugs outside of clinical trials. J Med Ethics 2018 Nov 07;44(11):761-767. [doi: 10.1136/medethics-2016-103917] [Medline: 29982174]
12. ClinicalTrials.gov Registration Data Element Definitions for Expanded Access. ClinicalTrials.gov PRS. 2020 Oct 1. URL: https://prsinfo.clinicaltrials.gov/expanded_access_definitions.html [accessed 2021-09-26]
13. ClinicalTrials.gov Protocol Registration Data Element Definitions for Interventional and Observational Studies. ClinicalTrials.gov PRS. 2020 Oct 1. URL: https://prsinfo.clinicaltrials.gov/definitions.html [accessed 2021-09-29]
14. Public Law 110–85. Food and Drug Administration Amendments Act of 2007. URL: https://www.govinfo.gov/content/pkg/PLAW-110publ85/pdf/PLAW-110publ85.pdf#page=82 [accessed 2021-09-29]

15. Folkers K, Leone S, Caplan A. Patient advocacy organizations' information for patients on pre-approval access to investigational treatments. BMC Res Notes 2019 Oct 28;12(1):706 [FREE Full text] [doi: 10.1186/s13104-019-4745-7] [Medline: 31661023]
16. Expanded Access Navigator. Reagan-Udall Foundation for the Food and Drug Administration. URL: https://navigator.reaganudall.org/expanded-access-navigator [accessed 2021-09-29]
17. Jung E, Zettler PJ, Kesselheim AS. Prevalence of Publicly Available Expanded Access Policies. Clin Pharmacol Ther 2018 Nov 23;104(5):1016-1021. [doi: 10.1002/cpt.996] [Medline: 29271014]
18. 21st Century Cures Act. United States Congress. 2016 Dec 13. URL: https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf [accessed 2021-09-29]
19. Jarow JP, Lemery S, Bugin K, Khozin S, Moscicki R. Expanded Access of Investigational Drugs: The Experience of the Center of Drug Evaluation and Research Over a 10-Year Period. Ther Innov Regul Sci 2016 Nov 30;50(6):705-709 [FREE Full text] [doi: 10.1177/2168479016656030] [Medline: 27917324]
20. Jarow JP, Lemery S, Bugin K, Lowy N. Ten-Year Experience for the Center for Drug Evaluation and Research, Part 2: FDA's Role in Ensuring Patient Safety. Ther Innov Regul Sci 2017 Mar 01;51(2):246-249 [FREE Full text] [doi: 10.1177/2168479016679214] [Medline: 28553566]
21. Jarow JP, Moscicki R. Impact of Expanded Access on FDA Regulatory Action and Product Labeling. Ther Innov Regul Sci 2017 Nov 30;51(6):787-789 [FREE Full text] [doi: 10.1177/2168479017707800] [Medline: 30079277]
22. Wu T, Gu X, Cui H. Emerging Roles of SKP2 in Cancer Drug Resistance. Cells 2021 May 10;10(5):1147 [FREE Full text] [doi: 10.3390/cells10051147] [Medline: 34068643]
23. Li J, Huang Q, Chen J, Qi H, Liu J, Chen Z, et al. Neuroprotective Potentials of Panax Ginseng Against Alzheimer's Disease: A Review of Preclinical and Clinical Evidences. Front Pharmacol 2021 Jun 2;12:688490 [FREE Full text] [doi: 10.3389/fphar.2021.688490] [Medline: 34149431]
24. Frieri M, Kumar K, Boutin A. Antibiotic resistance. J Infect Public Health 2017 Jul;10(4):369-378 [FREE Full text] [doi: 10.1016/j.jiph.2016.08.007] [Medline: 27616769]
25. Colpitts C, Baumert T. Addressing the Challenges of Hepatitis C Virus Resistance and Treatment Failure. Viruses 2016 Aug 16;8(8):226 [FREE Full text] [doi: 10.3390/v8080226] [Medline: 27537906]

Abbreviations

aOR: adjusted odds ratio
CT.gov: ClinicalTrials.gov
FDA: Food and Drug Administration
FDAAA: FDA Amendments Act
GVIF: generalized variance inflation factor
IND: investigational new drug
OR: odds ratio

Edited by R Kukafka; submitted 02.01.21; peer-reviewed by E Bunnik, H Belli; comments to author 24.01.21; revised version received 30.06.21; accepted 27.07.21; published 28.10.21.

Please cite as:
Borysowski J, Górski A
ClinicalTrials.gov as a Source of Information About Expanded Access Programs: Cohort Study
J Med Internet Res 2021;23(10):e26890
URL: https://www.jmir.org/2021/10/e26890
doi:10.2196/26890
PMID:34709189

Original Paper

Moderators of the Effect of a Self-directed Digitally Delivered Exercise Program for People With Knee Osteoarthritis: Exploratory Analysis of a Randomized Controlled Trial

Rachel K Nelligan[1], BPhysio, PGradCert(Physio); Rana S Hinman[1], BPhysio (Hons), PhD; Fiona McManus[2], BPhysio (Hons), MPH, MBiostat; Karen E Lamb[2,3], BA (Hons), PhD; Kim L Bennell[1], BAPPSci, PhD

[1]Centre for Health, Exercise and Sports Medicine, Department of Physiotherapy, School of Health Sciences, The University of Melbourne, Parkville, Australia

[2]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, Australia

[3]Methods and Implementation Support for Clinical Health research platform, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Parkville, Australia

Corresponding Author:
Rachel K Nelligan, BPhysio, PGradCert(Physio)
Centre for Health, Exercise and Sports Medicine, Department of Physiotherapy
School of Health Sciences
The University of Melbourne
161 Barry Street
Parkville, 3010
Australia
Phone: 61 403652115
Email: rachel.nelligan@unimelb.edu.au

Abstract

Background: A 24-week self-directed digitally delivered intervention was found to improve pain and function in people with knee osteoarthritis (OA). However, it is possible that this intervention may be better suited to certain subgroups of people with knee OA compared to others.

Objective: The aim of this study was to explore whether certain individual baseline characteristics moderate the effects of a self-directed digitally delivered intervention on changes in pain and function over 24 weeks in people with knee OA.

Methods: An exploratory analysis was conducted on data from a randomized controlled trial involving 206 people with a clinical diagnosis of knee OA. This trial compared a self-directed digitally delivered intervention comprising of web-based education, exercise, and physical activity program supported by automated exercise behavior change mobile phone text messages to web-based education alone (control). The primary outcomes were changes in overall knee pain (assessed on an 11-point numerical rating scale) and physical function (assessed using the Western Ontario and McMaster Universities Osteoarthritis Index function subscale [WOMAC]) at 24 weeks. Five baseline patient characteristics were selected as the potential moderators: (1) number of comorbidities, (2) number of other painful joints, (3) pain self-efficacy, (4) exercise self-efficacy, and (5) self-perceived importance of exercise. Separate linear regression models for each primary outcome and each potential moderator were fit, including treatment group, moderator, and interaction between treatment group and moderator, adjusting for the outcome at baseline.

Results: There was evidence that pain self-efficacy moderated the effect of the intervention on physical function compared to the control at 24 weeks (interaction P=.02). Posthoc assessment of the mean change in WOMAC function by treatment arm showed that each 1-unit increase in baseline pain self-efficacy was associated with a 1.52 (95% CI 0.27 to 2.78) unit improvement in the control group. In contrast, a reduction of 0.62 (95% CI –1.93 to 0.68) units was observed in the intervention group with each unit increase in pain self-efficacy. There was only weak evidence that pain self-efficacy moderated the effect of the intervention on pain and that number of comorbidities, number of other painful joints, exercise self-efficacy, or exercise importance moderated the effect of the intervention on pain or function.

Conclusions: With the exception of pain self-efficacy, which moderated changes in function but not pain, we found limited evidence that our selected baseline patient characteristics moderated intervention outcomes. This indicates that people with a range of baseline characteristics respond similarly to the unsupervised digitally delivered exercise intervention. As these findings are exploratory in nature, they require confirmation in future studies.

