Journal of Medical Internet Research

Contents

XSL•FO
RenderX

Short Paper

Letters to the Editor

Review

Trigger Tool–Based Automated Adverse Event Detection in Electronic Health Records: Systematic Review

Sarah N Musy[1,2], MSc; Dietmar Ausserhofer[1,3], RN, PhD; René Schwendimann[1,4], RN, PhD; Hans Ulrich Rothen[5], MD, PhD; Marie-Madlen Jeitziner[5], RN, PhD; Anne WS Rutjes[6,7], PhD; Michael Simon[1,2], RN, PhD

[1]Institute of Nursing Science, University of Basel, Basel, Switzerland

[2]Nursing & Midwifery Research Unit, Inselspital Bern University Hospital, Bern, Switzerland

[3]College for Health Care Professions, Claudiana, Bolzano, Italy

[4]University Hospital Basel, Patient Safety Office, Basel, Switzerland

[5]Department of Intensive Care Medicine, Inselspital Bern University Hospital, Bern, Switzerland

[6]Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland

[7]Institute of Primary Health Care (BIHAM), University of Bern, Bern, Switzerland

Corresponding Author:
Michael Simon, RN, PhD
Institute of Nursing Science
University of Basel
Bernoullistrasse 28
Basel, 4057
Switzerland
Phone: 41 61 267 09 12
Email: m.simon@unibas.ch

Abstract

Background: Adverse events in health care entail substantial burdens to health care systems, institutions, and patients. Retrospective trigger tools are often manually applied to detect AEs, although automated approaches using electronic health records may offer real-time adverse event detection, allowing timely corrective interventions.

Objective: The aim of this systematic review was to describe current study methods and challenges regarding the use of automatic trigger tool-based adverse event detection methods in electronic health records. In addition, we aimed to appraise the applied studies' designs and to synthesize estimates of adverse event prevalence and diagnostic test accuracy of automatic detection methods using manual trigger tool as a reference standard.

Methods: PubMed, EMBASE, CINAHL, and the Cochrane Library were queried. We included observational studies, applying trigger tools in acute care settings, and excluded studies using nonhospital and outpatient settings. Eligible articles were divided into *diagnostic test accuracy studies* and *prevalence studies*. We derived the study prevalence and estimates for the positive predictive value. We assessed bias risks and applicability concerns using Quality Assessment tool for Diagnostic Accuracy Studies-2 (QUADAS-2) for diagnostic test accuracy studies and an in-house developed tool for prevalence studies.

Results: A total of 11 studies met all criteria: 2 concerned *diagnostic test accuracy* and 9 *prevalence*. We judged several studies to be at high bias risks for their automated detection method, definition of outcomes, and type of statistical analyses. Across all the 11 studies, adverse event prevalence ranged from 0% to 17.9%, with a median of 0.8%. The positive predictive value of all triggers to detect adverse events ranged from 0% to 100% across studies, with a median of 40%. Some triggers had wide ranging positive predictive value values: (1) in 6 studies, hypoglycemia had a positive predictive value ranging from 15.8% to 60%; (2) in 5 studies, naloxone had a positive predictive value ranging from 20% to 91%; (3) in 4 studies, flumazenil had a positive predictive value ranging from 38.9% to 83.3%; and (4) in 4 studies, protamine had a positive predictive value ranging from 0% to 60%. We were unable to determine the adverse event prevalence, positive predictive value, preventability, and severity in 40.4%, 10.5%, 71.1%, and 68.4% of the studies, respectively. These studies did not report the overall number of records analyzed, triggers, or adverse events; or the studies did not conduct the analysis.

Conclusions: We observed broad interstudy variation in reported adverse event prevalence and positive predictive value. The lack of sufficiently described methods led to difficulties regarding interpretation. To improve quality, we see the need for a set of recommendations to endorse optimal use of research designs and adequate reporting of future adverse event detection studies.

XSL•FO
RenderX

Introduction

In recent decades, patient safety and quality of care have become a top priority in health care [1-3]. This has led to significant progress, especially regarding innovative use of electronic health records (EHRs). Adverse events (AEs), injuries attributed to medical care that are independent of the patient's underlying condition, nevertheless remain a persistent problem. Apart from the impact on patients, they entail large human and financial burdens at every health care system level [4]. Regarding patient health, AEs' negative consequences include extended hospital stays, higher readmission rates, and higher mortality [5]. Furthermore, AEs may lead to the patients' and their families' loss of trust in their health care professionals (HCPs), their health care system, or both [3]. The estimated prevalence of AEs in hospital inpatients ranges from 3% to 40% in acute care settings [2,6-10]. The wide range reflects the challenges involved in detecting and tracking AEs accurately [11].

To improve patient safety, health care organizations need valid and reliable tools to detect and assess AEs [12]. Several tools exist, but their ability to identify AEs is limited, and none of them are broadly accepted [13-15]. Currently, enumerating specific events that endanger patients depends mainly on voluntarily reporting by health care staff [16,17]. Systematic evaluations of this approach showed endemic underreporting, with only 2% to 8% of all harmful events being identified [18-20]. To depict the situation more robustly, the US Agency for Healthcare Research and Quality (AHRQ) published a set of Patient Safety Indicators (PSIs). Using administrative datasets, PSIs identify potential AEs, but are highly susceptible to variations in coding practice and are limited by many outcomes being easily concealed in the medical record [14]. Therefore, they miss a substantial fraction of AEs (low sensitivity), while producing a substantial fraction of false positive results (low specificity) [13,21].

One promising method is the Global Trigger Tool (GTT), developed by the Institute for Healthcare Improvement (IHI) [22]. Providing a structured method for identifying AEs from patient records [23,24], the GTT is a retrospective record review instrument that uses a list of triggers (or clues), ie, data elements within the health record, to alert reviewers to the potential presence of AEs [22,25]. By focusing on triggers within patient records, the GTT has demonstrated to identify up to ten times as many AEs as other detection methods [13]. Various studies have used the GTT, where some modified the methods, eg, by modifying the set of triggers, or by modifying the review process (eg, one reviewer instead of two for trigger identification). We refer to these modified versions as *trigger tool methodology,* reserving the term *GTT methodology* for the IHI's original procedures [24].

The trigger tool was developed as a manual approach, ie, for application by HCPs reviewing patient records. Recently, an increasing interest developed for semi or fully automated AE detection methods using EHRs where lesser time and personnel resources are required for the AE identification [25-28]. *Prospective AE detection* would supply real-time feedback to HCPs, allowing timely interventions. The development of automated surveillance systems using EHR data has greatly facilitated AEs' identification [28].

Semi or fully automated AEs detection methods show promise to efficiently measure AEs. Nevertheless, evidence need to be summarized based on the current literature to gather information for future development and implementation in a health care organization. As a variety of AEs' detection methods exist, we decided to focus on trigger tool–based AEs detection methods only, allowing comparisons between studies as suggested in a previous systematic review on automated detection of patient harm [29]. As trigger tool methodology has shown higher sensitivity compared with other detection methods, we considered the manual trigger tool as the *gold standard*. This systematic review aimed to describe current study methods and challenges regarding the use of automatic trigger tool–based AE detection methods in EHRs in acute care settings. In addition, we aimed to appraise the applied studies' designs and to synthesize estimates of AE prevalence and diagnostic test accuracy (DTA) of automatic detection methods using manual trigger tool as a reference standard.

Methods

Search Strategy and Study Selection

This systematic review followed the recommendations of the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy [30], the Cochrane Handbook for Systematic Reviews of Interventions [31], and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines for the reporting of systematic reviews [32].

Hausner et al's approach was applied to develop a robust search strategy (Multimedia Appendix 1) [33,34]. In PubMed's basic search mode, we entered the following medical subject headings (MeSH) and free-text terms for title and abstract fields: *(trigger OR triggers) AND (chart OR charts OR identif* OR record OR records) AND (adverse OR medical errors)*. The focus of the search was on "trigger" and not on GTT, as we aimed to include studies using variations of the original GTT. The search strategy was transposed to EMBASE, CINAHL, and the Cochrane Library, and terms were mapped to the appropriate keywords (eg, from MeSH to Emtree). Studies published in English, French, German, Italian, or Spanish were considered without restrictions concerning the years of the publication. In addition to searching the bibliographic databases, the team identified additional relevant literature from most common journals publishing in the field of trigger tool: *BMJ Quality & Safety*, *Journal of Patient Safety*, and *International Journal for Quality in Health*. For pragmatic reasons, the research team decided to

XSL•FO

RenderX

limit the hand search of the most common journals to the years 2014 to 2017. The search was conducted in November 2015 with updates in April 2016 and July 2017.

We included observational studies that applied a trigger-based tool to detect AEs in EHRs in any acute care setting. We defined the target population of interest as patients hospitalized for at least 48 hours for any reason. The evaluated trigger tool approach (index test) had to involve either semi (ie, one part of the process still manual) or fully automatic identification of AEs [29]. Regarding DTA studies, we opted for a reference standard that produced a relatively low rate of missed AEs alongside an acceptable rate of false positive test results (events flagged as AEs that, upon examination, did not qualify as AEs). As shown by Classen et al (2011), manual trigger-based tools met our target criteria [13]. We excluded studies from nonhospital settings (eg, long-term care), outpatient clinics, or that concerned with nonprimary research (eg, systematic reviews or editorials).

The eligible articles were divided into two sets: (1) an automated trigger tool in comparison with a manual trigger tool method for AE detection, potentially enabling the evaluation of the trigger tool's DTA (*diagnostic test accuracy studies*) and (2) application of an automated trigger tool without cross-verification with a manual trigger tool method, enabling us to synthesize the prevalence of AEs and the applied methods. We refer to the latter group as "*prevalence studies*" throughout this paper.

Data Extraction

Two main reviewers (SNM and MS) each screened half of the retrieved titles and abstracts for relevance according to the criteria outlined above. The other members of the research team each screened a quarter of the retrieved titles and abstracts, allowing double screening for all citations. Full-text screening was independently assessed by the main two reviewers, where disagreements were resolved by discussion, or by consulting the entire research team, if necessary.

Detailed study information was extracted into tables by SNM and a master student as part of her training. We used standardized piloted extraction sheets to tabulate variables related to design, sample population characteristics, applied trigger tool methodology, type and number of reviewers and triggers, and outcome data expressed as AE prevalence (overall and by AE type). To estimate DTA, we used 2x2 tables. Whenever possible, we derived the positive predictive value (PPV) of the triggers used. PPV is calculated by dividing the number of true positive triggers related to confirm AEs by the total number of positive triggers.

Unresolved disagreements or uncertainties between SNM and the master student were discussed and resolved in the research team, which included experienced systematic review and GTT methodologists, clinicians, and nurses, each with more than 10 years of experience in their specialty.

Quality Assessment

We assessed the risk of bias and the concerns regarding applicability of all included studies. With respect to DTA, we assessed the quality of the included studies with the QUADAS-2 instrument, which we adapted for use as recommended by its authors [35]. The tool consists of four domains: (1) patient selection, (2) index test, (3) reference standard, and (4) flow and timing. Each domain contains signaling questions for risk of bias and concerns regarding applicability except the domain flow and timing. Each signaling question has three answer options: yes, no, and unclear. On the basis of the overall rating of the reviewers, an assessment can be made in each domain concerning bias and applicability. For example, one signaling question in the domain patient selection is, "Was a consecutive or random sample of patients enrolled?" All adaptations of the instrument are explained in Multimedia Appendix 2.

For prevalence studies, the research team generated a new tool based on the structure of the QUADAS-2 instrument to assess the risk of bias and concerns regarding applicability. The new tool consisted of six domains: patient selection, reviewer and algorithm selection, automatic detection method, outcomes, and flow and timing. All but outcomes and flow and timing included a section on risk of bias and concern regarding applicability, including signaling questions (Multimedia Appendix 3). The goal of this quality assessment was to verify the quality of the semi or fully automated trigger tool studies by focusing specifically on algorithm development and the basis for choosing each trigger.

Quality assessment process was conducted by one reviewer (SNM), and each research team member received at least one study to compare the results with SNM. Members did each task individually; then all results were discussed jointly by the multidisciplinary research team.

Statistical Analyses

As we anticipated a paucity of evidence on DTA data, we deemed formal meta-analyses not feasible. For the total number of AEs and each type of AE, we present percentages with 95% CIs for prevalence, PPV, and rate of false negative test results. For the prevalence studies, we provided percentages for prevalence and PPV with 95% CIs, for AE overall and per type of trigger.

Results

Search Strategy and Study Selection

After removing duplicates, 2658 citations were identified via our search strategy, the core journals, and our personal library. Of these, 11 met all selection criteria: 2 concerned diagnostic test accuracy studies and and 9 prevalence studies. A detailed view of the included studies is provided with a flow diagram in Figure 1.

Figure 1. Flow diagram of the number of studies found with the search strategy, studies screened, and reasons for exclusions. Eleven studies fulfilled all inclusion and exclusion criteria.
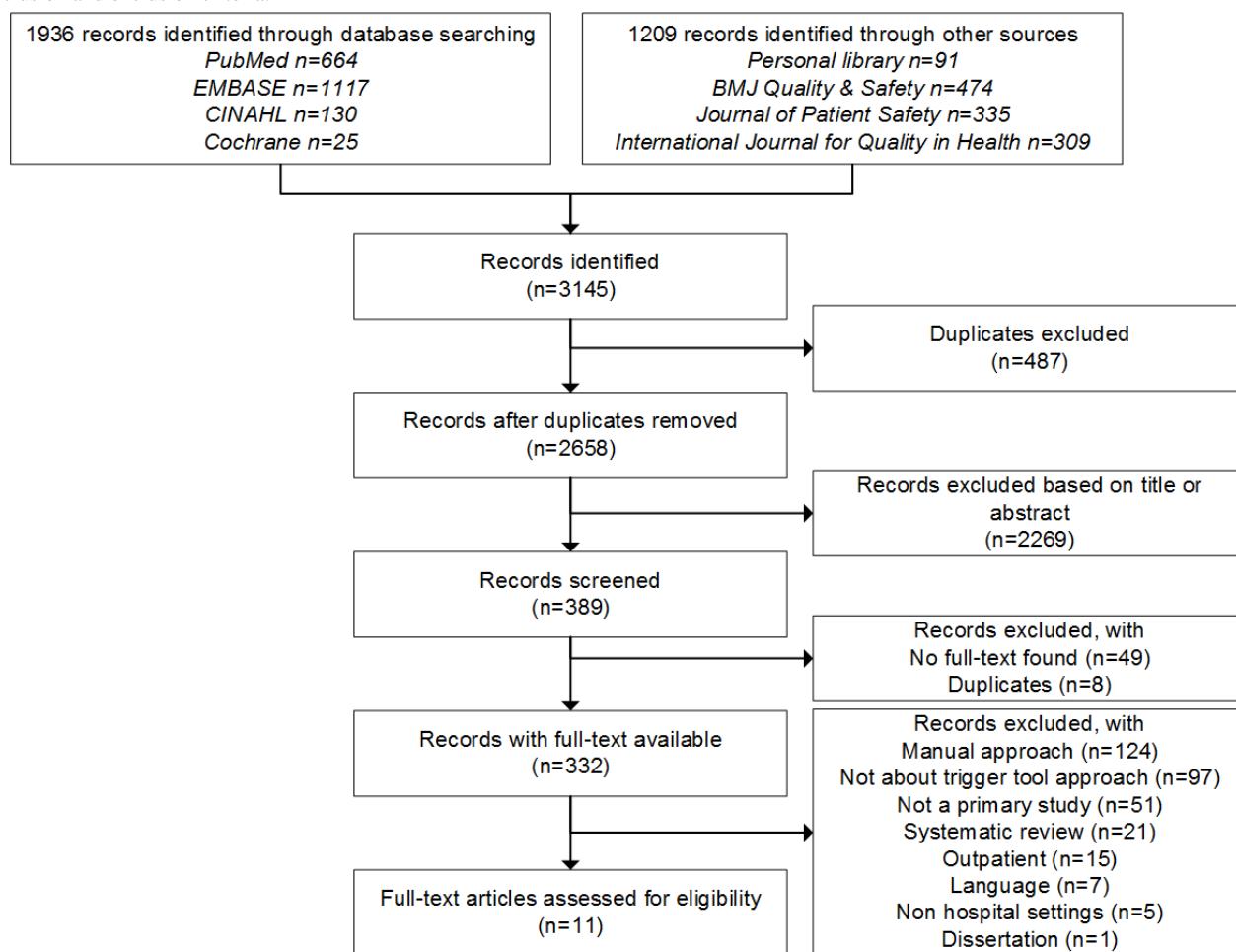
Description of Included Studies

The studies were published from 2005 to 2016. Of these, 9 were conducted in the United States [25,36-43], 1 in Denmark [44], and 1 in the United Kingdom [45]. Concerning the study populations, 7 were conducted among pediatric or neonatal patients [25,36,37,39,40,43,45], 3 among adults [38,41,42], and 1, published in abstract form only, provided no population information [44]. Two studies conducted multisite trigger tool research [25,42]. The bed capacity of the hospitals involved ranged from 26 to 1000 beds, with a median of 306. One study provided no information on bed capacity [36]. Further details concerning the design and the characteristics are summarized in Table 1.

One study explicitly followed the IHI guidelines for GTT [42], whereas the other studies used different trigger tool approaches or did not specify whether they followed the IHI guidelines. Concerning methodology, 6 studies addressed only adverse drug events (ADEs) [36,38,39,42,43,45], and only 1 used a fully automatic AE detection approach [44]. Definitions used to define and reference AEs varied [36-38,40-43,45] or were absent [25,39,44]. Seven studies used decision or agreement of the reviewer(s) to confirm an AE, without reference to the indexing method used [25,36-40,43]; and 2 studies omitted any explanation of the process [44,45]. Five studies did not consider preventability [38,39,42,44,45]; in the remainder, definitions

varied. Three studies used the definition of a preventable AE as an "event leading to action deviating from the local standard of care" [37,40,43]; one study used a scale from "1—virtually no evidence for management causation" to "6—virtually certain evidence for management causation," but without describing the applied cutoff [36]. One study used a 6-point confidence scale from "1—virtually no evidence of preventability" to "6—virtually certain evidence of preventability," with a score ≥4 denoting preventability [41]; and one study merely described that preventability was determined by the reviewers' decision [25]. Five studies assessed AEs' severity using the National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP) categories [25,37,39,40,43]; 4 studies did not state their assessment strategies [38,42,44,45]; one study categorized them according to four levels of severity (life-threatening, serious, clinically significant, or trivial) [41]; and one used NCC MERP categories together with the Common Terminology Criteria for Adverse Events version 4.03, ranging from mild (grade 1) to death (grade 5) [36]. Information concerning the data source, the triggers, and the reviewer(s) are detailed in Table 2.

Concerning the methodology, 5 studies came from the "Automated Adverse Event Detection Collaborative," which is a consortium to facilitate the use of automated triggers in pediatric hospitals [25,36,37,39,40,43]. They all used the same approach, where an EHR-driven surveillance system was used,

yet not providing detailed information on the software. Every night, trigger reports were automatically generated for laboratory results [37,40,43], medications levels in the blood [36,39], or both [25]. A clinical analyst examined every trigger by reviewing the EHR and interviewing care providers. The potential AEs were reviewed by specialists: pharmacists, physicians, endocrinologists, or anesthesiologists. The clinical analyst met with a multidisciplinary team, the "Automated Adverse Event Detection Steering Committee," monthly to present the results. Two studies used natural language processing (NLP) to extract information from EHRs [41,44]. NLP is defined as a technique extracting information from narrative text and transforming it into structured text [41,46]. NLP is able to deal

with synonyms, negations, and abbreviations used in narrative text. To build queries, SAS Text Miner tool [44] was used; however, no further details were provided. Structured Query Language [41] was used to identify AEs in the EHRs. The other studies used (1) Electronic trigger-detection messages that were sent automatically to two pharmacist reviewers [42], (2) Computerized trigger alert system that sent an alert to the project manager within 24 hours when conditions defined by the trigger algorithm were detected [38], and (3) electronic algorithms where triggers were identified automatically [45]. No further details concerning the development, the algorithms, or the tools used were given. Description of the methods are explained in Table 3.

Table 1. Design and characteristics of the sample and population of the included studies.

Study	Population	Time frame (months)	Sample size	Setting
Diagnostic test accuracy studies				
Gerdes and Hardahl, 2013 [44]	Not stated	26	500	Not stated
O'Leary et al, 2013 [41]	Adults, exclusion of patients admitted under observation status and cared for by either of the two medical record abstractors	12	250	General internal medicine
Prevalence studies				
Call et al, 2014 [36]	Children	48	390	Oncology Hematology
Dickermann et al, 2011 [37]	Children, exclusion weekend days for 5 months because of resource limitations	12	13,526	General internal medicine Surgical care Emergency department Intensive care unit (ICU) Psychiatric unit
Lim et al, 2016 [42]	Adults	3x1 month	Not stated	Not stated
Moore et al, 2009 [38]	Adults	5	456	General internal medicine Surgical care Obstetrics or gynecology
Muething et al, 2010 [39]	Children	21 for one trigger and 16 for another one	Not stated	Not stated
Nwulu et al, 2013 [45]	Not stated	12	54,244	Not stated
Patregnani et al, 2015 [43]	Children	52 for one trigger; 40 and 60 for the others	Not stated	Pediatric ICU Neonatal ICU Cardiac ICU Medical and surgical acute care areas
Shea et al, 2013 [40]	Children	36	6,872	Pediatric ICU Cardiac ICU
Stockwell et al, 2013 [25]	Children, exclusion of emergency departments and ambulatory clinics	36 for hospital 1 and 51 for hospital 2	Not stated	General internal medicine Surgical care Psychiatric unit Neonatal Cardiac ICU Pediatric ICU

Table 2. Data sources, triggers, and reviewers of included studies.

Study	Data source	Triggers	Reviewer(s)
Diagnostic test accuracy studies			
Gerdes and Hardahl, 2013 [44]	Unstructured and semistructured narrative texts in EHRs[a]	"Models," not defined, identifying the most common triggers and/or AEs[b]	Not stated
O'Leary et al, 2013 [41]	Enterprise Data Warehouse: EHRs or CPOES[c]; hospital and physician billing systems; incident reporting system; and admission or discharge or transfer with nightly updates from activities occurring in the preceding 24 h	Locally developed based on screening criteria from the Harvard Medical Practice Study and the IHI[d] GTT[e]	Experienced hospitalists and physician-researcher (prior experience with the research method)
Prevalence studies			
Call et al, 2014 [36]	EHR: laboratory, pharmacy, electronic medication administration record, CPOE, and documentation functions	Wide use in similar population and high likelihood to detect adverse drug events	Pharmacist and physician
Dickermann et al, 2011 [37]	EHRs	Increasing use in hospitals' protocols	CA[f] trained
Lim et al, 2016 [42]	EHR supports all inpatient and ambulatory care clinical and documentation activities	Review of literature and detectable in EHRs with reasonable PPV[g]	Pharmacists, medication safety pharmacist, and physician
Moore et al, 2009 [38]	CPOE with decision support, EHR, clinical event monitors	Most common inpatient adverse drug events	Study investigators
Muething et al, 2010 [39]	Clinical information system: computerized clinical order entry, clinical documentation, electronic medication administration record, data storage repository, and advanced clinical decision support	AEs steering committee	Endocrinologist, anesthesiologist, and frontline staff
Nwulu et al, 2013 [45]	Locally developed electronic health and prescription computer system (laboratory results and prescribing, except some chemotherapy regimens) has built-in checks to identify potential prescribing errors (flagged through warnings and alerts)	Test the usefulness of two medication module triggers from the GTT proposed by IHI	Not stated
Patregnani et al, 2015 [43]	EHRs	Clinical evidences	CA trained in the AE trigger process
Shea et al, 2013 [40]	EHRs and Laboratory Information System	Clinical evidences and risks of deaths	CA trained in the AE trigger process
Stockwell et al, 2013 [25]	EHRs	Multidisciplinary review process using several review criteria	CA

[a]EHRs: electronic health records.

[b]AE: adverse event.

[c]CPOES: computerized provider order entry system.

[d]IHI: Institute for Healthcare Improvement.

[e]GTT: Global Trigger Tool.

[f]CA: clinical analyst.

[g]PPV: positive predictive value.

XSL•FO
RenderX

Table 3. Overview of the automated trigger tool methodology.

Study	Description of the method
Diagnostic test accuracy studies	
Gerdes and Hardahl, 2013 [44]	(1) Extraction and preparation of all texts from the EHRs[a]; (2) Use of SAS Text Miner and the SAS Enterprise Content Categorization software to build query models (natural language processing algorithms)
O'Leary et al, 201 3[41]	(1) Leveraging of various information systems in the EDW[b]; (2) Write Structured Query Language queries to mimic work of a reviewer to identify potential AEs[c] based on trigger tool; (3) Two reviewers review the positive EDW screens; (4) Another reviewer reviews narrative summaries and determines presence of AEs
Prevalence studies	
Call et al, 2014 [36]	(1) Software program conducts an extensive search of patient records for any type of order containing specific medications and laboratory values; (2) Information generated into a report with patient-specific information; (3) Review by two reviewers
Dickermann et al, 2011 [37]	(1) Trigger reports automatically generated on a daily basis from the EHR by querying the Sunquest Laboratory Information System for laboratory results; (2) Reviewer examined every trigger by reading the EHRs and interviewing care providers
Lim et al, 2016 [42]	(1) Administration of a trigger drug to a patient automatically sent an electronic trigger-detection message to two reviewers; (2) Trigger-detection messages were evaluated immediately after or during the day by both reviewers (consensus if disagreement); (3) Event reviewed by a medication safety pharmacist and then by a physician for validation.
Moore et al, 2009 [38]	(1) The laboratory results and administered medications of each adult hospital patient were continuously monitored by the computerized trigger alert system; (2) If any of the conditions defined was satisfied (trigger algorithm), an alert was triggered, and data were collected by study investigators on the patient for a period of 72 hours after the initial trigger firing to determine whether an adverse drug event had occurred.
Muething et al, 2010 [39]	(1) Combination of trigger tool approach with the clinical information system; (2) Every evening, automatic detection of triggers are sent to the project manager (detection of event within 24 h); (3) Summary of the incident automatically generated and sent to the appropriate staff on the unit involved
Nwulu et al, 2013 [45]	(1) The triggers identified electronically were linked to the electronic prescription records; (2) Two or more positive triggers generated for the same patient, within a 24- or 72-hour interval (trigger-dependent) were treated as one trigger; (3) The paper-based case notes were reviewed to identify any documentation of interest
Patregnani et al, 2015 [43]	(1) Generation of a trigger report by querying the Laboratory Information System (2) Reviewer investigated the event by reading the patient's EHRs and interviewing the clinical care team
Shea et al, 2013 [40]	(1) Generation of a trigger report by querying the Laboratory Information System (2) Reviewer investigated the event by reading the patient's EHRs and interviewing the clinical care team
Stockwell et al, 2013 [25]	(1) Automated trigger reports are generated from hospital information systems on a nightly basis; (2) Each trigger report is examined by a reviewer and interviews conducted with care providers.

[a]EHRs: electronic health records.

[b]EDW: Enterprise Data Warehouse.

[c]AEs: adverse events.

Two studies compared results from automated trigger tool methodology with those obtained via the manual trigger tool method [41,44]. Six studies compared results with voluntary incident reports but did not cross-verify their results with those obtained via manually operated trigger-based tools; therefore, we considered these prevalence studies [25,36,37,39,43,45]. The remaining 3 studies did not compare their results with those obtained via any other method [38,40,42].

Quality Assessment

Figure 2 shows detailed results of the two quality assessments.

Diagnostic Test Accuracy Studies

In one of the 2 studies, the assessment of all domains was hampered by poor reporting, and we were unable to judge the risk of bias and concerns regarding applicability [44]. In the other, we judged the concern regarding applicability of "patient

selection" as low [41]. We judged a high risk of bias in the "flow and timing" domain and low bias risk and applicability concerns for the "index test" and "reference standard" domains [41].

Prevalence Studies

A total of 9 studies were selected as prevalence studies via our self-developed quality assessment tool (Multimedia Appendix 3). For "patient selection," the bias risk was deemed low in seven studies [25,36-38,40,42,43] and applicability concerns were also low in eight of them [25,36-40,42,43]. We had no concerns regarding applicability of the automatic detection methods. As Figure 2 shows, for the other domain, we judged low bias risk and concerns regarding applicability in a minority of studies. In 5 studies, we judged a high risk of bias in the "outcomes" domain, mainly because their AE definitions did not reference those of the IHI or Food and Drug Administration,

and they lacked clearly stated prevalence outcomes [37,38,42,43,45]. We judged high bias risks in 3 studies regarding the "automated detection method" domain [25,37,45]. In another, we judged the bias risk as high in the "reviewer or algorithm selection" domain [45] because the triggers were not consistently used or developed, and the reviewer lacked the required profile (eg, experience and training).

Estimates of Diagnostic Test Accuracy, Prevalence, and Reliability

Estimates of DTA can be found in Table 4, with additional information concerning prevalence, preventability, and false negative rates for AE categories. No information concerning 2x2 tables were available for all the triggers; thus, the decision was made to use the AE categories.

Figure 2. Risk of bias and concerns regarding applicability assessments for diagnostic test accuracy studies (upper panel) and prevalence studies (lower panel). Judgments are expressed as "low," "high," or "unclear" risk or concern for each of the domains (ie, "patient selection," "index test"). The percentages refer to the percentage of studies meeting the judgment low, high, or unclear risk of bias or concerns regarding applicability in each of the domains. Quality Assessment tool for Diagnostic Accuracy Studies-2 (QUADAS-2) was used for the two diagnostic test accuracy studies and an in-house developed tool was used to assess the 9 prevalence studies.
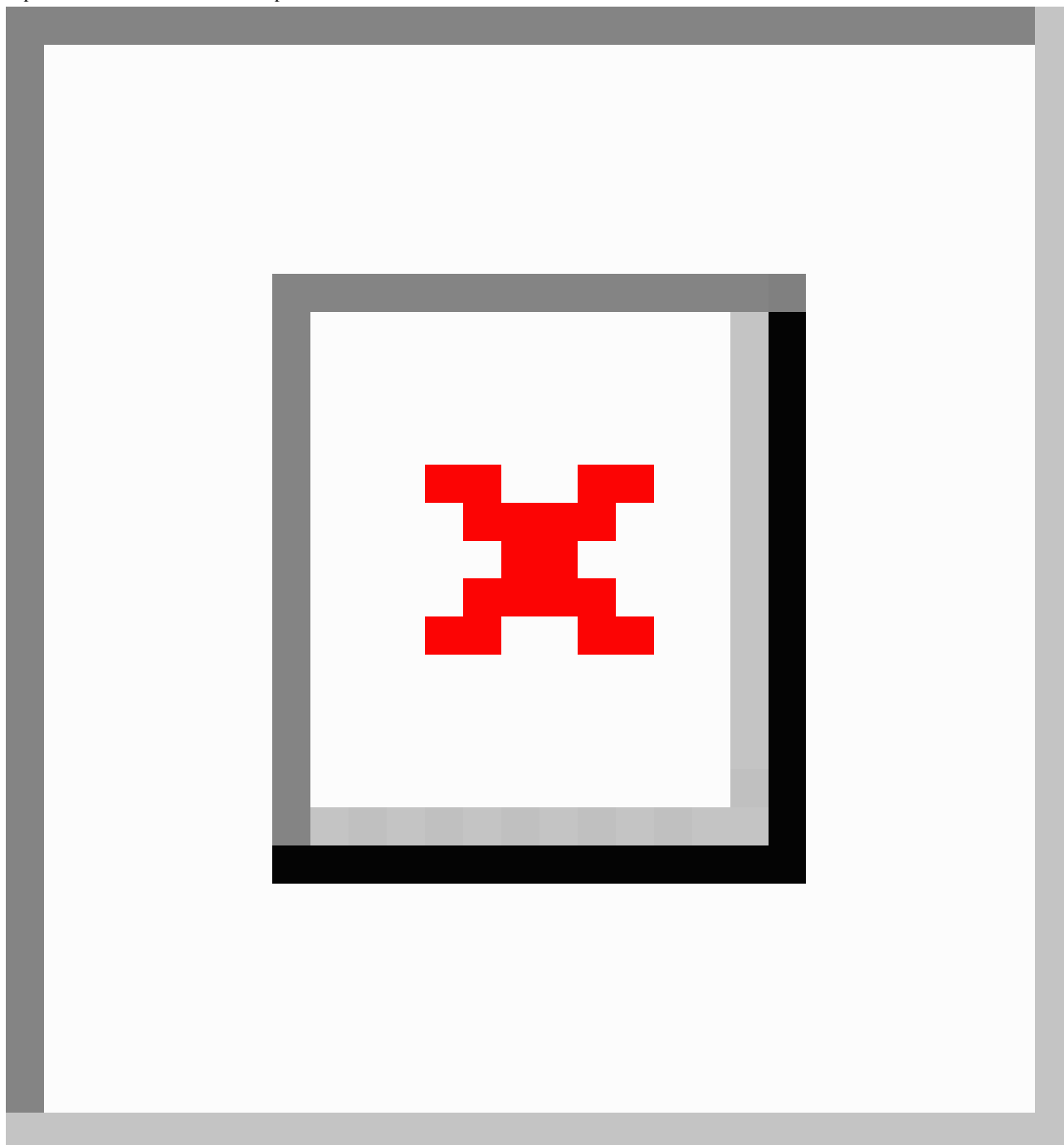
Table 4. The table displays the estimates of diagnostic test accuracy in 2 studies comparing automated trigger-based tools with a manual trigger-based tool as reference standard.

Study	Type of adverse events	2x2 table for adverse events (True positive / false positive / false negative / true negative)	Prevalence[a], % (95% CI)	Positive predictive value[b], % (95% CI)	False negative rate[c] (%)
Gerdes and Hardahl, 2013 [44]	Pressure ulcer	28 / 22 / 12 / 436	5.6 (3.6-7.6)	56 (42.2-69.8)	30
O'Leary et al, 2013 [41]	Adverse drug event	24 / 22 / 20 / N/A[d]	9.6 (5.9-13.3)	52.2 (37.7-66.6)	45.5
	Hospital acquired infection	7 / 11 / 4 / N/A	2.8 (0.8-4.9)	38.9 (16.4-61.4)	36.4
	Operative or procedural injury	5 / 4 / 4 / N/A	2 (0.3-3.7)	55.6 (23.1-88)	44.4
	Manifestation of poor glycemic control	3 / 2 / 5 / N/A	1.2 (–0.2 to 2.6)	60 (17.1-102.9)	62.5
	Pressure ulcer	0 / 8 / 2 / N/A	0 (0-0)	0 (0-0)	100
	Venous thromboembolism	5 / 1 / 0 / N/A	2 (0.3-3.7)	83.3 (53.5-113.2)	0
	Acute renal failure	2 / 1 / 0 / N/A	0.8 (–0.3 to 1.9)	66.7 (13.3-120)	0
	Delirium	0 / 0 / 0 / N/A	0 (0-0)	0 (0-0)	0
	Fall	0 / 0 / 0 / N/A	0 (0-0)	0 (0-0)	0
	Other	0 / 2 / 5 / N/A	0 (0-0)	0 (0-0)	100

[a]Prevalence is calculated by true positive/total number of patients.

[b]Calculated as triggers corresponding to an adverse event out of all triggers=true positive/(true positive+false positive).

[c]Calculated as false negative/(false negative+true positive).

[d]N/A: not applicable.

Across all the 11 studies, AE prevalence ranged from 0% to 17.9%, with a median of 0.8%. The PPV of all triggers to detect AEs ranged from 0% to 100% across studies, with a median of 40%. Some triggers are used in different studies and have different PPV values: (1) in 6 studies, hypoglycemia [25,37-41] had a PPV ranging from 15.8% to 60%; (2) in 5 studies, naloxone [25,36,41,42,45] had a PPV ranging from 20% to 91%; (3) in 4 studies, flumazenil [25,36,41,42] had a PPV ranging from 38.9% to 83.3%; and (4) in 4 studies, protamine [25,36,42,43] had a PPV ranging from 0% to 60%. We were unable to determine the AE prevalence, PPV, preventability, and severity in 40.4%, 10.5%, 71.1%, and 68.4% of the studies, respectively. These studies did not report the overall number of records analyzed, triggers, or AEs; or the studies did not conducted the analysis concerned. Detailed results for each trigger with prevalence, preventability, severity, and PPV are presented in Figures 3 and 4. Only 1 study supplied information

on interrater reliability, reporting a kappa value of .52 to .78 [41].

Challenges and Author Proposed Solutions

The challenges reported by authors concerned the relative large number of false alarms, the dependence of PPV on AE prevalence, and incomplete patient records leading to missed events. To reduce the fraction of false alarms, several authors suggested to use a threshold value for the triggers used based on patient characteristics [36,38,40,43]. For example, specific disease states of patients must have triggers with different threshold as the consequences might be stronger because of their disease. Authors from one study suggested to measure sensitivity and specificity instead of PPV, as latter is known to be largely affected by AE prevalence [38]. Another author group suggested to either improve completeness of patient information documentation or to combine different data sources to increase the chance to detect AEs as each data source contains different information type [42].

Figure 3. Prevalence, preventability, severity, and positive predictive value (PPV) for all the 11 studies. The figure begins with the results of all the triggers or adverse events (AEs) combined, then for each group of trigger order from the most studied to the least studied (part 1). Severity levels based on the National Coordinating Council for Medication Error Reporting and Prevention: D=an error that reached the patient and required monitoring or intervention to confirm that it resulted in no harm to the patient; E=temporary harm to the patient and required intervention; F=temporary harm to the patient and required initial or prolonged hospitalization; G=permanent patient harm; H=intervention required to sustain life; and I=patient death. H1: hospital 1; H2: hospital 2.

Authors and year	Prevalence (%) (95% CI)	Preventability (%)	Severity (%) D	E	F	G–I	Positive predictive value (PPV) (%) (95% CI)
All triggers							
Call et al 2014	29.2 (24.8–34)	NA	NA	NA	NA	NA	16.1 (13.5– 19.1)
Dickermann et al 2011	1.5 (1.3–1.7)	48	0	100	0	0	15.8 (13.8– 17.9)
Gerdes and Hardahl 2013	5.6 (3.8–8)	NA	NA	NA	NA	NA	56.0 (41.3– 70.0)
Lim et al 2016	NA (NA–NA)	NA	NA	NA	NA	NA	31.0 (21.5– 41.9)
Moore et al 2009	NA (NA–NA)	NA	NA	NA	NA	NA	NA (NA– NA)
Muething et al 2010	NA (NA–NA)	NA	0	92.2	6.2	1.6	58.7 (48.9– 68.1)
Nwulu et al 2013	0.1 (0.1–0.2)	NA	NA	NA	NA	NA	72.1 (62.5– 80.5)
O'Leary et al 2013	45.2 (38.9–51.6)	NA	NA	NA	NA	NA	30.9 (26.2– 35.9)
Patregnani et al 2015	NA (NA–NA)	NA	NA	NA	NA	NA	44.9 (42.5– 47.4)
Shea et al 2013	2.9 (2.5–3.3)	32	NA	NA	NA	NA	NA (NA– NA)
Stockwell et al (H1) 2013	NA (NA–NA)	27.3	25.4	66.5	8.1	0	23.5 (22.6– 24.5)
Stockwell et al (H2) 2013	NA (NA–NA)	57.3	0	94.2	0.3	5.5	86.2 (84.4– 87.8)
Hypoglycemia							
Dickermann et al 2011	1.5 (1.3–1.7)	48	0	100	0	0	15.8 (13.8– 17.9)
Moore et al 2009	NA (NA–NA)	NA	NA	NA	NA	NA	31.2 (18.2– 46.6)
Muething et al 2010	NA (NA–NA)	NA	0	93.3	6.7	0	60.0 (45.2– 73.6)
O'Leary et al 2013	1.2 (0.2–3.5)	NA	NA	NA	NA	NA	42.9 (9.9– 81.6)
Shea et al 2013	2.9 (2.5–3.3)	32	NA	NA	NA	NA	NA (NA– NA)
Stockwell et al (H1) 2013	NA (NA–NA)	26.9	0.4	99.6	0	0	22.0 (20.7– 23.4)
Naloxone							
Call et al 2014	8.7 (6.1–12)	NA	NA	NA	NA	NA	33.3 (24.3– 43.4)
Lim et al 2016	NA (NA–NA)	NA	NA	NA	NA	NA	41.9 (27.0– 57.9)
O'Leary et al 2013	0.4 (0–2.2)	NA	NA	NA	NA	NA	20.0 (0.5– 71.6)
Nwulu et al 2013	0.1 (0.1–0.1)	NA	NA	NA	NA	NA	91.0 (81.5– 96.6)
Stockwell et al (H1) 2013	NA (NA–NA)	54.2	0	86.4	13.6	0	81.9 (71.1– 90.0)
Stockwell et al (H2) 2013	NA (NA–NA)	81.6	0	0	0	100	46.9 (39.0– 54.9)
Flumazenil							
Call et al 2014	1.8 (0.7–3.7)	NA	NA	NA	NA	NA	38.9 (17.3– 64.3)
Lim et al 2016	NA (NA–NA)	NA	NA	NA	NA	NA	83.3 (35.9– 99.6)
O'Leary et al 2013	0 (0–1.5)	NA	NA	NA	NA	NA	NA (NA– NA)
Stockwell et al (H1) 2013	NA (NA–NA)	50	0	100	0	0	66.7 (29.9– 92.5)
Stockwell et al (H2) 2013	NA (NA–NA)	50	0	100	0	0	60.0 (26.2– 87.8)
Protamine							
Call et al 2014	0 (0–0.9)	NA	NA	NA	NA	NA	0.0 (0.0– 14.8)
Lim et al 2016	NA (NA–NA)	NA	NA	NA	NA	NA	60.0 (14.7– 94.7)
Patregnani et al 2015	NA (NA–NA)	NA	8.3	58.4	8.3	25	17.1 (9.2– 28.0)
Stockwell et al (H1) 2013	NA (NA–NA)	14.3	0	100	0	0	11.5 (4.7– 22.2)
Vitamin K							
Call et al 2014	0 (0–0.9)	NA	NA	NA	NA	NA	0.0 (0.0– 1.2)
Lim et al 2016	NA (NA–NA)	NA	NA	NA	NA	NA	5.9 (0.7– 19.7)
O'Leary et al 2013	0.8 (0.1–2.9)	NA	NA	NA	NA	NA	100.0 (15.8–100.0)
Stockwell et al (H2) 2013	NA (NA–NA)	0	0	100	0	0	50.0 (6.8– 93.2)
International normalized ratio (INR)							
Nwulu et al 2013 (>6)	0.1 (0–0)	NA	NA	NA	NA	NA	37.8 (22.5– 55.2)
O'Leary et al 2013 (>6)	0.8 (0.1–2.9)	NA	NA	NA	NA	NA	100.0 (15.8–100.0)
Patregnani et al 2015 (>4)	NA (NA–NA)	NA	NA	NA	NA	NA	17.9 (14.3– 22.1)
Stockwell et al (H1) 2013 (>4)	NA (NA–NA)	32.2	0	100	0	0	10.8 (7.4– 14.9)
Hyaluronidase							
Call et al 2014	0.8 (0.2–2.2)	NA	NA	NA	NA	NA	60.0 (14.7– 94.7)
Stockwell et al (H1) 2013	NA (NA–NA)	92.6	0	92.6	7.4	0	96.4 (81.7– 99.9)
Stockwell et al (H2) 2013	NA (NA–NA)	99.6	0	98.7	0.4	0.9	92.4 (88.4– 95.4)
Activated partial thromboplastin time (aPTT)							
O'Leary et al 2013 (>125s)	0.4 (0–2.2)	NA	NA	NA	NA	NA	7.7 (0.2– 36.0)
Patregnani et al 2015 (>100s)	NA (NA–NA)	7.7	98.1	1.6	0.3	0	52.4 (49.6– 55.3)
Stockwell et al (H1) 2013 (>100s)	NA (NA–NA)	29	97.4	2.6	0	0	47.9 (44.4– 51.4)
Hyperkalemia							
Moore et al 2009	NA (NA–NA)	NA	NA	NA	NA	NA	31.1 (25.1– 37.8)
O'Leary et al 2013	0.4 (0–2.2)	NA	NA	NA	NA	NA	100.0 (2.5–100.0)
Stockwell et al (H1) 2013	NA (NA–NA)	47.6	5	95	0	0	2.3 (1.4– 3.4)
Digoxine immune fab							
O'Leary et al 2013	0 (0–1.5)	NA	NA	NA	NA	NA	NA (NA– NA)
Stockwell et al (H1) 2013	NA (NA–NA)	0	0	100	0	0	50.0 (1.3– 98.7)
Stockwell et al (H2) 2013	NA (NA–NA)	0	0	0	0	0	0.0 (0.0– 97.5)

0   20   40   60   80   100

XSL•FO
RenderX

Figure 4. Prevalence, preventability, severity, and positive predictive value (PPV) for all the 11 studies. The figure begins with the results of all the triggers or adverse events (AEs) combined, then for each group of trigger order from the most studied to the least studied (part 2). Severity levels based on the National Coordinating Council for Medication Error Reporting and Prevention: D=an error that reached the patient and required monitoring or intervention to confirm that it resulted in no harm to the patient; E=temporary harm to the patient and required intervention; F=temporary harm to the patient and required initial or prolonged hospitalization; G=permanent patient harm; H=intervention required to sustain life; I=patient death. H1: hospital 1; H2: hospital 2; VT: venous thromboembolism; IR: incident report.
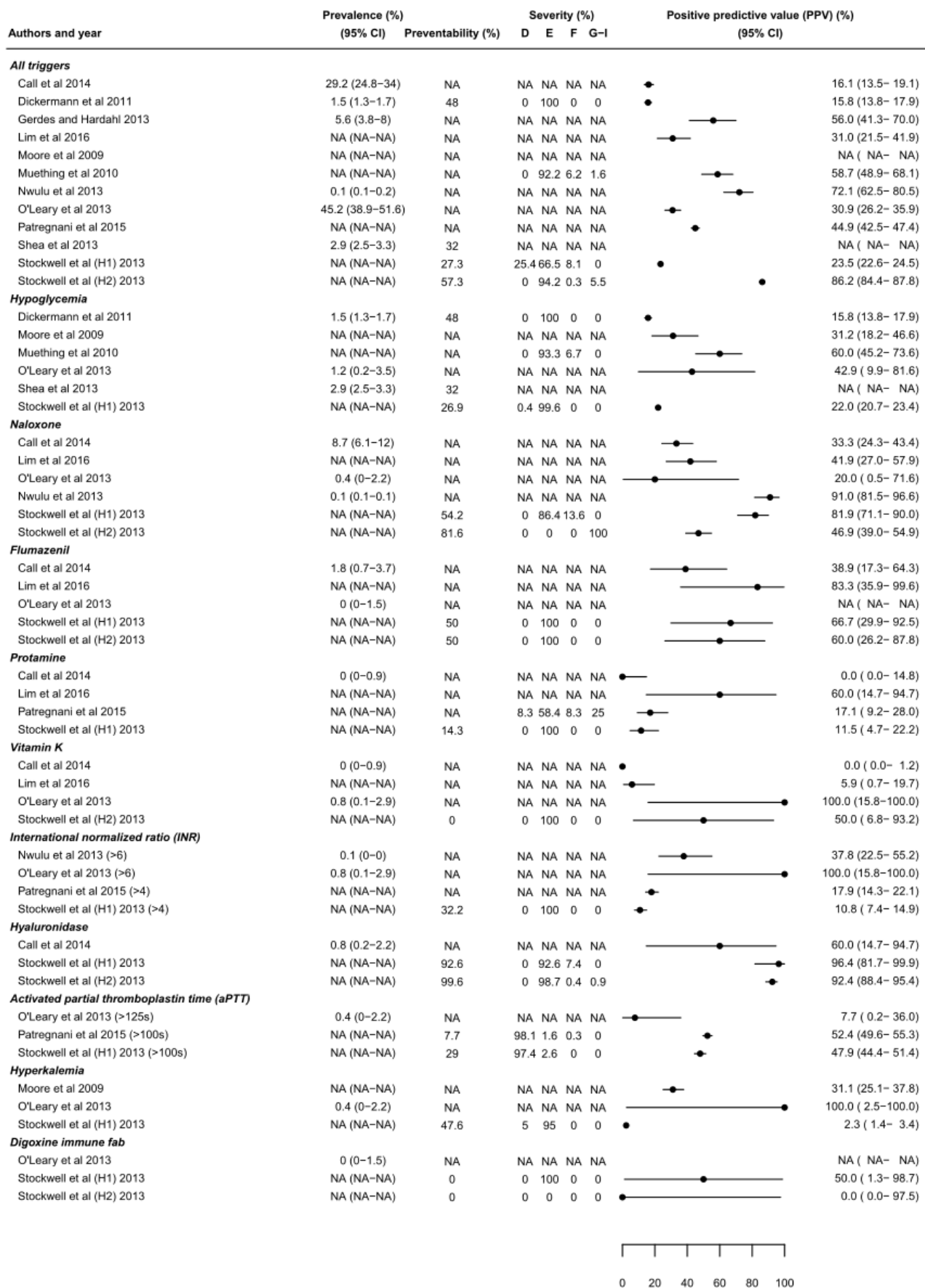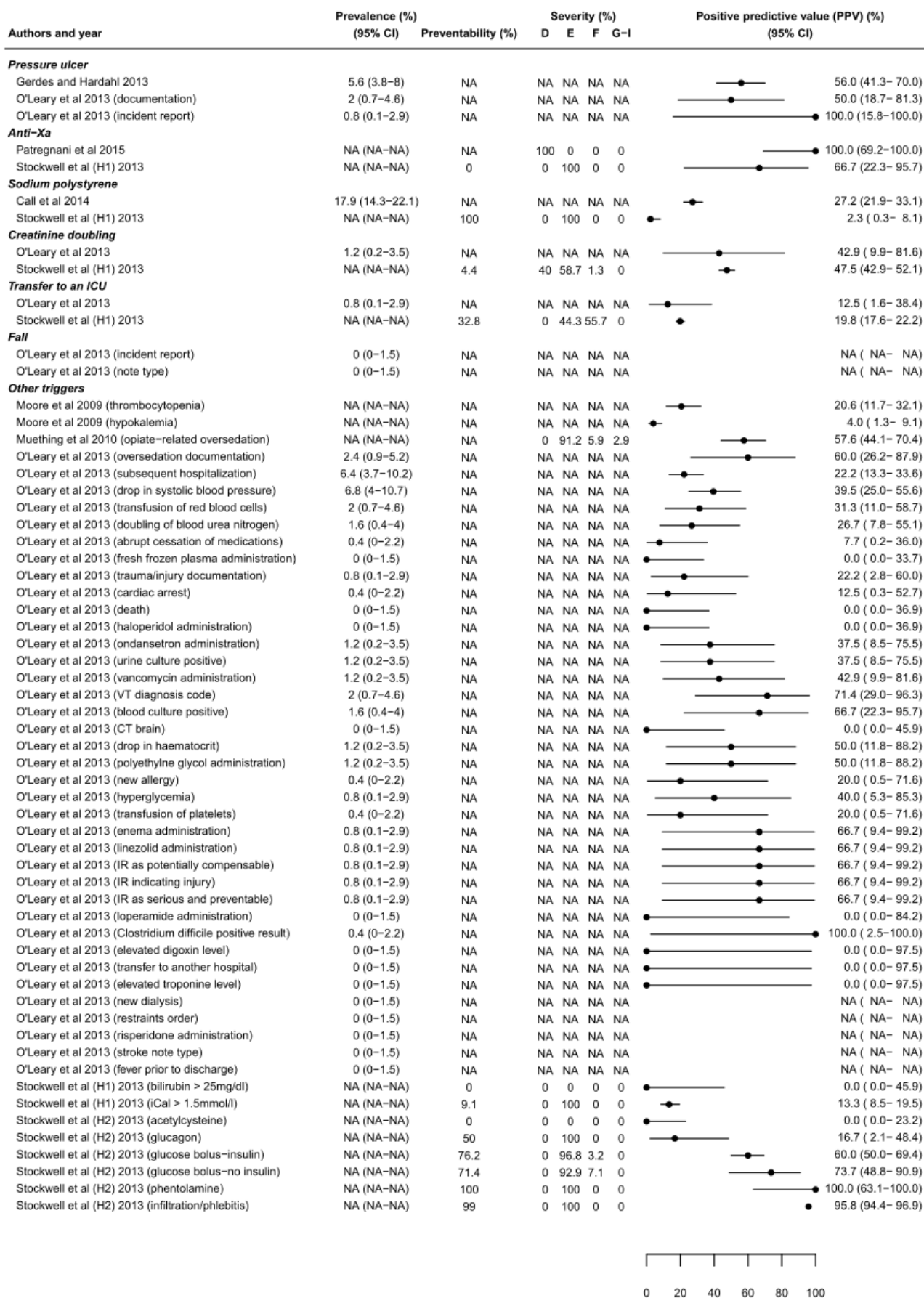
XSL•FO

RenderX

Authors and year	Prevalence (%) (95% CI)	Preventability (%)	Severity (%)				Positive predictive value (PPV) (%) (95% CI)
			D	E	F	G–I	
Pressure ulcer							
Gerdes and Hardahl 2013	5.6 (3.8–8)	NA	NA	NA	NA	NA	56.0 (41.3– 70.0)
O'Leary et al 2013 (documentation)	2 (0.7–4.6)	NA	NA	NA	NA	NA	50.0 (18.7– 81.3)
O'Leary et al 2013 (incident report)	0.8 (0.1–2.9)	NA	NA	NA	NA	NA	100.0 (15.8–100.0)
Anti–Xa							
Patregnani et al 2015	NA (NA–NA)	NA	100	0	0	0	100.0 (69.2–100.0)
Stockwell et al (H1) 2013	NA (NA–NA)	0	0	100	0	0	66.7 (22.3– 95.7)
Sodium polystyrene							
Call et al 2014	17.9 (14.3–22.1)	NA	NA	NA	NA	NA	27.2 (21.9– 33.1)
Stockwell et al (H1) 2013	NA (NA–NA)	100	0	100	0	0	2.3 (0.3– 8.1)
Creatinine doubling							
O'Leary et al 2013	1.2 (0.2–3.5)	NA	NA	NA	NA	NA	42.9 (9.9– 81.6)
Stockwell et al (H1) 2013	NA (NA–NA)	4.4	40	58.7	1.3	0	47.5 (42.9– 52.1)
Transfer to an ICU							
O'Leary et al 2013	0.8 (0.1–2.9)	NA	NA	NA	NA	NA	12.5 (1.6– 38.4)
Stockwell et al (H1) 2013	NA (NA–NA)	32.8	0	44.3	55.7	0	19.8 (17.6– 22.2)
Fall							
O'Leary et al 2013 (incident report)	0 (0–1.5)	NA	NA	NA	NA	NA	NA (NA– NA)
O'Leary et al 2013 (note type)	0 (0–1.5)	NA	NA	NA	NA	NA	NA (NA– NA)
Other triggers							
Moore et al 2009 (thrombocytopenia)	NA (NA–NA)	NA	NA	NA	NA	NA	20.6 (11.7– 32.1)
Moore et al 2009 (hypokalemia)	NA (NA–NA)	NA	NA	NA	NA	NA	4.0 (1.3– 9.1)
Muething et al 2010 (opiate–related oversedation)	NA (NA–NA)	NA	0	91.2	5.9	2.9	57.6 (44.1– 70.4)
O'Leary et al 2013 (oversedation documentation)	2.4 (0.9–5.2)	NA	NA	NA	NA	NA	60.0 (26.2– 87.9)
O'Leary et al 2013 (subsequent hospitalization)	6.4 (3.7–10.2)	NA	NA	NA	NA	NA	22.2 (13.3– 33.6)
O'Leary et al 2013 (drop in systolic blood pressure)	6.8 (4–10.7)	NA	NA	NA	NA	NA	39.5 (25.0– 55.6)
O'Leary et al 2013 (transfusion of red blood cells)	2 (0.7–4.6)	NA	NA	NA	NA	NA	31.3 (11.0– 58.7)
O'Leary et al 2013 (doubling of blood urea nitrogen)	1.6 (0.4–4)	NA	NA	NA	NA	NA	26.7 (7.8– 55.1)
O'Leary et al 2013 (abrupt cessation of medications)	0.4 (0–2.2)	NA	NA	NA	NA	NA	7.7 (0.2– 36.0)
O'Leary et al 2013 (fresh frozen plasma administration)	0 (0–1.5)	NA	NA	NA	NA	NA	0.0 (0.0– 33.7)
O'Leary et al 2013 (trauma/injury documentation)	0.8 (0.1–2.9)	NA	NA	NA	NA	NA	22.2 (2.8– 60.0)
O'Leary et al 2013 (cardiac arrest)	0.4 (0–2.2)	NA	NA	NA	NA	NA	12.5 (0.3– 52.7)
O'Leary et al 2013 (death)	0 (0–1.5)	NA	NA	NA	NA	NA	0.0 (0.0– 36.9)
O'Leary et al 2013 (haloperidol administration)	0 (0–1.5)	NA	NA	NA	NA	NA	0.0 (0.0– 36.9)
O'Leary et al 2013 (ondansetron administration)	1.2 (0.2–3.5)	NA	NA	NA	NA	NA	37.5 (8.5– 75.5)
O'Leary et al 2013 (urine culture positive)	1.2 (0.2–3.5)	NA	NA	NA	NA	NA	37.5 (8.5– 75.5)
O'Leary et al 2013 (vancomycin administration)	1.2 (0.2–3.5)	NA	NA	NA	NA	NA	42.9 (9.9– 81.6)
O'Leary et al 2013 (VT diagnosis code)	2 (0.7–4.6)	NA	NA	NA	NA	NA	71.4 (29.0– 96.3)
O'Leary et al 2013 (blood culture positive)	1.6 (0.4–4)	NA	NA	NA	NA	NA	66.7 (22.3– 95.7)
O'Leary et al 2013 (CT brain)	0 (0–1.5)	NA	NA	NA	NA	NA	0.0 (0.0– 45.9)
O'Leary et al 2013 (drop in haematocrit)	1.2 (0.2–3.5)	NA	NA	NA	NA	NA	50.0 (11.8– 88.2)
O'Leary et al 2013 (polyethylne glycol administration)	1.2 (0.2–3.5)	NA	NA	NA	NA	NA	50.0 (11.8– 88.2)
O'Leary et al 2013 (new allergy)	0.4 (0–2.2)	NA	NA	NA	NA	NA	20.0 (0.5– 71.6)
O'Leary et al 2013 (hyperglycemia)	0.8 (0.1–2.9)	NA	NA	NA	NA	NA	40.0 (5.3– 85.3)
O'Leary et al 2013 (transfusion of platelets)	0.4 (0–2.2)	NA	NA	NA	NA	NA	20.0 (0.5– 71.6)
O'Leary et al 2013 (enema administration)	0.8 (0.1–2.9)	NA	NA	NA	NA	NA	66.7 (9.4– 99.2)
O'Leary et al 2013 (linezolid administration)	0.8 (0.1–2.9)	NA	NA	NA	NA	NA	66.7 (9.4– 99.2)
O'Leary et al 2013 (IR as potentially compensable)	0.8 (0.1–2.9)	NA	NA	NA	NA	NA	66.7 (9.4– 99.2)
O'Leary et al 2013 (IR indicating injury)	0.8 (0.1–2.9)	NA	NA	NA	NA	NA	66.7 (9.4– 99.2)
O'Leary et al 2013 (IR as serious and preventable)	0.8 (0.1–2.9)	NA	NA	NA	NA	NA	66.7 (9.4– 99.2)
O'Leary et al 2013 (loperamide administration)	0 (0–1.5)	NA	NA	NA	NA	NA	0.0 (0.0– 84.2)
O'Leary et al 2013 (Clostridium difficile positive result)	0.4 (0–2.2)	NA	NA	NA	NA	NA	100.0 (2.5–100.0)
O'Leary et al 2013 (elevated digoxin level)	0 (0–1.5)	NA	NA	NA	NA	NA	0.0 (0.0– 97.5)
O'Leary et al 2013 (transfer to another hospital)	0 (0–1.5)	NA	NA	NA	NA	NA	0.0 (0.0– 97.5)
O'Leary et al 2013 (elevated troponine level)	0 (0–1.5)	NA	NA	NA	NA	NA	0.0 (0.0– 97.5)
O'Leary et al 2013 (new dialysis)	0 (0–1.5)	NA	NA	NA	NA	NA	NA (NA– NA)
O'Leary et al 2013 (restraints order)	0 (0–1.5)	NA	NA	NA	NA	NA	NA (NA– NA)
O'Leary et al 2013 (risperidone administration)	0 (0–1.5)	NA	NA	NA	NA	NA	NA (NA– NA)
O'Leary et al 2013 (stroke note type)	0 (0–1.5)	NA	NA	NA	NA	NA	NA (NA– NA)
O'Leary et al 2013 (fever prior to discharge)	0 (0–1.5)	NA	NA	NA	NA	NA	NA (NA– NA)
Stockwell et al (H1) 2013 (bilirubin > 25mg/dl)	NA (NA–NA)	0	0	0	0	0	0.0 (0.0– 45.9)
Stockwell et al (H1) 2013 (iCal > 1.5mmol/l)	NA (NA–NA)	9.1	0	100	0	0	13.3 (8.5– 19.5)
Stockwell et al (H2) 2013 (acetylcysteine)	NA (NA–NA)	0	0	0	0	0	0.0 (0.0– 23.2)
Stockwell et al (H2) 2013 (glucagon)	NA (NA–NA)	50	0	100	0	0	16.7 (2.1– 48.4)
Stockwell et al (H2) 2013 (glucose bolus–insulin)	NA (NA–NA)	76.2	0	96.8	3.2	0	60.0 (50.0– 69.4)
Stockwell et al (H2) 2013 (glucose bolus–no insulin)	NA (NA–NA)	71.4	0	92.9	7.1	0	73.7 (48.8– 90.9)
Stockwell et al (H2) 2013 (phentolamine)	NA (NA–NA)	100	0	100	0	0	100.0 (63.1–100.0)
Stockwell et al (H2) 2013 (infiltration/phlebitis)	NA (NA–NA)	99	0	100	0	0	95.8 (94.4– 96.9)

0   20   40   60   80   100

XSL•FO
RenderX

Discussion

Aim of This Review and Principal Findings

The goal of this systematic review was to synthesize the evidence concerning the development of a semi or fully automated method of AE detection in EHRs using trigger tools. The results show a broad variation in applied methods, selection of triggers, and estimates of AE prevalence and trigger-based PPVs. Insufficient reporting precluded full appreciation of the risk of bias and concerns regarding applicability.

Our findings are in line with another systematic review focusing on manual GTT [10]. Several interstudy differences can be hypothesized to explain the heterogeneity in the observed study estimates of AE prevalence and PPVs. These include the time frame (range: 1-51 months); the sample size (range: 250-54,244 records); the data sources and EHR system types; the automated approach; the review process; but also the differences in defining AEs, its severity, and preventability.

In addition, the choice and definition of triggers (eg, INR ≥ 6 [41,45] and INR >4 [25,43]) affect overall and per-trigger PPV. In some studies, only triggers identifying *unique* AEs are used for analysis, leading to varying results for similar triggers. Furthermore, the PPV is deemed to be strongly affected by the study's AE prevalence. We could not test for this because of the reviewed studies' heterogeneous definitions of prevalence. These included AEs per 100 patients admitted, AEs per 1000 patient days, or AEs per month. The missing information concerning the total number of patients included (36.4% of the 11 studies) hindered computation of prevalence estimates for these studies. The PPV of the triggers ranged from 0% to 100%. Even for the same trigger, high variability was observed.

Similar parallels can be observed regarding AEs' severity. Even when naloxone was used as a trigger in two hospitals in the same study [25], although one hospital had a severity level of 96.4%, falling into category "temporary harm to the patient and required intervention," the other had a severity level 100%, falling into the categories "permanent patient harm or patient death." Several included studies suggested improvement of PPV values by limiting triggers to specific patient subpopulations, or by modifying thresholds based on age or other patient characteristics [36,40]. Moore et al [38] suggested using sensitivity and specificity instead of PPV, as the former is less affected by changes in AE incidence. The studies did not address difficulties or opportunities regarding the implementation of trigger tool–based methods, but we do not exclude that such information can be found in more qualitative types of research that we did not collect for this review.

The decision of several of the reviewed studies' authors to employ semiautomatic approaches or to focus mainly on ADEs reflects the difficulty of detecting AEs fully automatically. The extracted high proportion of false alarms in various studies is likely explained in part by these difficulties. Although several studies showed that automatic tools can detect up to 69% of ADEs found manually, automatic detection of AEs overall remains problematic [47-50]. In our review, the sensitivity to detect AEs ranged from 0% to 100%. General AE detection

requires more advanced technology, such as NLP, which can read and process free-text narrative [51,52], addressing complex issues such as negation and lexical variation of terminology. Previous research has suggested that automated AE detection methods were superior to manual tools [26,53,54]. Automated AE detection methods have the potential to screen large numbers of patients to save valuable time, which would not be possible by doing manually by human reviewers with the same accuracy. Yet, timely intervention is an important factor to avoid complications and patient harm when an AE is detected. Even if PPVs are not high for all AEs, automated AE detection methods provide an excellent alternative to the manual approach by saving time and resources [14]. As shown by the systematic review of Wang et al, the use of NLP with EHRs is still at its infancy, and closer collaboration of NLP experts and clinicians is missing [46]. Nevertheless, automated AE detection methods are a promising approach for patient safety improvement.

Looking deeper into the individual studies via our quality assessment tools, we realized that, for most, their methodology, their results, or both were reported in ways that were unclear, inconsistent, or incomplete, which challenged our risk of bias and applicability assessments. Several studies failed to report the number of records screened or the type of patients sampled. These elements, however, are crucial for interpretation of the various estimates and also for its reproducibility. In other studies, the number of triggers or AEs remained unclear. AE severity and preventability were not always reported, and the variation in use of definitions for AEs, severity, and preventability further hampered interpretation of estimates across studies. Interestingly, the majority of studies lacked to report PPV CIs, which is essential for the swift interpretation of the estimate's precision.

Although not part of our quality assessment tools, we detected risk of selective outcome reprint in some studies. In these, severity and/or preventability assessments are reported in the Methods sections but not addressed in the Results section. Only one study checked for the presence of triggers at admission [42]. Such triggers or AEs should typically be excluded from the evaluation, as it cannot be targeted with interventions aiming at improving quality of care. Furthermore, although a reviewer's expertise plays an important role in the detection of AEs, information concerning their professional background, experience, or training was mostly absent. Overall, there is substantial room for improvement of the quality of reporting.

Limitations

Our decision to limit the inclusion criteria to studies concerning semi or fully automated trigger tool–like methodologies disqualified many studies, including those employing recommendations from the Harvard Medical Practice Study [28,55,56], machine learning [57], early warning systems [58-61], or other methods [62,63]. However, it allowed us to show that even within a narrow set of trigger-based tools, methods and outcomes varied considerably. The decision to exclude studies involving outpatients [62,64,65] or mixes of inpatients and outpatients [66] further decreased the number of eligible studies but increased the comparability of the patient population evaluated. Nevertheless, the overall low number of

eligible studies precluded statistical evaluation of the impact of sources of variation and bias. The evaluation of diagnostic test accuracy is generally hampered by the absence of a widely accepted reference standard.

Conclusions

This systematic review provides an overview about the application and outcomes of (semi) automatic trigger-based AE detection tools. We observed but could not formally explain the high degree of interstudy variation in reported estimates of prevalence and PPV, even in cases where similar triggers were tested. Although the AHRQ recently released common formats for event reporting [67], which supports the implementation of AE detection in the EHR, standards for the reporting of AE detection studies using trigger tools are lacking, yet urgently needed to overcome the methodological heterogeneity in future studies. We need better standards for reporting in this field of research to increase reproducibility, interpretation, and avoidance of research waste. A more standardized use of definitions of the types, severity, and preventability of AEs is desirable. We therefore call for a set of recommendations for the conduct and reporting of future studies and in the meantime, suggest authors, peer reviewers, and editors to pay special attention to complete reporting of study population, AE and trigger definitions, experience, training, and background of reviewers; methods employed to check for triggers and/or AEs at patient admission; and complete reporting of outcome data (numbers of triggers, nominators and denominators of the prevalence, and PPV).

Authors' Contributions

MS developed the idea for the study. MS, SNM, and AWSR contributed to the concept, design, analysis, and interpretation of data. SNM contributed to the drafting of the manuscript. AWSR contributed to the analysis and interpretation of the data. DA, RS, HUR, MMJ, and AWSR contributed to the screening and extraction phases, the interpretation of data, and the critical revision of the manuscript. All authors approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Hausner et al's approach.

[PDF File (Adobe PDF File), 16KB - jmir_v20i5e198_app1.pdf]

Multimedia Appendix 2

Quality Assessment tool for Diagnostic Accuracy Studies–2 (QUADAS–2) instrument.

[PDF File (Adobe PDF File), 42KB - jmir_v20i5e198_app2.pdf]

Multimedia Appendix 3

Trigger tool Quality Assessment Tool.

[PDF File (Adobe PDF File), 42KB - jmir_v20i5e198_app3.pdf]

References

1. Woloshynowych M, Neale G, Vincent C. Case record review of adverse events: a new approach. Qual Saf Health Care 2003 Dec;12(6):411-415 [FREE Full text] [Medline: 14645755]
2. Vincent C, Neale G, Woloshynowych M. Adverse events in British hospitals: preliminary retrospective record review. Br Med J 2001 Mar 3;322(7285):517-519 [FREE Full text] [Medline: 11230064]
3. Institute of Medicine. In: Kohn LT, Corrigan JM, Donaldson MS, editors. To Err is Human: Building a Safer Health System. Washington, DC: The National Academies Press; 2000.
4. Classen D. Medication safety: moving from illusion to reality. J Am Med Assoc 2003 Mar 05;289(9):1154-1156. [Medline: 12622587]
5. Adler L, Yi D, Li M, McBroom B, Hauck L, Sammer C, et al. Impact of Inpatient Harms on Hospital Finances and Patient Clinical Outcomes. J Patient Saf 2015 Mar 23:- (Forthcoming)(forthcoming). [doi: 10.1097/PTS.0000000000000171] [Medline: 25803176]

6.   Baker GR, Norton PG, Flintoft V, Blais R, Brown A, Cox J, et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. CMAJ 2004 May 25;170(11):1678-1686 [FREE Full text] [Medline: 15159366]

7.   Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, Harvard Medical Practice Study I. Incidence of adverse events and negligence in hospitalized patients: results of the Harvard Medical Practice Study I. 1991. Qual Saf Health Care 2004 Apr;13(2):145-51; discussion 151 [FREE Full text] [Medline: 15069223]

8.   Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA, et al. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. N Engl J Med 1991 Feb 7;324(6):377-384. [doi: 10.1056/NEJM199102073240605] [Medline: 1824793]

9.   Thomas EJ, Studdert DM, Burstin HR, Orav EJ, Zeena T, Williams EJ, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. Med Care 2000 Mar;38(3):261-271. [Medline: 10718351]

10.   Hibbert PD, Molloy CJ, Hooper TD, Wiles LK, Runciman WB, Lachman P, et al. The application of the Global Trigger Tool: a systematic review. Int J Qual Health Care 2016 Dec 01;28(6):640-649. [doi: 10.1093/intqhc/mzw115] [Medline: 27664822]

11.   Rosen AK, Mull HJ. Identifying adverse events after outpatient surgery: improving measurement of patient safety. BMJ Qual Saf 2016 Jan;25(1):3-5. [doi: 10.1136/bmjqs-2015-004752] [Medline: 26453634]

12.   Gandhi TK, Seger DL, Bates DW. Identifying drug safety issues: from research to practice. Int J Qual Health Care 2000 Feb;12(1):69-76. [Medline: 10733086]

13.   Classen DC, Resar R, Griffin F, Federico F, Frankel T, Kimmel N, et al. 'Global trigger tool' shows that adverse events in hospitals may be ten times greater than previously measured. Health Aff (Millwood) 2011 Apr;30(4):581-589 [FREE Full text] [doi: 10.1377/hlthaff.2011.0190] [Medline: 21471476]

14.   Murff HJ, Patel VL, Hripcsak G, Bates DW. Detecting adverse events for patient safety research: a review of current methodologies. J Biomed Inform 2003;36(1-2):131-143 [FREE Full text] [Medline: 14552854]

15.   Ausserhofer D, Zander B, Busse R, Schubert M, De Geest S, Rafferty AM, RN4CAST consortium. Prevalence, patterns and predictors of nursing care left undone in European hospitals: results from the multicountry cross-sectional RN4CAST study. BMJ Qual Saf 2014 Feb;23(2):126-135. [doi: 10.1136/bmjqs-2013-002318] [Medline: 24214796]

16.   Berry LL, Segal R, Sherrin TP, Fudge KA. Sensitivity and specificity of three methods of detecting adverse drug reactions. Am J Hosp Pharm 1988 Jul;45(7):1534-1539. [Medline: 3046347]

17.   Rogers AS, Israel E, Smith CR, Levine D, McBean AM, Valente C, et al. Physician knowledge, attitudes, and behavior related to reporting adverse drug events. Arch Intern Med 1988 Jul;148(7):1596-1600. [Medline: 3382304]

18.   Rozich JD, Haraden CR, Resar RK. Adverse drug event trigger tool: a practical methodology for measuring medication related harm. Qual Saf Health Care 2003 Jun;12(3):194-200 [FREE Full text] [Medline: 12792009]

19.   Resar RK, Rozich JD, Simmonds T, Haraden CR. A trigger tool to identify adverse events in the intensive care unit. Jt Comm J Qual Patient Saf 2006 Oct;32(10):585-590. [Medline: 17066996]

20.   Cullen DJ, Bates DW, Small SD, Cooper JB, Nemeskal AR, Leape LL. The incident reporting system does not detect adverse drug events: a problem for quality improvement. Jt Comm J Qual Improv 1995 Oct;21(10):541-548. [Medline: 8556111]

21.   West AN, Weeks WB, Bagian JP. Rare adverse medical events in VA inpatient care: reliability limits to using patient safety indicators as performance measures. Health Serv Res 2008 Feb;43(1 Pt 1):249-266 [FREE Full text] [doi: 10.1111/j.1475-6773.2007.00760.x] [Medline: 18211528]

22.   Griffin FA, Resar RK. IHI. 2009. IHI Global Trigger Tool for Measuring Adverse Events URL: http://www.ihi.org/resources/Pages/Tools/IHIGlobalTriggerToolforMeasuringAEs.aspx [accessed 2018-05-04] [WebCite Cache ID 6z9v8yzwv]

23.   Adler L, Denham CR, McKeever M, Purinton R, Guilloteau F, Moorhead JD, et al. Global trigger tool: implementation basics. J Patient Saf 2008;4(4):245-249. [doi: 10.1097/pts.0b013e31818e8a87]

24.   Doupi P. Using EHR data for monitoring and promoting patient safety: reviewing the evidence on trigger tools. Stud Health Technol Inform 2012;180:786-790. [Medline: 22874299]

25.   Stockwell DC, Kirkendall E, Muething SE, Kloppenborg E, Vinodrao H, Jacobs BR. Automated adverse event detection collaborative: electronic adverse event identification, classification, and corrective actions across academic pediatric institutions. J Patient Saf 2013 Dec;9(4):203-210. [doi: 10.1097/PTS.0000000000000055] [Medline: 24257063]

26.   Li Q, Melton K, Lingren T, Kirkendall ES, Hall E, Zhai H, et al. Phenotyping for patient safety: algorithm development for electronic health record based automated adverse event and medical error detection in neonatal intensive care. J Am Med Inform Assoc 2014;21(5):776-784 [FREE Full text] [doi: 10.1136/amiajnl-2013-001914] [Medline: 24401171]

27.   Menendez ME, Janssen SJ, Ring D. Electronic health record-based triggers to detect adverse events after outpatient orthopaedic surgery. BMJ Qual Saf 2016 Jan;25(1):25-30. [doi: 10.1136/bmjqs-2015-004332] [Medline: 26359272]

28.   Murff HJ, Forster AJ, Peterson JF, Fiskio JM, Heiman HL, Bates DW. Electronically screening discharge summaries for adverse medical events. J Am Med Inform Assoc 2003;10(4):339-350 [FREE Full text] [doi: 10.1197/jamia.M1201] [Medline: 12668691]

29. Govindan M, Van Citters AD, Nelson EC, Kelly-Cummings J, Suresh G. Automated detection of harm in healthcare with information technology: a systematic review. Qual Saf Health Care 2010 Oct;19(5):e11. [doi: 10.1136/qshc.2009.033027] [Medline: 20671081]

30. Reitsma JB, Rutjes AW, Whiting P, Vlassov VV, Leeflang MMG, Deeks JJ. Chapter 9: Assessing methodological quality. In: Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. London, England: The Cochrane Collaboration; 2009.