(J Med Internet Res 2021;23(10):e30768) doi:10.2196/30768

KEYWORDS

digital; text messaging; exercise; moderators; osteoarthritis; RCT; clinical trial; subgroups; pain; function; knee osteoarthritis; rehabilitation; digital health

Introduction

Osteoarthritis (OA) is a condition of the synovial joints [1], with the knee being commonly affected [2]. Knee OA causes joint pain and stiffness, which can often lead to reduced physical function and quality of life [3,4]. Knee OA is a highly prevalent condition and a leading contributor to disability globally [5]. Exercise is the core treatment in the management of knee OA, which is recommended in all clinical guidelines [6-9]. However, the effects of exercise on knee OA pain and function are modest overall [10]. This may be due to the existence of subgroups of people with certain baseline characteristics that cause them to respond in different ways to exercise [11]. Baseline patient characteristics that affect how a patient responds to treatment are called moderators [12]. A better understanding of potential moderators of the effects of exercise in knee OA will enable the identification of subgroups of people who respond more or less favorably to exercise treatments. This will facilitate the targeting of exercise treatments in knee OA and thus, may improve the effects of exercise on patient outcomes. The evaluation and identification of moderators of treatment effects has been named a major research priority in OA clinical guidelines [9,13].

The use of digitally delivered interventions to support chronic condition management is rapidly increasing as a means of improving access to evidence-based health care [14]. Recently, we developed and evaluated a 24-week self-directed digitally delivered intervention, designed to support people with knee OA to access and participate in an evidence-based exercise program [15]. In a randomized controlled trial (RCT), we found that this intervention led to greater improvements in pain and function compared to an education control at 24 weeks in people with knee OA. We also found that 72% and 68% of the participants in the intervention group (compared to 42% and 41% in the control group) experienced clinically meaningful improvements in pain and function, respectively. This demonstrates that most, but not all, people benefited from the intervention and could indicate the existence of subgroups of people who responded more (or less) favorably to the unsupervised, digitally delivered exercise intervention.

To our knowledge, no previous studies have conducted formal moderation analyses to explore baseline patient characteristics associated with the effect of self-directed digitally delivered exercise for people with OA although 2 studies have explored baseline characteristics as predictors of outcomes from self-directed exercise. One study identified that increased age and the presence of a comorbidity at baseline predicted nonusage of a self-directed web-based physical activity intervention for patients with knee and hip OA [16]. Another study found that the presence of comorbidity was associated with lower physical activity levels, while greater baseline arthritis self-efficacy was associated with greater physical activity, following a 12-week self-directed exercise program in adults with arthritis [17]. As these studies only examined associations between baseline characteristics and outcomes in the intervention arms with no control group comparisons, it is possible that these findings could have occurred, regardless of the interventions received. Therefore, these findings do not enable identification of potential subgroups of people who benefit (or not) from self-directed exercise interventions.

This exploratory study sought to identify moderators of the effect of a self-directed digitally delivered exercise intervention on changes in pain and physical function at 24 weeks relative to the control in people with knee OA. The findings of this study will address a key knee OA research priority and provide direction for future confirmatory studies.

Methods

Study Design

We conducted exploratory moderation analyses [18] by using data from a two-arm participant-blinded and assessor-blinded RCT [15]. The RCT evaluated the effects of self-directed digitally delivered exercise compared to an education control. Limited disclosure was used to blind the participants. All participants provided consent prior to enrolment into the RCT, which included the use of their deidentified data in secondary analysis. The RCT was approved by the University of Melbourne Human Research Ethics Committee (1851085) and prospectively registered (ACTRN12618001167257).

Participants

In the RCT, 206 people with knee OA were recruited from the Australia-wide community via internet sources (social media and web-based newspapers) and a volunteer database. Full RCT eligibility criteria are reported elsewhere [19] and included ≥45 years of age, a clinical diagnosis of knee OA, and internet access.

Intervention

Full details of the self-directed digitally delivered exercise intervention have previously been published [19,20]. To summarize, participants in the intervention received access to the same standardized custom-built website, “My Knee Exercise” and received a 24-week automated behavior change mobile phone text messaging program. The website was developed by the researchers (RKN, KLB, RSH) and feedback was provided by 3 people with knee OA. The website contained (1) educational information about OA, exercise, and sought to address common misconceptions about OA, (2) prescribed a 24-week lower limb strengthening exercise regimen, and (3) provided general physical activity guidance. The 24-week strengthening exercise regimen was divided into 3 programs, each of 8-weeks duration. The website advised that the 3 programs be completed consecutively. Each program contained

5-6 exercises. Participants were asked in the website to perform these exercises 3 times per week. The strengthening exercises focused on the hip, knee, and ankle (eg, sit-to-stand, seated knee extension, calf raise). Detailed exercise instructions, including when and how to increase an exercise challenge, were provided in both text and visual formats (photo, video) and were available to download. Exercise equipment (eg, ankle weights, resistance bands) was recommended, and suggestions about where to purchase equipment was provided. Exercise and physical activity logbooks were also provided and available to download.

Augmenting the strengthening exercise regimen was a 24-week automated exercise behavior change mobile phone text messaging program. The text messages were designed to monitor weekly exercise session completion and address exercise facilitators and barriers commonly encountered by people with knee OA. Program development systematically followed the Behavior Change Wheel Framework, which is a universally accepted approach to designing behavior change interventions [21]. The program functioned by prompting self-report of how many strengthening exercise sessions were completed in the previous week (each Monday initially, reducing to fortnightly by 24 weeks) and then provided tailored support depending on the level of exercise adherence (≥3 exercise sessions/week = adherent). Participants who self-reported ≥3 exercise sessions/week received a positive reinforcement message. Participants self-reporting <3 exercise sessions/week received a follow-up message asking them to select 1 reason (from a prespecified list of exercise barriers), which best explained the reason for <3 exercise sessions/week. Participants then received a message containing a behavior change suggestion linked to their selected barrier. Participants also received regular messages designed to facilitate ≥3 exercise sessions/week (twice weekly initially, reducing to fortnight by week 24). Dependent on weekly responses, participants received on average 2-5 messages per week. The frequency of messages sent to participants was designed to decline over the 24 weeks.

After randomization and enrolment into the study, intervention participants received an email containing website access, information about receiving text messages, the recommendation to access the website within a week to commence their exercise program, and were told they could continue to access the website at any time over the 24 weeks. Participants also received a text message encouraging website access. The control group received access to another custom-built website containing the same educational information as the intervention website. After randomization and enrolment into the study, control participants received an email containing website access, the recommendation to access the website within a week, and were told they could access the website at any time after the 24 weeks. Participants also received a single text message encouraging website access.

Dependent Variables

All RCT outcomes were participant-reported and collected via REDCap electronic surveys at baseline and at 24 weeks. The 2 primary outcomes were (1) overall pain in the last week, measured using an 11-point numerical rating scale (terminal descriptors, 0=no pain to 10=worst pain possible) and (2) physical function, measured using the physical function subscale of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) (score range 0=no dysfunction, 68=maximum dysfunction). These measures are reliable and valid measures recommended for knee OA clinical trials [22-24]. At 24 weeks, the change in pain and function was calculated as baseline minus 24-week values of each.

Selected Moderators

The selection of moderators was based on evidence and theoretical rationale and involved a review of the literature [16,17,25-29] and consensus by all authors (Multimedia Appendix 1). Five baseline variables were selected.

Number of Comorbidities

Number of comorbidities was collected via a question asking participants to select from a list of 13 comorbidities any of which were relevant to them (see Textbox 1). A participant's selected comorbidities were then added to create a continuous score of the total number of comorbidities per participant. This resulted in a score range of 0 to 4.

Textbox 1. List of comorbidities.

1. Heart disease (eg, angina, heart attack, heart failure)
2. High blood pressure
3. Problems caused by a stroke
4. Leg pain when walking due to poor circulation
5. Lung disease (eg, asthma, chronic bronchitis, emphysema)
6. Diabetes
7. Kidney disease
8. Diseases of the nervous system (eg, Parkinson disease, multiple sclerosis)
9. Liver disease
10. Cancer (within the last 5 years)
11. Depression
12. Arthritis in your back or other condition affecting your spine
13. Rheumatoid arthritis or another kind of arthritis in addition to osteoarthritis

Number of Other Joints With Pain

The number of other joints with pain was collected via a question asking participants to select any other joint they currently experience pain in, from a list of 9 joints, with responses converted into a continuous score of number of other joints with pain per participant. This resulted in a score range of 0 to 9.

Pain Self-efficacy

Pain self-efficacy relates to one's confidence in their ability to control or manage pain [30,31]. Pain self-efficacy was measured using the pain subscale of the Arthritis Self-Efficacy Scale. The score range of the pain subscale is 1 to 10, with higher scores indicating greater pain self-efficacy [31].

Exercise Self-efficacy

Exercise self-efficacy relates to one's ability to continue exercising in the face of barriers to exercise [32]. Exercise self-efficacy was measured using the Self-efficacy for Exercise Scale. The score range for this scale is 0 to 90, with higher scores indicating greater exercise self-efficacy [32].

Self-perceived Importance of Exercise

Exercise importance was measured in response to the question "How important is it to you to do regular exercise to manage your knee condition?" Responses were collected using a 7-point Likert scale with a score range of 1 to 7, with higher scores indicating greater importance.

Statistical Analysis

Separate linear regression models were fit for each primary outcome with the baseline of the relevant outcome, treatment group, and one of the 5 potential moderators as covariates, including an interaction between treatment group and the relevant potential moderator. Results were calculated as the estimated mean effect of a 1-unit increase in the potential moderator for each treatment group. Using complete case data, fractional polynomials were employed to determine if nonlinear interaction models were warranted. All other analyses were performed on complete case data and multiply imputed data, including assessing regression assumptions of linearity and homoscedasticity using standard diagnostic plots. These potential moderator interactions were also assessed visually via plots of the difference in the change in the primary outcomes between groups versus the potential moderator. Multiply imputed data were the primary analysis in all interpretations. All statistical analyses were performed using Stata version 16.1 (StataCorp LLC, College Station).

Results

Baseline Descriptive Information

In this study, 206 people with a clinical diagnosis of knee OA were recruited from all Australian states and territories and were enrolled into the study (126/206, 61.2% female, mean age 60 [SD 8.4] years). Baseline characteristics of participants in both groups were similar (Table 1). At the 24-week follow-up, 88.3% (91/103) and 87.3% (90/103) of the participants in the intervention group and 89.3% (92/103) and 87.3% (90/103) of the participants in the control group provided pain and function primary outcomes, respectively.

Table 1. Baseline descriptive characteristics by treatment group.

Baseline variable	Intervention group (n=103)	Control group (n=103)
Age (years), mean (SD)	60.3 (8.2)	59.0 (8.5)
Female, n (%)	60 (58.2)	66 (64.1)
Number of comorbidities[a], mean (SD)	0.8 (1)	0.8 (0.9)
Number of other joints with pain[b], mean (SD)	1.7 (1.5)	1.9 (1.7)
Arthritis self-efficacy pain subscale[c], mean (SD)	6.0 (1.7)	6.0 (1.7)
Self-efficacy for exercise[d], mean (SD)	60.6 (21.5)	58.8 (18.6)
Exercise importance[e], mean (SD)	6.1 (1.2)	6.1 (1.2)

[a]Collected via a question asking participants to select from a list of 13 comorbidities any of which were relevant to them. A participant's selected comorbidities were then added to create a continuous score of the total number of comorbidities per participant; this resulted in a range of 0 to 4.

[b]Collected via a question asking participants to select from a list of 9 joints any of which they currently experience pain in. Responses were converted into a continuous score of number of other joints with pain per participant, ranging from 0 to 9.

[c]Scores range from 1 to 10, with higher scores indicating greater self-efficacy for pain.

[d]Scores range from 0 to 90, with higher scores indicating greater self-efficacy for exercise.

[e]Measured via the response to the question "How important is it to you to do regular exercise to manage your knee condition?" Scores range from 1 to 7; higher score indicates higher importance.

Moderators of the Effect of the Intervention on Change in Physical Function

Findings from the fractional polynomial assessment indicated that the models assuming a linear relationship between each potential moderator and change in physical function provided the best fit; therefore, more complex models were not needed (refer to scatter plots in Multimedia Appendix 2). Results of linear models using multiply imputed data are presented in Table 2 and visually in Figure 1 and Figure 2. There was evidence

that pain self-efficacy moderated the effect of the intervention on physical function compared to the control at 24 weeks using multiply imputed data (estimated mean difference –2.14, 95% CI –3.96 to –0.33; *P*=.02). Posthoc assessment of the mean change in WOMAC function by treatment arm following identification of an interaction effect showed that each 1-unit increase in baseline pain self-efficacy was associated with a 1.52 (95% CI 0.27 to 2.78) WOMAC units improvement in the control group. In contrast, with each unit increase in pain self-efficacy, a reduction of 0.62 (95% CI –1.93 to 0.68) WOMAC units was observed in the intervention group. There was only weak evidence that the other selected baseline variables moderated the effect of the intervention on physical function compared to the control at 24 weeks (Table 2). Additionally, results show positive associations between each number of comorbidities, self-efficacy for exercise and exercise importance, and change in WOMAC function for both control and intervention groups (Figure 2). There appears to be a negative relationship between number of other joints with pain and change in WOMAC function for the intervention group but a positive relationship for the control group (Figure 2). Results using complete case data were similar (Multimedia Appendix 3 and Multimedia Appendix 4).

Table 2. Results of the moderation analysis presented in terms of the effect on change in Western Ontario and McMaster Universities Osteoarthritis Index function of a 1-unit increase in the potential moderators in each of the control and intervention groups using multiply imputed data.

Moderator (taken at baseline)	Estimated moderator coefficient (95% CI)		Interaction *P* value
	Intervention group	Control group	
Number of comorbidities[a]	0.49 (–1.68 to 2.66)	2.41 (0.01 to 4.81)	.24
Number of other joints with pain[b]	–0.68 (–2.21 to 0.84)	0.76 (–0.55 to 2.07)	.16
Arthritis self-efficacy pain subscale[c]	–0.62 (–1.93 to 0.68)	1.52 (0.27 to 2.78)	.02
Self-efficacy for exercise scale[d]	0.04 (–0.07 to 0.14)	0.09 (–0.03 to 0.20)	.54
Exercise importance[e]	0.33 (–1.48 to 2.13)	1.00 (–0.81 to 2.81)	.61

[a]Collected via a question asking participants to select from a list of 13 comorbidities any of which were relevant to them. A participant's selected comorbidities were then added to create a continuous score of the total number of comorbidities per participant; this resulted in a range of 0 to 4.

[b]Collected via a question asking participants to select from a list of 9 joints any of which they currently experience pain in. Responses were converted into a continuous score of number of other joints with pain per participant, ranging from 0 to 9.

[c]Scores range from 1 to 10, with higher scores indicating greater self-efficacy for pain.

[d]Scores range from 0 to 90, with higher scores indicating greater self-efficacy for exercise.

[e]Measured via the response to the question "How important is it to you to do regular exercise to manage your knee condition?" Scores range from 1 to 7; higher score indicates higher importance.

Figure 1. Differences in the mean change in Western Ontario and McMaster Universities Osteoarthritis Index function (baseline minus 24 weeks) between treatment groups (intervention minus control) for each potential continuous moderator by using multiply imputed data. Positive values favor the intervention. The solid line indicates the difference between the control and intervention arms. Dashed line indicates no difference between the control and intervention arms. Shaded areas indicate 95% confidence intervals. WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index.