31. Higgins JP, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. West Sussex, England: John Wiley & Sons Ltd; 2011.

32. 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]

33. Hausner E, Waffenschmidt S, Kaiser T, Simon M. Routine development of objectively derived search strategies. Syst Rev 2012 Feb 29;1:19 [FREE Full text] [doi: 10.1186/2046-4053-1-19] [Medline: 22587829]

34. Hausner E, Guddat C, Hermanns T, Lampert U, Waffenschmidt S. Development of search strategies for systematic reviews: validation showed the noninferiority of the objective approach. J Clin Epidemiol 2015 Feb;68(2):191-199 [FREE Full text] [doi: 10.1016/j.jclinepi.2014.09.016] [Medline: 25464826]

35. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011 Oct 18;155(8):529-536. [doi: 10.7326/0003-4819-155-8-201110180-00009] [Medline: 22007046]

36. Call RJ, Burlison JD, Robertson JJ, Scott JR, Baker DK, Rossi MG, et al. Adverse drug event detection in pediatric oncology and hematology patients: using medication triggers to identify patient harm in a specialized pediatric patient population. J Pediatr 2014 Sep;165(3):447-52.e4 [FREE Full text] [doi: 10.1016/j.jpeds.2014.03.033] [Medline: 24768254]

37. Dickerman MJ, Jacobs BR, Vinodrao H, Stockwell DC. Recognizing hypoglycemia in children through automated adverse-event detection. Pediatrics 2011 Apr;127(4):e1035-e1041. [doi: 10.1542/peds.2009-3432] [Medline: 21402631]

38. Moore C, Li J, Hung C, Downs J, Nebeker JR. Predictive value of alert triggers for identification of developing adverse drug events. J Patient Saf 2009 Dec;5(4):223-228. [doi: 10.1097/PTS.0b013e3181bc05e5] [Medline: 22130215]

39. Muething SE, Conway PH, Kloppenborg E, Lesko A, Schoettker PJ, Seid M, et al. Identifying causes of adverse events detected by an automated trigger tool through in-depth analysis. Qual Saf Health Care 2010 Oct;19(5):435-439. [doi: 10.1136/qshc.2008.031393] [Medline: 20798069]

40. Shea PL, Spaeder MC, Sayal P, Jacobs BR, Stockwell DC. Hypoglycemia adverse events in PICUs and cardiac ICUs: differentiating preventable and nonpreventable events*. Pediatr Crit Care Med 2013 Oct;14(8):741-746. [doi: 10.1097/PCC.0b013e3182975f0f] [Medline: 23863820]

41. O'Leary KJ, Devisetty VK, Patel AR, Malkenson D, Sama P, Thompson WK, et al. Comparison of traditional trigger tool to data warehouse based screening for identifying hospital adverse events. BMJ Qual Saf 2013 Feb;22(2):130-138. [doi: 10.1136/bmjqs-2012-001102] [Medline: 23038408]

42. Lim D, Melucci J, Rizer MK, Prier BE, Weber RJ. Detection of adverse drug events using an electronic trigger tool. Am J Health Syst Pharm 2016 Sep 01;73(17 Suppl 4):S112-S120. [doi: 10.2146/ajhp150481] [Medline: 27543596]

43. Patregnani JT, Spaeder MC, Lemon V, Diab Y, Klugman D, Stockwell DC. Monitoring the harm associated with use of anticoagulants in pediatric populations through trigger-based automated adverse-event detection. Jt Comm J Qual Patient Saf 2015 Mar;41(3):108-114. [Medline: 25977126]

44. Gerdes LU, Hardahl C. Text mining electronic health records to identify hospital adverse events. Stud Health Technol Inform 2013;192:1145. [Medline: 23920919]

45. Nwulu U, Nirantharakumar K, Odesanya R, McDowell SE, Coleman JJ. Improvement in the detection of adverse drug events by the use of electronic health and prescription records: an evaluation of two trigger tools. Eur J Clin Pharmacol 2013 Feb;69(2):255-259. [doi: 10.1007/s00228-012-1327-1] [Medline: 22706621]

46. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: a literature review. J Biomed Inform 2018 Jan;77:34-49 [FREE Full text] [doi: 10.1016/j.jbi.2017.11.011] [Medline: 29162496]

47. Jha AK, Kuperman GJ, Teich JM, Leape L, Shea B, Rittenberg E, et al. Identifying adverse drug events: development of a computer-based monitor and comparison with chart review and stimulated voluntary report. J Am Med Inform Assoc 1998;5(3):305-314 [FREE Full text] [Medline: 9609500]

48. Bates DW, Leape LL, Cullen DJ, Laird N, Petersen LA, Teich JM, et al. Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. J Am Med Assoc 1998 Oct 21;280(15):1311-1316. [Medline: 9794308]

49. Classen DC, Pestotnik SL, Evans RS, Burke JP. Computerized surveillance of adverse drug events in hospital patients. J Am Med Assoc 1991 Nov 27;266(20):2847-2851. [Medline: 1942452]

50. Raschke RA, Gollihare B, Wunderlich TA, Guidry JR, Leibowitz AI, Peirce JC, et al. A computer alert system to prevent injury from adverse drug events: development and evaluation in a community teaching hospital. J Am Med Assoc 1998 Oct 21;280(15):1317-1320. [Medline: 9794309]

51. Hripcsak G, Friedman C, Alderson PO, DuMouchel W, Johnson SB, Clayton PD. Unlocking clinical data from narrative reports: a study of natural language processing. Ann Intern Med 1995 May 01;122(9):681-688. [Medline: 7702231]

52. Friedman C. Towards a comprehensive medical language processing system: methods and issues. Proc AMIA Annu Fall Symp 1997:595-599 [FREE Full text] [Medline: 9357695]

53. Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. J Am Med Assoc 2011 Aug 24;306(8):848-855. [doi: 10.1001/jama.2011.1204] [Medline: 21862746]

54. Melton GB, Hripcsak G. Automated detection of adverse events using natural language processing of discharge summaries. J Am Med Inform Assoc 2005;12(4):448-457 [FREE Full text] [doi: 10.1197/jamia.M1794] [Medline: 15802475]

55. Szekendi MK, Sullivan C, Bobb A, Feinglass J, Rooney D, Barnard C, et al. Active surveillance using electronic triggers to detect adverse events in hospitalized patients. Qual Saf Health Care 2006 Jun;15(3):184-190 [FREE Full text] [doi: 10.1136/qshc.2005.014589] [Medline: 16751468]

56. Kane-Gill SL, MacLasco AM, Saul MI, Politz ST, Kloet MA, Kim C, et al. Use of text searching for trigger words in medical records to identify adverse drug reactions within an intensive care unit discharge summary. Appl Clin Inform 2016 Jul 13;7(3):660-671 [FREE Full text] [doi: 10.4338/ACI-2016-03-RA-0031] [Medline: 27453336]

57. Marella WM, Sparnon E, Finley E. Screening electronic health record-related patient safety reports using machine learning. J Patient Saf 2017 Mar;13(1):31-36. [doi: 10.1097/PTS.0000000000000104] [Medline: 24721977]

58. Kavanaugh MJ, So JD, Park PJ, Davis KL. Validation of the intensive care unit early warning dashboard: quality improvement utilizing a retrospective case-control evaluation. Telemed J E Health 2017 Feb;23(2):88-95. [doi: 10.1089/tmj.2016.0073] [Medline: 27391204]

59. Le Lagadec MD, Dwyer T. Scoping review: the use of early warning systems for the identification of in-hospital patients at risk of deterioration. Aust Crit Care 2017 Jul;30(4):211-218. [doi: 10.1016/j.aucc.2016.10.003] [Medline: 27863876]

60. Paliwoda M, New K, Bogossian F. Neonatal Early Warning Tools for recognising and responding to clinical deterioration in neonates cared for in the maternity setting: a retrospective case-control study. Int J Nurs Stud 2016 Sep;61:125-135. [doi: 10.1016/j.ijnurstu.2016.06.006] [Medline: 27348358]

61. Sefton G, Lane S, Killen R, Black S, Lyon M, Sproule C, et al. Accuracy and efficiency of recording paediatric early warning scores (PEWS) using electronic hand-held devices compared with traditional paper-based documentation. Comput Inform Nurs 2017 May;35(5):228-236 [FREE Full text] [doi: 10.1097/CIN.0000000000000305]

62. Long AL, Horvath MM, Cozart H, Eckstrand J, Whitehurst J, Ferranti J. Tailoring adverse drug event surveillance to the paediatric inpatient. Qual Saf Health Care 2010 Oct;19(5):e40 [FREE Full text] [doi: 10.1136/qshc.2009.032680] [Medline: 20511599]

63. Haffner S, von LN, Wirth S, Thürmann PA. Detecting adverse drug reactions on paediatric wards: intensified surveillance versus computerised screening of laboratory values. Drug Saf 2005;28(5):453-464. [Medline: 15853446]

64. Mull HJ, Rosen AK, Shimada SL, Rivard PE, Nordberg B, Long B, et al. Assessing the potential adoption and usefulness of concurrent, action-oriented, electronic adverse drug event triggers designed for the outpatient setting. EGEMS (Wash DC) 2015;3(1):1116 [FREE Full text] [doi: 10.13063/2327-9214.1116] [Medline: 25992386]

65. Rosen AK, Mull HJ, Kaafarani H, Nebeker J, Shimada S, Helwig A, et al. Applying trigger tools to detect adverse events associated with outpatient surgery. J Patient Saf 2011 Mar;7(1):45-59. [doi: 10.1097/PTS.0b013e31820d164b] [Medline: 21921867]

66. Lemon V, Stockwell DC. Automated detection of adverse events in children. Pediatr Clin North Am 2012 Dec;59(6):1269-1278. [doi: 10.1016/j.pcl.2012.08.007] [Medline: 23116524]

67. Patient Safety Organization Privacy Protection. Common Formats for Event Reporting - Hospital Version 2 URL: https://www.psoppc.org/psoppc_web/publicpages/commonFormatsHV2.0 [accessed 2018-03-28] [WebCite Cache ID 6yFTMJEMN]

Abbreviations

ADE: adverse drug event
AE: adverse event
AHRQ: Agency for Healthcare Research and Quality
DTA: diagnostic test accuracy
EHR: electronic health record
GTT: Global Trigger Tool
HCP: heath care provider
IHI: Institute for Healthcare Improvement
MeSH: medical subject headings
NCC MERP: National Coordinating Council for Medication Error Reporting and Prevention
NLP: natural language processing
PPV: positive predictive value
PSI: Patient Safety Indicator

QUADAS-2: Quality Assessment tool for Diagnostic Accuracy Studies-2

<u>Viewpoint</u>

Too Important to Ignore: Leveraging Digital Technology to Improve Chronic Illness Management Among Black Men

Stuart W Grande[1*], MPA, PhD; Ledric D Sherman[2*], MA, PhD

[1]The Geisel School of Medicine at Dartmouth, The Dartmouth Institute for Health Policy and Clinical Practice, Dartmouth College, Hanover, NH, United States

[2]College of Education and Human Development, Department of Health and Kinesiology, Texas A&M University, College Station, TX, United States

[*]all authors contributed equally

Corresponding Author:
Stuart W Grande, MPA, PhD
The Geisel School of Medicine at Dartmouth
The Dartmouth Institute for Health Policy and Clinical Practice
Dartmouth College
37 Dewey Field Rd
HB 7256
Hanover, NH, 03755
United States
Phone: 1 603 646 6601
Email: stuart.w.grande@dartmouth.edu

Abstract

Health disparities associated with chronic illness experiences of black men demonstrate widespread, systematic failures to meet an urgent need. Well-established social and behavioral determinants that have led to health disparities among black men include racism, discrimination, and stress. While advocacy work that includes community-engagement and tailoring health promotion strategies have shown local impact, evidence shows the gaps are increasing. We suspect that failure to reduce current disparities may be due to conventional public health interventions and programs; therefore, we submit that innovative interventions, ones that embrace digital technologies and their ability to harness naturally occurring social networks within groups, like black men, have particular importance and deserve attention. This commentary characterizes the current literature on chronic illness among black men as well as health interventions that use digital technology, to build a case for expanding research in this area to reduce the overwhelming burden of chronic illness among black men.

Prioritizing Chronic Illness Management Among Black Men

There is no shortage of data on the debilitating effects of health disparities, particularly on the effects of chronic illness among black men. If these data are so compelling, then why hasn't the story changed? Many have argued this is a moral issue that is centered on the principle that suffering due to race, gender, or genetics is unconscionable. Despite how often these injustices are professed or predicted, these moral commitments and righteous indignations fall short. In this commentary, we take the position that failures to reduce health disparities among black men are the result of more traditional community-engagement approaches that reflect passive education and one-off health promotion activities. We apply the term black men inclusively to refer to any Hispanic or non-Hispanic man who may or may not self-identify as African American. Further, borrowing from the world of health services research, we contend that patient engagement strategies—those promoted by digital technology and where the key ingredient is partnership and collaboration—offer promising advances to mitigate health disparities experienced by black men [1,2]. Simply put, the mounting evidence on the potential of these technologies is too great to ignore. An innovation of this caliber is necessary to refocus attention to enduring gaps in health service delivery systems and to support black men in ways that are the most meaningful to them. In this regard, digital innovation must emerge from foundations of partnership and

collaboration, where preferences and needs drive technology that guides self-management of chronic illness.

One of the leading contributors to morbidity among black males is chronic illness, particularly heart disease and high rates of hypertension (about 43% of black men have hypertension compared with about 30% of white men) [3]. The chronic illness burden among black men in the United States is disproportionate compared to other male ethnic and racial groups [4]. Unmanaged chronic illness, where illness symptoms go untreated and advance in severity, also adversely impact the health of black men. According to the Centers for Disease Control and Prevention (CDC), 7 of the top 10 causes of death among black men in 2014 were related to chronic illness, two of which—heart disease and cancer—comprise nearly 46% of all deaths [5]. As unmanaged chronic illness occurs based on several factors, we suspect that black men are more often unsupported to navigate the healthcare system in a way that makes sense for them. We are aware that black men face a wide set of systematic barriers, mostly out of their control, related to accessing and engaging in supportive and beneficial health promotion activities. This suggests that alternative approaches are necessary.

Since the publication of *Unequal Treatment in 2003*, the causal pathways of health disparities are well known, and yet, insufficient change has occurred. Over the last several years in the United States, while access and quality have improved, health disparities remain consistent across groups [6]. A recent review of black male health examines how the healthcare system affects their lived experiences, making the point that social and behavioral determinants that would positively impact their lives are under-supported and neglected [7]. There is a trend of decreasing life expectancy in black vs whites; and among men, this trend remains persistent. One of the most credible factors is higher rates of uncontrolled hypertension or high-blood pressure, often reflected as an individual with high blood pressure not receiving treatment. While control of hypertension is linked to higher quality care, there are reports that better communication strategies with doctors are vital, especially for developing successful treatment strategies [8]. Knowing that discrimination, racism, and stigma negatively impact black men and their engagement with the healthcare system, building communication strategies that counteract these biases is critical [9]. We take the view here that the goal of overcoming barriers to engagement must be the centerpiece of novel approaches. These approaches may not be practical at the system-level but may be possible at the individual level. Therefore, designing solutions at the individual level so that systems can better support how black men engage in healthcare may have the highest likelihood of success.

Much of the current literature on black men and health reflect outcomes associated with poverty, education, discrimination (racism), and poorly managed stress [7]. It is unreasonable to think that these can be attenuated by health education or one-off health promotion events alone. Understandably, these experiences are highly personal and further complicated by chronic illness management, which requires informed medical support. Scholars and practitioners alike see little benefit to simply identifying these disparities and failing to build platforms aimed toward changing current systems. While

community-based participatory methods have emerged as a leading strategy for engaging racial and ethnic groups, fostering empowerment through trust building and transparency remains a major hurdle [10]. Given these gaps and opportunities, might building digital technologies that are designed to meaningfully engage black men, prioritize the contribution of brotherhood and existing social networks among black men [11]? Here we adopt Franklin's definition of brotherhood as a "community of African American or black men connected through tacit understanding, common experiences, traditions, and identity." It is this level of connection that has been until recently, mostly unexplored, and offers exciting areas co-designing interventions to reduce disparities [12].

Integrating health interventions within trusted social systems like a brotherhood, built around shared beliefs, attitudes, and authentic relationships may overcome many recognized barriers of current strategies [13]. To this end, we argue that digital technologies present a unique potential for building meaningful solutions to current challenges associated with managing chronic illness among black men.

Technology as a Means of Engagement

Mobile digital devices and apps, including websites and social media platforms, offer immediate access to medical and health information on the internet and also offer new ways of monitoring, measuring, and visualizing the human body while sharing personal information and experiences with others [14]. Electronic or digital health encompasses a wide range of technologies that are used for healthcare, health informatics, health education, and health promotion. Mobile health (mHealth)—the application of wireless technology to promote wellness and information provision between healthcare and individuals—is rapidly gaining attention in both preventive medicine and chronic illness management [15]. One of mHealth's strengths is its ability to leverage the existing mobile technology infrastructure and the ubiquity of the mobile phone across populations and subgroups. Previous research has corroborated the high prevalence of smartphone usage among blacks as well as Hispanics and indicated the value of mHealth in managing chronic illness [14].

Smartphone and tablet apps can educate, assist with decision-making, and promote adherence to lifestyle and medication regimens. Consumer-wearable technology comprises devices equipped with sensors and wireless connectivity that can assist with a wide range of health-related applications such as monitoring blood sugar levels, personalizing treatment, connecting with health-care providers, and even delivering medication into the body [15]. Recent work in diabetes and obesity shows technological advancement in the use of wrist and hip accelerometers, global positioning system (GPS), mobile technology apps for weight loss, apps for dietary self-monitoring, internet weight loss programs, and behavioral intervention technologies. While there is some fear regarding information security and potential Health Insurance Portability and Accountability Act (HIPAA) concerns, potentially opt-in rather than opt-out features offer individuals needed flexibility to assume or reject risk based on preference [16].

XSL•FO

RenderX

New Opportunities with Digital Technologies

Developments in new technologies go beyond self-monitoring. Some include prevention of illness complications such as diabetic neuropathy, and others, like prototype socks and shoes with thermal and pressure sensors, can identify areas of the feet with insufficient blood supply [17,18]. Potentially, a physician could use this type of technology to routinely inspect minor tissue damage and greatly minimize the risk of amputations. In the more immediate future, technologies like these will offer practical approaches to managing chronic illnesses like diabetes. *The OneTouch Reveal* mobile app, for example, can test a drop of blood and tell whether sugar levels are within normal range as well as provide a summary of overall health performance [19]. Another app, called *Diabetik,* provides quick and interactive data entry to monitor diet, blood glucose levels, and medication [20]. The user can set medication or activity reminders according to time or location. *Fooducate* has an extensive database of food information that uses barcode scanning to search nutritional value and suggest healthier alternatives [21]. Uniquely, this app also creates a community in which users can share progress and healthy recipes.

Although fewer in number, apps that meaningfully connect patients to doctors, like *Glooko*, that aggregates biometric data-glucose monitoring and fitness information - offer physicians novel ways to engage with vital patient information [22]. These data are critical and often overlooked during face-to-face consultations. By presenting this data in easy-to-manage formats, these types of apps support patient and physician engagement that contribute to robust dialogue and more efficient clinical visits. These apps also provide the opportunity to foster new conversations that both patients and doctors may have been too intimidated or preoccupied to engage.

Weight management apps are available that focus primarily on tracking and require a certain amount of user-input to manage. These act more like calendars and reminders to record the food and drink intake. Apps also let users track your exercise using information that you enter manually or by tracking your activity automatically using the GPS or accelerometer in your phone. Compared to using paper diaries, tracking via smartphone may be more convenient because the device does some of the work for you (eg, calories, movement tabulations), but these are only for self-monitoring purposes. These and other wearable devices typically display the information collected and provide feedback via a smartphone app and website, allowing you to track changes over time. Importantly, one trial showed strong evidence for the use of smartphone technology for managing weight, giving us hope for broader interventions in communities where managing chronic illness are a challenge [23].

Tailoring Digital Technology for Black Men

Tailoring interventions for black men makes good sense given existing communication networks within subgroups. Digital technology is uniquely positioned to attend to subgroup differences as they are, by nature, personal and self-managed. Others point to the importance of system-level identification and intervention using structural factors, but these, while necessary, are too general and are less accessible than digital technology [24]. To date, there is good evidence for targeted and tailored interventions in black communities [25]. In this sense, a tailored intervention would be more appropriate than simply modifying language or applying a "one size fits all" approach to effectively meet black men where they are. According to Hawkins and colleagues, this would include personalization, gaining feedback from black men, and matching content to their personal data [26]. As identified elsewhere, community-based participatory research (CBPR) strategies are effective evidence-based methods for engaging, co-designing and tailoring mHealth interventions to meet preference-driven solutions [27].

Despite the limited amount of published mHealth interventions to improve chronic illness among black men, the existing research on computer-tailored and social media interventions offer some insight. Consensus seems to suggest that tailored applications should be: theory-based, contextual, modifiable to individual preferences, able to connect black men to other black men, include a preference-based dashboard, and be supported by popular culture references [28]. Informed by CBPR, mHealth interventions must be developed and optimized for black men. Taken a step further, the perspective supports a more intersectional approach to building more impactful and effective interventions. Research that targets the intersection of social, health, context, and identity overcomes the myopic view of black men's health. This complex and nuanced viewpoint challenges this view by suggesting that an app that mediates naturally occurring social networks of black men with type 2 diabetes will consider the overlap of both gender and the marginalization of race [29].

Further, we see that tailoring can and should mean developing interventions (in this case mHealth apps) in partnership with black men. For example, if developers choose a deep tailoring approach rather than a surface tailoring one, the mHealth apps will integrate cultural values, norms, and religious or spiritual beliefs of the population of interest (black males) [30]. Examples of these tailoring messages might include age-appropriate material to engage churches, ministry teams, and fellowship groups in the development and integratration of popular culture messaging/references. In addition, using content specific (chronic illness) framing along with connectivity platforms, enhances relationship building and help seeking around chronic illness management [31]. Such tailoring would also support the existing social networks and relationships aligned with brotherhood, as described earlier.

There is also thinking that bias and preference play a role in the perpetuation of race-based health disparities, but often these are superseded by communication challenges [32]. The systematic bias in health delivery and its contribution to a perpetuation of fear and apprehension among black men is widely recognized [24]. As shown, attending to individual fears about talking with physicians is necessary for improving confidence and readiness (activation) for engaging in open and honest dialog [33]. And while the bias issue may overstate the role of the clinician in

this relationship, an argument can be made that provided the right support platform, black men offer each other the best strategies.

Health services research has demonstrated that individuals who feel more empowered and confident to make decisions are healthier [34]. These broad approaches have shown efficacy in other areas, but when it comes to tailoring interventions for black men, there is a need to leverage platforms for social support. By providing access to health resources, important health information that relies on and utilizes recognized social support can be more effective particularly with self-management and adherence [35]. This strengthens the argument that innovative technologies designed to leverage social ties have the greatest likelihood of success. The challenge will be determining best practices. What is needed is efficient processes and communication to improve self-efficacy, the confidence to participate meaningfully in one's health.

Feasibility of Digital Technology for Advancing Black Men's Health

There have been many efforts to mitigate disparities among black men. Given the scope and trends of current disparities, we suggest that technology, and particularly digital technology focused on engagement and activation, can address some of these gaps.

There is currently a wide gap in exploring the use of these technologies among black men to empower and encourage health behaviors and self-management strategies. In 2000, a survey of internet usage among blacks showed they were 38% more likely than whites to seek information about jobs on the web, and 45% of African-Americans vs 35% of whites say that the internet helps them find health care information [36]. A 2018 Pew fact sheet shows that 98% of blacks own a cell phone, and 75% own a smartphone, suggesting that opportunities to download and utilize online content are wide and far-reaching [37]. We also see that young African-Americans have high use of Twitter, and 96% of those aged 18–29 use some social networking platform [38]. The high frequency of socializing using these new technologies supports our efforts to capitalize on social norms in the community.

There are multiple initiatives that support person–centered design, and tailoring of messaging in ways that have been minimally examined among black men. For example, health and medical support and information on websites, apps, and social media sites have proliferated, facilitating the access of lay people to health-related information and providing them with the opportunity to share experiences of their illnesses or health-promoting activities. Health promotion researchers have sought to investigate how websites and social media sites operate in generating and disseminating information about strategies for promoting health among lay people who do this voluntarily

as part of social interactions and support systems. Further, social support systems that foster pre-existing social networks, characteristic of the brotherhood as defined earlier have been unexplored and offer exciting areas for mitigating disparities [12].

Embracing Novel Technologies to Support Engagement and Chronic Illness Management

The health disparities that describe the population health outcomes for black men in the United States has garnered much attention, and yet the story remains the same. Failure to reduce current disparities may be due to conventional public health interventions; therefore, we submit that innovative interventions, ones that embrace digital technologies and their ability to harness inherent social trends within groups, like black men, have particular importance and deserve attention. Further, we outline alternative interventions to mitigate chronic illness in black men as a way to break the dominance of current deficit-based approaches that frame behaviors as negative or abnormal.

We recognize that medical anthropologists have also weighed in on the debate, suggesting that technologies that seemingly enhance the patient experience may also foster a false sense of hope—what Mary-Jo DelVecchio Good refers to as the "medical imaginary" [39]. The hope of technology seems to be conflated in concepts of innovation and improvement, where hope sits more with consumers than providers. This debate on the benefits of user-centered technologies for patient outcomes seems much less controversial when we reflect on non-hospital based approaches. In other words, the potential of these technologies seems most impactful in communities, where engaging individuals through tailored approaches have the most likely benefit.

According to Forbes one of the top five most important innovations in health care is the use of digital technology for empowering consumers [40]. While many have suggested that black men are members of hard-to-reach populations, with a history of not participating in much research, there is literature debunking these myths. In fact, when researchers make an effort to engage black men in meaningful ways, the evidence is strong that positive behavior change is not only possible but dynamic. What is needed are interventions that show promise across other groups, based on co-design strategies that prove to engage and empower black men to make healthy choices around the management of their chronic illness. We believe that if designed properly these technologies offer needed ways to more fully connect communities to health care resources and professionals. We have hopefully built a case for expanding research in this area to reduce the overwhelming burden of chronic illness among black men.

Acknowledgments

Authors' Contributions

Each author contributed equally to the conception and writing of this commentary.

Conflicts of Interest

None declared.

References

1. Graffigna G, Barello S, Bonanomi A, Menichetti J. The motivating function of healthcare professional in eHealth and mHealth interventions for type 2 diabetes patients and the mediating role of patient engagement. J Diabetes Res 2016;2016:2974521-2974510 [FREE Full text] [doi: 10.1155/2016/2974521] [Medline: 26881243]
2. Weiner JP. Doctor-patient communication in the e-health era. Isr J Health Policy Res 2012;1(1):33 [FREE Full text] [doi: 10.1186/2045-4015-1-33] [Medline: 22929000]
3. CDC. Health, United States, 2015: With special feature on racial and ethnic health disparities.: National Center for Health Statistics URL: https://www.cdc.gov/nchs/data/hus/hus15.pdf [accessed 2017-11-16] [WebCite Cache ID 6v1e4bedR]
4. Thorpe RJ, Wilson-Frederick SM, Bowie JV, Coa K, Clay OJ, LaVeist TA, et al. Health behaviors and all-cause mortality in African American men. Am J Mens Health 2013 Jul;7(4 Suppl):8S-18S [FREE Full text] [doi: 10.1177/1557988313487552] [Medline: 23649171]
5. Centers for Disease Control and Prevention. 2014. Leading Causes of Death in Males, 2014 URL: https://www.cdc.gov/healthequity/lcod/men/2014/black/index.htm [WebCite Cache ID 6v1e94guQ]
6. Agency for Healthcare Research and Quality. Rockville, MD 2014 National Healthcare Quality and Disparities Report URL: https://www.ahrq.gov/research/findings/nhqrdr/nhqdr14/index.html [WebCite Cache ID 6v1dc16Jt]
7. Noonan A, Velasco-Mondragon H, Wagner F. Improving the health of African Americans in the USA: an overdue opportunity for social justice. Public Health Rev 2016;37:12 [FREE Full text] [doi: 10.1186/s40985-016-0025-4] [Medline: 29450054]
8. Centers for Disease Control and Prevention. Atlanta: U.S. Department of Health and Human Services; 2010. A closer look at African American men and high blood pressure control: a review of psychosocial factors and systems-level interventions URL: https://www.cdc.gov/bloodpressure/docs/african_american_sourcebook.pdf [WebCite Cache ID 6v1dtn04Z]
9. Hammond WP, Matthews D, Mohottige D, Agyemang A, Corbie-Smith G. Masculinity, medical mistrust, and preventive health services delays among community-dwelling African-American men. J Gen Intern Med 2010 Dec;25(12):1300-1308 [FREE Full text] [doi: 10.1007/s11606-010-1481-z] [Medline: 20714819]
10. Thompson B, Molina Y, Viswanath K, Warnecke R, Prelip ML. Strategies To Empower Communities To Reduce Health Disparities. Health Aff (Millwood) 2016 Dec 01;35(8):1424-1428 [FREE Full text] [doi: 10.1377/hlthaff.2015.1364] [Medline: 27503967]
11. Franklin AJ. From brotherhood to manhood: How Black men rescue their relationships and dreams from the invisibility syndrome. Hoboken, New Jersey: Wiley; 2004:p236.
12. Grande SW, Sherman L, Shaw-Ridley M. A brotherhood perspective: how African American male relationships may improve trust and utilization of health care. Am J Mens Health 2013;7(6):494-503. [doi: 10.1177/1557988313485783] [Medline: 23620540]
13. Lupton D. Health promotion in the digital era: a critical commentary. Health Promotion International 2014;30(1):174-183. [doi: 10.1093/heapro/dau091] [Medline: 25320120]
14. Davidson T, McGillicuddy J, Mueller M, Brunner-Jackson B, Favella A, Anderson A, et al. Evaluation of an mHealth medication regimen self-management program for African American and Hispanic uncontrolled hypertensives. J Pers Med 2015;5(4):389-405 [FREE Full text] [doi: 10.3390/jpm5040389] [Medline: 26593951]
15. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med 2016;13(2):e1001953 [FREE Full text] [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]
16. Simon SR, Evans JS, Benjamin A, Delano D, Bates DW. Patients' attitudes toward electronic health information exchange: qualitative study. J Med Internet Res 2009 Aug 06;11(3):e30 [FREE Full text] [doi: 10.2196/jmir.1164] [Medline: 19674960]
17. Sousa P, Felizardo V, Oliveira D, Couto R, Garcia NM. A review of thermal methods and technologies for diabetic foot assessment. Expert Rev Med Devices 2015;12(4):439-448. [doi: 10.1586/17434440.2015.1032251] [Medline: 25896130]
18. Bu D, Pan E, Walker J, Adler-Milstein J, Kendrick D, Hook JM, et al. Benefits of information technology-enabled diabetes management. Diabetes Care 2007 May;30(5):1137-1142. [doi: 10.2337/dc06-2101] [Medline: 17322483]
19. OneTouch. OneTouch Reveal mobile and web apps. 2018. URL: https://www.onetouch.com/softwareandapps/Reveal [accessed 2018-02-17] [WebCite Cache ID 6xIyc0pHp]

XSL·FO

RenderX

20.   Agarwal A. The Diabetik app provides patients a way to manage their diabetes. 2018. URL: https://www.imedicalapps.com/2014/04/diabetik-app-patients-diabetes/ [accessed 2018-02-17] [WebCite Cache ID 6xIq441BS]

21.   Fooducate. Eat better, lose weight, get healthy. 2018. URL: https://www.fooducate.com/ [accessed 2018-02-17] [WebCite Cache ID 6xIx7No27]

22.   Glooko, Inc. Insights that improve diabetes. 2018. URL: https://www.glooko.com/ [accessed 2018-02-17] [WebCite Cache ID 6xIz8W2Tz]

23.   Allen JK, Stephens J, Dennison HCR, Stewart KJ, Hauck S. Randomized controlled pilot study testing use of smartphone technology for obesity treatment. J Obes 2013;2013:151597 [FREE Full text] [doi: 10.1155/2013/151597] [Medline: 24392223]

24.   Williams DR. The health of men: structured inequalities and opportunities. Am J Public Health 2008 Sep;98(9 Suppl):S150-S157. [Medline: 18687602]

25.   Bharmal N, Kennedy D, Jones L, Lee-Johnson C, Morris D, Caldwell B, et al. Through our eyes: exploring African-American men's perspective on factors affecting transition to manhood. J Gen Intern Med 2012 Feb;27(2):153-159 [FREE Full text] [doi: 10.1007/s11606-011-1836-0] [Medline: 21910088]

26.   Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]

27.   Unertl KM, Schaefbauer CL, Campbell TR, Senteio C, Siek KA, Bakken S, et al. Integrating community-based participatory research and informatics approaches to improve the engagement and health of underserved populations. J Am Med Inform Assoc 2016;23(1):60-73 [FREE Full text] [doi: 10.1093/jamia/ocv094] [Medline: 26228766]

28.   Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016;60:243-251. [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]

29.   Griffith DM. An intersectional approach to men's health. J Mens Health 2012;9(2):106-112. [doi: 10.1016/j.jomh.2012.03.003]

30.   Huang Y, Shen F. Effects of cultural tailoring on persuasion in cancer communication: a meta-analysis. J Commun 2016;66(4):694-715. [doi: 10.1111/jcom.12243]

31.   James DCS, Harville C, Efunbumi O, Babazadeh I, Ali S. "You have to approach us right": a qualitative framework analysis for recruiting african americans into mHealth research. Health Educ Behav 2017;44(5):781-790. [doi: 10.1177/1090198117727324] [Medline: 28863726]

32.   Ashton CM, Haidet P, Paterniti DA, Collins TC, Gordon HS, O'Malley K, et al. Racial and ethnic disparities in the use of health services: bias, preferences, or poor communication? J Gen Intern Med 2003 Feb;18(2):146-152 [FREE Full text] [Medline: 12542590]

33.   Hibbard JH, Greene J, Tusler M. Improving the outcomes of disease management by tailoring care to the patient's level of activation. Am J Manag Care 2009 Jun;15(6):353-360 [FREE Full text] [Medline: 19514801]

34.   Greene J, Hibbard JH, Sacks R, Overton V. When seeing the same physician, highly activated patients have better care experiences than less activated patients. Health Aff 2013;32(7):1299-1305. [doi: 10.1377/hlthaff.2012.1409] [Medline: 23836747]

35.   Gallant MP. The influence of social support on chronic illness self-management: a review and directions for research. Health Educ Behav 2003;30(2):170-195. [doi: 10.1177/1090198102251030] [Medline: 12693522]

36.   Spooner T, Rainie L. African-Americans and the internet.: Pew Research Center; 2000. URL: http://www.pewinternet.org/2000/10/22/african-americans-and-the-internet/ [accessed 2017-11-16] [WebCite Cache ID 6v1fdz57B]

37.   Pew Research Center. Mobile Fact Sheet. 2018 Mar 05. URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2018-02-17] [WebCite Cache ID 6xJ3V8SuV]

38.   Smith A. African Americans and technology use: a demographic report.: Pew Research Center; 2014 Jan 06. URL: http://www.pewinternet.org/2014/01/06/african-americans-and-technology-use/ [accessed 2017-11-16] [WebCite Cache ID 6v1fjYOug]

39.   Good MJD. The biotechnical embrace. Cult Med Psychiatry 2001;25(4):395-410. [Medline: 11800072]

40.   Newman D. Top five digital transformation trends in health care.: Forbes URL: https://www.forbes.com/sites/danielnewman/2017/03/07/top-five-digital-transformation-trends-in-healthcare/#3a4a52c12561 [accessed 2017-11-16] [WebCite Cache ID 6v1fwhA36]

Abbreviations

CBPR: Community-Based Participatory Research
CDC: Centers for Disease Control and Prevention
GPS: global positioning system
HIPAA: Health Insurance Portability and Accountability Act
mHealth: mobile health

XSL•FO

RenderX

XSL•FO
RenderX

Viewpoint

The Biopsychosocial-Digital Approach to Health and Disease: Call for a Paradigm Expansion

Alireza Ahmadvand[1*], MD; Robert Gatchel[2*], PhD, ABPP; John Brownstein[3,4*], MD, PhD; Lisa Nissen[1*], BPharm, PhD

[1]School of Clinical Sciences, Faculty of Health, Queensland University of Technology, Brisbane, Australia

[2]Department of Psychology, College of Science, University of Texas at Arlington, Arlington, TX, United States

[3]Department of Pediatrics, Harvard Medical School, Boston, MA, United States

[4]Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, United States

[*]all authors contributed equally

Corresponding Author:
Lisa Nissen, BPharm, PhD
School of Clinical Sciences
Faculty of Health
Queensland University of Technology
2 George Street
Brisbane, 4000
Australia
Phone: 61 7 3138 4404
Fax: 61 7 3138 1534
Email: l.nissen@qut.edu.au

Abstract

Digital health is an advancing phenomenon in modern health care systems. Currently, numerous stakeholders in various countries are evaluating the potential benefits of digital health solutions at the individual, population, and/or organizational levels. Additionally, driving factors are being created from the customer-side of the health care systems to push health care providers, policymakers, or researchers to embrace digital health solutions. However, health care providers may differ in their approach to adopt these solutions. Health care providers are not assumed to be appropriately trained to address the requirements of integrating digital health solutions into daily everyday practices and procedures. To adapt to the changing demands of health care systems, it is necessary to expand relevant paradigms and to train human resources as required. In this article, a more comprehensive paradigm will be proposed, based on the 'biopsychosocial model' of assessing health and disease, originally introduced by George L Engel. The "biopsychosocial model" must be leveraged to include a "digital" component, thus suggesting a 'biopsychosocial-digital' approach to health and disease. Modifications to the "biopsychosocial" model and transition to the "biopsychosocial-digital" model are explained. Furthermore, the emerging implications of understanding health and disease are clarified pertaining to their relevance in training human resources for health care provision and research.

Introduction

Digital health is an advancing phenomenon in modern societies, and their health care systems. The US Food and Drug Administration identifies "digital health" as, "a broad scope which includes mobile health (mHealth), health information technology, wearable devices, telehealth and telemedicine, and personalized medicine." The US Food and Drug Administration perceives benefits of digital health to include expanded access to health care services, reduced health care system inefficiencies and costs, improved quality of care, enhanced self-management, and more personalized approaches towards medicine [1].

Driving Factors of Digital Health and the "Digitally-Engaged Patient"

Currently, numerous stakeholders in developed and developing countries are continuously exploring and evaluating the potential

XSL·FO
RenderX

benefits of digital health solutions at the individual, population, and/or organizational levels. The advancement of the digital health space has formed important driving factors including a progressive desire for innovation by industry players and developers of digital health solutions. Aspects of consideration is the increasing demand from societies to embrace more individualized, yet engaging, continuous health care services; the empowerment of health care system clients; and the growing diversity, affordability, and efficiency of digital health solutions.

Interestingly, these driving factors are creating "pull" effects from the customer-side, to 'push' health care providers, policymakers, or researchers to re-examine the traditional paradigms to ultimately address and embrace digital health solutions. Recent examples are the joint efforts by the US Department of Health and Human Services and the UK's National Health Service for the adoption of digital health, and the adoption of principles that promote safe and effective mHealth applications by the American Medical Association [2,3]. Another implementation avenue of digital technologies has been in the management of chronic conditions such as pain, which has caused an influx of scientific literature on digital health solutions [4,5]. Additionally, the notion of 'the digitally-engaged patient' is emerging across health care policy, service provision, and research domains.

Digitally-engaged patients are demanding more independence in their choices, voices, and shared decision-making. They may demand more—sometimes radical—digital innovations in the delivery modes of health care services, evaluating diseases, psychological conditions and social behaviors, and analyzing generated health care data. However, health care providers may differ in their approaches to adopting digital health solutions. Specifically, health care providers' knowledge, attitudes, skills, and practices regarding digital health may vary depending on the context or condition [6]. To maximize the benefits of modern digital technologies in improving patient outcomes, England has reported the need for clinicians to use their expertise to improve health care and redesigning care utilizing digital health [7]. Additionally, in the UK, studies have shown that investment in provision of additional training for both professionals and the public would help strengthen the normalization, uptake, and use of digital health and wellness services [8]. The World Health Organization's mHealth Technical Evidence Review Group, provides another interesting example, by proposing the mHealth Evidence Reporting and Assessment checklist to boost both comprehensiveness and quality in the reporting by health care practitioners, on the effectiveness of digital health programs [9].