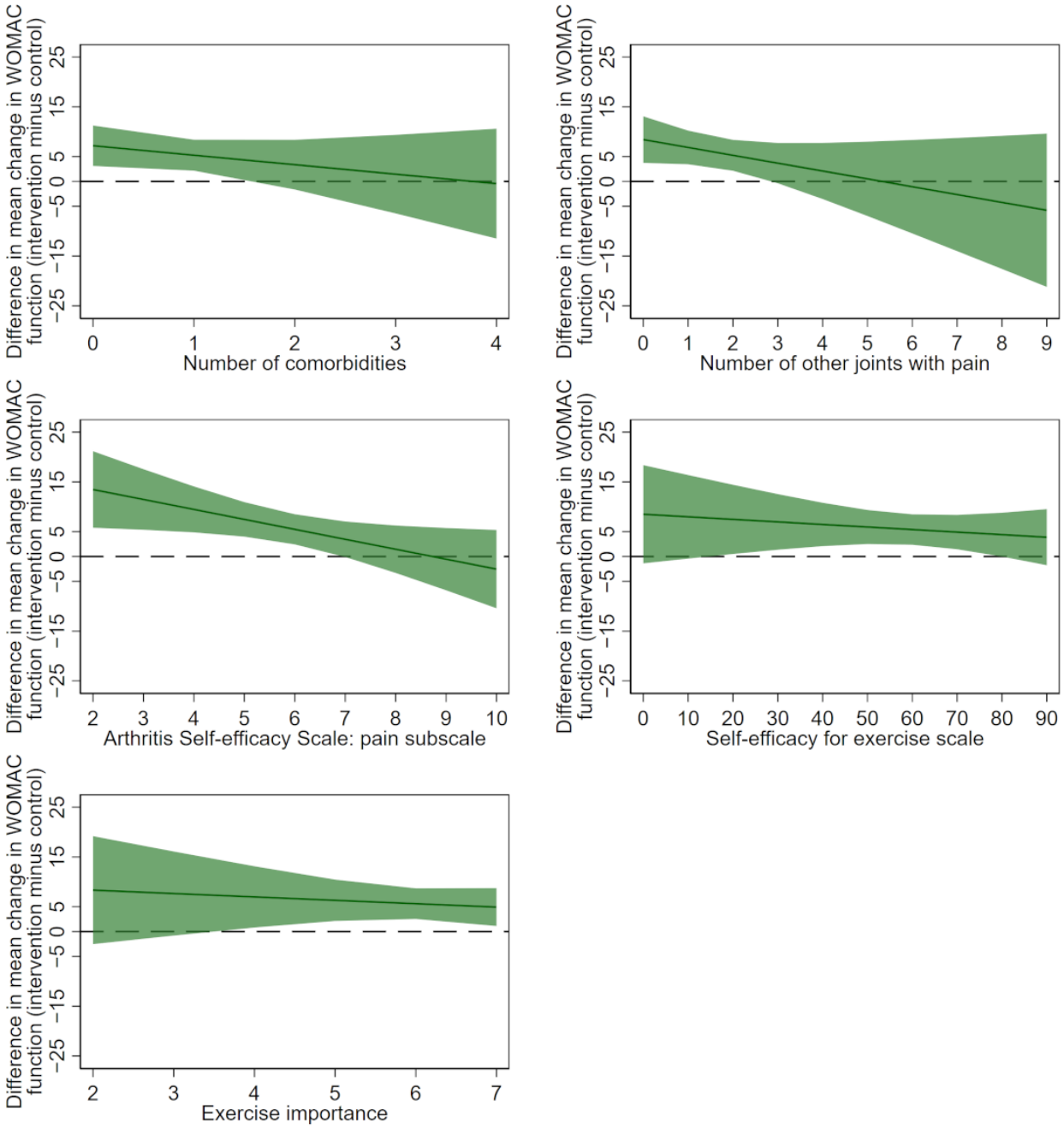
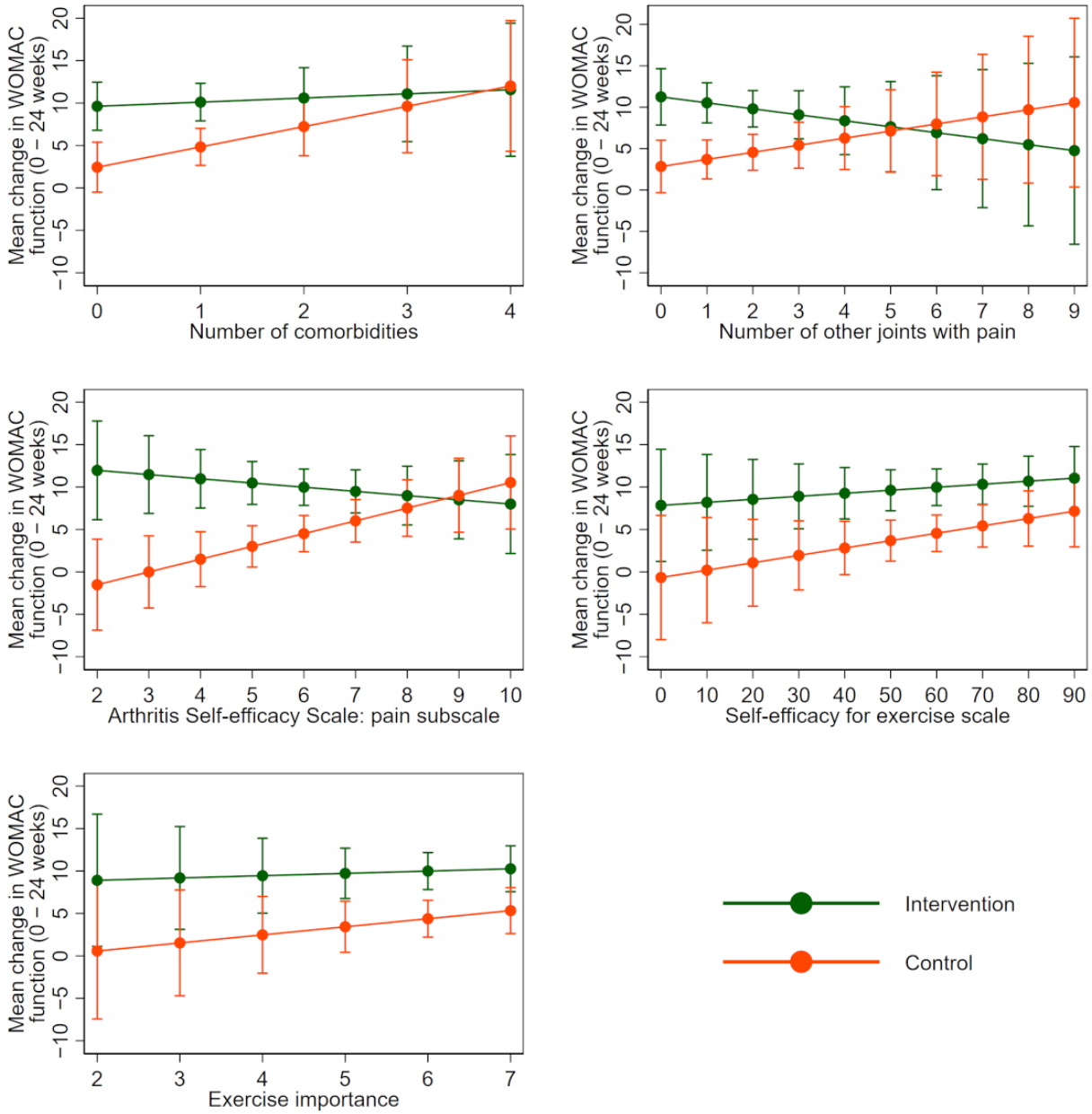


Figure 2. Mean change in Western Ontario and McMaster Universities Osteoarthritis Index function (baseline minus 24 weeks) in each treatment group for each potential continuous moderator by using multiply imputed data. Positive changes indicate improvement. The solid line indicates the average change in each treatment group. Bars indicate 95% confidence intervals. WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index.



Moderators of the Effect of the Intervention on Change in Overall Pain

The models assuming a linear relationship between each potential moderator and change in overall pain provided the best fit; therefore, more complex models were not needed (refer to scatter plots in Multimedia Appendix 5). The results of linear models using multiply imputed data are presented in Table 3 and visually in Figure 3 and Figure 4. There was only weak evidence that any of the investigated variables moderated the effect of the intervention on pain compared to the control at 24 weeks. In general, positive relationships were observed between each of the baseline characteristics and change in overall pain for both intervention and control arms, with little difference in the magnitude of the slope for each group (Figure 4). Results using complete case data were similar (Multimedia Appendix 3 and Multimedia Appendix 6).

Table 3. Results of the moderation analysis, presented in terms of the effect on change in the numerical rating scale of overall knee pain of a 1-unit increase in the potential moderators in each of the control and intervention groups using multiply imputed data.

Moderator (taken at baseline)	Estimated moderator coefficient (95% CI)		Interaction *P* value
	Intervention group	Control group	
Number of comorbidities[a]	0.23 (–0.21 to 0.68)	0.13 (–0.35 to 0.62)	.76
Number of other joints with pain[b]	–0.02 (–0.32 to 0.28)	0.10 (–0.16 to 0.36)	.56
Arthritis self-efficacy pain subscale[c]	0.23 (–0.02 to 0.49)	0.14 (–0.10 to 0.38)	.60
Self-efficacy for exercise scale[d]	0.00 (–0.02 to 0.02)	0.02 (0.00 to 0.05)	.13
Exercise importance[e]	0.11 (–0.23 to 0.45)	0.47 (0.12 to 0.83)	.15

[a]Collected via a question asking participants to select from a list of 13 comorbidities any of which were relevant to them. A participant's selected comorbidities were then added to create a continuous score of the total number of comorbidities per participant; this resulted in a range of 0 to 4.

[b]Collected via a question asking participants to select from a list of 9 joints, any of which they currently experience pain in. Responses were converted into a continuous score of number of other joints with pain per participant, ranging from 0 to 9.

[c]Scores range from 1 to 10, with higher scores indicating greater self-efficacy for pain.

[d]Scores range from 0 to 90, with higher scores indicating greater self-efficacy for exercise.

[e]Measured via the response to the question "How important is it to you to do regular exercise to manage your knee condition?" Scores range from 1 to 7; higher scores indicate higher importance.

Figure 3. Differences in the mean change in the numerical rating scale for overall knee pain (baseline minus 24 weeks) between treatment groups (intervention minus control) for each potential continuous moderator by using multiply imputed data. Positive values favor the intervention. The solid line indicates the difference between the control and intervention arms. The dashed line indicates no difference between the control and intervention arms. Shaded areas indicate 95% confidence intervals. NRS: numerical rating scale.

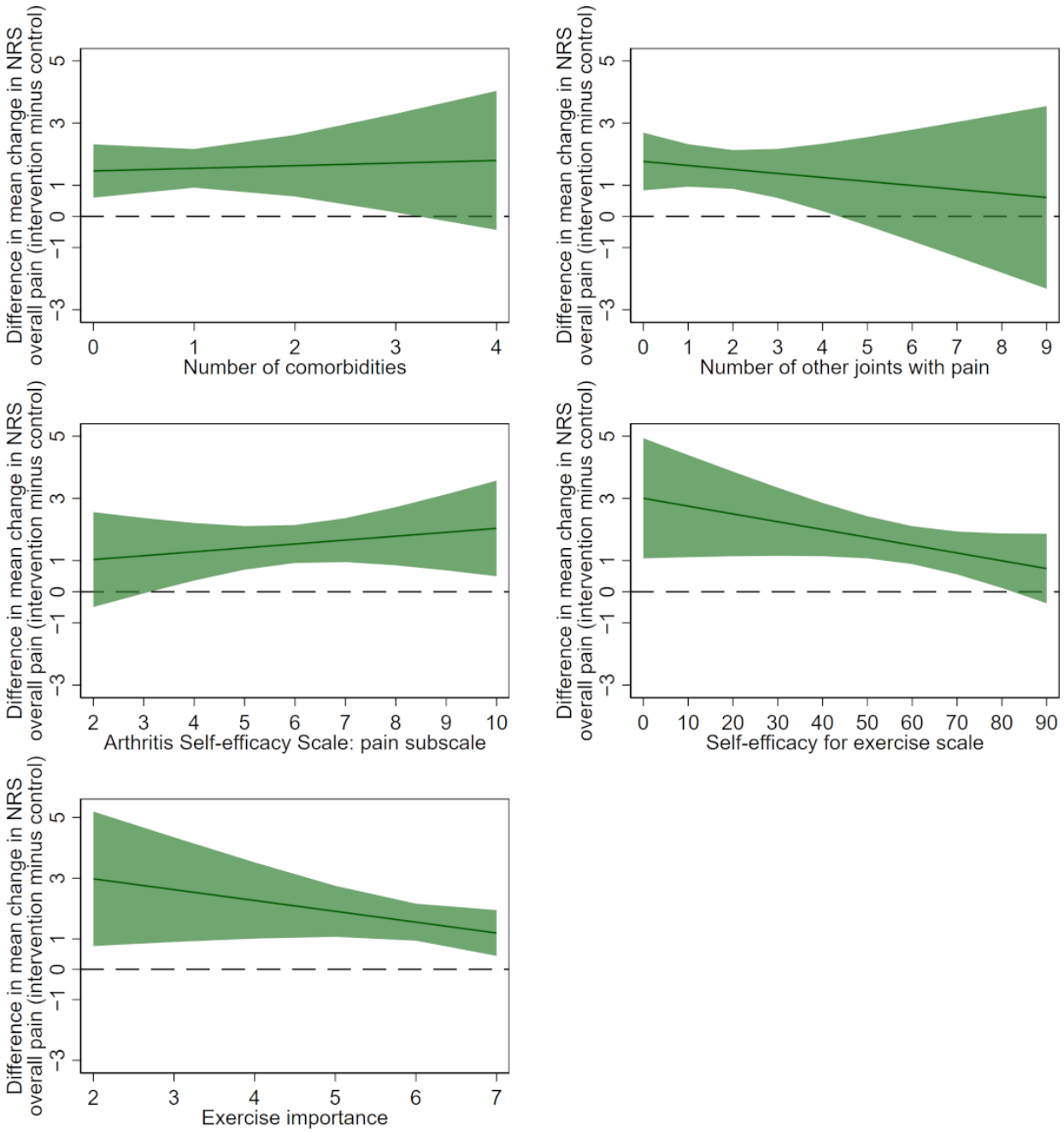
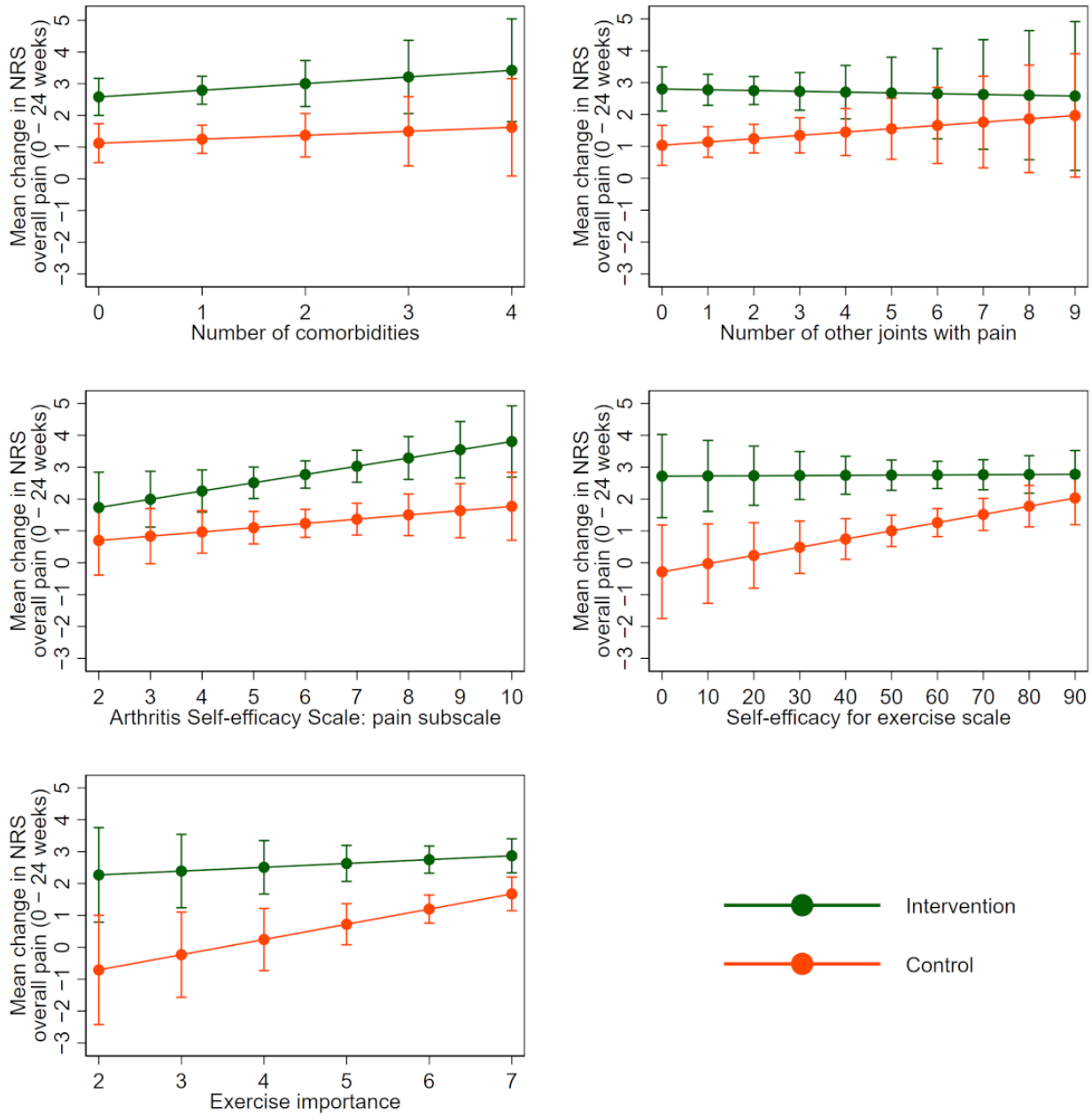


Figure 4. Mean change in the numerical rating scale for overall knee pain (baseline minus 24 weeks) in each treatment group for each potential continuous moderator by using multiply imputed data. Positive changes indicate improvement. The solid line indicates the average change in each treatment group. Bars indicate 95% confidence intervals. NRS: numerical rating scale.