Approaches to Digital Empowerment of Patients

To address the effectiveness, safety, and security of digital health solutions, it is still logical to expect that most of the information exchange and decision support at the individual level may essentially happen between health care providers and their patients during in-person or virtual visits. Nevertheless, no one can assume that health care providers are equipped or trained appropriately to responsibly react to the new demands for the

usage of digital health solutions, and the requirements of their integration into usual daily practices. Due to these technological advances, it is necessary to re-evaluate and expand relevant paradigms used for training human resources for health care systems and service provision.

Scientific literature focused on the effects of digital empowerment paradigms on patients and their effects on health care provision highlights two strong mainstream approaches: "techno-utopian," and "techno-critical." "Techno-utopian" idealistically approaches the revolutionary effects of digital health with an optimistic perspective. Contrastingly, the "techno-critical" approach is more pragmatic, emphasizing the inherent complexities, both for patients and for health care providers, in managing health and disease conditions. The latter approach focuses on the possible associations between psychological and sociocultural dimensions of patients' engagement in their own health care via digital health. The main dimensions may include: provoked emotions and possible resistance, regulatory and disciplinary issues, perceived contribution towards the burden of care, and the requirements for operation of "unseen work" on the part of health care workers, attributed to digital technology [10].

Need for More Comprehensiveness About Digital Health

To suggest a more comprehensive approach, the biopsychosocial (BPS) model, introduced by George L Engel [11] has been adapted to include technology. This individual-level model of assessing health and disease, previously expanded in a heuristic manner by other scientists [12-14], must include a digital component. The proposed, "biopsychosocial-digital" model expands the traditional three-tier domains of understanding health and diseases (ie, biological, psychological, and social), in addition to training future human resources of health care systems. Furthermore, this model may act as a generic, neutral, and extended basis for the assessment of the possible effects and interactions of digital health technologies. Therefore, the hopes are that the "biopsychosocial-digital" model will bridge the techno-utopian and techno-critical approaches.

The BPS model emphasizes that the causes and outcomes of health and disease conditions should be considered in lieu with biological, psychological, and social factors. The model implies that for optimal management of conditions, the health care team should address all three influences on the patient. Subsequently, health care providers must be directly, or indirectly, familiar with applying this heuristic model for optimal care. Health care providers must acknowledge that the patients' engagement in health and disease management is influenced by their concurrent medical conditions, psychological factors, and sociocultural barriers in the environment [11]. The BPS model has been widely used in different health-related disciplines to understand the nature of various disease conditions and the provision of training to professionals. Even though limitations exist with this model, most have been shown to be unfounded [15]. Limitations include the dichotomization between biology, psychology, and society, as well as the masking of an underlying biomedical approach [16]. Digital health technologies are

expected to address these limitations by the provision of more accessible, scalable, and comprehensive data from various aspects of health and disease. Furthermore, the availability of various technologies which can continue to capture biological, psychological, and social information provide a non-dichotomized understanding of an individual's medical condition and help in unmasking the biomedical focus of the original model.

Reasons Behind the Call and Methodology for its Justification

In summary, to justify the BPS expansion, the fundamental effects of various digital health technologies on individual components of the original BPS model (ie, biological, psychological, and social) are discussed and supported by the provision of a number of pertinent examples. The potential mechanisms of how digital health may directly or indirectly affect the above-mentioned domains are beyond the scope of this current viewpoint manuscript. Nevertheless, one of the potential mechanisms involves the rapid changes in health care resulting from the adaptation of digital health technologies in various aspects of health and disease. In recent years, there have been increased efforts for the documentation, reflection, and analysis of the changes caused by digital health in the academic, critical psychosocial scientific literature [10].

A summary of the important reasons behind the paradigm expansion, from "biopsychosocial" to "biopsychosocial-digital" is provided below.

Firstly, individual-level digital biological data, both provider-generated and patient-generated, are becoming increasingly available and accessible to people and patients. As a fundamental example, DNA as a digital molecule is becoming the cornerstone of digital genomics services. Interestingly, portable biosensors have proven useful in providing information to support managing health and enabling affordable access to populations in low socioeconomic situations and/or remote geographical environments [17,18]. Additionally, the integration of personal self-tracking data with electronic medical records is already being implemented in some American hospitals [19]. Consequently, the digital "expansion" of one's "biological self," has become an expected outcome, as people are able to explore their biology through new methods.

Secondly, evidence shows that digital health solutions may directly or indirectly affect psychosocial processes that are part of the complex interactions of the individuals' environments [20]. Digital health solutions may be additional sources of both positive and negative psychological reactions at the individual level (eg, personal reassurance from supported decisions by apps, or escalated anxiety induced by lost health data). Overall, digital health solutions have produced greater patient satisfaction [21]. Some authors have highlighted a paradigm shift, through digital technologies, in delivering psychological behavior change interventions [22]. As a result, health care providers should be systematically equipped to address these new psychological aspects of digital health solutions.

Thirdly, socialization around health, both at individual and group levels, has become increasingly popular in the digital space. The popularity of this trend shows both favorable and unfavorable effects as there is improved access to individuals in hard-to-reach sociocultural environments. Also, improving quality-of-care by patients' social media inputs has been considered a promising area in US hospitals [23]. Conversely, social isolation because of breached privacy and hacked health data has become an increasing concern. Figure 1 compares the conventional biopsychosocial model (panel A) with the proposed biopsychosocial-digital model (panel B).

A three-step methodology provides justification of the biopsychosocial model expansion. The details of the methodology used are beyond the scope of this manuscript. In summary, a systematic review of literature was conducted to determine supporting evidence about the relationship between digital health technologies and independent or interrelated components of the biopsychosocial model. A gap analysis was performed to identify potential domains that could be incorporated to the original model. Lastly, an expert consensus report was conducted on the new model to provide feedback and critique through discussion.

Implications of the Paradigm Expansion

The biopsychosocial-digital paradigm could have emerging implications in understanding health and disease and therefore, in training human resources for health care provision and research. Furthermore, the new biopsychosocial-digital paradigm offers a systematically-expanded means of addressing patients' experiences of digital health solutions with their respective biological, psychological, and social aspects of health and disease.

Additionally, this expanded paradigm emphasizes why and how the knowledge of new digital health solutions, as they appear, is critical for training health care providers. Training is essential for improving health care providers' understanding of the possible effects of digital health solutions on the three conventional aspects of health and disease. Subsequently, such effects can be successfully taught to health care providers [24]. Moreover, this paradigm goes beyond assessing acceptability, usability, and satisfaction of digital health solutions, allowing a deeper understanding of their usage. Insights from the biopsychosocial-digital paradigm helps anticipate and interpret the usage and outcomes of digital health solutions towards individualized health solutions. Interestingly, even legal frameworks such as Legal Challenges in Digital Health, support the development and evaluation of digital health services [25] and the proposed model further strengthens those frameworks.

The biopsychosocial-digital model further complements the growing trend of integrating digital health services through joined interdisciplinary teams, to provide better care and to address patients' needs. Therefore, health care providers will consider the complex interactions among digital solutions and biological, psychological, and social aspects of health and disease, rather than the traditionally overly-simplistic biomedical-espoused causal and associated processes.

Figure 1. Comparison between the conventional biopsychosocial model (panel A) with our proposed paradigm (panel B), along with relevance of the proposed paradigm to six core digital health categories that operate within business to patient/consumer market contexts. B: biological; P: psychological; S: social; D: digital.
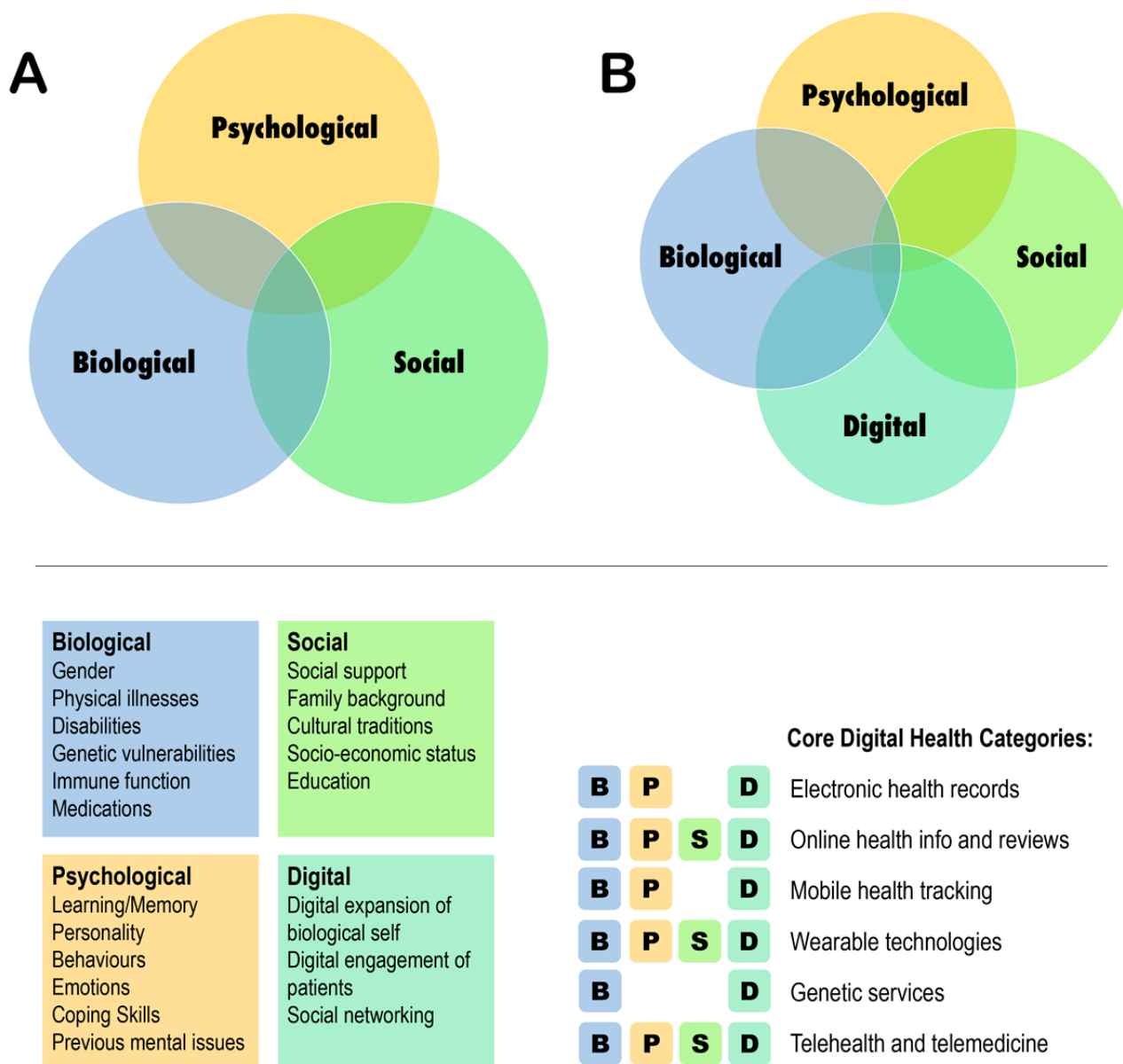
Influencing Medical Education and Clinical Practice

The scientific literature is being enriched continuously with the possible implications of digital health technologies for medical education. Two important aspects are worth noticing: educating future health care practitioners about current and foreseeable technological innovations and enabling practitioners to adjust accordingly to future changes in their fields of specialty [26]. Specifically, in some disciplines, education is focused on the impact and advancements of modern digital technologies on patient care [27].

Various case studies discuss the incorporation of digital health technologies such as Blockchain use in health care, digital health device assessment, mobile health apps, pharmacogenomics and personalized medicine, use of electronic health records in practice, virtual reality and augmented reality, and wearable devices in clinical practice [26]. The importance of integrating digital technologies into medical education results in the improvement of interprofessional team dynamics, extended opportunities for inter-professional team education, and increased possibilities for interprofessional team practice. Moreover, health knowledge brokerage as another fundamental example of interprofessional collaboration and teaching health info-mediary activities from a patient-centered perspective are novel aspects of digital health incorporation into medical education [28]. The biopsychosocial-digital approach to health and disease benefits medical education initiatives by providing a more inclusive paradigm.

Textbox 1. Summary points from the biopsychosocial-digital model.

- Digitally-engaged patients are demanding more independence, in their choices, voices, and shared decision-making.

- Health care providers' knowledge, attitudes, skills, and practices regarding digital health may vary notably, depending on the context or condition.

- A newer biopsychosocial-digital paradigm offers a systematically-expanded means of addressing patients' experiences of intended or unintended interactions of digital health solutions with their health and disease conditions.

- The new paradigm also helps in expanding the research and dialogue approaches to the perspectives of the digitally-engaged patients and to address the pros and cons of digital health solutions by health care providers.

The proposed paradigm has implications for the education of patients, as an important part of clinical practice. Recent efforts focused on the development of more comprehensive instruments to assess digital health literacy or eHealth literacy (eg, Digital Health Literacy Instrument) [29] will be better aligned with the biopsychosocial-digital model.

Certain methods might be used to support the integration of digital health concepts into medical education and clinical practice. Methods include systematic and regular revision of the educational curricula by medical universities, updating clinical practice guidelines and/or learning outcomes by professional medical colleges and providing examples, information, advice, and assistance about relevant digital health technologies. Moreover, the expansion of research efforts in different medical disciplines will increase the understanding of digital health utilization across the health and disease spectra whilst benefitting medical education and clinical practice. Textbox 1 summarizes key information from the biopsychosocial-digital model.

Conclusions

Limitations of the New Viewpoint

The original intention for proposing the new biopsychosocial-digital approach to health and disease was to generate further discussion and feedback. However, the ultimate expansion of the paradigm, requires future research beyond the scope of this manuscript to generate empirical data for potential substantiations of the proposed model. The new model provides value to future studies which explore the interrelationships between the different domains of the paradigm.

Conclusion and Call for Action

The proposed biopsychosocial-digital paradigm helps expand the research and dialogue that accommodates the digitally-engaged perspective of patients around their health and disease. This model plausibly helps in the training of health care professionals by providing them a combination of clinical and digital skills. Professionals are able to benefit from the positive aspects of digital health solutions, address their challenges, and ultimately overcome negative aspects.

In the near future, an extension of research efforts to validate and provide feedback on this new biopsychosocial-digital paradigm, using data from real-world settings will be required. Furthermore, designing applied projects which address the integration of biopsychosocial-digital paradigm into the educational curricula across health and medical disciplines will be necessary for health care providers.

Authors' Contributions

AA and LN created the first idea of the viewpoint manuscript. RG and JB contributed to expanding the conceptualization and methodology. AA and LN wrote the first draft of the manuscript. AA, RG, JB, and LN contributed equally to the editing and revising of the manuscript in four rounds. LN is the corresponding author of the manuscript.

Conflicts of Interest

None declared.

References

1. US Food and Drug Administration. 2017. Digital Health URL: http://www.fda.gov/medicaldevices/digitalhealth/ [accessed 2018-03-21] [WebCite Cache ID 6vyUZplO1]
2. NHS England. Harnessing the Information Revolution URL: https://www.england.nhs.uk/digitaltechnology/info-revolution/ [accessed 2018-03-21] [WebCite Cache ID 6vyUZplO1]
3. American Medical Association. 2017. AMA Adopts Principles to Promote Safe, Effective mHealth Applications URL: http://www.ama-assn.org/ama-adopts-principles-promote-safe-effective-mhealth-applications [accessed 2018-03-21] [WebCite Cache ID 6vyVF255XYY]

XSL•FO

RenderX

4.   Heapy AA, Higgins DM, Cervone D, Wandner L, Fenton BT, Kerns RD. A Systematic Review of Technology-assisted Self-Management Interventions for Chronic Pain: Looking Across Treatment Modalities. Clin J Pain 2015 Jun;31(6):470-492. [doi: 10.1097/AJP.0000000000000185] [Medline: 25411862]

5.   McGeary DD, McGeary CA, Gatchel RJ. A comprehensive review of telehealth for pain management: where we are and the way ahead. Pain Pract 2012 Sep;12(7):570-577. [doi: 10.1111/j.1533-2500.2012.00534.x] [Medline: 22303839]

6.   Topooco N, Riper H, Araya R, Berking M, Brunn M, Chevreul K. Attitudes towards digital treatment for depression: a European stakeholder survey. Internet Interv 2017 Jun;8:1-9. [doi: 10.1016/j.invent.2017.01.001]

7.   Sood H, NcNeil K, Keogh B. Chief clinical information officers: clinical leadership for a digital age. BMJ 2017 Dec 10;358:j3295. [Medline: 28694288] [doi: 10.1136/bmj.j3295]

8.   Lennon MR, Bouamrane M, Devlin AM, O'Connor S, O'Donnell C, Chetty U, et al. Readiness for Delivering Digital Health at Scale: Lessons From a Longitudinal Qualitative Evaluation of a National Digital Health Innovation Program in the United Kingdom. J Med Internet Res 2017 Feb 16;19(2):e42 [FREE Full text] [doi: 10.2196/jmir.6900] [Medline: 28209558]

9.   Agarwal S, Lefevre AE, Labrique AB. A Call to Digital Health Practitioners: New Guidelines Can Help Improve the Quality of Digital Health Evidence. JMIR Mhealth Uhealth 2017 Oct 06;5(10):e136 [FREE Full text] [doi: 10.2196/mhealth.6640] [Medline: 28986340]

10. Lupton D. The digitally engaged patient-monitoring and self-care in the digital health era. Soc Theory Health 2013 Aug;11(3):256-270. [doi: 10.1057/sth.2013.10]

11. Engel GL. The biopsychosocial model and the education of health professionals. Ann N Y Acad Sci 1978 Jun 21;310:169-187. [Medline: 290321]

12. Gatchel RJ, McGeary DD, McGeary CA, Lippe B. Interdisciplinary chronic pain management: past, present, and future. Am Psychol 2014;69(2):119-130. [doi: 10.1037/a0035514] [Medline: 24547798]

13. Gatchel RJ, Peng YB, Peters ML, Fuchs PN, Turk DC. The biopsychosocial approach to chronic pain: scientific advances and future directions. Psychol Bull 2007 Jul;133(4):581-624. [doi: 10.1037/0033-2909.133.4.581] [Medline: 17592957]

14. Borrell-Carrió F, Suchman AL, Epstein RM. The biopsychosocial model 25 years later: principles, practice, and scientific inquiry. Ann Fam Med 2004;2(6):576-582 [FREE Full text] [doi: 10.1370/afm.245] [Medline: 15576544]

15. Gatchel RJ, Turk DC. Criticisms of the biopsychosocial model in spine care: creating and then attacking a straw person. Spine (Phila Pa 1976) 2008 Dec 01;33(25):2831-2836. [doi: 10.1097/BRS.0b013e31817d24ad] [Medline: 19050589]

16. Benning TB. Limitations of the biopsychosocial model in psychiatry. Adv Med Educ Pract 2015 May;6:347-352. [doi: 10.2147/AMEP.S82937] [Medline: 25999775]

17. Li X, Dunn J, Salins D, Zhou G, Zhou W, Schüssler-Fiorenza RSM, 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]

18. Guendelman S, Broderick A, Mlo H, Gemmill A, Lindeman D. Listening to Communities: Mixed-Method Study of the Engagement of Disadvantaged Mothers and Pregnant Women With Digital Health Technologies. J Med Internet Res 2017 Jul 05;19(7):e240 [FREE Full text] [doi: 10.2196/jmir.7736] [Medline: 28679489]

19. Lupton D. Towards critical digital health studies: Reflections on two decades of research in health and the way forward. Health (London) 2016 Jan;20(1):49-61. [doi: 10.1177/1363459315611940] [Medline: 26487686]

20. 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]

21. Fisher E, Law E, Palermo TM, Eccleston C. Psychological therapies (remotely delivered) for the management of chronic and recurrent pain in children and adolescents. Cochrane Database Syst Rev 2015 Mar 23(3):CD011118 [FREE Full text] [doi: 10.1002/14651858.CD011118.pub2] [Medline: 25803793]

22. Moller AC, Merchant G, Conroy DE, West R, Hekler E, Kugler KC, et al. Applying and advancing behavior change theories and techniques in the context of a digital health revolution: proposals for more effectively realizing untapped potential. J Behav Med 2017 Feb;40(1):85-98. [doi: 10.1007/s10865-016-9818-7] [Medline: 28058516]

23. Ranney ML, Genes N. Social media and healthcare quality improvement: a nascent field. BMJ Qual Saf 2016 Jun;25(6):389-391. [doi: 10.1136/bmjqs-2015-004827] [Medline: 26658773]

24. Dubin RE, Flannery J, Taenzer P, Smith A, Smith K, Fabico R, et al. ECHO Ontario Chronic Pain & Opioid Stewardship: Providing Access and Building Capacity for Primary Care Providers in Underserviced, Rural, and Remote Communities. Stud Health Technol Inform 2015;209:15-22. [Medline: 25980700]

25. Garell C, Svedberg P, Nygren JM. A Legal Framework to Support Development and Assessment of Digital Health Services. JMIR Med Inform 2016 May 25;4(2):e17 [FREE Full text] [doi: 10.2196/medinform.5401] [Medline: 27226391]

26. Clauson K, Aungst T, Simard R, Fox B, Breeden E. Health Professionals' Education in the Age of Clinical Information Systems, Mobile Computing and Social Networks. UK: Elsevier; 2017:181-199.

27. Aungst TD. Integrating mHealth and mobile technology education into the pharmacy curriculum. Am J Pharm Educ 2014 Feb 12;78(1):19 [FREE Full text] [doi: 10.5688/ajpe78119] [Medline: 24558287]

28.   Van ED, Newman K, DeForge R, Urquhart R, Cornelissen E, Dainty KN. Knowledge brokering for healthy aging: a scoping review of potential approaches. Implement Sci 2016 Dec 19;11(1):140 [FREE Full text] [doi: 10.1186/s13012-016-0504-5] [Medline: 27756358]

29.   van DVR, Drossaert C. Development of the Digital Health Literacy Instrument: Measuring a Broad Spectrum of Health 1.0 and Health 2.0 Skills. J Med Internet Res 2017 Jan 24;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]

Abbreviations

BPS: biopsychosocial
mHealth: mobile health

XSL•FO
RenderX

<u>Original Paper</u>

Psychobehavioral Profiles to Assist Tailoring of Interventions for Patients With Hypertension: Latent Profile Analysis

Rika Tanaka[1], PhD; Robert P Nolan[1,2], PhD

[1]Cardiac eHealth and Behavioural Cardiology Research Unit, Peter Munk Cardiac Centre, University Health Network, Toronto, ON, Canada
[2]Department of Psychiatry and Institute of Medical Sciences, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Rika Tanaka, PhD
Cardiac eHealth and Behavioural Cardiology Research Unit
Peter Munk Cardiac Centre
University Health Network
585 University Ave., 6NU-6N617
Toronto, ON, M5G 2N2
Canada
Phone: 1 416 340 4800 ext 6038
Email: rika.tanaka@uhnresearch.ca

Abstract

Background: Practice guidelines advocate combining pharmacotherapy with lifestyle counseling for patients with hypertension. To allow for appropriate tailoring of interventions to meet individual patient needs, a comprehensive understanding of baseline patient characteristics is essential. However, few studies have empirically assessed behavioral profiles of hypertensive patients in Web-based lifestyle counseling programs.

Objective: The objectives of this study were to (1) specify baseline psychobehavioral profiles of patients with hypertension who were enrolled in a Web-based lifestyle counseling trial, and (2) examine mean differences among the identified profile groups in demographics, psychological distress, self-reported self-care behaviors, physiological outcomes, and program engagement to determine prognostic implications.

Methods: Participants (N=264; mean age 57.5 years; 154/264, 58.3% female; 193/264, 73.1% white) were recruited into a longitudinal, double-blind, randomized controlled trial, designed to evaluate an online lifestyle intervention for hypertensive patients. A series of latent profile analyses identified psychobehavioral profiles, indicated by baseline measures of mood, motivation, and health behaviors. Mean differences between profile groups were then explored.

Results: A 2-class solution provided the best model fit (the Bayesian information criterion (BIC) is 10,133.11; sample-size adjusted BIC is 10,006.54; Lo-Mendell-Rubin likelihood ratio test is 65.56, P=.001). The 2 profile groups were (1) adaptive adjustment, marked by low distress, high motivation, and somewhat satisfactory engagement in health behaviors and (2) affectively distressed, marked by clinically significant distress. At baseline, on average, affectively distressed patients had lower income, higher body mass index, and endorsed higher stress compared with their adaptive adjustment counterparts. At 12-months post intervention, treatment effects were sustained for systolic blood pressure and Framingham risk index in the adaptive adjustment group, and those in the adaptive adjustment group were 2.4 times more likely to complete the 12-month intervention study, compared with their affectively distressed counterparts.

Conclusions: Interventions for patients who are adaptively adjusted may differ in focus from those designed for the affectively distressed patients. As such, this study underscores the importance of identifying psychobehavioral profiles, as they allow for evidence-based tailoring of lifestyle counseling programs for patients with hypertension.

Trial Registration: ClinicalTrials.gov NCT01541540; https://clinicaltrials.gov/ct2/show/NCT01541540 (Archived by WebCite at http://www.webcitation.org/6yzZYZcWF)

XSL•FO
RenderX

Introduction

Background

Elevated systolic blood pressure (SBP) places individuals at increased risk of cardiovascular disease (CVD), stroke, coronary heart disease, heart failure, and CVD mortality [1]. Nevertheless, blood pressure is treated and controlled in only two-thirds of those diagnosed with hypertension [2,3], leaving a large proportion of this population at significant risk for the development of further cardiac risk. Current guidelines consider the combination of pharmacotherapy with lifestyle counseling as best practice for the management of hypertension [4,5].

As the feasibility and clinical utility of large-scale motivational interviewing and cognitive behavioral therapy-based lifestyle counseling programs continue to be established and further disseminated for patients with hypertension [6-10], it has become increasingly important to develop a systematic, comprehensive, and efficient way to assess the psychological and behavioral characteristics of the growing number of patients enrolling in such programs. A comprehensive understanding of baseline patient characteristics is essential to allow for more appropriate tailoring of interventions to meet individual patient needs. In keeping with this objective, psychometric instruments have been developed to identify adaptive and maladaptive patterns of adjustment among patients with chronic pain, and preliminary work has been reported for patients with cardiovascular conditions [11,12]. Few studies have empirically assessed psychological or behavioral profiles of hypertensive patients enrolled in Web-based lifestyle counseling programs.

Objectives

The goals of this study were to (1) assess and specify baseline psychobehavioral profiles of patients with hypertension who were enrolled in a Web-based lifestyle counseling trial, (2) examine mean baseline differences among the identified profile groups in demographics, psychological distress, and self-reported self-care behaviors, and (3) assess differences in physiological outcomes (SBP; diastolic blood pressure, DBP; pulse pressure, PP; and Framingham risk index, FRI) and program engagement across these profile groups over 12 months to determine prognostic implications.

Methods

Overview

This is a substudy of the Reducing risk with E-based support for Adherence to lifestyle Change in Hypertension (REACH) trial [13], which was a multicenter, longitudinal, double-blind, randomized controlled trial of e-counseling for persons with stage 1 or 2 hypertension (Clinicaltrial.gov: NCT01541540). REACH was designed to evaluate the efficacy of a standardized, evidence-based e-counseling protocol that promoted adherence to recommended guidelines for exercise, diet, prescribed medications, and smoke-free living over 12 months. In the parent study, eligible hypertensive participants were recruited across 5 Canadian sites: Toronto (n=174), Vancouver (n=39), Grey Bruce (n=19), London (n=15), and Prince Edward Island (n=17). All participants were randomly assigned to treatment or control

interventions after their eligibility was confirmed at their baseline visit. Primary endpoints of the REACH study included SBP and DBP, PP, non-high-density lipoprotein cholesterol, and the FRI of 10-year absolute risk of CVD.

Study Interventions and Assessments

Both the control and e-counseling arms of REACH were organized by sessions that included a URL that linked participants to their session content. For controls, each session included content representative of the e-based support provided by heart health organizations at the inception of the study [13]. In addition to the materials made available to participants assigned to the control condition, the e-counseling intervention used key components from motivational interviewing [14] and cognitive behavioral therapy [15] to promote adherence to self-care behaviors. In keeping with guidelines from motivational interviewing [14], e-counseling sessions in the early phase of the e-counseling intervention were designed to resolve ambivalence about behavior change and to help participants feel connected to a salient personal goal. Subsequent sessions provided videos, online handouts, and monitoring forms to guide and reinforce skills to sustain positive changes in targeted self-care behaviors.

As a part of the parent study, all participants were asked to complete in-person study assessments at baseline, 4-month, and 12-month follow-up. All in-person study assessments were conducted by a trained nurse or research assistant and included the collection of both questionnaire and physiological data.

Participants

This study included 263 participants (mean age 57.5 years; 154/264, 58.3% female; 193/264, 73.1% white), who completed the baseline assessment. Inclusion criteria for the larger longitudinal study included the following: age: 35-74 years, hypertension diagnosis, baseline blood pressure measured at baseline study session: ≥140/90 (if no meds); ≥130/85 (if on meds); if on medications, an unchanged prescription for ≥2 months, and comprehension of written and oral English [13]. Exclusion criteria for the larger study included the following: diagnosis of clinically significant arrhythmia, sleep apnea, kidney disease, major psychiatric illness (eg, psychosis), alcohol or drug dependence in the previous year, institutional residence, or little to no English comprehension [13]. One participant from the larger sample was excluded from this study because of missing data on all indicators used for the analysis of psychobehavioral profiles. See Multimedia Appendix 1 for background characteristics, health behavior, and cardiovascular risk factors for total sample of parents study at baseline.

Measures

Psychological Distress

Four well-established self-reported measures of psychological distress were used in this study. The Patient Health Questionnaire-9 (PHQ-9) [16] is a 9-item diagnostic screening tool that assesses symptoms of depression. The clinical interpretations of scores on this measure are as follows: 0-4=minimal depression; 5-9=mild depression; 10-14=moderate depression; 15-19=moderately severe depression; and

20-27=severe depression [16]. Both the Anxiety and Depression subscales of the Hospital Anxiety and Depression Scale (HADS) [17] also assessed symptoms of anxiety and depression, whereas the Perceived Stress Scale [18] was used to assess overall psychological distress. Only baseline measures of self-reported psychological distress were used this study [19].

Health Behaviors

Baseline measures of physical activity and dietary behaviors were used in this study. Physical activity was measured by calculating the 4-day number of steps recorded on a triaxial pedometer (4-day step count, LifeSource/A&D XL-18CN Activity Monitor, China) [13]. In a study of healthy adults, individuals have been classified into various lifestyle groups based on average daily step counts: (1) <5000 steps per day=sedentary lifestyle; (2) 5000-7499 steps per day=low active; (3) 7500-9999 steps per day=somewhat active; (4) ≥10,000 steps per day=active; and (5) >12,500 steps per day=highly active [20]. Moreover, according to the Canadian Hypertension Education Program (CHEP) 2016 guidelines, it is recommended that patients with hypertension engage in 30 to 60 min of moderate-intensity exercise 4 to 7 days/week, including walking, in addition to their activities of daily living [5].

Dietary behaviors were monitored by a 24-hour urinary sodium analysis (mmol/day) and the National Institute of Health/National Cancer Institute Dietary Health Questionnaire, a self-reported measure of fruit and vegetable intake, which has established validity and has been successfully adapted for a Canadian population [21]. According to the CHEP 2016 guidelines, it is recommended that patients with hypertension consume no more than 87 mmol of sodium daily and a diet high in fruits and vegetables [5].

Motivational Readiness to Change

Prochaska transtheoretical algorithm [22] was used as a proxy measure to assess baseline motivation to initiate or maintain self-management behaviors, including planned exercise, daily activities for active living, fruit and vegetable intake, and salt use [22]. The stages of change are conventionally defined as follows: 1=precontemplation (not ready to adhere to the target behavior in the next 6 months); 2=contemplation (ready to adhere to the target behavior in the next 6 months); 3=preparation (ready to adhere to the target behavior in the next 4 weeks); 4=action (adherence to the behavior but for less than 6 months); and 5=maintenance (adherence to the behavior for 6 months or more).

Physiological Measures

SBP, DBP, and PP, as well as the FRI for 10-year absolute risk of CVD [23], were assessed at baseline and 12-month endpoint. All assessments took place at the participating hospital or clinic, by a trained nurse or research assistant as noted for the REACH trial [13]. Blood pressure was assessed using a validated protocol

for automated office blood pressure measurement with the BpTRU device [5]. Blood samples were taken by the hospital or clinic laboratory using conventional procedures to assess the 12-hour fasting lipoprotein cholesterol profile. The FRI was obtained from the previously noted data along with a questionnaire [23]. In addition, baseline body mass index (BMI) was calculated (kg/m^2).

Program Engagement

Study completion was used as a measure of participant engagement in their assigned treatment programs. Study completion was coded as a binary measure such that 0=Incomplete, assigned to participants who did not complete the in-person 12-month study assessment, and 1=Complete, assigned to participants who completed the in-person 12-month study assessment.

Data Analyses

Latent Profile Analysis

A series of latent profile analyses (LPA) were conducted within a structural equation modeling framework to obtain psychobehavioral profiles of patients with hypertension. Psychobehavioral profiles were indicated by baseline measures of mood (PHQ-9), motivation (readiness to change: exercise and diet), and health behaviors (4-day Step Count and Urinary Sodium, see Figure 1). To determine the appropriate number of profiles, a series of models with sequentially increasing number of classes were tested for overall model fit. Each model was compared with the previous model by examining multiple indices of model fit (eg, the Bayesian Information Criterion, BIC; the sample-size adjusted BIC, ABIC; and the Lo-Mendell-Rubin test, LMR [24]). Improved model fit was indicated by smaller BIC and ABIC and a significant LMR test, and worsened model fit was evidenced by larger BIC and ABIC and nonsignificant LMR test. Patients with hypertension were then categorized into different psychobehavioral profile groups, based on saved profile classifications resulting from the LPA.

Analysis of Baseline and 12-Month Differences Between Profiles

Once psychobehavioral profile groups were established, a series of t tests and Pearson chi-square tests were performed to test the differences between psychobehavioral profile groups on baseline demographic, physiological, psychological, and behavioral variables. A series of analysis of covariance (ANCOVA) analyses were also conducted to test mean differences in baseline to 12-month assessment change in SBP, DBP, PP, and FRI between treatment groups within each of the psychobehavioral profile groups. Assessments of mean differences were made controlling for age, sex, BMI, and baseline outcome variable. Logistic regression was then used to determine differences in program engagement between profile groups.

Figure 1. Theoretical model of latent psychobehavioral profiles of patients with hypertension.
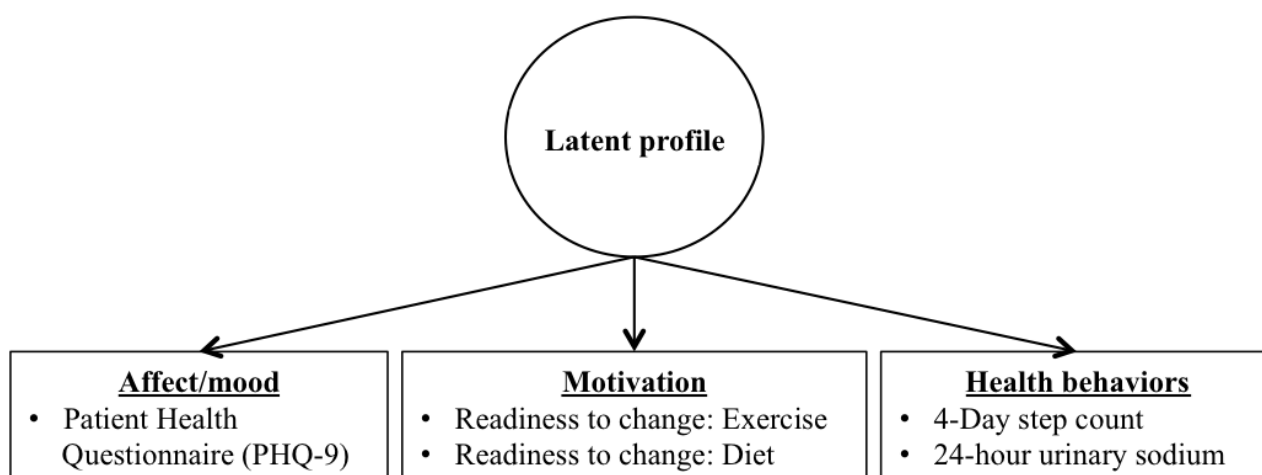
Results

Latent Profile Analysis

Model Fit

A series of models with sequentially increasing number of psychobehavioral profiles of patients with hypertension were tested for overall model fit. First, indicators of psychobehavioral profiles (eg, baseline measures of mood, motivation, and health behaviors) were assessed in a 1-class LPA (BIC=10,166.92; ABIC=10,119.36). Second, a 2-class LPA using indicators of psychobehavioral profiles was tested (BIC=10,133.11; ABIC=10,006.54; LMR=65.56; P=.001). This analysis was followed by a 3-class LPA using these same indicators (BIC=10,116.72; ABIC=10,031.11; LMR=48.59; P=.59). It was determined that the 2-class solution provided the best model fit because the 3-class model of psychobehavioral profiles did not significantly improve overall model fit from the 2-class model, and the 2-class model was found to have significantly improved model fit over the 1-class model (Figure 2).

Psychobehavioral Profile Groups

The 2-class solution revealed 2 main psychobehavioral profiles of hypertensive patients. Most patients (228/263, 86.7%) were found to be adaptively adjusting to their hypertension diagnosis. The psychobehavioral profile of the adaptive adjustment group was marked by symptoms of depression in the minimal range of clinical severity (mean$_{PHQ}$ 3.69, SD 0.23), relatively high motivation to adhere to guidelines for both diet and exercise (mean$_{readiness diet}$ 3.98, SD 0.05; mean$_{readiness exercise}$ 3.98, SD 0.07, both approaching action stage), and somewhat active engagement in physical activity (mean$_{steps}$ 7900.38, SD 222.25). Nevertheless, on average, the adaptive adjustment group showed very poor adherence to a low-sodium diet as indicated by 24-hour urinary sodium (mean$_{sodium}$ 130.81, SD 4.47; Figure 2).

A minority of patients (13.3%, 35/263) were classified in the second psychobehavioral profile group, which was marked by moderately elevated levels of depression, indicative of clinically significant distress (mean$_{PHQ}$ 13.39, SD 0.83). This affectively distressed group also demonstrated relatively lower levels of motivation to exercise (mean$_{readiness exercise}$ 3.34, SD 0.25, preparation stage) and physical activity (mean$_{steps}$ 7165.43, SD 562.44, low active). Although this group indicated motivation to change dietary behavior indicative of individuals approaching the action stage (mean$_{readiness diet}$ 3.85, SD 0.13), adherence to sodium intake guidelines was poor (mean$_{sodium}$ 121.44, SD 9.74; Figure 2).