Discussion

Principal Findings

This study explored potential moderators of the effect of a self-directed digitally delivered exercise intervention for people with knee OA on changes in pain and physical function over 24 weeks. Except for pain self-efficacy, we found little evidence that our selected patient characteristics moderated treatment outcomes. Regarding pain self-efficacy, the hypothesis generated from this exploratory study is that participants with higher pain self-efficacy at baseline experienced smaller improvements in function with the self-directed digitally delivered exercise intervention compared to the control. Conversely, participants with lower pain self-efficacy at baseline experienced greater improvements in function from the intervention compared to the control. After identifying an interaction effect between pain self-efficacy and function, we further explored our finding by treatment group. We identified that the interaction effect was driven by changes in the control group. Specifically, each unit increase in pain self-efficacy was associated with a 1.52-unit (95% CI 0.27 to 2.78) improvement in function in the control group. In contrast, there was a reduction in function of 0.62 units (95% CI –1.93 to 0.68) for each unit increase in pain self-efficacy in the intervention group, although this finding is equivocal, given the confidence interval crossed zero. Pain self-efficacy refers to one's confidence in their ability to control their pain and function while in pain [30] and determines how much effort people will expend and how long they will persist

while experiencing pain [33]. Therefore, it may be possible that participants in the control group who had higher baseline pain self-efficacy were more confident and may have felt more willing to act upon the general knee exercise information offered in the control website while experiencing pain. Hence, this may be why people in the control group with greater baseline pain self-efficacy (over 4/10) experienced improvements in function while those with lower pain self-efficacy (less than 4/10) experienced worsening function. This suggests that for people with knee OA and low pain self-efficacy, interventions that combine exercise with strategies to improve pain self-efficacy such as evidence-based pain education and psychological interventions (eg, pain coping skills training) [34] may be more appropriate than education only.

There is limited research investigating the moderating effects of self-efficacy on exercise outcomes in knee OA. One exploratory study found evidence that pain self-efficacy moderated the effect of a telehealth-delivered physiotherapist-prescribed exercise program combined with an automated pain coping skills training program for people with knee OA on changes in pain but not function when compared to an education control [25]. This contrasts our findings where pain self-efficacy moderated changes in function but not pain. The reason for the variation in the findings between the 2 studies is unclear; however, it may relate to the vast differences between the interventions. For example, the former study included physiotherapist-prescribed exercise and psychological treatment addressing pain coping, while the intervention in our study comprised fully self-directed exercise and automated exercise adherence support. Even so, these differences in findings highlight the need for future, adequately powered studies to rule out chance findings and rigorously explore the moderating role of pain self-efficacy on exercise effects and the related mechanisms that may be at play in people with knee OA. Regarding self-efficacy for exercise, one study has conducted a secondary analysis of data from an RCT comparing exercise, self-management, and active coping strategies (individual-delivered and group-delivered) to usual care on function in people with chronic knee pain [35]. Similar to our findings, this study found no evidence that baseline self-efficacy for exercise (measured in this study via a subscale of an exercise health beliefs and self-efficacy questionnaire) moderated the effect of the intervention on function at 6 months compared to the usual care control. A possible explanation for these findings may relate to uncertainties about the effect of exercise adherence on knee OA outcomes. For example, although self-efficacy for exercise is associated with higher levels of physical activity/exercise participation in people with knee pain [26], greater exercise adherence may not actually translate to improved knee OA outcomes [36]. As these findings are all exploratory in nature, further studies are required to explore the potential of self-efficacy to moderate the effects of exercise on OA symptoms. Nonetheless, our findings do demonstrate that patients with all levels of self-efficacy experienced improvements in function and reductions in knee pain from the self-directed digitally delivered exercise intervention, which supports the use of this intervention for people with knee OA and all levels of self-efficacy at baseline.

To our knowledge, this is the first study to investigate multi-joint pain and comorbidities as moderators of the effect of self-directed exercise on outcomes in people with OA. We found that neither of these baseline characteristics moderated the effect of the intervention on changes in pain or function relative to the control. Despite this, an interesting observation was a negative relationship between number of other joints with pain and changes in physical function for the intervention group (ie, those with a greater number of joints with pain experienced smaller improvements in function). Although, this was supported by only weak evidence of an interaction effect (P=.16), this observation may indicate that the self-directed digitally delivered intervention may be less beneficial for people with pain in multiple joints. Multi-joint pain can be indicative of complex pain presentations such as widespread pain and fibromyalgia [37], and people with knee OA and widespread pain have been found to have poorer self-reported function when compared to people with knee OA only [27,38]. Furthermore, personalized exercise prescription and monitoring (eg, via a health professional) and psychological treatments such as cognitive behavior therapy are recommended for people with widespread pain [39]. Therefore, people with knee OA and pain in multiple other joints may be better suited to a tailored approach to management over a self-directed digitally delivered exercise intervention. Regarding comorbidity, 2 studies have explored the potential for the number of comorbidities present at baseline to moderate the effect of therapist-led exercise on pain and function in people with knee OA and support our findings [40,41]. A systematic review of subgroup analyses from 14 RCTs found that the number of comorbidities present at baseline did not moderate the effects of exercise interventions on pain or function compared to nonexercise controls for people with knee or hip OA [40]. Similarly, a recent secondary analysis of RCT data found that the number of comorbidities present at baseline did not moderate the effect of a course of physiotherapist-led exercise on pain and function at 6 months compared with a nonexercise control [41]. Although these findings are exploratory and require confirmation in future studies, they do indicate that people with knee OA and multiple comorbid conditions may respond to exercise in similar ways as those without comorbidity. This may be unsurprising, considering exercise is safe and effective for people with multimorbidity and is recommended for a broad range of chronic conditions [42].

There appears to be currently no research exploring patient characteristics as potential moderators of the effects of digitally delivered exercise in populations with OA and limited research in adult populations more broadly. We found 2 studies that conducted subgroup analyses of RCT data exploring potential moderators of the effect of digitally delivered physical activity interventions in adults. One study explored potential moderators of the effect of computer-tailored physical activity on changes in total weekly minutes of physical activity at 12 months in people aged 50 years and older [43]. Moderation analysis found participants with a higher age, lower body mass index, and higher self-reported intention to be physically active at baseline were not responsive to computer-tailored physical activity relative to a waitlist control. Another study explored potential moderators of the effect of computer-tailored physical activity

intervention on changes in physical activity behavior (accelerometry measured moderate-to-vigorous physical activity per week and steps per day) at 3 months in inactive adults [44]. This study found that the intervention was more effective for women than men, relative to the usual care control. Owing to the differences in the selected moderators and outcomes, direct comparisons cannot be made with our findings. However, our findings add to the scant literature regarding the role of patient characteristics as moderators of the effect of digitally delivered exercise in adult populations and may provide direction for the selection of potential moderators in future research.

Overall, our findings suggest that our self-directed digitally delivered exercise intervention is similarly effective for pain for a range of people with knee OA. With respect to function, it may be that our 24-week intervention is less beneficial when compared to control for people with high baseline pain self-efficacy (eg, a score of over 9/10; see Figure 2). Conversely, the intervention may be more beneficial when compared to control for people with lower levels of pain self-efficacy. Therefore, it may be a novel treatment approach for people with knee OA and low pain self-efficacy who typically report greater levels of disability [45].

Strengths and Limitations

The strengths of this study include its robust design using reliable and valid clinical outcome measures [22-24] and appropriate methods of statistical analyses of interaction/moderation [18]. Further, our selection of potential moderator variables was evidence-informed and explicitly described *a priori* while the selection of only a few variables aimed to minimize the risk of potentially erroneous findings. Our findings and the hypothesis generated can also be considered generalizable to the broader knee OA population owing to the limited inclusion criteria and nationwide community recruitment of participants in the RCT. Several limitations must also be acknowledged. This study was exploratory as the original RCT was powered to detect changes in pain and function and not to detect moderator effects. As this was exploratory, no adjustment for multiple testing was conducted. It is possible that a lack of power prevented the identification of potential moderators rather than the absence of an effect or that our finding occurred by chance, although a chance finding was controlled for by limiting the number of potential moderators to analyze. As such, further confirmatory studies are required. Additionally, despite our selection of variables being evidence-informed, it is possible that baseline variables other than those selected could be potential moderators.

Conclusion

With the exception of pain self-efficacy, we found little evidence that our selected patient characteristics moderated treatment outcomes. Although these findings are exploratory in nature, they do contribute to the sparse literature regarding moderators of the effect of digitally delivered exercise in adult populations and may inform future research aiming to improve the targeting of exercise treatments.

Acknowledgments

Funding was provided by the National Health & Medical Research Council (NHMRC) (Grant 1091302). KLB is supported by an NHMRC Fellowship (1058440). RSH is supported by an NHMRC Fellowship (1154217). RKN is supported by an Australian Government Research Training Program Scholarship.