Baseline Mean Differences Between Profile Groups
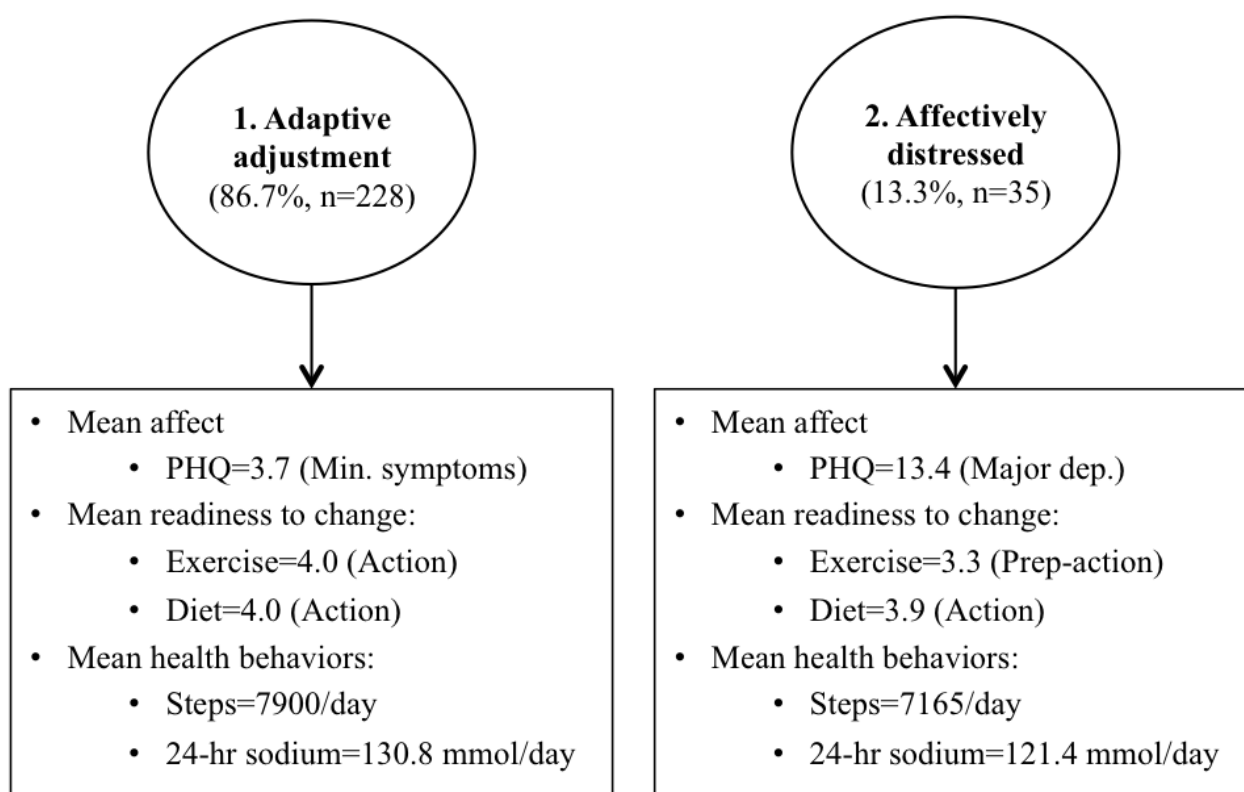
Demographic Differences

The t tests revealed a significant difference in mean baseline income between the psychobehavioral profile groups (t_{237}=−3.73, P<.001), with the adaptive adjustment group having a higher average income compared with the affectively distressed group. No significant baseline differences were found between the 2 profile groups on baseline age, gender, or level of education (Table 1).

Physiological Differences

The t tests revealed a significant difference in mean baseline BMI between the psychobehavioral profile groups (t_{261}=2.95, P=.003), with the adaptive adjustment group having a lower BMI (mean$_{BMI}$ 30.63, SD 6.10), compared with the affectively distressed group (mean$_{BMI}$ 33.97, SD 7.14). No significant baseline differences were found between the 2 profile groups on SBP, DBP, or PP (Table 1).

Psychological Differences

The t tests revealed significant differences between the psychobehavioral profile groups across several measures of psychological distress, including HADS-anxiety (t_{249}=8.10, P<.001), HADS-depression (t_{251}=11.15, P<.001), and Perceived Stress Scale (t_{247}=6.66, P<.001). As expected, across all 3 of these measures, the adaptive adjustment group had a lower mean level of distress compared with the affectively distressed group (Table 1).

XSL•FO
RenderX

Figure 2. Two psychobehavioral profiles of patients with hypertension. PHQ: Patient Health Questionnaire.

Behavioral Differences

The *t* tests revealed a difference in fruit and vegetable intake between the psychobehavioral profile groups that was approaching significance ($t_{241}=-1.87$, $P=.06$) at baseline. The adaptive adjustment group reported, on average, higher levels of fruit and vegetable intake ($mean_{servings}$ 8.36, SD 5.55), compared with their affectively distressed counterparts ($mean_{servings}$ 6.43, SD 4.16; Table 1).

Twelve-Month Treatment Effects Within Profile Groups

Adaptive Adjustment Group

In keeping with the therapeutic changes at 12 months reported for the REACH trial [25], ANCOVA analyses revealed significant differences in change in SBP, PP, and FRI from baseline to 12-month follow-up, ($F_1=5.80$, $P=.02$; $F_1=4.27$, $P=.04$; $F_1=5.39$, $P=.02$, respectively), even when controlling for age, sex, baseline BMI, and baseline values. Adaptively adjusted participants assigned to the treatment group showed, on average, a greater decrease in change in SBP, PP, and FRI compared with their control group counterparts. Nevertheless,

for adaptively adjusted participants, mean change in DBP from baseline to 12-month follow-up did not differ significantly between treatment and control groups ($F_1=2.28$, $P=.13$; Table 2).

Affectively Distressed Group

ANCOVA analyses were not conducted for the affectively distressed group because of the small number of participants assigned to each group ($n_{control}=16$, $n_{treatment}=4$) and the wide variability in change in SBP, PP, and FRI, from baseline to 12-month follow-up (Table 3).

Differences in 12-Month Program Engagement Between Profile Groups

Logistic regression found a significant positive effect of profile group on program engagement, as assessed by completion of the in-person 12-month study assessment (beta=0.88, SE=0.386, $P=.02$), even when controlling for age, sex, baseline SBP, and treatment group. Adaptively adjusted participants were 2.41 times more likely than the affectively distressed participants to complete the in-person study assessments, Exp(B)=2.41, regardless of age, gender, baseline SBP, or assigned treatment group.

Table 1. Assessing baseline outcome mean differences between profile groups. BMI: body mass index; BP: blood pressure; HADS: Hospital Anxiety and Depression Scale.

Outcome variable	Profiles		t test (df)	P value
	Adaptive adjustment (n=228)	Affectively distressed (n=35)		
Demographics				
Females, n	131	23	N/A[a]	N/A
Age, mean (SD)	57.88 (9.53)	55.51 (9.32)	−1.37 (261)	.17
Income[b], mean (SD)	7.46 (2.80)	5.47 (2.86)	−3.73 (237)	<.001
Education, mean (SD)	16.22 (2.62)	15.28 (3.01)	2.03 (250)	.06
Physiological, mean (SD)				
Systolic BP	141.02 (11.46)	140.17 (11.45)	−0.41 (261)	.69
Diastolic BP	87.26 (8.67)	87.26 (8.34)	−0.001 (261)	.99
Pulse pressure	53.76 (12.51)	52.91 (12.53)	−0.37 (261)	.71
Framingham risk index	16.24 (10.71)	15.11 (12.03)	−0.56 (257)	.58
BMI	30.63 (6.10)	33.97 (7.14)	2.95 (261)	.003
Psychological, mean (SD)				
Patient Health Questionnaire	3.69 (.23)	13.39 (.83)	N/A	N/A
HADS: Anxiety	5.40 (3.43)	10.59 (3.74)	8.1 (249)	<.001
HADS: Depression	2.96 (2.64)	8.53 (3.13)	11.15 (251)	<.001
Perceived stress	14.75 (5.04)	20.97 (4.70)	6.66 (247)	<.001
Behavioral, mean (SD)				
4-day step count	7900.38 (222.25)	7165.43 (562.44)	N/A	N/A
24-hour urinary sodium[c]	130.81 (4.47)	121.44 (9.74)	N/A	N/A
Fruit and vegetable intake	8.36 (5.55)	6.43 (4.16)	−1.89 (241)	.06
Motivational, mean (SD)				
Exercise	3.98 (0.05)	3.34 (0.25)	N/A	N/A
Diet	3.98 (0.07)	3.85 (0.13)	N/A	N/A

[a]N/A: not applicable.

[b]Self-reported income per family (Can $), 1=≤$19,000, 10≥$100,000.

[c]89.4% (236/264) of all participants completed the baseline 24-hour urinary sodium assessment.

Table 2. Change in main outcomes at 12 months in adaptive adjustment group by treatment allocation. DBP: diastolic blood pressure; FRI: Framingham risk index; PP: pulse pressure; SBP: systolic blood pressure.

Adaptive adjustment	Treatment group		F test (df)	P value
	e-Counseling (n=96), mean (SD)	Control (n=80), mean (SD)		
ΔSBP	−10.51 (1.31)	−5.81 (1.44)	5.80 (1)	.02
ΔDBP	−5.15 (0.78)	−3.40 (0.86)	2.28 (1)	.13
ΔPP	−5.40 (0.90)	−2.63 (0.99)	4.27 (1)	.04
ΔFRI	−2.24 (0.64)	−0.02 (0.70)	5.39 (1)	.02

Table 3. Change in main outcomes at 12 months in affectively distressed group. DBP: diastolic blood pressure; FRI: Framingham risk index; PP: pulse pressure; SBP: systolic blood pressure.

Affectively distressed[a]	Treatment group	
	e-Counseling (n=4), mean (SD)	Control (n=16), mean (SD)
ΔSBP	−12.00 (8.25)	−6.63 (15.01)
ΔDBP	−11.00 (8.29)	−4.25 (8.51)
ΔPP	−1.00 (15.94)	−2.37 (11.10)
ΔFRI	−2.04 (1.34)	−0.93 (6.05)

[a]ANCOVA analyses were not conducted for the affectively distressed group because of small sample size; however, raw means and standard deviation are reported.

Discussion

Principal Findings

This study identified 2 baseline psychobehavioral profiles for hypertensive patients: adaptive adjustment and affectively distressed. The affectively distressed group had significantly lower income and significantly elevated baseline BMI and levels of distress (eg, anxiety, depression, and perceived stress) and engaged in the Web-based counseling program less than their adaptively adjusted counterparts. Moreover, treatment effects on SBP, PP, and FRI were statistically significant in the adaptive adjustment group but failed to reach statistical significance for DBP. These findings indicate that a large majority of patients with hypertension are likely to benefit greatly from interventions designed to provide practical support regarding adherence to lifestyle recommendations for the management of hypertension. Nevertheless, a minority of patients may also benefit from additional support to help manage psychological symptoms and associated stressors that may interfere with a patient's ability to adhere to Web-based interventions and suggested lifestyle changes.

Psychobehavioral Profiles

Most participants (86.7%, 228/263) in this study were found to be psychologically well-adjusted to their diagnosis and indicated relatively high motivation to engage in both healthy diet and exercise behaviors. Participants categorized in the adaptive adjustment group reported levels of motivation and physical activity within expected ranges for a cohort of patients seeking help in lifestyle behavior changes to manage their blood pressure. For example, the mean readiness for change in exercise and the somewhat active range of physical activity found in the adaptive adjustment group was comparable to a previous report of both motivation to increase physical activity and engagement in physical activity in a large sample of individuals diagnosed with CVD and/or diabetes or who were at high risk of CVD [26]. It is important to recognize that, although this group is well adjusted to their hypertension diagnosis, there is still a need for this group to engage in a comprehensive lifestyle counseling program. For instance, although this group reported high motivation for change in dietary behaviors, their urinary sodium excretions indicated difficulty in adhering to low-sodium diet recommendations [5]. Similarly, although motivation to increase exercise is relatively high in this group, average 4-day step counts indicate suboptimal level of activity [20].

A small proportion of the participants in this study (13.3%, 35/263) was identified as being affectively distressed, as their profile was marked by clinically significant elevations in depressive symptoms. The rate of clinically elevated depressive symptoms is comparable to rates of diagnosis of anxiety or depression previously reported in a large sample of hypertensive patients [27]. Although the affectively distressed group reported comparable levels of motivation for change in exercise as their adaptively adjusted counterparts, this group engaged in relatively lower amounts of physical activity. In addition, the affectively distressed group reported lower motivation to engage in dietary changes, and their urinary sodium excretion was somewhat lower than the adaptively adjusted group. Discrepancies between self-reported ratings of motivation to change and more objective health behavior indices may be an important clinical manifestation of the overall distress experienced by individuals in this psychobehavioral profiles group and may be important to address in interventions tailored for this affectively distressed group of patients with hypertension.

Differences Between Adaptive Adjustment and Affectively Distressed Profile Groups

Baseline Characteristics

Examination of baseline differences between the profile groups worked to further validate the 2 psychobehavioral groups identified in this study. As would be expected, the 2 profile groups differed significantly in their baseline endorsement of psychological distress. On average, the affectively distressed group was more anxious, depressed, and stressed compared with their adaptive adjustment counterparts. The affectively distressed group also reported, on average, a lower household income compared with their adaptively adjusted counterparts. These results indicate that patients who are classified in the affectively distressed group may not only experience clinically elevated symptoms of depression but also the elevated symptoms of a wide range of other psychological symptoms and associated stressors that are also important to acknowledge and address when planning interventions for this subgroup of patients with hypertension.

Although there were no significant differences between these groups on baseline measures of SBP, DBP, or PP, the 2 groups did differ in baseline BMI. The affectively distressed group had significantly higher mean BMI than the adaptive adjustment group. The strong reciprocal association between depressive symptoms and obesity has been well established, and it has been

hypothesized that multiple biological, psychosocial, and behavioral pathways likely account for this association [28]. For example, from a behavioral perspective, the adaptive adjustment group had a higher mean 4-day step count compared with the affectively distressed group. Similarly, the results of this study indicated that, on average, adaptive adjustment group may consume higher levels of fruit and vegetable compared with their affectively distressed counterparts. Because symptoms of depression often include amotivation and decreased engagement in activities, elevations in depressive symptoms can place patients with hypertension at higher risk for noncompliance to self-care recommendations. These differences in health behaviors, in addition to differences in metabolic processes, may contribute to the higher mean BMI in the affectively distressed group. It is important to note, however, that both the adaptively adjusted and affectively distressed groups had, on average, BMIs in the obese range. Because of the reciprocal relation between obesity and depression, it would be important for interventions designed for adaptively adjusted groups to provide basic psychoeducation regarding stress and coping and their potential impact on healthy lifestyle maintenance.

Twelve-Month Outcomes

Analysis of 12-month outcomes within the adaptive adjustment group indicated that the e-counseling program effectively reduced SBP, PP, and FRI for this profile group. Although there was no treatment effect for DBP overall, these outcomes are similar to those found in the primary outcomes paper for the larger study (unpublished data, 2018, [29]) and indicated that for most hypertensive patients, the current e-counseling program is likely effective in lowering risk for the development of CVD. Due to our limited sample size, we were unable to statistically examine these same 12-month outcomes within the affectively distressed group. It is important to note, however, that there was a pattern of greater improvements in outcome measures (SBP, DBP, PP, and FRI) for distressed patients receiving the e-counseling intervention compared with their control counterparts. This pattern of results indicates that, the current e-counseling program may be effective in reducing risk for these patients. However, because of our lack of power to assess this effect, no such conclusions can be made at this time.

Program Engagement

This study also found that the adaptively adjusted participants were 2.4 times more likely than the affectively distressed participants to complete both baseline and 12-month follow-up assessments. This finding is important, as it highlights a potentially significant difference in program engagement between profile groups. This finding is consistent with other studies that have found that psychological symptom severity is an important predictor of adherence to Web-based interventions [30]. Therefore, early identification of those who are significantly distressed and tailoring of interventions to address potential psychosocial barriers to program engagement is likely an important aspect of designing longitudinally effective person-centered internet-based intervention programs for patients with hypertension.

Clinical Implications

As noted previously, psychobehavioral profiles identified in this study are important to consider from a clinical perspective when looking to implement large-scale Web-based lifestyle intervention programs for hypertensive patients, such as the program tested in the parent REACH study [13]. The psychobehavioral profiles found in this study are promising, as they show that most patients with hypertension who enroll in a Web-based behavioral counseling program are likely to be highly motivated and already on the road to effective engagement in recommended self-care behaviors. Therefore, most patients will likely benefit from programs focused on behavioral adjustments to achieve a heart-healthy lifestyle, basic psychoeducation regarding the association between stress and the maintenance of a healthy lifestyle, and strategies to maintain such changes over time.

Nevertheless, this study also identified a minority of patients who indicated that they experienced clinically significant elevations in low mood. Identification of these distressed patients is likely critical. Previous studies have indicated that patient distress or depression impairs their ability to adhere to self-care behavior and to engage in programs focused on promoting therapeutic change in lifestyle [31-33]. Our findings suggest that among patients who endorse significantly higher levels of distress, programs may prioritize supplementary self-help counseling to reinforce cognitive behavioral skills aimed at reducing psychological symptoms and managing associated stressors, before presenting strategies to maintain or adhere to recommended self-care behaviors. It is also interesting to note that a diagnosis of anxiety or depression has been found to be associated with greater health care utilization and faster blood pressure control in patients diagnosed with hypertension [27]. Such studies highlight the complexities of the pathways by which psychological symptoms ultimately influence blood pressure in patients with hypertension, and they are important to keep in mind when developing content for a wide-reaching program likely to recruit a diverse population of patients.

Limitations and Future Directions

Although this study provides a promising new way to tailor Web-based health behavior counseling interventions for patients with hypertension, there are limitations with regard to these results. First, it is important to consider limits to the generalizability of these findings. This study sample represents a cohort of patients with hypertension who actively sought information regarding a Web-based program for self-care adherence. The relatively high levels of motivation and moderate levels of engagement in physical exercise seen in the sample may be a reflection of sampling bias introduced in the recruitment strategy. Our sample comprised health information seekers who had initially landed on the website of a public heart health organization and responded to our invitation about participating in a research project [13]. The overall population of hypertensive patients may include a wider range of psychobehavioral profiles. For example, there may be a subgroup of patients who are both highly distressed and highly unmotivated to change their lifestyle or a subgroup of patients who are highly motivated and who have been highly effective

XSL•FO

RenderX

in adhering to recommended self-care behaviors. Second, this study was underpowered to detect treatment effects within the affectively distressed group, making it difficult to make any definitive conclusions regarding the effects of the e-counseling program on those participants who are most distressed on entering the program. Similarly, because of the greatly different sample sizes between the 2 profile groups, we were unable to conduct a more comprehensive analysis comparing treatment effects directly across these groups. This study was also limited in its ability to assess program engagement. Completion of the in-person 12-month study assessment was used as a proxy for program engagement, as more detailed information regarding program engagement (eg, number of emails opened or online resources accessed) was not available at the time of this study.

Future studies may aim to identify psychobehavioral profiles across a wider range of patients with hypertension to get a more accurate estimate of whether the 2 profiles reported here are replicable. Furthermore, it would be important to examine whether individuals with varying profiles respond differently to interventions aimed at promoting self-care adherence. Future randomized controlled trials may aim to oversample patients who are particularly distressed at baseline to directly examine how treatment effects may differ across varying psychobehavioral profiles. Moreover, future studies would benefit from working to eliminate in-person assessments and collecting detailed information regarding the degree to which participants engage in online programs to better define and tailor such interventions for a heterogeneous population of patients. Nevertheless, this study underscores the importance of identifying and understanding psychobehavioral profiles, as they allow for efficient evidence-based tailoring of lifestyle counseling programs for patients with hypertension.

Conclusions

This study identified 2 latent psychobehavioral profiles for hypertensive patients based on an analysis of baseline characteristics: adaptive adjustment (86.7%, 228/263) and affectively distressed (13.3%, 35/263). Those in the affectively distressed group had significantly lower self-reported household income, elevated BMI, higher levels of distress (eg, anxiety, depression, and perceived stress) and significantly lower program adherence compared with their adaptively adjusted counterparts. The adaptively adjusted patients enrolled in an e-counseling intervention also showed significant improvements in SBP, PP, and FRI compared with their control counterparts. Accordingly, a large majority of patients with hypertension are likely to respond well to Web-based interventions designed to provide practical support regarding adherence to lifestyle recommendations for the management of hypertension. It also indicates that a minority of patients may require additional support to help manage psychological symptoms and associated stressors that may interfere with their ability to implement and adhere to suggested changes. The establishment of such psychobehavioral profiles provides an evidence-based strategy to understand the variability in patients with hypertension interested in enrolling in a Web-based intervention for lifestyle change. Such information is imperative in the development of effective person-centered Web-based interventions for a broad sample of patients with hypertension.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Background Characteristics, Health Behavior and Cardiovascular Risk Factors for Total Sample at Baseline.

[PDF File (Adobe PDF File), 34KB - jmir_v20i5e149_app1.pdf]

References

1. Sagie A, Larson MG, Levy D. The natural history of borderline isolated systolic hypertension. N Engl J Med 1993 Dec 23;329(26):1912-1917. [doi: 10.1056/NEJM199312233292602] [Medline: 8247055]
2. Leenen FH, Dumais J, McInnis NH, Turton P, Stratychuk L, Nemeth K, et al. Results of the Ontario survey on the prevalence and control of hypertension. CMAJ 2008 May 20;178(11):1441-1449 [FREE Full text] [doi: 10.1503/cmaj.071340] [Medline: 18490640]
3. Wilkins K, Campbell NR, Joffres MR, McAlister FA, Nichol M, Quach S, et al. Blood pressure in Canadian adults. Health Rep 2010 Mar;21(1):37-46 [FREE Full text] [Medline: 20426225]
4. James PA, Oparil S, Carter BL, Cushman WC, Dennison-Himmelfarb C, Handler J, et al. 2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). J Am Med Assoc 2014 Feb 5;311(5):507-520. [doi: 10.1001/jama.2013.284427] [Medline: 24352797]
5. Leung AA, Nerenberg K, Daskalopoulou SS, McBrien K, Zarnke KB, Dasgupta K, CHEP Guidelines Task Force. Hypertension Canada's 2016 Canadian hypertension education program guidelines for blood pressure measurement,

XSL•FO

RenderX

diagnosis, assessment of risk, prevention, and treatment of hypertension. Can J Cardiol 2016 Dec;32(5):569-588. [doi: 10.1016/j.cjca.2016.02.066] [Medline: 27118291]

6.   Friedberg JP, Rodriguez MA, Watsula ME, Lin I, Wylie-Rosett J, Allegrante JP, et al. Effectiveness of a tailored behavioral intervention to improve hypertension control: primary outcomes of a randomized controlled trial. Hypertension 2015 Feb;65(2):440-446 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.114.03483] [Medline: 25403606]

7.   Nolan RP, Upshur RE, Lynn H, Crichton T, Rukholm E, Stewart DE, et al. Therapeutic benefit of preventive telehealth counseling in the Community Outreach Heart Health and Risk Reduction Trial. Am J Cardiol 2011 Mar 1;107(5):690-696. [doi: 10.1016/j.amjcard.2010.10.050] [Medline: 21215382]

8.   Fahey T, Schroeder K, Ebrahim S. Interventions used to improve control of blood pressure in patients with hypertension. Cochrane Database Syst Rev 2005 Jan 25(1):CD005182. [doi: 10.1002/14651858.CD005182] [Medline: 15654709]

9.   Boulware LE, Daumit GL, Frick KD, Minkovitz CS, Lawrence RS, Powe NR. An evidence-based review of patient-centered behavioral interventions for hypertension. Am J Prev Med 2001 Oct;21(3):221-232. [Medline: 11567845]

10. Neubeck L, Redfern J, Fernandez R, Briffa T, Bauman A, Freedman SB. Telehealth interventions for the secondary prevention of coronary heart disease: a systematic review. Eur J Cardiovasc Prev Rehabil 2009 Jun;16(3):281-289. [doi: 10.1097/HJR.0b013e32832a4e7a] [Medline: 19407659]

11. Junghaenel DU, Broderick JE. Validation of the MPI patient profiles by partners and healthcare providers. Pain 2009 Jul;144(1-2):130-138 [FREE Full text] [doi: 10.1016/j.pain.2009.03.025] [Medline: 19410367]

12. Klein DM, Turvey CL, Pies CJ. Relationship of coping styles with quality of life and depressive symptoms in older heart failure patients. J Aging Health 2007 Feb;19(1):22-38. [doi: 10.1177/0898264306296398] [Medline: 17215200]

13. Nolan RP, Liu S, Feldman R, Dawes M, Barr S, Lynn H, et al. Reducing risk with e-based support for adherence to lifestyle change in hypertension (REACH): protocol for a multicentred randomised controlled trial. Br Med J Open 2013;3(8):e003547 [FREE Full text] [doi: 10.1136/bmjopen-2013-003547] [Medline: 23965936]

14. Miller WR, Rollnick S. Motivational Interviewing: Preparing People for Change. 3rd ed. New York: Guilford Press; 2013.

15. Dobson K. Handbook of Cognitive-Behavioral Therapies. 3rd ed. New York: Guilford Press; 2010.

16. 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] [Medline: 11556941]

17. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [Medline: 6880820]

18. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]

19. Payne AY, Surikova J, Liu S, Ross H, Mechetiuc T, Nolan RP. Usability testing of an internet-based e-counseling platform for adults with chronic heart failure. JMIR Hum Factors 2015 May 08;2(1):e7 [FREE Full text] [doi: 10.2196/humanfactors.4125] [Medline: 27026267]

20. Tudor-Locke C, Bassett Jr DR. How many steps/day are enough? Preliminary pedometer indices for public health. Sports Med 2004;34(1):1-8. [Medline: 14715035]

21. Csizmadi I, Kahle L, Ullman R, Dawe U, Zimmerman TP, Friedenreich CM, et al. Adaptation and evaluation of the National Cancer Institute's Diet History Questionnaire and nutrient database for Canadian populations. Public Health Nutr 2007 Jan;10(1):88-96. [doi: 10.1017/S1368980007184287] [Medline: 17212847]

22. Prochaska JO, Velicer WF, Rossi JS, Goldstein MG, Marcus BH, Rakowski W, et al. Stages of change and decisional balance for 12 problem behaviors. Health Psychol 1994 Jan;13(1):39-46. [Medline: 8168470]

23. Grundy SM, Pasternak R, Greenland P, Smith Jr S, Fuster V. Assessment of cardiovascular risk by use of multiple-risk-factor assessment equations: a statement for healthcare professionals from the American Heart Association and the American College of Cardiology. Circulation 1999 Sep 28;100(13):1481-1492 [FREE Full text] [Medline: 10500053]

24. Nylund KL, Asparouhov T, Muthén B. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. Struct Equ Modeling 2007 Oct 23;14(4):535-569. [doi: 10.1080/10705510701575396]

25. Nolan RP, Feldman RF, Dawes M, Kaczorowski J, Lynn H, Barr SI, et al. e-Counseling for self-care adherence adds therapeutic benefit for hypertension: the REACH Trial. Presented at: American College of Cardiology Congress; March 17-19, 2017; Washington, DC, United States URL: http://www.clinicaltrialresults.org/Slides/ACC2017/REACH_Nolan.pdf

26. Grace SL, Barry-Bianchi S, Stewart DE, Rukholm E, Nolan RP. Physical activity behavior, motivational readiness and self-efficacy among Ontarians with cardiovascular disease and diabetes. J Behav Med 2007 Feb;30(1):21-29. [doi: 10.1007/s10865-006-9080-5] [Medline: 17109217]

27. Ho AK, Thorpe CT, Pandhi N, Palta M, Smith MA, Johnson HM. Association of anxiety and depression with hypertension control: a US multidisciplinary group practice observational study. J Hypertens 2015 Nov;33(11):2215-2222 [FREE Full text] [doi: 10.1097/HJH.0000000000000693] [Medline: 26259121]

28. Luppino FS, de Wit LM, Bouvy PF, Stijnen T, Cuijpers P, Penninx BW, et al. Overweight, obesity, and depression: a systematic review and meta-analysis of longitudinal studies. Arch Gen Psychiatry 2010 Mar;67(3):220-229. [doi: 10.1001/archgenpsychiatry.2010.2] [Medline: 20194822]

29.   Nolan R, Feldman R, Dawes M, Kaczorowski J, Lynn H, Barr S. A randomized controlled trial of e-counseling for hypertension: REACH. J Am Coll Cardiol 2018:xx-xx Under Review with Revisions.

30.   Wangberg SC, Bergmo TS, Johnsen JK. Adherence in internet-based interventions. Patient Prefer Adherence 2008;2:57-65 [FREE Full text] [Medline: 19920945]

31.   Durrani S, Irvine J, Nolan RP. Psychosocial determinants of health behaviour change in an e-counseling intervention for hypertension. Int J Hypertens 2012;2012:191789 [FREE Full text] [doi: 10.1155/2012/191789] [Medline: 22229083]

32.   Aggarwal B, Liao M, Mosca L. Predictors of physical activity at 1 year in a randomized controlled trial of family members of patients with cardiovascular disease. J Cardiovasc Nurs 2010;25(6):444-449 [FREE Full text] [doi: 10.1097/JCN.0b013e3181defd3e] [Medline: 20856131]

33.   Gonzalez JS, Safren SA, Delahanty LM, Cagliero E, Wexler DJ, Meigs JB, et al. Symptoms of depression prospectively predict poorer self-care in patients with Type 2 diabetes. Diabet Med 2008 Sep;25(9):1102-1107 [FREE Full text] [doi: 10.1111/j.1464-5491.2008.02535.x] [Medline: 19183315]

Abbreviations

ABIC: sample-size adjusted Bayesian information criterion
ANCOVA: analysis of covariance
BIC: Bayesian information criterion
BMI: body mass index
CHEP: Canadian Hypertension Education Program
CVD: cardiovascular disease
DBP: diastolic blood pressure
FRI: Framingham risk index
HADS: Hospital Anxiety and Depression Scale
LMR: Lo-Mendell-Rubin likelihood ratio test
LPA: latent profile analysis
PHQ-9: Patient Health Questionnaire
PP: pulse pressure
REACH: Reducing risk with E-based support for Adherence to lifestyle Change in Hypertension
SBP: systolic blood pressure

XSL•FO
RenderX

Original Paper

Web-Based Intervention for Women With Type 1 Diabetes in Pregnancy and Early Motherhood: Critical Analysis of Adherence to Technological Elements and Study Design

Marie Berg[1*], RNM, PhD; Karolina Linden[1*], RNM, PhD; Annsofie Adolfsson[2], RNM, PhD; Carina Sparud Lundin[1], RN, PhD; Agneta Ranerup[3], PhD

[1]Centre for Person-Centred Care (GPCC), Institute of Health and Care Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
[2]School of Health Sciences, Örebro University, Örebro, Sweden
[3]Department of Applied Information Technology, University of Gothenburg, Gothenburg, Sweden
[*]these authors contributed equally

Corresponding Author:
Marie Berg, RNM, PhD
Centre for Person-Centred Care (GPCC), Institute of Health and Care Sciences
Sahlgrenska Academy
University of Gothenburg
Box 457
Gothenburg, 405 30
Sweden
Phone: 46 31 7866084
Fax: 46 31 786 6050
Email: marie.berg@fhs.gu.se

Abstract

Background: Numerous Web-based interventions have been implemented to promote health and health-related behaviors in persons with chronic conditions. Using randomized controlled trials to evaluate such interventions creates a range of challenges, which in turn can influence the study outcome. Applying a critical perspective when evaluating Web-based health interventions is important.

Objective: The objective of this study was to critically analyze and discuss the challenges of conducting a Web-based health intervention as a randomized controlled trial.

Method: The MODIAB-Web study was critically examined using an exploratory case study methodology and the framework for analysis offered through the Persuasive Systems Design model. Focus was on technology, study design, and Web-based support usage, with special focus on the forum for peer support. Descriptive statistics and qualitative content analysis were used.

Results: The persuasive content and technological elements in the design of the randomized controlled trial included all four categories of the Persuasive Systems Design model, but not all design principles were implemented. The study duration was extended to a period of four and a half years. Of 81 active participants in the intervention group, a maximum of 36 women were simultaneously active. User adherence varied greatly with a median of 91 individual log-ins. The forum for peer support was used by 63 participants. Although only about one-third of the participants interacted in the forum, there was a fairly rich exchange of experiences and advice between them. Thus, adherence in terms of social interactions was negatively affected by limited active participation due to prolonged recruitment process and randomization effects. Lessons learned from this critical analysis are that technology and study design matter and might mutually influence each other. In Web-based interventions, the use of design theories enables utilization of the full potential of technology and promotes adherence. The randomization element in a randomized controlled trial design can become a barrier to achieving a critical mass of user interactions in Web-based interventions, especially when social support is included. For extended study periods, the technology used may need to be adapted in line with newly available technical options to avoid the risk of becoming outdated in the user realm, which in turn might jeopardize study validity in terms of randomized controlled trial designs.

Conclusions: On the basis of lessons learned in this randomized controlled trial, we give recommendations to consider when designing and evaluating Web-based health interventions.

XSL•FO
RenderX

Introduction

Background

One important component of eHealth research is the evaluation of the use of information and communication technology. This was stated by Eysenbach more than 15 years ago [1] and further critically discussed, focusing on emerging Web-based interventions for patients with chronic illnesses. The issue of using randomized controlled trials (RCTs) as the "gold-standard" research methodology was also highlighted, as was the risk that the control group could become involved with other Web-based applications with similar objectives, thus "contaminating" the design [2]. Since then, the global use of mobile phones, the Internet, and a plethora of related technologies and applications has increased exponentially, and numerous interventions for promoting health and health-related behaviors have been carried out.

Thus, applying a critical perspective to Web-based interventions for persons with long-term and chronic illnesses is of specific importance. However, several methodological challenges exist, for example, issues of recruitment, randomization, fidelity, retention data quality, and degree of adherence [3]. Another matter is how the used technology and its design influence a Web-based intervention. To investigate whether intervention characteristics and persuasive design affect adherence, 83 Web-based health interventions were reviewed by Kelders et al [4]. Through the use of a framework for Persuasive System Design (PSD) [5] and its design principles, they coded persuasive technology elements and analyzed intervention characteristics and adherence. In a multiple regression analysis, they found that users' adherence to an intervention was predicted by differences in technology and interaction with a counselor. Providing persuasive technology for social support did not affect adherence, although the need to further explore this was stressed [4].

Objectives

In this paper, we further critically analyze and discuss the challenges of conducting an RCT using Web-based technology, including devices for social support, with special focus on adherence. According to Kelders et al, adherence is defined as "the intended usage in line with the therapeutic regime" [4] in which social support is a prominent component. The basis is a recently finalized RCT in which Web-based support was offered to women with type 1 diabetes mellitus (T1DM) in pregnancy and early motherhood: the MODIAB-Web (MOtherhood and DIABetes) study.

The following research questions have been elaborated upon in this paper:

- What persuasive content and technological elements were used in the design of the Web-based support?
- How was the Web-based support used?
- What was the content of the social peer support and how did the peers support each other?
- What were the main challenges in relation to adherence to the RCT?

Methods

Research Design

Using an exploratory case study methodology [6] and the framework for analysis offered through the PSD model [5], we have critically examined the MODIAB-Web study. The analytical focus was on technology and its design, as well as the usage of the different parts of the Web-based support with special focus on the forum for peer support.

The Motherhood and Diabetes (MODIAB)-Web Study

The MODIAB-Web RCT study was directed at women with T1DM registered at 6 hospital-based antenatal care units in Sweden. Complementary Web-based support was offered to the intervention group, with support starting in early pregnancy and persisting up to 6 months after the child was born [7,8]. The rationale for the study was that T1DM women face particular demands in relation to pregnancy, childbirth, and early motherhood. During pregnancy, they struggle to achieve normal blood glucose levels, which optimize the probability of giving birth to a healthy child [9-11]. In the early months after childbirth, the women have to deal with unstable blood glucose levels, whereas at the same time having to meet all the challenges of early motherhood, and especially those related to breastfeeding [9,12,13]. Their need of Web-based support including social peer support has been identified in earlier studies [14-16].

The Web-based support was developed in line with the ideal of participatory design [17] and includes three parts: (1) information based on scientific evidence, (2) a self-care diary, and (3) a forum for peer support divided into three topics: pregnancy, childbirth, and life as a new mother. In addition, the Web-based support comprised a frequently asked question section where participants could ask questions and receive answers from experts in the field, and a collection of links to other recommended resources. The intention was to offer complementary Web-based support as an add-on to regular pregnancy care and use it especially after childbirth to fill the gap of healthcare support between the different healthcare providers, that is, the maternity care professionals and the diabetes care professionals. If they wished, the women could share their self-care diaries with healthcare professionals during their visits, as the diaries were part of the Web-support structure [7]. The hypothesis was that the Web-based support should strengthen the women's personal capacity and autonomy, thereby leading to improved self-management of diabetes and overall well-being. Ethical approval was attained from the Ethics Committee of Gothenburg, Sweden (No. 659-09), and the trial was registered at clinicaltrials.gov (ID: NCT015665824).

Eligible study participants were successively recruited by an appointed study midwife in early pregnancy. The study was performed over a period of more than 4 years. The first study participant was included on November 22, 2011, and the last participant ended participation on January 25, 2016, that is, 6 months after the last childbirth. Further details about the study design are described elsewhere [7,8].

The findings showed that the Web-based support plus standard care was not superior to standard care alone in terms of general well-being and self-efficacy of diabetes management. Details on this are reported elsewhere [18].

Analysis

Research question 1, "*what persuasive content and technological elements were used in the design of the Web-based support?,*" has been answered using the PSD model for coding design principles in different categories, each one comprising 7 design principles. The categories are: primary task support, dialogue support, system credibility support, and social support [5]. A deductive analysis consisted of identifying and describing which PSD principles and technological elements were used in the MODIAB-Web intervention. Research question 2, "*how was the Web-based support used?*", was answered using descriptive statistics and qualitative content analysis [19]. Research question 3, "*the content of the social peer support and how the peers supported each other*", was answered using descriptive statistics and qualitative content analysis [19]. Initially, the forum posts were read as a whole several times to get an overall sense of the data. In the next step, new readings followed in which the data were organized. This process included open coding, in which notes and headings were written in the margins. The headings were transferred to a separate coding sheet and were grouped into categories. The categories were presented under the preset topics of pregnancy, childbirth, and life as a new mother. The quotations used to illustrate the dialogue in the forum for peer support were professionally translated into English. The analysis was performed by MB in close collaboration with AA. Research question 4, "*what were the main challenges in relation to adherence to the RCT?*", was answered through a critical analysis based on the results of the first three research questions and the intervention as a whole.

Results

Persuasive Content and Technological Elements in the Design of the Motherhood and Diabetes -Web Support

The implemented categories with used design principles [5] in the MODIAB-Web study are presented in Table 1. In the category *Primary Task Support* the focus is on providing technological elements to manage the targeted behavior [5]. The MODIAB-Web study implemented 3 out of 7 design principles: tailoring, reduction, and self-monitoring*,* whereas tunneling, personalization, simulation, and rehearsal were not used. In *Dialogue Support* the focus is on providing various kinds of feedback between the human and the system [5]. In

the MODIAB-Web study reminders, liking and social role were implemented, whereas praise, rewards, suggestions, and similarity were not. In *System Credibility Support*, the design principles aim to increase credibility and consequently to be more persuasive [5]. Five design principles were used in the MODIAB-Web study: trustworthiness, expertise, surface credibility, real-world feel, and authority, whereas third-party endorsement and variability were not. The category *Social Support* motivates usage by leveraging social influence [5]. The MODIAB-Web study implemented the principles social learning and social comparison while the principles normative influence, social facilitation, cooperation, competition, and recognition were not used.

Use of the Motherhood and Diabetes Web-Based Support

In total, 83 women were randomized to the MODIAB Web-based support, and of these, 81 received a log-in. As the women were successively randomized to either control group or intervention, and as the intervention group could use the Web-based support from randomization to 6 months after birth of the child, a maximum of 36 women could be active simultaneously. This is illustrated in Figure 1.

Of the 81 women with log-in, 69 women were classified as "active users," that is, they had logged in at least once after the introduction session. How the participants used the Web-based support is presented in Table 2.

The peer support forum was used by 63 women of whom one was a facilitator appointed by the researchers in the first year of the intervention. This facilitator was an experienced mother with T1DM. After that period, it was presumed that the included women could collectively run their activities. The forum was moderated by a member of the research team to have some kind of control over what was discussed and to make sure it did not include advice that was contrary to current scientific evidence. Almost all forum activity occurred in the first 3 years (2011-2014). Of the 63 forum users, 39 participants were readers and 24 were active writers including the study facilitator.

Forum Content and Type of Peer Support

There were 109 written posts (range: 1-20) divided into 19 threads, of which 84 posts (16 threads) were related to the topic *pregnancy*, 15 posts (1 thread) to *childbirth*, and 10 posts (2 threads) to *life as a new mother*. No inappropriate advice was given during the study period, so the moderator did not have to act. Some women who asked questions in their posts had to wait some time before receiving an answer, while others did not receive an answer at all. This happened after the study facilitator had left the forum.

Forum Content

The design of the forum for peer support was intended to help users sort experiences in relation to the different phases of pregnancy, childbirth, and life as a new mother. To some extent, it became apparent that the topics overlapped in each category.

Table 1. Categories and design principles in the Motherhood and Diabetes (MODIAB)-Web intervention according to the Persuasive Systems Design (PSD) framework model.

Category and design principle[a]	Implementation in the MODIAB-Web intervention
Primary task support	
Tailoring	
Information provided by the system will be more persuasive if it is tailored to the potential needs, interests, personality, usage context, or other factors relevant to a user group [5].	The evidence-based information was tailored to three themes [20], adapted to reflect the unique aspects that type 1 diabetes adds: being pregnant; labor and childbirth; and life as a new mother
Reduction	
A system that reduces complex behavior into simple tasks, helps users perform the target behavior, and may increase the benefit/cost ratio of a behavior [5].	The evidence-based information, presented with headlines and clickable subheadings to scan for quick access, contained simple task helps for the participants such as what a healthy breakfast might consist of and how to adjust insulin doses in the first days after childbirth [20].
Self-monitoring	
A system that keeps track of one's own performance or status and supports the user in achieving goals [5].	In the self-care diary, the women could use either a smartphone or computer to register daily life information, such as blood glucose levels, insulin and food intake, and overall well-being status. The registered data were presented in tables and diagrams, with the intention of supporting the woman in analyzing and managing her daily life to accomplish optimal blood glucose levels. This information could also be presented to health care professionals if the women consented.
Dialogue support	
Reminders	
If a system reminds users of their target behavior, the users are more likely to achieve their goals [5].	Text messages to inactive users were sent to the participants as reminders every 2 weeks consisting of a greeting and contact information in case of technical difficulties. There was no flagging of new posts in the forum for peer support.
Liking	
A system that is visually attractive to its users is likely to be more persuasive [5].	The Web-based support was developed in collaboration between the research group and Web designers and evaluated by a group of mothers with type 1 diabetes [21]. Pictures of pregnant women and babies were used to illustrate the content.
Social role	
If a system adopts a social role, users are more likely to use it for persuasive purposes [5].	The system was designed to enable the women to coach each other. During the first year of the intervention when there were very few participants, a "coach," a woman with type 1 diabetes who had given birth to a child, was used as a social facilitator in the forum. She initiated contact and replied to new users' posts in the forum.
System credibility support	
Expertise	
A system that is viewed as incorporating expertise will have increased powers of persuasion [5].	All parts of the Web-based support were developed in a collaborative developmental process by researchers and experts; including experienced mothers with diabetes and Web-designers. The professional experts were: nurse-midwives specializing in diabetes care and human lactation; gynecologist/obstetricians; diabetologists; neonatal nurse; dietician [7,20]. The support's design was reviewed by experienced mothers with type 1 diabetes [21].
Trustworthiness	
A system that is viewed as trustworthy will have increased powers of persuasion [5].	The names and titles of the healthcare professionals who had contributed to the design were clearly stated in the information section of the Web-based support.
Surface credibility	
People make initial assessments of the system credibility based on a firsthand inspection [5].	The Web-based support was test-piloted by a focus group consisting of women with diabetes with experience of pregnancy and childbirth. The Web-based support was adapted in line with their comments [21].

Category and design principle[a]	Implementation in the MODIAB-Web intervention
Real-world feel	
A system that highlights people or organization behind its content or services will have more credibility [5].	The frequently asked questions section contained the participants' questions and the anonymized expert answers, which were published for all to read. There was a delay of up to 2 weeks before answers were received and later published.
Authority	
A system that leverages roles of authority will have enhanced powers of persuasion [5].	The physician-in-chief of the respective antenatal care units supported the intervention. This was clearly stated in the Web-based support. The system did not include health care professionals as users.
Social support	
Social learning	
A person will be more motivated to perform a target behavior if he can use a system to observe others performing the behavior [5].	In the forum for peer support, the users could start up their own threads as an opportunity for learning from each other's experiences under the three defined themes: pregnancy, childbirth, and life as a new mother.
Social comparison	
System users will have a greater motivation to perform the target behavior if they can compare their performance with the performance of others [5].	The system partly supported social comparison between participants, specifically in the forum for peer support in which experiences were shared.

[a]Cited from the description of the PSD model by Oinas-Kukkonen and Harjumaa [5].

Figure 1. Possible active users during the study period of the MODIAB-Web intervention.
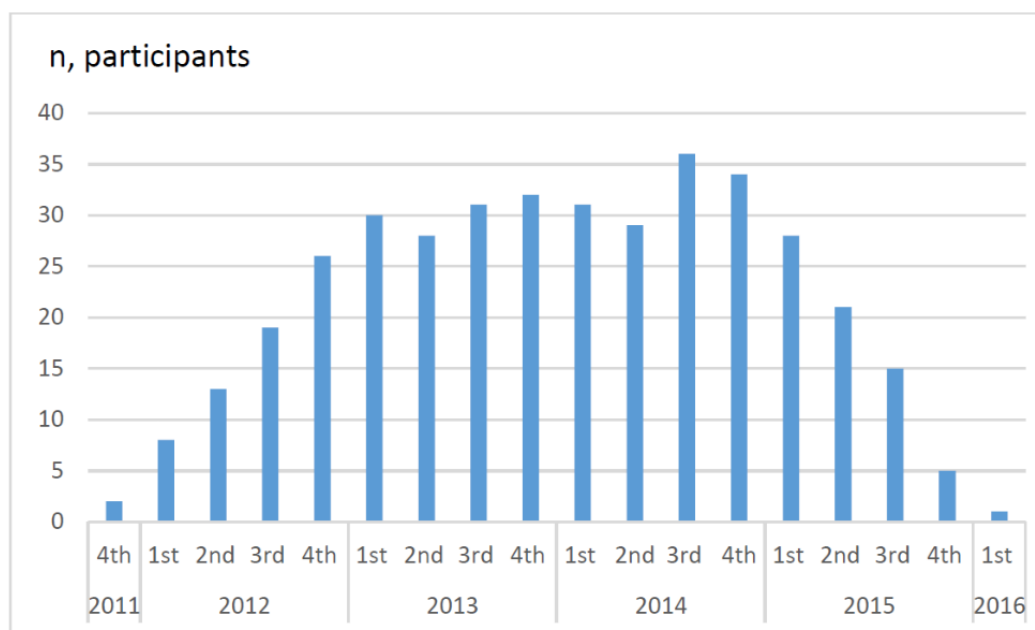
Table 2. Web-based support usage.

Activities	N	Median	Minimum	Maximum
Total log-ins by active users	69	91	2	6413
Evidence-based information	61	10	1	508
Forum for peer support	62[a]	60	1	703
Self-care diary	37	21	1	5850
Frequently asked questions	53	2	1	17

[a]A study facilitator participated in the first year, giving 63 active users.

Pregnancy

This topic had the largest number of posts. Its content related to overly focusing on diabetes as a disease; the first ultrasound; how to handle glycemia and specifically hypoglycemic episodes; insulin dosages, and the estimated weight of the expectant child. Minor discussions concerned managing attendance at the many checkups combined with working and how to plan for parenthood after the birth of the child, including how to get in contact with the Swedish social insurance agency for the parental benefit.

Childbirth

The posts related to inducing labor and labor routines, including how the insulin was administered and if an insulin pump was permitted during labor. Other posts discussed wounds and wound healing. One woman who had given birth 3 times shared her story of all these childbirths, one having been vaginal, one vaginal vacuum extraction, and one elective cesarean section.

Life as a New Mother

Participants shared posts on this topic for a period of only 6 months (December 2011 to July 2012). Content mostly concerned postpartum care at the hospital—that there was insufficient professional support or insufficient knowledge about diabetes. Another main concern was breastfeeding-related issues. One woman expressed her disappointment that the ward ignored her wish to avoid giving her baby cows' milk-based nutrition supplements. Another woman concluded that in her situation with diabetes, it was best to combine breastfeeding and a milk supplement. One woman asked for experiences of breastfeeding and how to integrate it in daily life as a mother with diabetes. Unfortunately, no one responded. In addition, there were many diverse posts, for example, about a baby who had got an infection, and another baby who had a clavicle fracture. Several questions were asked, for example, did you get preeclampsia? What type of pain relief did you choose? One woman, who was active at the beginning of the intervention period (December 2012), advertised her own blog and its weblink.

Type of Support

The analysis of the peer support in the forum shows that, although there were few participants there was a rich exchange between them. The dialogues comprised sharing of personal experiences, giving concrete advice, affirmations, and words of reassurance. Examples of two dialogues between the participating women are given in Textbox 1.

Textbox 1. Examples of dialogues in the peer forum between participating women.

Dialogue—Example 1

Woman 1: "As a diabetic, having kids isn't a walk in the park but I've managed twice before with completely perfect kids…this pregnancy has been really tough – with high blood pressure and low sugar but after a lot of messing about, it's sorted itself out now."

Woman 2: "I'm so inspired and motivated when I hear people like you – congratulations on pregnancy number three! I'm in my first but we're almost the same age and I've been a diabetic for 22 years. My biggest worry is the hypos. I hate them and I worry the baby can feel them. Once all its organs are I'm s' and I know it can produce insulin and has a liver with a glycogen store, I'll be calmer. My levels aren't low all the time – my HBA1C is 7.9 so it has to be lowered. One side effect of my managing to lower it is that I'm more insulin sensitive. But I know that changes in different stages of pregnancy. Was it obvious in your other pregnancies – in which case, when might you have become more insulin resistant?

I'm noticing I'm actually pregnant now, week 10+4, and I've done the CUB (Combined Ultrasound and Biochemistry screening) and one ultrasound. So it's weird and wonderful to see that a fetus is living inside me. Of course, it's early yet but it feels real now. I won't tell them at work until after week 12. How do you feel about telling people, when did you do that?

Woman 1: "Yes the hypos suck, I have to say but don't worry during the first weeks, during this pregnancy I had such low levels an ambulance had to come lots of times but those little ones are tough – nothing happened to him. Now I'm in week 20+. What happens when you have high levels is that the baby has to work harder to produce more insulin but at low levels, not much happens, as long as you don't go into a coma, that is – don't push yourself or you'll go mad. I told work about my first pregnancy straight away but that was because I had heavy lifting and stuff to do. I've always felt you should say when you're ready to say you're going to become a mum. And it takes time, I promise you – I've been a mum for almost eight years now and I still can't see myself as one. Everything will be alright, you'll see. Everything's tough at the start but with time it gets better and calmer."

Woman 2: "Thanks for such a good answer – it's calming what you write about having hypos."

Dialogue—Example 2

Woman 3: "I wonder if there are others like me who think it feels like all the focus is on the disease/s and that you never talk about the pregnancy. I'm in week 29 and had diabetes for 22 years and I also have trouble with my thyroid. My blood sugar and thyroid tests all have great results but still the midwife just keeps talking about the diseases. Have you experienced that? Do you get enough support around pregnancy issues or is the focus only on the diabetes?"

Woman 4: "I think that's varied a bit with each of my different pregnancies. It's had a lot to do with the midwife I had. Sure, there's been a lot of focus on diabetes and blood sugar levels but the midwife I have now makes a point of also talking about the baby and pregnancy-related things. But my last one wasn't like that…I think you should raise it with your midwife again – say you need her help with some pregnancy issues. After all, she's there for you and your baby! Good luck!"

Woman 3: "Thanks! I'll try again tomorrow!!"

Woman 5: "I agree the focus is mostly on the diabetes when you visit the midwife – you almost forget what you've been wondering about and want to discuss with the midwife but the only thing they want to talk about is test results. Of course, it's good to know they're really on the ball with the diabetes during the pregnancy but you've never felt this ill as a diabetic after living with it for so many years."

XSL•FO

RenderX

What Were the Main Challenges in Relation to Adherence to the Web-Based Randomized Controlled Trial?

This section comprises a critical analysis of both significant and minor issues in relation to adherence (defined as "the intended usage in line with the therapeutic regime" [4]) to the Web-based intervention. It is also based on the fact that the presumed effect of the RCT, in terms of increased well-being and diabetes management, was not achieved [18]. The MODIAB-Web study was typically designed according to the gold standard of an RCT [22]. How did this come about and why?

Adherence in Relation to Study Design

There were few women available to include in this study. In Sweden, around 0.5% of all pregnant women have T1DM, corresponding to 500 per year. At the beginning, two study centers were included based on an estimate of few eligible women. It became apparent that the recruitment rate of study participants was slower than expected. In addition, four study centers were included to avoid a more prolonged data collection period. Furthermore, the contact frequency between the researchers and the healthcare professionals at the study centers was increased. Prolonging clinical trials because of slow recruitment pace and retention of study participants is a common issue. One reason for this might be poor engagement from the healthcare professionals who are inviting the patients to take part in the study [23].

The slow inclusion pace prolonged the study duration, which in turn meant only a few women (about 30) could be active at the same time (Figure 1). Thus, the critical mass of simultaneous active users was not sufficient and undermined the extent of interactions in the Forum. It has been concluded that many of the participants in a Web-based study may be lost to follow-up or end up not adhering to the intervention [2]. This was not taken into account when calculating the sample size. One way to get around this could have been to have started the study at more study centers from the beginning to gain a greater critical mass faster.

The use of a facilitator who was active in the peer support forum the first year was fruitful, as it increased the dialogues in the peer support forum. A shortfall in the study design was that the facilitator was not engaged during the whole study period, which led to the activity level in the forum dropping.

The quality of the support shared between the participants in the MODIAB-Web study was reasonably good, as exemplified in the dialogues (Textbox 1). The active writers in the forum for peer support coached each other and shared their experiences in a generous way. The majority of participants were "readers"; they did not actively participate in the dialogues. However, in line with recent research, these are probably also "passive actors"—meaning only readers can receive social support related to their specific needs [24].

The forum user who advertised the weblink to her blog in the MODIAB-Web study at least potentially created an alternative Web-based support. This makes it hard to truly evaluate the usefulness of the intervention, as meaningful interaction might happen outside the study platform. It also proposes a threat to adherence to the intervention, if participants choose to communicate in alternative ways, such as through social media. It is almost impossible to control for this, and it is a methodological weakness in studies evaluating Web-based social support.

A main problem arising due to the prolonged duration of the MODIAB-Web RCT study was the simultaneous explosion of more advanced general mobile phone technology [24] as well as similar technology in diabetes care contexts. The developed Web-support serving as a platform for self-management consequently became outdated with time as few available alternatives grew.

Adherence in Relation to Technology

With regard to the PSD model [5], four categories were identified in the MODIAB-Web intervention: *Primary Task Support*, *Dialogue Support*, *System Credibility Support,* and *Social Support* (see Table 1). This shows that several but not all of the design principles were implemented.

To accomplish adherence, Kelder et al argue that it is essential to actually plan for adherence when designing Web-based interventions [4]. The MODIAB-Web intervention was developed using participatory design [17] and grounded on previous empirical and theoretical investigations of the needs of the user groups, as well as ongoing user participation of lay and professional experts [7]. However, there was no theoretical basis for the technological design or the use of design principles of a more general kind, such as the PSD model. A theoretical grounding would have generated several advantages. One prominent advantage is insights into a greater repertoire of design elements than those available through participating lay and professional experts. In this manner, design is based on practical and theoretical knowledge outside of the local project.

Moreover, a theoretical grounding such as the PSD model was launched in 2009, meaning that it existed well before the investigations of general user needs [14] and the design of the MODIAB-Web intervention in particular [7]. However, it must be recognized that the recent increasing importance and subtleties of research processes focusing on design to accomplish behavior change are still treated in specific research fields such as design studies [25]. A theoretical grounding when it comes to behavior change itself would enable a repertoire of more passive and active strategies for influencing people that, in turn, should be brought into the technology design part of the intervention. Such strategies might involve using models of how to accomplish a feeling of social connectedness [26], which is relevant in interventions aiming to increase the social support between peers. A theoretically more conscious design might be an alternative to the recent suggestion of generating predictive models of potential dropouts in online health communities by Big Data analysis to inform design [27].

The use of design theory in the analysis of the MODIAB-Web intervention enabled a discussion of the rationale behind its emphasis on *System Credibility Support* and its significantly less emphasis on *Dialogue Support* and *Social Support*. Quite likely, potential was lost in unutilized design principles. For

example, in the *Dialogue Support* category, the design principle "Rewards" that aims to give credit for performing the target behavior [5], that is, using the Web-support, could have been implemented to increase adherence. For instance, this could entail rewarding participants once they had used the support 5 times, or when at least 5 other participants had read a person's posts in the forum.

A further technical element related to *Social Support*, not used in the MODIAB-Web forum for peer support, was "flagging" of new posts. Such flagging had probably increased the participants' "adherence," in terms of active communication. Another issue that probably reduced adherence was that the forum was divided into 3 parts. This division meant, there was a risk, that new posts were not seen by the other participants. Another shortfall was that the participants were not obliged to write a post to introduce themselves upon first entering the peer support forum. Admittedly, certain PSD design principles related to *Social Support* would have been positive in relation to adherence, for example, "Recognition," which would include success stories. Other design principles, such as "Normative influence," might be less relevant in this context due to the extreme demand for optimal diabetes control for childbearing women.

Discussion

In this paper, we used case study methodology to enable a critical discussion of RCTs involving Web-based support and adherence in general, and social support in particular. In other words, this methodology has been used as a basis for our critical endeavor in line with Baxter and Jack, who state: "it enables the researcher to answer 'how' and 'why' type questions, while taking into consideration how a phenomenon is influenced by the context within which it is situated" [6].

In this paper adherence is, in line with Kelders et al [4], defined as "use in accordance with identified intentions," and in which the need for relevant information about the specific condition as well as social support from peers is emphasized. Our analysis supports the conclusion that PSD does matter in general, by positively influencing adherence [4]. We argue that there is a particular value in basing technology design on theoretical design principles, as it increases the repertoire of options, thus enabling adaption to suit the target population for the intervention and creating opportunities to utilize the full design potential of social support. The value of applying theory ties in with recent acknowledgment of theoretical consciousness regarding adaptation and implementation of health information technology innovations [28]. When it comes to health information technology such as Web-based support, we argue that theories related to technology use and design, such as the ones discussed above, are equally important. This notwithstanding, the essential rationale of Web-based health interventions remains the same but becomes significantly better theoretically informed. A systematic review published post completion of our analysis showed that justifications for intended use, and adherence to eHealth technology is often underdeveloped and improperly used. The authors believe that adherence can be standardized, and this will improve comparison

of adherence rates to different technologies with the same goals [29]. Our critical analysis thus contributes to ongoing discussion about adherence.

Moreover, we conclude that study design matters and that technology and study design might mutually influence each other. With a limited target group, the comparative element that constitutes an RCT design becomes a barrier to achieving a critical mass of user interactions. Extended project periods mean that the used technology must either be changed or tailored according to new technical options that become available during the process, which in turn might jeopardize study validity. Not doing so implies a risk that the technology will become somewhat outdated in the world of the users. In relation to providing social support, a further conclusion is that using theory enables a multifaceted repertoire in technology design. However, we also note that apparently low user activity might still provide social support for the individual, irrespective of whether they are an active or passive user.

On the basis of critical analysis of how study design and technology interacts, we recommend fellow researchers to consider the following aspects when designing and evaluating Web-based health interventions:

- When designing a Web-based intervention, use existing design theories to utilize the full potential of the technology and increase adherence, especially with regard to social support.
- Be realistic when calculating your sample size. Take the risk of losing participants and poor adherence to the intervention into account.
- Be aware of limited target populations. If the number of possible participants is low, consider starting the intervention in multiple study centers simultaneously to avoid prolonged study periods and thereby outdated technology.
- Bear in mind that the majority of participants will not actively engage in discussions in peer support forums. Therefore, a larger critical mass of participants is needed in interventions based on effects of social support.
- Keep the design of the forum for peer support simple and avoid divisions. Web-design is often costly if the researchers do not possess the skills themselves or team up with researchers from other fields. Consider integrating your intervention into existing social media (but be aware of ethical pitfalls if you cannot guarantee the security and confidentiality of data).
- If your intervention consists of a forum for peer support, consider engaging a study facilitator from the target group for the entire study period to boost activity.
- Health interventions are often complex in their nature. RCT design is one way of evaluating the effects of an intervention but it is in itself inadequate for truly capturing complex interactions. Design your study with multiple methods of analysis and consider the contamination of the control group as technology evolves during the study period.
- Adjust per-protocol criteria after intended usage or conduct a dose-response analysis to properly evaluate the effect of the intervention within the RCT design.

Finally, we want to return to the initial critical questioning of whether the RCT design really is appropriate as the gold standard for Web-based interventions [2]. Our analysis does not fully answer this question, nor does it reveal what could be a better research design, but we would like to encourage further constructive discussion on these issues within the scientific community.

Conflicts of Interest

None declared.

References

1. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun;3(2):E20. [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
2. Eysenbach G. Issues in evaluating health websites in an Internet-based randomized controlled trial. J Med Internet Res 2002 Dec;4(3):E17 [FREE Full text] [doi: 10.2196/jmir.4.3.e17] [Medline: 12554548]
3. Murray E, Khadjesari Z, White IR, Kalaitzaki E, Godfrey C, McCambridge J, et al. Methodological challenges in online trials. J Med Internet Res 2009;11(2):e9 [FREE Full text] [doi: 10.2196/jmir.1052] [Medline: 19403465]
4. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
5. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system feautures. Commun Assoc Inf Syst 2009;24(1):485--4500 [FREE Full text]
6. Baxter P, Jack S. Qualitative case study methodology: study design and implementation for novice researchers. Qual Rep 2008;13(4):544-559 [FREE Full text]
7. Berg M, Adolfsson A, Ranerup A, Sparud-Lundin C, University of Gothenburg Centre for Person-Centred Care. Person-centered Web support to women with type 1 diabetes in pregnancy and early motherhood--the development process. Diabetes Technol Ther 2013 Jan;15(1):20-25. [doi: 10.1089/dia.2012.0217] [Medline: 23297670]
8. Adolfsson A, Linden K, Sparud-Lundin C, Larsson PG, Berg M. A web-based support for pregnant women and new mothers with type 1 diabetes mellitus in Sweden (MODIAB-Web): study protocol for a randomized controlled trial. Trials 2014 Dec 29;15:513 [FREE Full text] [doi: 10.1186/1745-6215-15-513] [Medline: 25543854]
9. Rasmussen B, Hendrieckx C, Clarke B, Botti M, Dunning T, Jenkins A, et al. Psychosocial issues of women with type 1 diabetes transitioning to motherhood: a structured literature review. BMC Pregnancy Childbirth 2013 Nov 23;13:218 [FREE Full text] [doi: 10.1186/1471-2393-13-218] [Medline: 24267919]
10. Berg M, Honkasalo ML. Pregnancy and diabetes--a hermeneutic phenomenological study of women's experiences. J Psychosom Obstet Gynaecol 2000 Mar;21(1):39-48. [Medline: 10907214]
11. Berg M. Pregnancy and diabetes: how women handle the challenges. J Perinat Educ 2005;14(3):23-32 [FREE Full text] [doi: 10.1624/105812405X57552] [Medline: 17273439]
12. Sparud-Lundin C, Berg M. Extraordinary exposed in early motherhood - a qualitative study exploring experiences of mothers with type 1 diabetes. BMC Womens Health 2011 Apr 07;11:10 [FREE Full text] [doi: 10.1186/1472-6874-11-10] [Medline: 21473755]
13. Berg M, Erlandsson L, Sparud-Lundin C. Breastfeeding and its impact on daily life in women with type 1 diabetes during the first six months after childbirth: a prospective cohort study. Int Breastfeed J 2012 Dec 21;7(1):20 [FREE Full text] [doi: 10.1186/1746-4358-7-20] [Medline: 23259843]
14. Sparud-Lundin C, Ranerup A, Berg M. Internet use, needs and expectations of web-based information and communication in childbearing women with type 1 diabetes. BMC Med Inform Decis Mak 2011 Jul 07;11:49 [FREE Full text] [doi: 10.1186/1472-6947-11-49] [Medline: 21736713]
15. Edwards H, Speight J, Bridgman H, Skinner TC. The pregnancy journey for women with type 1 diabetes: a qualitative model from contemplation to motherhood. Pract Diab 2016 Aug 03;33(6):194-199a. [doi: 10.1002/pdi.2036]
16. Carlsson IM, Berg M, Adolfsson A, Sparud-Lundin C. Reprioritizing life: a conceptual model of how women with type 1 diabetes deal with main concerns in early motherhood. Int J Qual Stud Health Well-being 2017;12(sup2):1394147 [FREE Full text] [doi: 10.1080/17482631.2017.1394147] [Medline: 29092688]
17. Spinuzzi C. The methodology of participatory design. Tech Commun 2005 May;52(2):163-174.
18. Linden K, Berg M, Adolfsson A, Sparud-Lundin C. Person-centred, web-based support in pregnancy and early motherhood for women with type 1 diabetes mellitus: a randomized controlled trial. Diabet Med 2018 Feb;35(2):232-241 [FREE Full text] [doi: 10.1111/dme.13552] [Medline: 29171071]

XSL•FO
RenderX

19. 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]

20. Linden K, Berg M, Sparud-Lundin C. Web-based information for pregnant women and new mothers with type 1 diabetes--a description of the development process. BMC Med Inform Decis Mak 2012 Nov 20;12:134 [FREE Full text] [doi: 10.1186/1472-6947-12-134] [Medline: 23167552]

21. Adolfsson A, Jansson M. Prototype for Internet support of pregnant women and mothers with type 1 diabetes: focus group testing. Psychol Res Behav Manag 2012;5:97-103 [FREE Full text] [doi: 10.2147/PRBM.S32799] [Medline: 22915948]

22. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Br Med J 2010 Mar;340:c869 [FREE Full text] [Medline: 20332511]

23. Tufts Center for the Study of Drug Development. Poor physician and nurse engagement driving low patient recruitment. Tufts Center for the Study of Drug Development Impact Report 2017;19(1):1-5.

24. Kao CW, Chuang HW, Chen TY. The utilization of health-related applications in chronic disease self-management. Hu Li Za Zhi Aug 2017;64(4):19-25. [doi: 10.6224/JN.000050] [Medline: 28762221]

25. Cash PJ, Hartlev CG, Durazo CB. Behavioural design: a process for integrating behaviour change and design. Des Stud 2017 Jan;48:96-128. [doi: 10.1016/j.destud.2016.10.001]

26. Visser T, Vastenburg MH, Keyson DV. Designing to Support Social Connectedness: The Case of SnowGlobe. Int J Des 2011 Dec;5(3):129-142 [FREE Full text]

27. Wang X, Zhao K, Street N. Analyzing and predicting user participations in online health communities: a social support perspective. J Med Internet Res 2017 Apr 24;19(4):e130 [FREE Full text] [doi: 10.2196/jmir.6834] [Medline: 28438725]

28. 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]

29. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]

Abbreviations

MODIAB-Web: Motherhood and Diabetes Web
PSD: persuasive systems design
RCT: randomized controlled trial
T1DM: type 1 diabetes mellitus

XSL•FO
RenderX

Original Paper

Behind the Scenes of Online Therapeutic Feedback in Blended Therapy for Depression: Mixed-Methods Observational Study

Mayke Mol[1,2], MSc; Els Dozeman[1,2], PhD; Simon Provoost[3], MSc; Anneke van Schaik[1,2], MD, PhD; Heleen Riper[1,3,4], PhD; Johannes H Smit[1,2], PhD

[1]Department of Research and Innovation, GGZ inGeest, Specialized Mental Health Care, Amsterdam, Netherlands

[2]Department of Psychiatry, Amsterdam Public Health Research Institute, VU University Medical Center, Amsterdam, Netherlands

[3]Department of Clinical, Neuro and Developmental Psychology, Clinical Psychology Section, Vrije Universiteit Amsterdam and the Amsterdam Public Health Research Institute, Amsterdam, Netherlands

[4]Institute of Telepsychiatry, University of Southern Denmark, Odense, Denmark

Corresponding Author:
Mayke Mol, MSc
Department of Psychiatry
Amsterdam Public Health Research Institute
VU University Medical Center
Van der Boechorststraat 7
Amsterdam, 1081 BT
Netherlands
Phone: 31 204448180
Email: m.mol@ggzingeest.nl

Abstract

Background: In Internet-delivered cognitive behavioral therapies (iCBT), written feedback by therapists is a substantial part of therapy. However, it is not yet known how this feedback should be given best and which specific therapist behaviors and content are most beneficial for patients. General instructions for written feedback are available, but the uptake and effectiveness of these instructions in iCBT have not been studied yet.

Objective: This study aimed to identify therapist behaviors in written online communication with patients in blended CBT for adult depression in routine secondary mental health care, to identify the extent to which the therapists adhere to feedback instructions, and to explore whether therapist behaviors and adherence to feedback instructions are associated with patient outcome.

Methods: Adults receiving blended CBT (10 online sessions in combination with 5 face-to-face sessions) for depression in routine mental health care were recruited in the context of the European implementation project MasterMind. A qualitative content analysis was used to identify therapist behaviors in online written feedback messages, and a checklist for the feedback instruction adherence of the therapists was developed. Correlations were explored between the therapist behaviors, therapist instruction adherence, and patient outcomes (number of completed online sessions and symptom change scores).

Results: A total of 45 patients (73%, 33/45 female, mean age 35.9 years) received 219 feedback messages given by 19 therapists (84%, 16/19 female). The most frequently used therapist behaviors were informing, encouraging, and affirming. However, these were not related to patient outcomes. Although infrequently used, confronting was positively correlated with session completion (ρ=.342, *P*=.02). Therapists adhered to most of the feedback instructions. Only 2 feedback aspects were correlated with session completion: the more therapists adhere to instructions containing structure (limiting to 2 subjects and sending feedback within 3 working days) and readability (short sentences and short paragraphs), the less online sessions were completed (ρ=−.340, *P*=.02 and ρ=−.361, *P*=.02, respectively). No associations were found with depression symptom change scores.

Conclusions: The therapist behaviors found in this study are comparable to previous research. The findings suggest that online feedback instructions for therapists provide sufficient guidance to communicate in a supportive and positive manner with patients. However, the instructions might be improved by adding more *therapeutic* techniques besides the focus on style and form.

XSL•FO
RenderX

Introduction

Internet-Delivered Cognitive Behavioral Therapy

There is considerable evidence that Internet-based interventions are effective for the treatment of mild, moderate, and major depression [1-3]. Therapist-guided, Internet-delivered cognitive behavioral therapy (iCBT) has been found to be more effective than unguided iCBT [1,4] and has also been found to be equally effective compared with face-to-face-delivered CBT [5]. Beyond these findings, a number of studies focused on nonspecific factors that might be effective in iCBT (eg, therapeutic alliance, therapist competence, and placebo-expectancy effects) and especially showed interest in the role of therapist guidance in iCBT [6]. So far, mixed results have been found. In a systematic review, Richards and Richardson, eg, found that the way guidance is given has an impact on treatment adherence in depressed patients [4]. Therapist-guided iCBT had a 72% completion rate, iCBT interventions with administrative support (support by staff to guide patients through the program in a nontherapeutic way) 65%, and interventions with no support at all 26%. For Internet-delivered problem-solving treatment (PST), there also is evidence that the level of support is important in reaching effects for patients with depression [7]. Patients who received PST with weekly support from a coach improved significantly more than the waitlist control group. In the group that received no support, completion rates were lowest (22%), and the completion rates were highest in the group that received nonspecific support (60%). Patients who received weekly support had comparable completion rates with patients who received "support on request" (33% and 31%, respectively). In a study by Titov et al, patients with depression showed significant clinical improvement after receiving iCBT, regardless of whether the support came from a therapist or a technician [8]. Findings of a recent American study indicate that iCBT with 5 hours of therapeutic face-to-face contact was noninferior to CBT that provided over 8 additional hours of therapist contact for patients with depression [9].

Therapist Behaviors

However, there is much more to discover about online guidance. One point of interest is how therapists give online feedback to their patients. This can be done by looking at the communication strategies and content they use in their written support. For example, looking at therapist behaviors such as validating what patients write (eg, "That must be very difficult for you...") and stimulating patients to come up with their own solution (eg, "When was the last time you felt that way? What did you think and what did you do differently?"). Written feedback is a substantial part of Internet-based treatments and requires specific skills of therapists. It is therefore interesting to further explore such therapeutic microprocesses in online feedback because this part of therapy may be very relevant in the adherence and also the effectiveness of iCBT [10].

Therapist Behaviors in Face-to-Face Therapy

The content of feedback and its impact on treatment results have been studied in face-to-face–delivered psychotherapies, and especially in CBT. Studies have identified different therapist behaviors that are frequently used in CBT sessions with patients. These behaviors range from expressing empathy, making supportive communications (eg, encourage, praise, or guide the patient), asking directive questions, and confronting patients with different points of view [11-13]. Self-disclosures by therapists appear to be infrequently used [11], although these are generally considered helpful by patients in the therapeutic process [14,15]. In addition, research shows that therapist behaviors such as expressing empathy, giving positive regards, and confronting patients can have a positive impact on treatment outcome in CBT for various patients such as people with depression [11].

Therapist Behaviors in Internet-Delivered Cognitive Behavioral Therapy

Therapist behaviors in iCBT have also been studied. This was done for several psychiatric diseases such as eating disorders [16], insomnia [17], anxiety [18], and depression [19,20]. Comparable to the behaviors in face-to-face therapies, the most frequently used therapist behaviors were encouraging, reinforcing, and supporting patients. When looking at the association between these therapist behaviors, patient treatment outcome, and patient online session completion, mixed results were found. Holländare et al found that encouraging, guiding, and affirming were strongly associated with session completion [19]. Encouraging, affirming, and self-disclosure were weakly to moderately associated with an improvement in depressive symptoms. The most important finding by Paxling et al was the effect of therapists' task reinforcement (eg, reinforcing completed assignments) on session completion as well as treatment outcome [18]. Interestingly, a negative association was found between deadline flexibility of therapists and treatment outcome. Thus, the more lenient therapists were with homework assignment deadlines, the fewer patients improved. In a replication study of Schneider et al, the same type of therapist behaviors were found with the addition of a few more behavior categories (eg, asking questions) [20]. However, a different distribution of the therapist behavior frequencies was found, and the outcomes were different for patients with depression than for patients with anxiety. Thus, the way online feedback is provided by therapists differs across studies, patients, interventions, and possibly also the instructions used for feedback.

Online Feedback Instructions

In addition to more general communicative behaviors of therapists, the extent to which they follow instructions for online feedback may also influence treatment effectiveness. Research on written feedback predominantly stems from the field of education. Some of the main principles can be applied to online therapeutic feedback as well. Overall, research shows that effective written feedback is timely (provided in time), selective

(commenting only on 2 or 3 things that someone can change), balanced (pointing out positive aspects as well as areas in need of improvement), forward-looking (suggesting how to improve), and understandable (written in a language that someone will understand) [21]. Instructions for training therapists in written feedback are adapted to the therapeutic process but also comparable to those used in education (eg, beginning with a compliment, responding within 3 working days, or being careful with giving solutions). The elements in these instructions are primarily based on expert opinion rather than theory and mainly aim to motivate and support the patients, respond to the content of homework, and structure the feedback.

Study Objectives

In this study, written feedback will be studied in blended CBT, in the context of the European implementation project MasterMind [22-24], with a focus on therapist behaviors and on the extent to which feedback instructions are followed. In the Netherlands, iCBT for depression is slowly but increasingly adopted in routine mental health care mostly in a blended format. Blended CBT entails one integrated, standardized CBT treatment protocol that combines face-to-face sessions and digital modules to the best clinical benefit for patients and therapists [25]. The evidence of blended CBT over iCBT is unfortunately still scarce. Some first studies indicate that potential benefits of blended CBT are saving therapist time without reducing therapeutic outcome, lower treatment dropout rates, more emphasis on patient self-management, more face-to-face therapy time for deepening the CBT elements, and targeting another (often more complicated) population than iCBT [6,9,26]. The objectives of this study were to (1) identify therapist behaviors in written online communication from therapists to patients in blended CBT for adult depression in routine secondary mental health care, (2) identify the extent to which therapists adhere to feedback instructions, and (3) investigate whether therapist behaviors and therapist adherence to feedback instructions are associated with patient outcome (symptom change scores and number of completed online sessions).

Methods

Design

For the purpose of this observational study, the feedback messages of 45 Dutch patients that were offered blended CBT for depression by 19 therapists in routine mental health care were recruited between April 2015 and February 2017 from one outpatient clinic. This clinic was one of the participating MasterMind sites and was selected for this study because it offered a blended treatment protocol to patients within secondary health care, and the online usage information was made available for research. Patients received 219 feedback messages through a secure Web-based platform [27]. A mixed-method design was chosen to explore the content of the feedback messages: a directed qualitative content analysis [28] was used to identify therapist behaviors, and a checklist for the feedback instruction adherence of therapists was developed. To explore correlations between the frequency of therapist behaviors, scores on the checklist, and patients' outcomes, an explorative quantitative approach was used.