Authors' Contributions

All authors contributed to the conception and design of this study. RKN was responsible for participant recruitment, data collection, and drafting of the manuscript. FM and KL were responsible for data analysis. All authors contributed to the interpretation of data and critical revision of the manuscript. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of the selected moderators and their rationale for inclusion.
[PDF File (Adobe PDF File), 76 KB - jmir_v23i10e30768_app1.pdf]

Multimedia Appendix 2
Change in function (baseline minus 24 weeks) against each continuous potential moderator by treatment group by using multiply imputed data.
[PDF File (Adobe PDF File), 345 KB - jmir_v23i10e30768_app2.pdf]

Multimedia Appendix 3
Results of the moderation analysis, presented in terms of the effect on the primary outcomes of a 1-unit increase in the moderators in each of the control and intervention groups by using complete case data.
[PDF File (Adobe PDF File), 134 KB - jmir_v23i10e30768_app3.pdf]

Multimedia Appendix 4
Difference in mean change in Western Ontario and McMaster Universities Osteoarthritis Index function (baseline minus 24 weeks) between treatment groups (intervention minus control) for each potential continuous moderator by using complete case data.
[PDF File (Adobe PDF File), 270 KB - jmir_v23i10e30768_app4.pdf]

Multimedia Appendix 5
Change in overall pain (baseline minus 24 weeks) against each continuous potential moderator, by treatment group using multiply imputed data.
[PDF File (Adobe PDF File), 263 KB - jmir_v23i10e30768_app5.pdf]

Multimedia Appendix 6
Difference in mean change in overall knee pain (baseline minus 24 weeks) between treatment groups (intervention minus control) for each potential continuous moderator by using complete case data.
[PDF File (Adobe PDF File), 250 KB - jmir_v23i10e30768_app6.pdf]

References

1. Nuki G. Osteoarthritis: a problem of joint failure. Z Rheumatol 1999 Jun;58(3):142-147. [doi: 10.1007/s003930050164] [Medline: 10441841]
2. Pereira D, Peleteiro B, Araújo J, Branco J, Santos RA, Ramos E. The effect of osteoarthritis definition on prevalence and incidence estimates: a systematic review. Osteoarthritis Cartilage 2011 Nov;19(11):1270-1285 [FREE Full text] [doi: 10.1016/j.joca.2011.08.009] [Medline: 21907813]
3. Dieppe P, Cushnaghan J, Tucker M, Browning S, Shepstone L. The Bristol 'OA500 study': progression and impact of the disease after 8 years. Osteoarthritis Cartilage 2000 Mar;8(2):63-68 [FREE Full text] [doi: 10.1053/joca.1999.0272] [Medline: 10772234]
4. Sharma L, Cahue S, Song J, Hayes K, Pai Y, Dunlop D. Physical functioning over three years in knee osteoarthritis: role of psychosocial, local mechanical, and neuromuscular factors. Arthritis Rheum 2003 Dec;48(12):3359-3370 [FREE Full text] [doi: 10.1002/art.11420] [Medline: 14673987]
5. Safiri S, Kolahi A, Smith E, Hill C, Bettampadi D, Mansournia MA, et al. Global, regional and national burden of osteoarthritis 1990-2017: a systematic analysis of the Global Burden of Disease Study 2017. Ann Rheum Dis 2020 Jun;79(6):819-828. [doi: 10.1136/annrheumdis-2019-216515] [Medline: 32398285]
6. Kolasinski SL, Neogi T, Hochberg MC, Oatis C, Guyatt G, Block J, et al. 2019 American College of Rheumatology/Arthritis Foundation Guideline for the Management of Osteoarthritis of the Hand, Hip, and Knee. Arthritis Rheumatol 2020 Feb;72(2):220-233 [FREE Full text] [doi: 10.1002/art.41142] [Medline: 31908163]
7. Bannuru RR, Osani MC, Vaysbrot EE, Arden NK, Bennell K, Bierma-Zeinstra SMA, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthritis Cartilage 2019 Nov;27(11):1578-1589 [FREE Full text] [doi: 10.1016/j.joca.2019.06.011] [Medline: 31278997]
8. The Royal Australian College of General Practitioners. Guideline for the Management of Knee and Hip Osteoarthritis, Second edition. Victoria: RACGP; 2018.
9. Osteoarthritis: care and management. National Institute for Health Clinical Excellence Clinical Guideline No. 177. 2020 Dec 11. URL: https://www.nice.org.uk/guidance/cg177 [accessed 2021-05-01]
10. Fransen M, McConnell S, Harmer AR, Van der Esch M, Simic M, Bennell KL. Exercise for osteoarthritis of the knee: a Cochrane systematic review. Br J Sports Med 2015 Dec;49(24):1554-1557. [doi: 10.1136/bjsports-2015-095424] [Medline: 26405113]
11. Holden MA, Burke DL, Runhaar J, van Der Windt D, Riley RD, Dziedzic K, et al. Subgrouping and TargetEd Exercise pRogrammes for knee and hip OsteoArthritis (STEER OA): a systematic review update and individual participant data meta-analysis protocol. BMJ Open 2017 Dec 22;7(12):e018971 [FREE Full text] [doi: 10.1136/bmjopen-2017-018971] [Medline: 29275348]
12. Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and moderators of treatment effects in randomized clinical trials. Arch Gen Psychiatry 2002 Oct;59(10):877-883. [doi: 10.1001/archpsyc.59.10.877] [Medline: 12365874]
13. Fernandes L, Hagen KB, Bijlsma JWJ, Andreassen O, Christensen P, Conaghan PG, European League Against Rheumatism (EULAR). EULAR recommendations for the non-pharmacological core management of hip and knee osteoarthritis. Ann Rheum Dis 2013 Jul;72(7):1125-1135 [FREE Full text] [doi: 10.1136/annrheumdis-2012-202745] [Medline: 23595142]
14. Kvedar JC, Fogel AL, Elenko E, Zohar D. Digital medicine's march on chronic disease. Nat Biotechnol 2016 Mar;34(3):239-246. [doi: 10.1038/nbt.3495] [Medline: 26963544]
15. Nelligan RK, Hinman RS, Kasza J, Crofts SJC, Bennell KL. Effects of a Self-directed Web-Based Strengthening Exercise and Physical Activity Program Supported by Automated Text Messages for People With Knee Osteoarthritis: A Randomized Clinical Trial. JAMA Intern Med 2021 Jun 01;181(6):776-785. [doi: 10.1001/jamainternmed.2021.0991] [Medline: 33843948]