The study was approved by a Medical Ethics Committee. They confirmed that the "Medical Research Involving Human Subjects Act" does not apply (registration number 2014.580) because the patients in this study are not required to follow certain procedures on behalf of the research (no randomization) and routine practice was followed. An internal scientific research committee approved the research proposal (CWO 2015-005).

Participants

Patients

Patients were recruited through their therapists. Eligible patients received study information and an information leaflet from their therapist. After approval for telephone contact with researchers for additional information, patients received an informed consent. Patients were invited for participation in MasterMind if they (1) were aged 18 years or older; (2) had a mild, moderate, or severe depression as a primary diagnosis according to the therapist; and (3) were indicated for cognitive behavioral treatment for depression following routine secondary mental health care procedures. All patients needed to explicitly consent to take part in the study. Patients were excluded from the study if they (1) did not have a valid email address and did not have a computer with Internet access and (2) did not have adequate Dutch language skills (both verbal and written).

Therapists

Therapists who were trained in iCBT or who were motived for iCBT were invited to participate in the MasterMind study. They were recruited through team managers and eHealth attention officers of the different therapist teams. The iCBT training consisted of a 4-hour group training, provided by the outpatient clinic. During the training, the functionalities on the online platform were shown, and therapists got the chance to practice with a fictional patient. The therapists received individual instructions, access to the blended CBT treatment protocol online, and the feedback instructions. In addition, monthly 1-hour group sessions were organized where the therapists could exchange their experiences with each other.

The feedback instructions for therapists comprised general and specific elements that go in to the structure of the messages (eg, correct greeting, limiting to 2 subjects), readability (short sentences and paragraphs), writing style (eg, limiting abbreviations and misspellings, use of emoticons), referring to parts of the treatment (eg, filling in the diary, referring to the next online session), and communication skills (eg summarizing, not providing solutions).

Intervention

In the blended CBT treatment for depression of the outpatient clinic, it was agreed upon in advance that patients would receive 10 sessions online and meet with their therapist in 5 face-to-face sessions biweekly. In practice, therapists could deviate from the protocol by repeating online sessions. The online and individual face-to-face sessions were based on evidence-based treatment protocols for face-to-face CBT and are in agreement with multidisciplinary instructions for depression [29]. There were 4 core components: (1) psychoeducation, (2) cognitive restructuring, (3) behavioral activation, and (4) relapse

prevention. Besides the online sessions on the treatment platform, patients were given online access to a diary and filled out questionnaires to monitor their symptoms. After each completed online session, the therapist (the same therapist as in the face-to-face sessions) wrote a feedback message to the patient. Patient and therapist could additionally communicate through a message function about practical issues (eg, about upcoming appointments and reminders or questions about assignments).

Measures

Patient information on selected demographics (eg, age, gender, employment status) and clinical data (eg, use of medication) were obtained by an online self-report questionnaire at baseline. Demographic and background information (eg, treatment and iCBT experience) of the therapists were obtained by an online self-report questionnaire at the end of the study. Usage information (eg, number of online sessions followed and number of feedback messages) was obtained from the online platform.

Session completion was defined as the number of completed online sessions per patient. Symptom improvement was measured with the 16-item Quick Inventory of Depressive Symptomatology (QIDS) [30]. The total score varies from 0 to 27, with higher scores being indicative of a higher severity of depressive symptoms. The QIDS was administered weekly on the online platform during the course of the treatment. The number of QIDS measures can vary, with up to 30 weekly measures. Of each patient, the baseline scores were included, and the last known value was used as a posttreatment score. The change score on the QIDS was calculated by subtracting the baseline measurement from the final measurement.

Coding of Therapist Behavior and Adherence

To subtract therapist behaviors from the 219 online feedback messages, a coding matrix was developed, with 9 main categories and 13 subcategories (see Multimedia Appendix 1). The coding categories were based on the directed content analysis; categories from prior research [18,19,31] were used to develop the initial coding scheme before analyzing the data [25]. To score the therapists' adherence to the feedback instructions, a coding checklist was created based on the instructions that the therapists received. In total, there were 6 main categories and 19 different subcategories with a dichotomous scale (present or not present, see Multimedia Appendix 2).

The coding matrix and checklist were first tested by researcher ED by coding 4 feedback messages from 2 randomly selected patients. Each of the included feedback messages was then anonymously coded and scored by researchers MM and SP. For the coding of therapist behaviors, qualitative data analysis software, ATLAS.ti 7.5.18 (ATLAS.ti Scientific Software Development GmBH, Berlin, Germany), was used.

To investigate interrater reliability, both researchers (MM and SP) coded 60 transcripts of therapeutic feedback from 10 randomly selected patients. The intraclass correlation coefficient for the therapist behaviors was .83 (95% CI 0.82-0.85) indicating good interrater reliability, based on 2-way mixed-effects

agreement model [32,33]. The interrater reliability for the feedback instruction adherence categories was found to be kappa (κ)=.84 (P<.001, 95% CI 0.80-0.87), indicating a good agreement between the raters as well. After reaching agreement, the remaining messages (n=159) were equally divided between the 2 coders. As analysis proceeded, additional codes were developed, and the initial coding matrix and checklist were revised, discussed, and refined.

The total frequency of therapist behaviors was calculated with a query tool in ATLAS.ti. A frequency score represented the total number of times the therapist displayed a behavior in the feedback messages sent to the patient (eg, total number of informing the patient about the assignments). To correct for the number of received feedback messages (eg, some patients received 4 messages, and the others received 8 messages), relative frequencies were used (frequency of one category divided by the total number of frequencies of all categories per patient). The percentage of therapists' adherence to the instructions for each patient was calculated by the frequency of the adherence (eg, the total number of times a therapist started with giving a compliment) divided by the total number of messages received by a patient.

Analyses

Statistics were conducted using IBM SPSS (SPSS Inc., Chicago IL), version 22. First, descriptive statistics (means, SDs, percentages) were used to describe the patient and therapist sample, number of online sessions, and symptom improvement. Descriptive statistics were then used to examine the frequencies of therapist behaviors and percentages of therapist instruction adherence in the messages to the patients. Spearman correlation analyses, 2-sided, were conducted to assess the relationship between the therapist behaviors, feedback adherence scores and session completion, and symptom improvement. Spearman rho was used to avoid violation of assumptions of normality. Due to the small sample size, only explorative analysis, no missing values imputation techniques and no post-hoc correction for multiple testing (ie, Bonferroni), were applied.

Results

Patients' and Therapists' Characteristics

A total of 45 patients (73%, 33/45 female, mean age 35.9 years) were given blended CBT in routine care by 19 therapists. Patients' characteristics can be found in Table 1. Of the 19 therapists (84%, 16/19 female), most were licensed psychologists (53%, 10/19), others were psychologists in training under supervision for health care psychologists (26%, 5/19) or mental health nurses (21%, 4/19). Moreover, 11% (2/19) of the therapists had less than 3 years of professional experience, 26% (5/19) had between 3 and 5 years of experience, 37% (7/19) had between 5 and 10 years of experience, and 21% (4/19) had more than 10 years of experience. The experience with iCBT treatments varied among the therapists: 32% (6/19) had given less than 5 iCBT treatments, 26% (5/19) had given between 5 and 10 treatments, 21% (4/19) had given between 10 and 15 treatments, and 16% (3/19) had given more than 15 treatments.

Table 1. Characteristics of patients at baseline.

Patients' characteristics	Statistics (n=45)
Gender, female, n (%)	33 (73)
Age in years, mean (SD; range)	35.9 (12.3; 21-64)
Education level, n (%)	
Secondary education level	15 (37)
Higher education level	25 (61)
Employment, yes, n (%)	21 (51)
Antidepressant use, yes, n (%)	10 (24)
Duration of current depression symptoms, n (%)	
Duration of current depression symptoms less than 3 months	8 (20)
Duration of current depression symptoms between 3 and 12 months	22 (54)
Duration of current depression symptoms more than 1 year	10 (24)

Frequencies of Therapist Behaviors and Percentages of Therapist Instruction Adherence

Multimedia Appendix 1 lists the categories, definitions, and examples of the therapist behaviors. In total, 1825 therapist behaviors were coded. The most frequently used therapist behaviors were informing (27.56%, 503/1825; eg, informing the patient about the next session or specific assignments), encouraging (23.56%, 430/1825; eg, praising past behavior), and affirming (22.25%, 406/1825; eg, normalizing behavior, summarizing what the patient has written or said). Making self-disclosures, confronting, and emphasizing the responsibility of the patient were never or rarely used.

An overview of the percentages of the categories, descriptions, and examples of adherence to the feedback instructions can be found in Multimedia Appendix 2. The therapists adhered in most cases to correct greeting and ending of messages (95.9%, 210/219). They also scored high on writing style (93.6%, 205/219; eg, limiting of abbreviations and misspellings) and structure (87.7%, 192/219; eg, limiting to 2 subjects and sending the feedback within 3 working days). Therapists scored the lowest on referring (34%, 74.5/219; eg, referring to monitoring of the symptoms or reflecting on the dairy). Within the category communication skills, therapists were very often careful with giving solutions (95.6%, 209/219) and regularly showed in their writing that they read the patients' homework (88.6%, 194/219). Formulating sentences as hypotheses is something the therapists did not often apply (10.5%, 23/219).

Session Completion and Symptom Improvement

The 45 patients completed, on average, 6.3 online sessions (Table 2). On average, patients received 4.9 feedback messages (SD 2.7; range 1-10). One feedback message contained an average of 139 words (SD 95.4; range 1-504), 14.2 words in one sentence (SD 4.0; range 1-26), 3.5 sentences in a paragraph (SD 1.9; 1-17), and 2.9 paragraphs (SD 1.6; 1-10).

From 7 patients, all QIDS data were missing because their therapists did not activate the online monitoring, leaving 38

patients for this exploration. Results on depressive symptoms showed that at baseline, the patients scored, on average, 15.8 points (SD 3.8) on the QIDS, and at postmeasurement, the patients scored, on average, 11.0 points (SD 6.0), so there was an average reduction of 4.8 points (SD 6.4). Looking at symptom severity at baseline, 8% (3/38) of the patients had mild symptoms, 34% (13/38) had moderate symptoms, and 58% (22/38) had (very) severe symptoms (Table 3). At posttreatment, 21% (8/38) of the patients had no symptoms, 29% (11/38) had mild symptoms, 24% (9/38) had moderate symptoms, and 26% (10/38) had (very) severe symptoms. In total, 63% (24/38) of the patients improved on one or more categories (Table 4). Moreover, 24% (9/38) of the patients showed no change, and 13% (5/38) deteriorated in a category.

Multimedia Appendix 3 contains case descriptions of 3 patients.

Correlations of Therapist Behaviors With Session Completion and Symptom Improvement

One correlation between therapist behaviors and session completion was found (Table 5): the therapist behavior confronting was positively correlated with online session completion (ρ=.342, P=.02). This indicates that more confrontations were related to completing more online sessions. No significant correlations were found with symptom improvement.

Correlations of Therapist Instruction Adherence With Session Completion and Symptom Improvement

In Table 6, correlations of therapist instruction adherence with session completion and symptom improvement are shown. Statistically significant negative medium correlations were found between therapist instruction adherence and completed online sessions for structure (ρ=−.340, P=.02) and readability (ρ=−.361, P=.02). Meaning that the more therapists adhered to instructions containing structure (limiting to 2 subjects and sending feedback within 3 working days) and readability (short sentences and short paragraphs), the less online sessions were completed. No significant correlations were found with symptom improvement.

Table 2. Treatment completion and duration (n=45).

Treatment completion and duration	Mean (SD; range)
Completed online sessions,	6.3 (2.6; 2-11)
Completed face-to-face sessions	7.1 (2.7; 2-13)
Completed face-to-face + online sessions	13.4 (4.4; 5-23)
Treatment duration in weeks	26.2 (11.2; 8-52)
Period of online activity in weeks	17.8 (10.9; 2-45)

Table 3. Severity Quick Inventory of Depressive Symptomatology scores at baseline and postmeasurement.

Severity Quick Inventory of Depressive Symptomatology	Quick Inventory of Depressive Symptomatology (n=38), n (%)	
	Baseline	Postmeasurement
None	0 (0)	8 (21)
Mild	3 (8)	11 (29)
Moderate	13 (34)	9 (24)
Severe	20 (53)	5 (13)
Very severe	2 (5)	5 (13)

Table 4. Changes in symptom severity (n=38).

Change in depressive symptom severity	n (%)
Reduction in 1 category	11 (29)
Reduction in 2 categories	10 (26)
Reduction in 3 categories	3 (8)
Deterioration	5 (13)
No change	9 (24)

Table 5. Correlations of therapist behaviors with session completion and symptom improvement.

Therapist behavior	Session completion (n=45)	Change score Quick Inventory of Depressive Symptomatology (n=38)
Emphasizing responsibility	.094	.278
Affirming	.074	.035
Clarifying the framework	.232	.069
Self-disclosure	—[a]	—
Informing	−.087	−.249
Confronting	.342[b]	.184
Urging	.258	.310
Encouraging	−.054	−.008
Guiding	−.055	.146
Questions	.066	.115

[a]Indicates "not applicable"; self-disclosures did not occur.

[b]$P<.05$, a positive correlation indicates more session completion.

Table 6. Correlations of therapist instruction adherence with session completion and symptom improvement.

Therapist instruction adherence	Session completion (n=45)	Change score Quick Inventory of Depressive Symptomatology (n=38)
Greeting and ending	−.277	−.064
Communication skills	−.146	−.212
Structure	−.340[a]	−.214
Referring	.170	−.085
Readability	−.361[a]	−.185
Writing style	−.150	−.139

[a]$P<.05$, a positive correlation indicates more session completion.

Discussion

Aim of This Study

This observational study has uncovered several important factors in the content of online feedback messages in blended iCBT for depression. We further explored therapist behaviors and the extent to which therapists wrote their feedback according to their instructions. In addition, we wanted to know if therapist behaviors and adherence to the feedback instructions could be linked to patient adherence and treatment outcome. The study was carried out in a Dutch sample of participants of the MasterMind study, in routine practice, in a patient population with mild to (very) severe depressive symptoms and with a diverse group of trained and skilled therapists.

Principal Findings

Results show that therapist behaviors in relation to the online guidance are informing the patient about the functionalities on the platform, encouraging the patient by praising past behavior or inciting future behavior, and affirming by showing interest in the thoughts, emotions, and behaviors of the patient. Making self-disclosures, confronting, and emphasizing the responsibility of the patient are never or infrequently used. This is largely in line with the frequencies of the categories found by Holländare et al and may indicate that therapists use the same CBT principles in their written communication as in their face-to-face communication with the patient [19]. Previous research also found that more supportive therapists' behaviors are used frequently in iCBT and that behaviors such as confronting and self-disclosures are seldom used [18,28]. However, in contrast to the findings of Holländare et al, we found that one and also a different therapist behavior correlated with module completion, and we also found that none of the therapist behaviors were related to symptom improvement. A possible explanation for this difference can be found in the patient group; in Holländare et al's study, patients with partially remitted depression were included within the context of a randomized controlled trial.

Although therapists applied confronting in limited cases (<1%), this was positively correlated with online session completion. In face-to-face CBT, the occurrence of confrontations has been found to be somewhat higher (6%-14.3%), but is also significantly correlated with therapy outcomes [11]. Hill et al argued that "confrontation often interrupts the client's thinking by presenting discrepancies and another point of view […].

Although confrontation feels negative at the time, such disruption may be a necessary foundation for change" [15].

Furthermore, therapists followed the feedback instructions that were used in this study on most of the defined elements, such as beginning with a compliment and being careful about providing solutions too soon. Different than expected, only half of the therapists formulated their sentences as hypotheses, and did so in only 10% of the feedback messages (eg, "It sounds like you are not sure, is that correct?"). Misspellings occurred regularly: in 21.5% of the feedback messages, therapists made more than 3 spelling mistakes. One of the possible explanations for this is that the treatment platform did not contain a spelling corrector, and it may have taken therapists more time to correct their own writing. Emoticons were not used often, as only 3 therapists sometimes used an (positive) emoticon. In the "Supportive Accountability" model by Mohr and Cuijpers, it is argued that therapists may mirror the content, style, tense, and cues (eg, emoticons) in online communication by patients to create mutual trust [31]. In this model, it is also pointed out that people pay attention to the timing and date stamps of the responses. This means responses should be timely because delays may be perceived as expressing lack of affection. In our study, the therapists sent their feedback within the limit of 3 working days in almost 80% of the cases.

Only negative associations were found with therapist instruction adherence and session completion. Providing structure and the readability was significantly negatively associated with session completion. This means that if the therapists adhered more to writing short sentences and paragraphs and the more they limited their feedback to 2 different subjects and sent the feedback back within 3 working days, the less online sessions were completed. These findings might be explained by the adaptive, and also reactive, style of the therapists to the behavior of the patient. When patients are doing well on the online platform, they are more flexible with certain elements of the instructions. On the other hand, when patients display more difficulties or when the therapist gets the feeling that he or she is losing contact with the patient, therapists may be more inclined to adhere more to some parts of the instructions. Schneider et al also found that therapists were responsive in their online feedback and that they increased some behaviors during the course of treatment when patient depressive symptoms worsened. There are similar indications in psychotherapy, where more flexibility of therapists

XSL•FO
RenderX

was found related to better treatment outcomes than therapists who were less flexible [34].

Strengths and Limitations

The study took place in a naturalistic setting, with routine care patients and therapists and without the restrictions of a randomized controlled trial. Patient demographic characteristics in the study sample are comparable to blended CBT research, also in routine care [9,26]. Previous studies were carried out in small samples of therapists (3-5), often trained students, who delivered treatment in a research setting [16-19]. With the use of a directed approach of the content analysis, the findings of the previous research were supported and extended. We found the same proportions of categories as Holländare et al with the addition of the category "asking questions" [19]. This was also found by Schneider et al when they replicated the study conducted by Paxling et al [18,20].

In addition, there are several limitations to this study. The generalizability of the results is limited because of the small sample size. With a greater sample size, it would have been possible to explore initial symptom severity as a predictor of the use of different therapist behaviors. The exploration of this association would be interesting for further research. Second, although this study was able to capture a group of experienced professionals, the distribution of patients over the therapist was slightly skewed. Half of the therapists treated 3 to 6 patients, and the other half treated 1 or 2 patients. Due to the small sample size, it was not possible to explore potential differences in writing style or skills between the therapists. Furthermore, in face-to-face treatment, therapist characteristics such as age, gender, and ethnicity of the therapist seemed not to be related to patient treatment outcomes [35], but therapist facilitative interpersonal skills were found to be a successful predictor of treatment outcome [1]. To further explore therapist online feedback, it would be interesting to look at therapeutic skills as well. In this study, there was a high variability in the number of words in the feedback messages, and it could also be interesting to further explore this. Third, within this observational study, only prepost data and explorative analyses and, no post-hoc corrections, were used. The found correlations should be interpreted with care. Previous research showed that it is possible that therapist behaviors change over the course of treatment, with more focus on certain categories at the beginning of treatment versus the end of treatment [19,20]. Finally, this study only focused on the online part of the blended treatment and not on the content in the face-to-face sessions.

Conclusions

In sum, this study showed that in blended CBT for depression, therapists primarily used supportive and positive communications like informing, encouraging, and affirming patient behavior. Therapists refrained from using therapeutic techniques, such as making self-disclosures, urging, and confronting. This can be explained by the way the online feedback instructions were constructed. They provided the therapists guidelines that concentrate on style and form instructions, and this is also reflected in the adherence of the therapists to most of these instructions. It can be suggested that the instructions should also focus more on "disruptive" therapeutic techniques that can foster patients to address their symptoms. The blended format can give the therapist more flexibility in writing feedback because of the combination with face-to-face contact, meaning that therapists can check the interpretation of their online feedback with the patients in the face-to-face sessions. The combination with online contact gives the therapist the possibility to incorporate elements and reflect on issues that were discussed in the face-to-face sessions. On the other hand, therapists are aware that online communications can emotionally positively and also negatively affect the patient, without them being there, and are therefore careful in their communications. The therapists may miss nonverbal cues such as facial expressions and are not able to respond immediately. Writing feedback requires the therapist to assess whether the patient can correctly understand it. The extent to which this calls for specific competencies of the "online" therapist is assumed and requires further exploration. Additional research is needed to further explore the content of online feedback. With an experimental design, more causal explanations can, eg, be made about the amount of certain therapist behaviors, the interaction with the written content of the patients, patient expectations or the timing of feedback, and also the interaction with the contact of the face-to-face sessions. With more knowledge, instructions on feedback can be enriched, and therapists can be offered more guidance in giving feedback.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Main and subcategories therapist behaviors, definitions, examples and percentages out of 219 feedback messages.

[PDF File (Adobe PDF File), 42KB - jmir_v20i5e174_app1.pdf]

Multimedia Appendix 2

Main and feedback instructions, definitions, examples and percentages out of 219 feedback messages.

[PDF File (Adobe PDF File), 46KB - jmir_v20i5e174_app2.pdf]

Multimedia Appendix 3

Case descriptions of 3 patients.

[PDF File (Adobe PDF File), 42KB - jmir_v20i5e174_app3.pdf]

References

1. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
2. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
3. Josephine K, Josefine L, Philipp D, David E, Harald B. Internet- and mobile-based depression interventions for people with diagnosed depression: a systematic review and meta-analysis. J Affect Disord 2017 Dec 01;223:28-40. [doi: 10.1016/j.jad.2017.07.021] [Medline: 28715726]
4. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
5. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
6. Schueller SM, Tomasino KN, Mohr DC. Integrating human support into behavioral intervention technologies: the efficiency model of support. Clin Psychol Sci Pract 2016 Nov 17;24(1):27-45. [doi: 10.1111/cpsp.12173]
7. Kleiboer A, Donker T, Seekles W, van Straten A, Riper H, Cuijpers P. A randomized controlled trial on the role of support in Internet-based problem solving therapy for depression and anxiety. Behav Res Ther 2015 Sep;72:63-71. [doi: 10.1016/j.brat.2015.06.013] [Medline: 26188373]
8. Titov N, Andrews G, Davies M, McIntyre K, Robinson E, Solley K. Internet treatment for depression: a randomized controlled trial comparing clinician vs. technician assistance. PLoS One 2010 Jun 08;5(6):e10939 [FREE Full text] [doi: 10.1371/journal.pone.0010939] [Medline: 20544030]
9. Thase ME, Wright JH, Eells TD, Barrett MS, Wisniewski SR, Balasubramani GK, et al. Improving the efficiency of psychotherapy for depression: computer-assisted versus standard CBT. Am J Psychiatry 2018 Mar 01;175(3):242-250. [doi: 10.1176/appi.ajp.2017.17010089] [Medline: 28969439]
10. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012 Jul;12(7):861-870. [doi: 10.1586/ern.12.63]
11. Keijsers GP, Schaap CP, Hoogduin CA. The impact of interpersonal patient and therapist behavior on outcome in cognitive-behavior therapy. A review of empirical studies. Behav Modif 2000 Apr;24(2):264-297. [doi: 10.1177/0145445500242006] [Medline: 10804683]
12. Watson JC, McMullen EJ. An examination of therapist and client behavior in high- and low-alliance sessions in cognitive-behavioral therapy and process experiential therapy. Psychother Theor Res Pract Train 2005;42(3):297-310. [doi: 10.1037/0033-3204.42.3.297]
13. Watzke B, Rueddel H, Koch U, Rudolph M, Schulz H. Comparison of therapeutic action, style and content in cognitive-behavioural and psychodynamic group therapy under clinically representative conditions. Clin Psychol Psychother 2008 Nov;15(6):404-417. [doi: 10.1002/cpp.595] [Medline: 19115459]
14. Knox S, Hill CE. Therapist self-disclosure: research-based suggestions for practitioners. J Clin Psychol 2003 May;59(5):529-539. [doi: 10.1002/jclp.10157] [Medline: 12696129]
15. Hill CE, Helms JE, Tichenor V, Spiegel SB, O'Grady KE, Perry ES. Effects of therapist response modes in brief psychotherapy. J Couns Psychol 1988;35(3):222-233. [doi: 10.1037/0022-0167.35.3.222]
16. Sánchez-Ortiz VC, Munro C, Startup H, Treasure J, Schmidt U. The role of email guidance in internet-based cognitive-behavioural self-care treatment for bulimia nervosa. Eur Eat Disord Rev 2011 Jul;19(4):342-348. [doi: 10.1002/erv.1074] [Medline: 21394832]
17. de Bruin EJ, Meijer AM. The impact of online therapeutic feedback on outcome measures in Internet-CBTI for adolescents with insomnia. Sleep Med 2017 Jan;29:68-75. [doi: 10.1016/j.sleep.2016.05.017] [Medline: 27866826]
18. Paxling B, Lundgren S, Norman A, Almlöv J, Carlbring P, Cuijpers P, et al. Therapist behaviours in internet-delivered cognitive behaviour therapy: analyses of e-mail correspondence in the treatment of generalized anxiety disorder. Behav Cogn Psychother 2013 May;41(3):280-289. [doi: 10.1017/S1352465812000240] [Medline: 22717145]

19. Holländare F, Gustafsson SA, Berglind M, Grape F, Carlbring P, Andersson G, et al. Therapist behaviours in internet-based cognitive behaviour therapy (ICBT) for depressive symptoms. Internet Interv 2016 Mar;3:1-7. [doi: 10.1016/j.invent.2015.11.002]

20. Schneider LH, Hadjistavropoulos HD, Faller YN. Internet-delivered cognitive behaviour therapy for depressive symptoms: an exploratory examination of therapist behaviours and their relationship to outcome and therapeutic alliance. Behav Cogn Psychother 2016 Nov;44(6):625-639. [doi: 10.1017/S1352465816000254] [Medline: 27302220]

21. McKeachie WJ, Svinicki MD. McKeachie's Teaching Tips: Strategies, Research, and Theory for College and University Teachers (College Teaching Series). Belmont: Wadsworth; 2010.

22. Vis C, Kleiboer A, Prior R, Bønes E, Cavallo M, Clark SA, et al. Implementing and up-scaling evidence-based eMental health in Europe: the study protocol for the MasterMind project. Internet Interv 2015 Nov;2(4):399-409. [doi: 10.1016/j.invent.2015.10.002] [Medline: 26188373]

23. Mol M, Dozeman E, van Schaik DJ, Vis CP, Riper H, Smit JH. The therapist's role in the implementation of internet-based cognitive behavioural therapy for patients with depression: study protocol. BMC Psychiatry 2016 Sep 30;16(1):338 [FREE Full text] [doi: 10.1186/s12888-016-1045-9] [Medline: 27716108]

24. MasterMind. 2018. MasterMind Homepage URL: http://mastermind-project.eu/ [WebCite Cache ID 6wF4PBAQj]

25. Triple-E. Definition blended treatment Heleen Riper, 2017 URL: https://www.triple-ehealth.nl/en/projects-cat/depressie/ [WebCite Cache ID 6wZFBY9pX]

26. Kenter RM, van de Ven PM, Cuijpers P, Koole G, Niamat S, Gerrits RS, et al. Costs and effects of Internet cognitive behavioral treatment blended with face-to-face treatment: results from a naturalistic study. Internet Interv 2015 Mar;2(1):77-83. [doi: 10.1016/j.invent.2015.01.001]

27. Minddistrict. Minddistrict platform URL: https://www.minddistrict.com/ [WebCite Cache ID 6wF4sBjHm]

28. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]

29. Spijker J, Bockting C, Meeuwissen JA, Van Vliet IM, Emmelkamp PM, Hermens ML, et al. Multidisciplinaire richtlijn Depressie (Derde revisie): Richtlijn voor de diagnostiek, behandeling en begeleiding van volwassen patiënten met een depressieve stoornis [Multidisciplinary guideline Depression (Third revision): Guideline for the diagnosis, treatment and counseling of adult patients with a depressive disorder]. Utrecht: Trimbos Institute; 2013.

30. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biol Psychiatry 2003 Sep 01;54(5):573-583. [Medline: 12946886]

31. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]

32. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159. [doi: 10.2307/2529310]

33. Portney L, Watkins M. Foundations of Clinical Research: Applications to Practice. Michigan: Prentice Hall Health; 2000.

34. Owen J, Hilsenroth MJ. Treatment adherence: the importance of therapist flexibility in relation to therapy outcomes. J Couns Psychol 2014 Apr;61(2):280-288. [doi: 10.1037/a0035753] [Medline: 24635591]

35. Beutler LE, Malik ML, Alimohamed S, Harwood TM, Talebi H, Noble S, et al. Therapist variables. In: Lambert MJ, editor. Bergin and Garfield's Handbook of Psychotherapy and Behavior Change. New York: Wiley; 2004:227-306.

Abbreviations

CBT: cognitive behavioral therapy
iCBT: Internet-delivered cognitive behavioral therapy
PST: problem-solving treatment
QIDS: Quick Inventory of Depressive Symptomatology

XSL·FO

RenderX

Effectiveness of Internet-Based Interventions on Glycemic Control in Patients With Type 2 Diabetes: Meta-Analysis of Randomized Controlled Trials

Ying Shen[1], BS; Fengbin Wang[1,2], MM; Xing Zhang[1], MB; Xiaorou Zhu[1], MB; Qiudan Sun[3], MA; Edwin Fisher[4], BA, PhD; Xinying Sun[1], PhD

[1]School of Public Health, Peking University Health Science Center, Beijing, China

[2]Department of Health and Education, Beijing Municipal Commission of Health and Family Planning, Beijing, China

[3]Institute for Medical Humanities, Peking University Health Science Center, Beijing, China

[4]Gillings School of Global Public Health, University of North Carolina, North Carolina, NC, United States

Corresponding Author:
Xinying Sun, PhD
School of Public Health
Peking University Health Science Center
38, Xueyuan Road
Haidian District, Beijing, China
Beijing, 100191
China
Phone: 86 86 13691212050
Fax: 86 86 1082801743
Email: xysun@bjmu.edu.cn

Abstract

Background: The popularity of internet as an area of research has grown manifold over the years. Given its rapid development and increasing coverage worldwide, internet-based interventions seem to offer a promising option to ameliorate huge burdens brought by type 2 diabetes mellitus. However, studies conducted by different researchers have provided contradictory results on the effect of internet-based interventions in glycemic control.

Objective: This meta-analysis aims to summarize currently available evidence and evaluate the overall impact of internet-based interventions on glycemic management of type 2 diabetic patients.

Methods: A systematic literature search was performed in PubMed, ScienceDirect, and Web of Science. Randomized controlled trials that used glycosylated hemoglobin values as the outcome measure of glycemic control were considered. Risk of bias and publication bias were evaluated.

Results: Of the 492 studies, 35 were included in meta-analysis, and results indicated that the weighted mean difference (WMD) between usual care and internet-based interventions at endpoint was –0.426% (95% CI –0.540 to –0.312; $P<.001$). Subgroup analyses revealed that intervention duration ≤3 months yielded optimal performance (WMD –0.51%; 95% CI –0.71 to –0.31; $P<.001$). Combined mobile and website interventions were substantially superior to solely Web-based and mobile-based interventions in glycemic control (combined WMD –0.77%, 95% CI –1.07 to –0.47; $P<.001$; Web only: WMD –0.48%; 95% CI –0.71 to –0.24, $P<.001$; mobile only WMD –0.31%, 95% CI –0.49 to –0.14; $P<.001$). Furthermore, the effect of interventions with automated feedbacks was similar to those with manual feedbacks, and studies with internet-based educational contents were more effective in glycemic control. The assessment revealed a low risk of bias.

Conclusions: In conclusion, utilization of internet-based intervention is beneficial for patients with type 2 diabetes mellitus, and taking full advantage of this type of intervention may substantially reduce the incidence of complications and improve quality of life.

Trial Registration: International Prospective Register of Systematic Reviews (PROSPERO): CRD42017058032; https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=58032 (Archived by WebCite at http://www.webcitation.org/6yY7eQNHr)

XSL•FO

RenderX

Introduction

With increasing prevalence and serious chronic and acute complications, diabetes has brought enormous burden to people's living and production. Glycemic control is vital in disease management as hyperglycemia and hypoglycemia can both lead to serious clinical consequences. However, in some areas, health care providers can only obtain patients' information on glycemic control and lifestyle during their consultation with a doctor, which is probably infrequent because of various reasons, such as patients' economic issues, educational levels, health awareness, and remote distance to health care centers. Hence, information obtained by doctors cannot comprehensively reflect patients' health status and may lead to suboptimal health decisions [1]. Therefore, most of the impetus of disease management is on the patients as diabetes is a life-course chronic condition, and health care providers cannot possibly accompany patients all the time [2]. As a result, to achieve wide-scale promotion of diabetes self-management, it is essential to explore an efficient, easily available, and cost-effective approach to bridge geographical and economic gaps and reach broader populations. Given the rapid development of new technology and penetration of internet in people's daily life, integrating diabetic inventions into existing technologies may have important implications and accord with clinical guidelines as they have widely referred to the significance of self-management and patients' education in diabetes control [3-5]. In this aspect, internet-based interventions perfectly meet the requirement, as they can not only easily reach the global population but also serve as a medium of mass information exchange [6].

Currently, many studies have explored the efficacy of using internet-based interventions in diabetes management. However, there is not a universally acknowledged verdict on whether internet-based interventions could make a significant difference in patients' glycemic level. Some research reported that compared with usual care, internet-based interventions could significantly reduce glycated hemoglobin (HbA_{1c}) levels in patients with type 2 diabetes mellitus (T2DM) [7-11], whereas others reported otherwise [12-16]. Therefore, to comprehensively evaluate the overall impact of internet-based interventions, a meta-analysis is needed to achieve the pooled result. Furthermore, to our best knowledge, although there were some meta-analyses studying the effect of internet-based interventions on diabetes management, most of them just looked at a single form such as mobile apps and few conducted a comprehensive comparison between different forms of internet-based interventions. As a result, by analyzing relevant randomized controlled trials (RCTs) on this issue, this meta-analysis aims to evaluate the feasibility and efficacy of internet-based interventions on diabetes management and hopes to discover the most effective model of intervention that will benefit as many patients as possible.

This meta-analysis was registered at the international prospective register of systematic reviews (no. CRD42017058032).

Methods

Data Sources and Searches

For this meta-analysis, we have conducted a search in PubMed, Web of Science, and ScienceDirect to identify studies relevant to this topic from their beginning to January 16, 2017. Keywords used in this search were "T2DM," "type 2 diabetes," "social media," "e-learning," "new media," "m-health," "mHealth," "internet-based," "web-based," "twitter," "eHealth," "e-health," "Facebook," "computer-delivered," "social web," "social software," "online case-based learning," "mobile learning," "digital game-based learning," "serious games," "wearable devices," "mobile app," "mobile application," and "smart phone-based." To achieve a searching scope that was broad enough, the above-mentioned keywords were not enclosed. In PubMed, we searched in all fields and found 291 papers. In both of Web of Science and ScienceDirect, we searched in title, abstracts and keywords and retrieved 228 and 65 papers, respectively. The language of studies was not limited.

To extend our search scope and include more studies in this field, we also manually searched the reference lists of several systematic reviews and meta-analyses that might be related to our own study [6,17-23]. The number of studies identified in this way was 76. Furthermore, in the process of records screening, we found that 1 study was built on a previously published research conducted by the same research group, which examined the efficacy of improvement made to the previous program and evaluated people's acceptance. As a result, the study of the previous research was also found through the reference list and was included. Therefore, a total of 77 studies were retrieved from the references lists of other research. After the removal of duplicates, 492 records were identified in total. Multimedia Appendix 1 shows details of search syntax.

Study Selection

Two reviewers, Shen and Wang, independently reviewed abstracts and full-text papers. Study selection was strictly based on uniform inclusion and exclusion criteria. The exclusion reason for each study was recorded, and in the case of disagreement, Sun was consulted, and discussion was held until a consensus was reached. Studies were included if they were RCTs, assessed patients aged older than 18 years with T2DM, compared the effect of internet-based interventions in glycemic management with a control group (CG) of usual care, and reported means and SDs of HbA_{1c} values for intervention group (IG) and CG both at baseline and endpoint or other relevant data from which mean and SD of HbA_{1c} could be calculated. If information on diabetes type was not clear, studies with the mean age of patients older than 30 years were included as there is a high possibility that they have T2DM [24]. Studies were

excluded if they were nonrandomized, reviews, protocols, case reports, or commentaries; did not use HbA_{1c} as the outcome measure or there was an incomplete report of HbA_{1c}; or included patients with type 1 diabetes or gestational diabetes. Furthermore, we also excluded studies that were not predominantly internet-based or the internet was only used as a supporting or additional tool to intensify the effect of other interventions, as it was difficult to tell whether the study effect was attributable to the internet or the other intervention. Interventions that purely used the mobile phone as a tool to send text messages were also excluded, as this was not internet-based. Moreover, to compare the effect of internet-based interventions with usual care, the CG should use usual care rather than internet-based interventions.

On the basis of retrieved abstracts, we excluded 398 records. We further examined the full texts of the remaining 94 studies, and 59 studies were excluded. A total of 25 studies did not provide data on HbA_{1c} or provided incomplete data on HbA_{1c}; 15 studies also studied patients with T1DM; 7 studies used internet-based intervention in CG; interventions in 5 studies

were designed for health professionals and still depended on face-to-face interactions; 12 studies reported results of the same 6 interventions, respectively, and thus 6 were excluded; and in 1 study, only 46.1% of patients (n=511) used the internet at least monthly; therefore, the research group decided to send participants paper versions of their medical record and ask them to bring it during the next hospital visit. Therefore, as this research was largely not internet-based, it was also excluded (Figure 1).

Data Extraction and Quality Assessment

The following data in 35 studies were extracted using a standardized spreadsheet: the first author, year of publication, participants included (CG and IG, respectively), patients' characteristics (including attrition, completion rate, mean age, body mass index [BMI], and gender ratio), study design, intervention duration, study location, intervention method, HbA_{1c} data (including values at baseline and endpoint), and other useful information. Furthermore, studies included were numbered from 1 to 35. One study had 2 IGs that met the eligibility criteria and was numbered 12a and 12b.