16. Bossen D, Buskermolen M, Veenhof C, de Bakker D, Dekker J. Adherence to a web-based physical activity intervention for patients with knee and/or hip osteoarthritis: a mixed method study. J Med Internet Res 2013 Oct 16;15(10):e223 [FREE Full text] [doi: 10.2196/jmir.2742] [Medline: 24132044]
17. Baruth M, Wilcox S, Sharpe P, Schoffman D, Becofsky K. Baseline predictors of physical activity in a sample of adults with arthritis participating in a self-directed exercise program. Public Health 2014 Sep;128(9):834-841 [FREE Full text] [doi: 10.1016/j.puhe.2014.06.018] [Medline: 25204768]
18. Royston P, Sauerbrei W. Two Techniques for Investigating Interactions between Treatment and Continuous Covariates in Clinical Trials. The Stata Journal 2009 Aug 01;9(2):230-251. [doi: 10.1177/1536867x0900900204]
19. Nelligan RK, Hinman RS, Kasza J, Bennell KL. Effectiveness of internet-delivered education and home exercise supported by behaviour change SMS on pain and function for people with knee osteoarthritis: a randomised controlled trial protocol. BMC Musculoskelet Disord 2019 Jul 27;20(1):342 [FREE Full text] [doi: 10.1186/s12891-019-2714-x] [Medline: 31351449]
20. Nelligan RK, Hinman RS, Atkins L, Bennell KL. A Short Message Service Intervention to Support Adherence to Home-Based Strengthening Exercise for People With Knee Osteoarthritis: Intervention Design Applying the Behavior Change Wheel. JMIR Mhealth Uhealth 2019 Oct 18;7(10):e14619 [FREE Full text] [doi: 10.2196/14619] [Medline: 31628786]
21. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide To Designing Interventions. London: Silverback Publishing; May 31, 2014.
22. McAlindon TE, Driban JB, Henrotin Y, Hunter DJ, Jiang G, Skou ST, et al. OARSI Clinical Trials Recommendations: Design, conduct, and reporting of clinical trials for knee osteoarthritis. Osteoarthritis Cartilage 2015 May;23(5):747-760 [FREE Full text] [doi: 10.1016/j.joca.2015.03.005] [Medline: 25952346]
23. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988 Dec;15(12):1833-1840. [Medline: 3068365]
24. Rolfson O, Wissig S, van Maasakkers L, Stowell C, Ackerman I, Ayers D, et al. Defining an International Standard Set of Outcome Measures for Patients With Hip or Knee Osteoarthritis: Consensus of the International Consortium for Health Outcomes Measurement Hip and Knee Osteoarthritis Working Group. Arthritis Care Res (Hoboken) 2016 Nov;68(11):1631-1639 [FREE Full text] [doi: 10.1002/acr.22868] [Medline: 26881821]
25. Lawford BJ, Hinman RS, Kasza J, Nelligan R, Keefe F, Rini C, et al. Moderators of Effects of Internet-Delivered Exercise and Pain Coping Skills Training for People With Knee Osteoarthritis: Exploratory Analysis of the IMPACT Randomized Controlled Trial. J Med Internet Res 2018 May 09;20(5):e10021 [FREE Full text] [doi: 10.2196/10021] [Medline: 29743149]
26. Quicke JG, Foster NE, Ogollah RO, Croft PR, Holden MA. Relationship Between Attitudes and Beliefs and Physical Activity in Older Adults With Knee Pain: Secondary Analysis of a Randomized Controlled Trial. Arthritis Care Res (Hoboken) 2017 Aug;69(8):1192-1200 [FREE Full text] [doi: 10.1002/acr.23104] [Medline: 27696795]
27. Bergman S, Thorstensson C, Andersson MLE. Chronic widespread pain and its associations with quality of life and function at a 20- year follow-up of individuals with chronic knee pain at inclusion. BMC Musculoskelet Disord 2019 Dec 09;20(1):592 [FREE Full text] [doi: 10.1186/s12891-019-2976-3] [Medline: 31818282]
28. Mahgoub MY, Elnady BM, et al. Erratum: Comorbidity of Fibromyalgia in Primary Knee Osteoarthritis: Potential Impact on Functional Status and Quality of Life [Corrigendum]. Open Access Rheumatol 2020;12:125-126 [FREE Full text] [doi: 10.2147/OARRR.S267971] [Medline: 32617027]
29. Dobson F, Bennell KL, French SD, Nicolson PJA, Klaasman RN, Holden MA, et al. Barriers and Facilitators to Exercise Participation in People with Hip and/or Knee Osteoarthritis: Synthesis of the Literature Using Behavior Change Theory. Am J Phys Med Rehabil 2016 May;95(5):372-389. [doi: 10.1097/PHM.0000000000000448] [Medline: 26945211]
30. Nicholas MK. The pain self-efficacy questionnaire: Taking pain into account. Eur J Pain 2007 Feb;11(2):153-163. [doi: 10.1016/j.ejpain.2005.12.008] [Medline: 16446108]
31. Lorig K, Chastain RL, Ung E, Shoor S, Holman HR. Development and evaluation of a scale to measure perceived self-efficacy in people with arthritis. Arthritis Rheum 1989 Jan;32(1):37-44. [doi: 10.1002/anr.1780320107] [Medline: 2912463]
32. Resnick B, Jenkins LS. Testing the reliability and validity of the Self-Efficacy for Exercise scale. Nurs Res 2000;49(3):154-159. [doi: 10.1097/00006199-200005000-00007] [Medline: 10882320]
33. Bandura A, Freeman WH, Lightsey R. Self-Efficacy: The Exercise of Control. J Cogn Psychother 1999 Jan 01;13(2):158-166. [doi: 10.1891/0889-8391.13.2.158]
34. Keefe F, Caldwell D, Williams D, Gil K, Mitchell D, Robertson C. Pain coping skills training in the management of osteoarthritic knee pain: A comparative study. Behavior Therapy 1990;21(1):49-62 [FREE Full text] [doi: 10.1016/s0005-7894(05)80188-1]
35. Hurley MV, Walsh NE, Mitchell HL, Pimm TJ, Patel A, Williamson E, et al. Clinical effectiveness of a rehabilitation program integrating exercise, self-management, and active coping strategies for chronic knee pain: a cluster randomized trial. Arthritis Rheum 2007 Oct 15;57(7):1211-1219 [FREE Full text] [doi: 10.1002/art.22995] [Medline: 17907147]
36. Nicolson PJ, Hinman RS, Wrigley TV, Stratford PW, Bennell KL. Effects of Covertly Measured Home Exercise Adherence on Patient Outcomes Among Older Adults With Chronic Knee Pain. J Orthop Sports Phys Ther 2019 Jul;49(7):548-556. [doi: 10.2519/jospt.2019.8843] [Medline: 31258045]

37. Wolfe F, Smythe HA, Yunus MB, Bennett RM, Bombardier C, Goldenberg DL, et al. The American College of Rheumatology 1990 Criteria for the Classification of Fibromyalgia. Report of the Multicenter Criteria Committee. Arthritis Rheum 1990 Feb;33(2):160-172. [doi: 10.1002/art.1780330203] [Medline: 2306288]
38. Guérard O, Dufort S, Forget Besnard L, Gougeon A, Carlesso L. Comparing the association of widespread pain, multi-joint pain and low back pain with measures of pain sensitization and function in people with knee osteoarthritis. Clin Rheumatol 2020 Mar;39(3):873-879. [doi: 10.1007/s10067-019-04828-3] [Medline: 31713734]
39. Macfarlane GJ, Kronisch C, Dean LE, Atzeni F, Häuser W, Fluß E, et al. EULAR revised recommendations for the management of fibromyalgia. Ann Rheum Dis 2017 Feb;76(2):318-328. [doi: 10.1136/annrheumdis-2016-209724] [Medline: 27377815]
40. Quicke J, Runhaar J, van der Windt D, Healey E, Foster N, Holden M. Moderators of the effects of therapeutic exercise for people with knee and hip osteoarthritis: A systematic review of sub-group analyses from randomised controlled trials. Osteoarthritis and Cartilage Open 2020 Dec;2(4):100113. [doi: 10.1016/j.ocarto.2020.100113]
41. Legha A, Burke DL, Foster NE, van der Windt DA, Quicke JG, Healey EL, et al. Do comorbidities predict pain and function in knee osteoarthritis following an exercise intervention, and do they moderate the effect of exercise? Analyses of data from three randomized controlled trials. Musculoskeletal Care 2020 Mar;18(1):3-11. [doi: 10.1002/msc.1425] [Medline: 31837126]
42. Bricca A, Harris L, Jäger M, Smith S, Juhl C, Skou S. Benefits and harms of exercise therapy in people with multimorbidity: A systematic review and meta-analysis of randomised controlled trials. Ageing Res Rev 2020 Nov;63:101166 [FREE Full text] [doi: 10.1016/j.arr.2020.101166] [Medline: 32896665]
43. van Stralen MM, de Vries H, Bolman C, Mudde AN, Lechner L. Exploring the efficacy and moderators of two computer-tailored physical activity interventions for older adults: a randomized controlled trial. Ann Behav Med 2010 May;39(2):139-150. [doi: 10.1007/s12160-010-9166-8] [Medline: 20182833]
44. To Q, Duncan M, Short C, Plotnikoff R, Kerry Mummery W, Alley S, et al. Examining moderators of the effectiveness of a web- and video-based computer-tailored physical activity intervention. Prev Med Rep 2021 Jun;22:101336 [FREE Full text] [doi: 10.1016/j.pmedr.2021.101336] [Medline: 33732607]
45. Marks R. Self-efficacy and arthritis disability: An updated synthesis of the evidence base and its relevance to optimal patient care. Health Psychol Open 2014 Jul;1(1):2055102914564582 [FREE Full text] [doi: 10.1177/2055102914564582] [Medline: 28070346]

Abbreviations

OA: osteoarthritis
RCT: randomized controlled trial
WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index

Edited by R Kukafka; submitted 01.06.21; peer-reviewed by J Eyles, S Wissig, D White; comments to author 30.07.21; revised version received 27.08.21; accepted 12.09.21; published 29.10.21.

Please cite as:
Nelligan RK, Hinman RS, McManus F, Lamb KE, Bennell KL
Moderators of the Effect of a Self-directed Digitally Delivered Exercise Program for People With Knee Osteoarthritis: Exploratory Analysis of a Randomized Controlled Trial
J Med Internet Res 2021;23(10):e30768
URL: https://www.jmir.org/2021/10/e30768
doi:10.2196/30768
PMID:

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