Figure 1. Flowchart of study selection. RCT: randomized controlled trials; T2DM: type 2 diabetes mellitus.
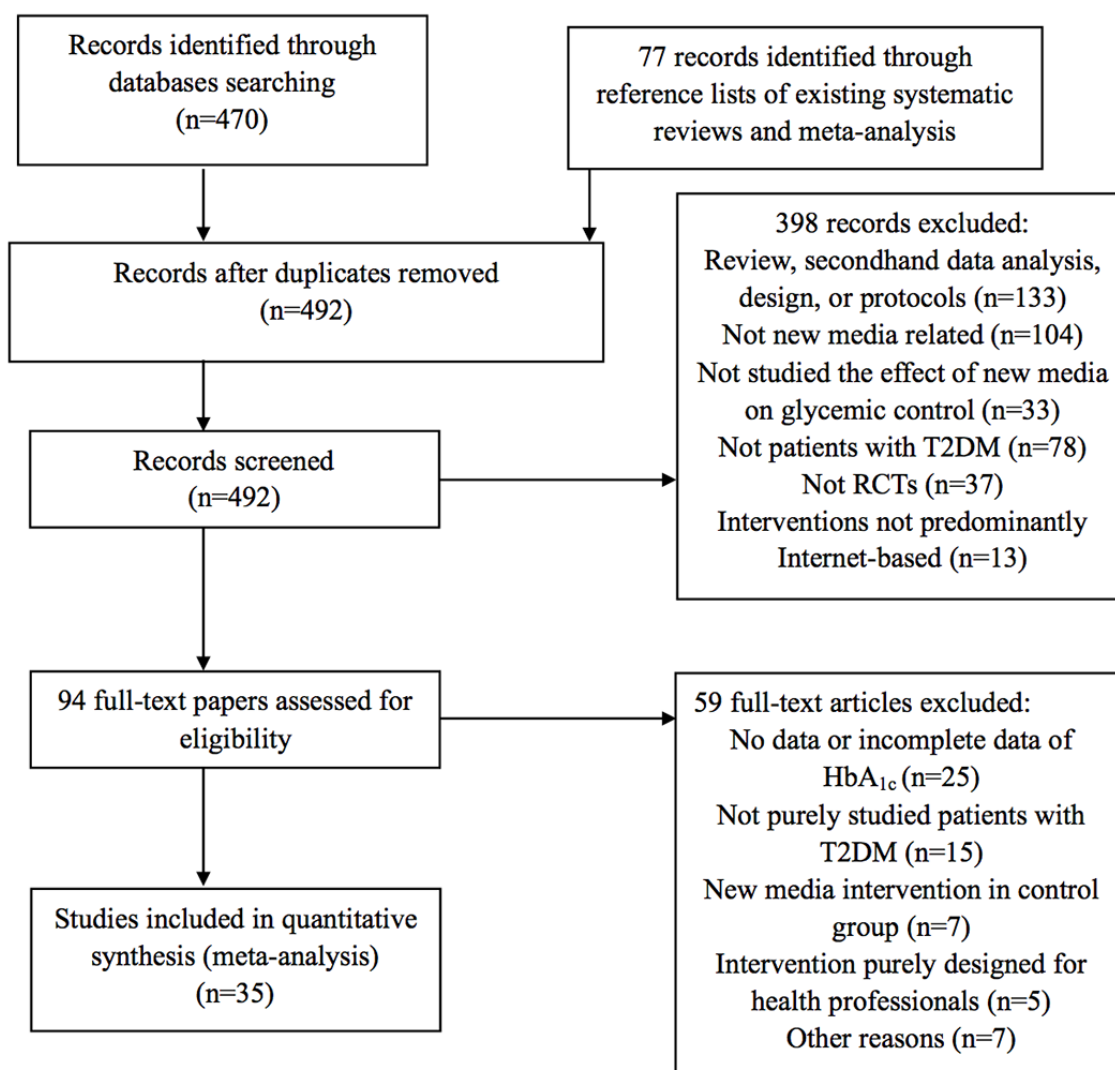
Mean values and SD of HbA_{1c} at baseline and endpoint were extracted separately for IGs and CGs. Some studies only presented changes of HbA_{1c}, and in this case, HbA_{1c} values at endpoint were calculated correspondingly based on baseline and changes of values [8,16,25-27]. For studies that provided SE rather than SD, SD value was calculated based on the SE value [12,28-30]. Some studies gave 95% CI of mean HbA_{1c}, and SD was derived according to 95% CI [14,25,31]. If neither endpoint SD nor other information that could be used to calculate SD was provided, baseline SD was adopted as endpoint SD [26,32].

Cochrane collaboration's tool for assessing the risk of bias was adopted to evaluate the quality of studies included. Six domains were assessed, including selection bias, performance bias, detection bias, attrition bias, reporting bias, and other bias.

Data Synthesis and Analysis

Interstudy heterogeneity was evaluated by the I^2 test. If $I^2 \leq 50\%$, heterogeneity across studies was acceptable, and fixed effect model could be used to achieve the pooled result. If $I^2 > 50\%$, random effect model would be used, and source of heterogeneity would be explored. Meta-analysis with continuous outcome variables was performed, and weighted mean difference (WMD) was adopted as the effect indicator. If $P < .05$ and 95% CI did not include zero, the point estimate of WMD was considered statistically significant. Publication bias was explored by Begg and Egger tests. Duval and Tweedie's nonparametric "trim-and-fill" procedure was also performed to further assess the possible effect of publication bias [33,34]. Sensitivity analysis was conducted to assess the stability of studies. Statistical analyses were performed with STATA/SE 14.1 (StataCorp LP, College Station, Texas) and risk of bias was evaluated with Review Manager (RevMan 5.3; Cochrane Collaboration).

Results

Characteristics of Studies

This meta-analysis included 35 studies with a pooled dataset of 6475 participants, of which 3338 were allocated to the IG and 3137 to CG [7-12,14,16,25-32,35-53]. The total number of patients in each study ranged from 30 to 1665. The characteristics of eligible studies are presented in Multimedia Appendix 2.

Of the 35 studies included, 11 were conducted in the United States [11,12,26,27,29,32,37,39,40,48,50]; 10 in the Republic of Korea [7,8,28,30,35,36,46,49,52,53]; 2 in Poland [38,45], Italy [16,44], and China [43,47] each; 1 in Japan [42], Finland [25], Turkey [10], Canada [9], Spain [31], Norway [14], England [41], and Congo[51] each. Included studies were published between 2004 and 2016, and the intervention duration ranged from 6 weeks to 5 years. The mean age of the participants ranged from 42.3 to 79.9 years and BMI from 22.8 to 36.9. All studies, except 2, clearly stated that all participants had T2DM. However, in these 2 studies, the mean age of the patients was

above 60 years; therefore, it was assumed that the included participants were all patients with T2DM [29,50].

Characteristics of Interventions

A total of 8 studies used Web-based interventions solely [7,9,10,12,28,38,43], 8 used mobile technology [14,25,31,36,42,45,48,52], and 9 used a combination of both [8,35,39,40,46,47,49,51,53]. In addition, 10 studies used other forms of internet-based interventions, including portable digital assistant, tablet computer, and data transmission equipment only [11,16,26,27,29,30,32,37,41,44]. Moreover, 28 studies clearly stated ways of providing feedbacks, among which 5 used automated algorithm [25,36,39,46,52] and 22 provided feedbacks manually via health care providers [7,9-11,14,27-32, 35,37,40,41,43,44,48-51,53]. In the automated algorithm, patient-specific messages were automatically generated and sent to patients based on the data entered into the system. Manual feedback was accomplished through telephone, videoconferencing, and short message services by cellular phone and the internet. A total of 14 studies provided internet-based education [8,10,27,29,32,37,39-41,43,47,49,50,53], and only 3 studies did not incorporate transmission function of glucose measurements [8,12,47]. A detailed description of intervention characteristics is presented in Multimedia Appendix 3.

Publication bias was detected by Egger test but not by Begg test. Trim-and-fill method was further conducted to account for publication bias. In trim-and-fill method, although the strength of intervention was slightly attenuated, the pooled analysis incorporating the hypothetical studies continued to show a statistically significant positive effect of internet-based interventions (WMD –0.332, 95% CI –0.456 to –0.209; $P < .001$; Multimedia Appendix 4), which suggested that the mean difference of HbA_{1c} between IG and CG was not an artifact of unpublished negative studies, but the effect of internet-based interventions. However, the possibility was not altogether excluded by this method. Detailed results on risk of bias, publication bias, and sensitivity analysis can be found in Multimedia Appendices 4-6. Multimedia Appendix 7 describes the HbA_{1c} value and related data of IG and CG at baseline and endpoint.

Overall Results of Internet-Based Intervention

All 35 studies provided mean values and SDs of HbA_{1c} as the outcome measure of internet-based interventions. At baseline, there was no significant difference of HbA_{1c} levels between the CG and IG as 95% CI including zero (Figure 2).

Measurements made immediately after interventions were used to achieve the overall pooled effect of the internet-based intervention. In one study, 2 IGs met eligibility criteria, and therefore, information was extracted twice and treated as from 2 distinct studies [39]. They were marked as Quinn, C.C., 2011a and Quinn, C.C., 2011b in the analysis. Figure 3 presented the combined results of the 35 studies in this meta-analysis. The result showed that the WMD between CG and IG was –0.426 (95% CI –0.540 to –0.312; $P < .001$), favoring the IG. Heterogeneity was 60.1% across studies, and random effect model was adopted.
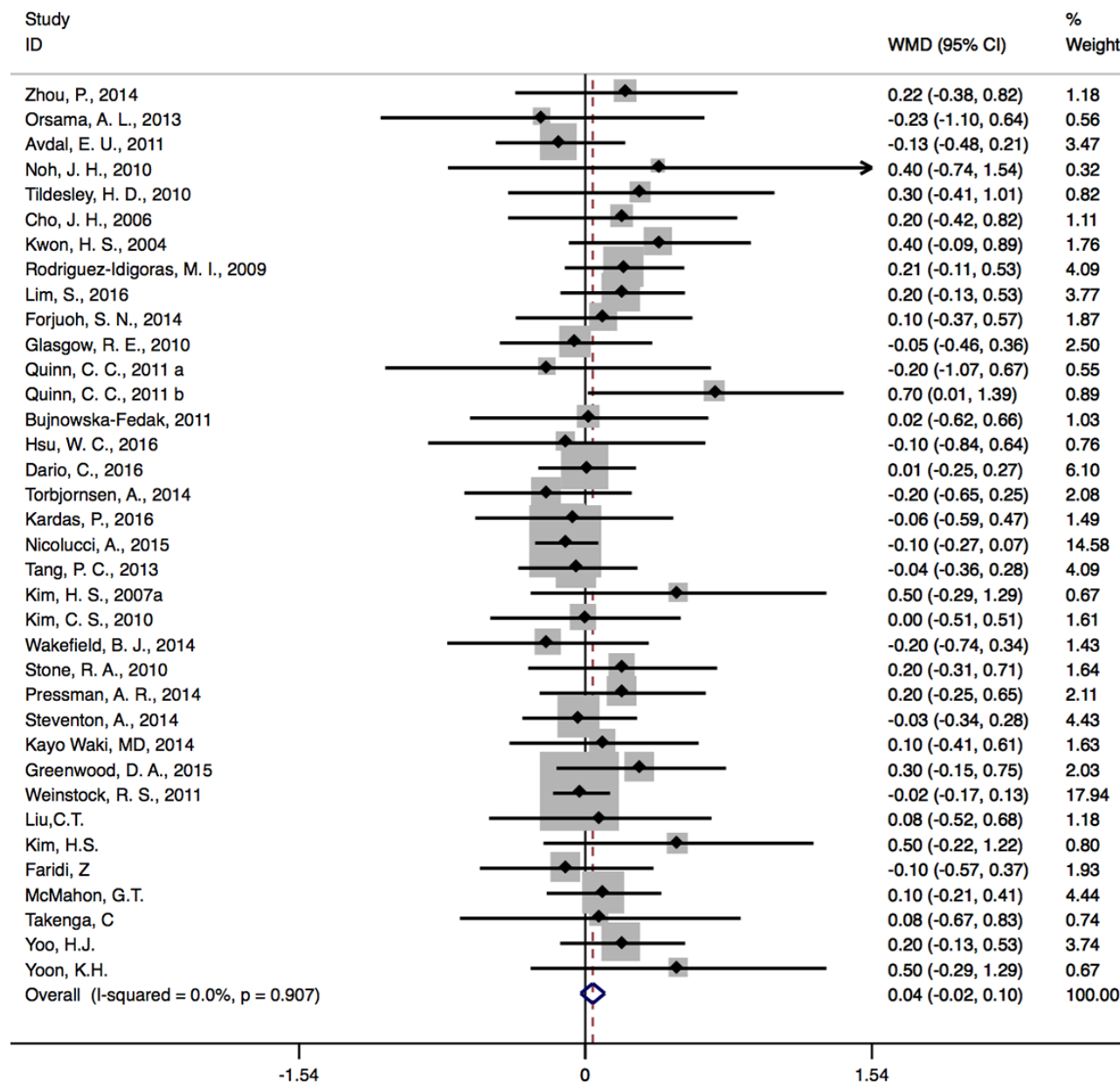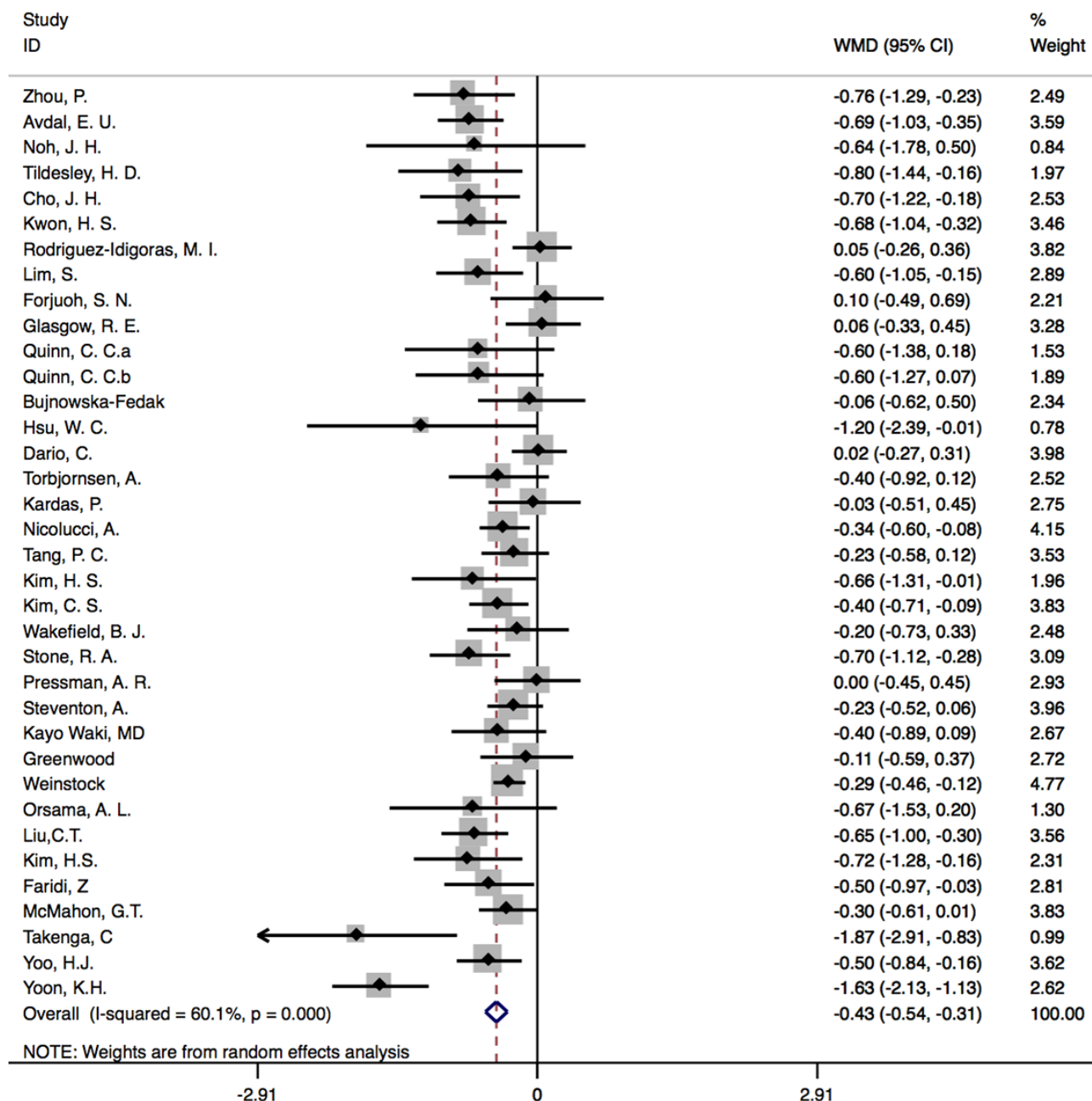
Figure 2. Forest plots of baseline results. WMD: weighted mean difference.

**Figure 3.** Forest plots of endpoint results. WMD: weighted mean difference.

Study ID	WMD (95% CI)	% Weight
Zhou, P.	-0.76 (-1.29, -0.23)	2.49
Avdal, E. U.	-0.69 (-1.03, -0.35)	3.59
Noh, J. H.	-0.64 (-1.78, 0.50)	0.84
Tildesley, H. D.	-0.80 (-1.44, -0.16)	1.97
Cho, J. H.	-0.70 (-1.22, -0.18)	2.53
Kwon, H. S.	-0.68 (-1.04, -0.32)	3.46
Rodriguez-Idigoras, M. I.	0.05 (-0.26, 0.36)	3.82
Lim, S.	-0.60 (-1.05, -0.15)	2.89
Forjuoh, S. N.	0.10 (-0.49, 0.69)	2.21
Glasgow, R. E.	0.06 (-0.33, 0.45)	3.28
Quinn, C. C.a	-0.60 (-1.38, 0.18)	1.53
Quinn, C. C.b	-0.60 (-1.27, 0.07)	1.89
Bujnowska-Fedak	-0.06 (-0.62, 0.50)	2.34
Hsu, W. C.	-1.20 (-2.39, -0.01)	0.78
Dario, C.	0.02 (-0.27, 0.31)	3.98
Torbjornsen, A.	-0.40 (-0.92, 0.12)	2.52
Kardas, P.	-0.03 (-0.51, 0.45)	2.75
Nicolucci, A.	-0.34 (-0.60, -0.08)	4.15
Tang, P. C.	-0.23 (-0.58, 0.12)	3.53
Kim, H. S.	-0.66 (-1.31, -0.01)	1.96
Kim, C. S.	-0.40 (-0.71, -0.09)	3.83
Wakefield, B. J.	-0.20 (-0.73, 0.33)	2.48
Stone, R. A.	-0.70 (-1.12, -0.28)	3.09
Pressman, A. R.	0.00 (-0.45, 0.45)	2.93
Steventon, A.	-0.23 (-0.52, 0.06)	3.96
Kayo Waki, MD	-0.40 (-0.89, 0.09)	2.67
Greenwood	-0.11 (-0.59, 0.37)	2.72
Weinstock	-0.29 (-0.46, -0.12)	4.77
Orsama, A. L.	-0.67 (-1.53, 0.20)	1.30
Liu,C.T.	-0.65 (-1.00, -0.30)	3.56
Kim, H.S.	-0.72 (-1.28, -0.16)	2.31
Faridi, Z	-0.50 (-0.97, -0.03)	2.81
McMahon, G.T.	-0.30 (-0.61, 0.01)	3.83
Takenga, C	-1.87 (-2.91, -0.83)	0.99
Yoo, H.J.	-0.50 (-0.84, -0.16)	3.62
Yoon, K.H.	-1.63 (-2.13, -1.13)	2.62
Overall (I-squared = 60.1%, p = 0.000)	-0.43 (-0.54, -0.31)	100.00

NOTE: Weights are from random effects analysis

Subgroup Analysis

Subgroup analyses were conducted based on intervention duration, types of internet-based interventions, ways of feedback provision, and whether or not internet-based educational content was provided.

Results of Subgroup Analysis Based on Intervention Duration

As intervention duration in the 35 studies ranged from 12 weeks to 5 years, we divided included research into 3 mutually exclusive groups (x≤3 months: n=10, 3 months<x≤6 months: n=10, and x>6 months: n=16) to perform subgroup analysis. In the group of x≤3 months [11,28,30,36,42,43,45,48,51,52], heterogeneity across studies was 41.8% and the pooled WMD was –0.51% (95% CI –0.71 to –0.31; *P*<.001). Significant difference between IG and CG was also observed in the group of duration >3 months and ≤6 months [8-10,27,32,35,37,38,46,49], with a pooled WMD of –0.48% (95% CI –0.68 to –0.28; *P*<.001). The heterogeneity across study was 34.5%. In the group of duration >6 months [7,12,14,16,25,26,29,31,39-41,44,47,50,53], difference in HbA_{1c} outcomes between IG and CG groups decreased with a WMD of –0.35% (95% CI –0.53 to –0.18; *P*<.001) and heterogeneity increased to 70.5% (Figure 4).

Figure 4. Subgroup analysis based on intervention duration. WMD: weighted mean difference.
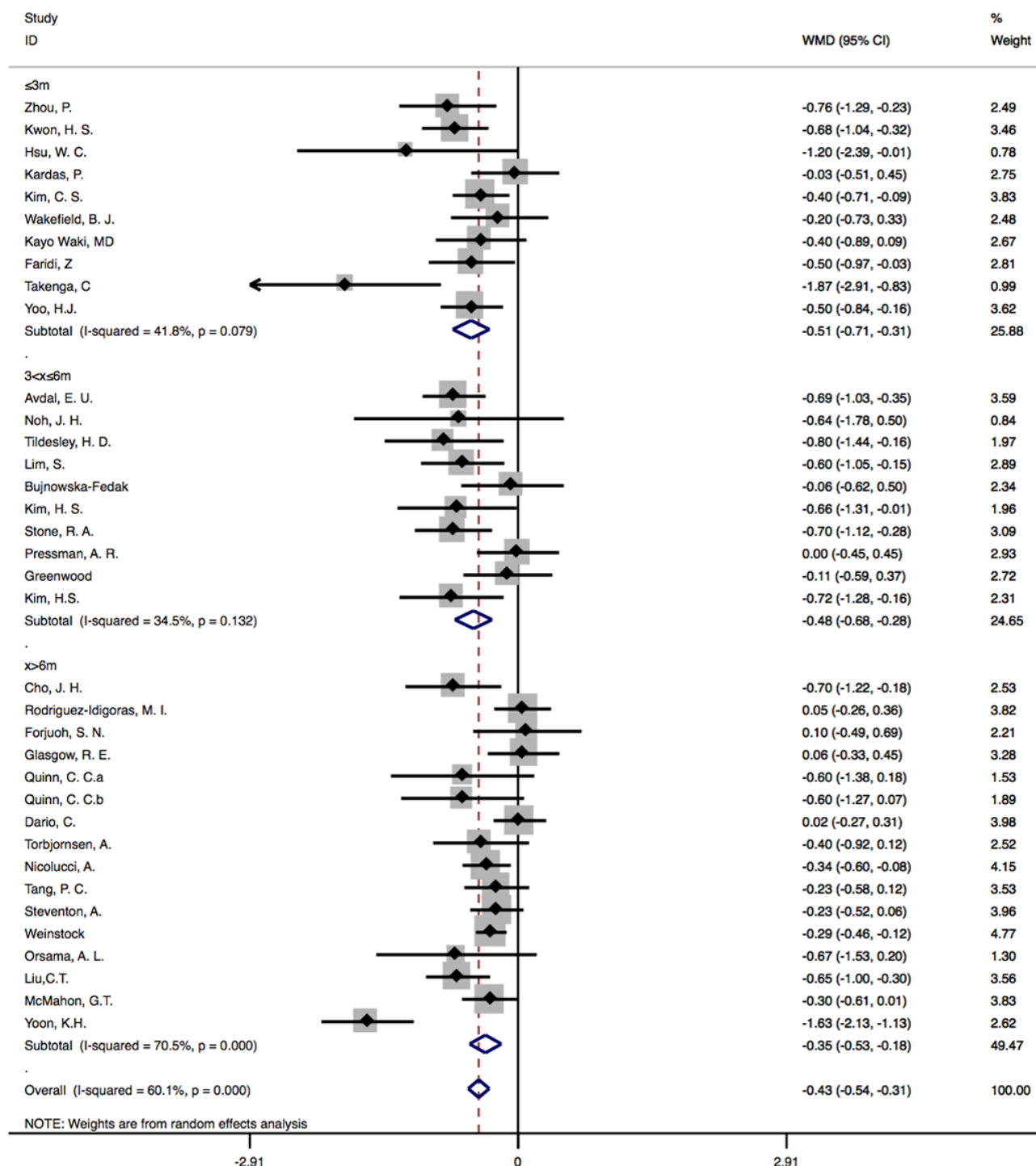
Results of Subgroup Analysis Based on the Type of Internet-Based interventions

In this analysis, we divided 35 studies into 4 subgroups based on types of internet-based interventions (Web-based solely: n=8, mobile-based solely: n=8, combined website and mobile technology: n=10, and others: n=10). Forms of intervention that did not fall in the above 3 categories were classified as others. In the group of website-only interventions [7,9,10,12,28,38,43,50], there was statistically significant difference in HbA$_{1c}$ outcomes between IG and CG, and WMD was –0.48% (95% CI –0.71 to –0.24; P<.001). However, a moderate heterogeneity of 57% was also observed across studies in this subgroup. In the group of mobile-only interventions [14,25,31,36,42,45,48,52], difference in HbA$_{1c}$ between IG and CG decreased, with a pooled WMD of –0.31% (95% CI –0.49 to –0.14; P<.001) and heterogeneity of 27.2%. In combined mobile and website group [8,35,39,40,46,47,49,51,53], a heterogeneity of 64.6% was observed, and the pooled effect was –0.77% (95% CI –1.07 to –0.47; P<.001). In the group of other intervention types, heterogeneity across studies was 35.5%, with a pooled WMD of –0.23 (95% CI –0.38 to –0.09; P=.001; Figure 5).

XSL•FO
RenderX

Figure 5. Subgroup analysis based on the type of internet-based interventions. WMD: weighted mean difference.
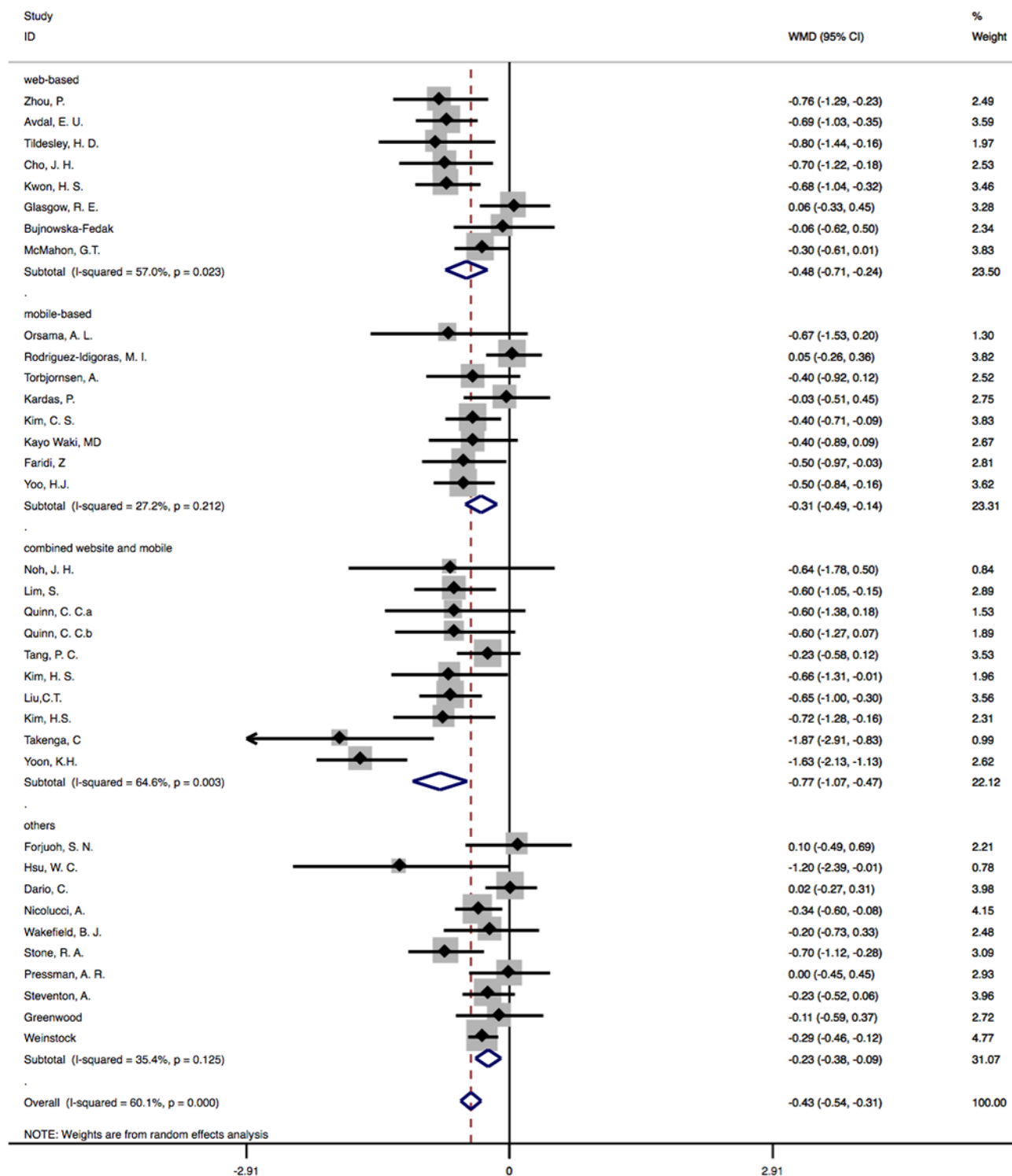
Results of Subgroup Analysis Based on Feedback Provision

We divided included studies into 3 groups (manual: n=22, automated: n=6, and unclear: n=8) to perform subgroup analysis based on feedback provision. A total of 8 studies did not specifically mention ways of providing feedback and thus were classified as unclear. In the group of manual feedback [7,9-11,14,27-32,35,37,40,41,43,44,48-51,53], statistically significant difference in HbA_{1c} between CG and IG was observed, and heterogeneity across studies was 67.2% (WMD –0.50%, 95% CI –0.65 to –0.34; $P<.001$). In studies that provided automated feedbacks [25,36,39,46,52], heterogeneity was reduced to 0.0%, and difference between groups was –0.50% (95% CI –0.69 to –0.32; $P<.001$). In unclear group [8,12,16,26,38,42,45,47], difference in HbA_{1c} outcomes was not significant as $P=.14$, which is greater than .05 (WMD –0.17%, 95% CI –0.39 to –0.06), and heterogeneity across studies was 46.9% (Figure 6).

Figure 6. Subgroup analysis based on feedback provision. WMD: weighted mean difference.
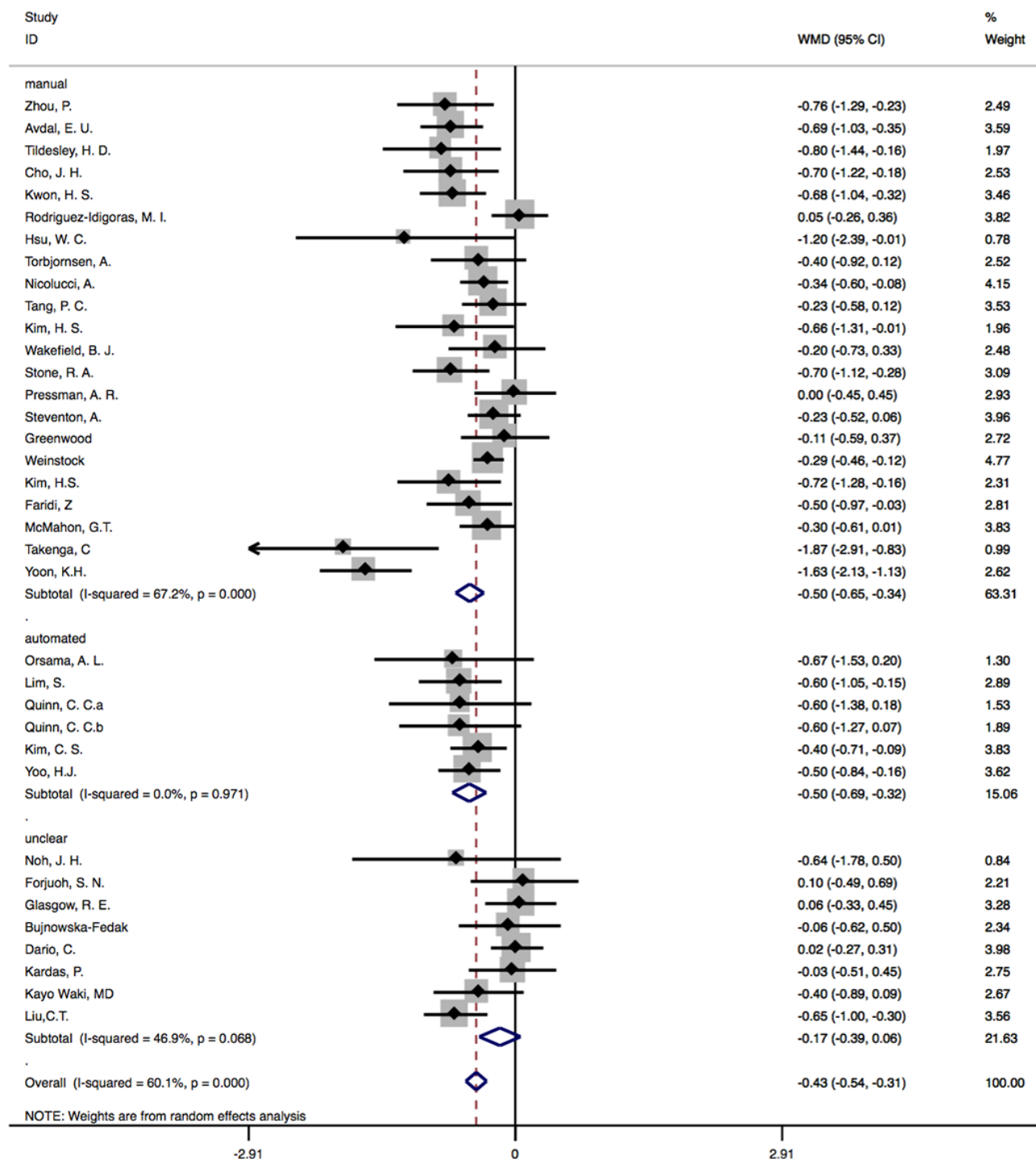
Results of Subgroup Analysis Based on Internet-Based Education Provision

On the basis of internet-based education provision, 35 studies were divided into 2 subgroups (yes: n=15; no: n=21). In the group that provided internet-based diabetes education [8,10,27,29,32,37,39-41,43,47,49,50,53], there was significant difference between CG and IG, with a pooled WMD of –0.51% (95% CI –0.69 to –0.32; *P*<.001), and heterogeneity across studies was 66.9%. In the group without internet-based education [7,9,11,12,14,16,25,26,28,30,31,35,36,38,42,44-46,48,51,52], heterogeneity decreased to 53.6%, and WMD was –0.36% (95% CI –0.51 to –0.22; *P*<.001; Figure 7).

Figure 7. Subgroup analysis based on internet-based education provision. WMD: weighted mean difference.
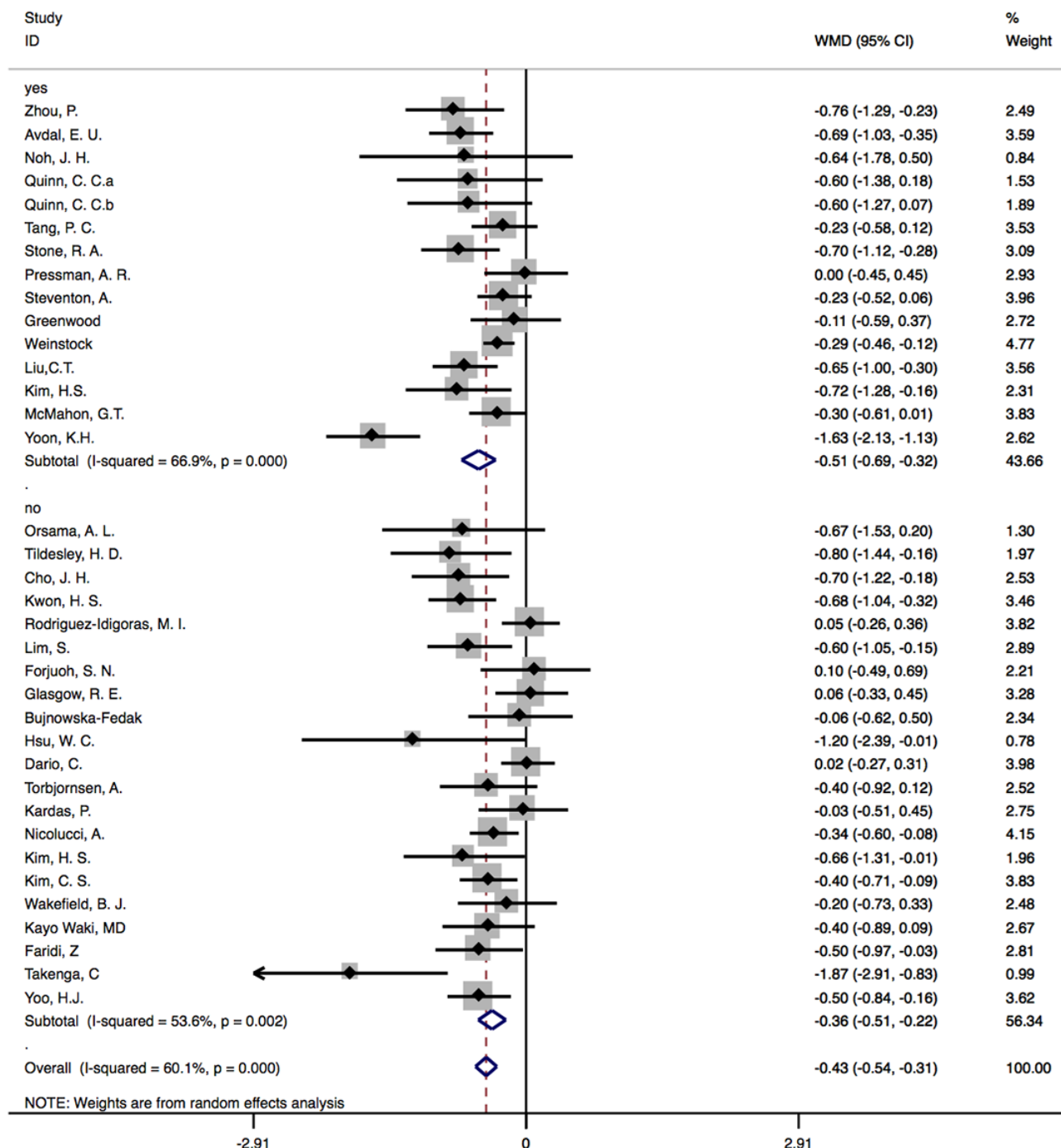
Discussion

Principal Findings

This meta-analysis included 35 RCT studies to evaluate the effect of internet-based interventions on patients with T2DM. WMD of HbA_{1c} value was adopted as the effect indicator of glucose control, and results indicated that the mean difference of HbA_{1c} values between the CG and internet-based interventions was 0.43%, favoring the internet-based interventions. It is reported in research that each 1% decrease of mean HbA_{1c} values will lead to 21% reduction of risk for any endpoint related to diabetes, including death, myocardial infarction, and microvascular complications [54]. Therefore, optimizing the utilization of internet-based interventions in the management of T2DM has great importance.

Intervention Duration

Duration has a significant influence on the effect of internet-based interventions. Results of this meta-analysis indicate a downward trend of intervention effect with an increasing duration, which is similar to that in other systematic reviews and meta-analyses [19,23]. There are some possible rationales that may serve to explain such differences over time. First, the decline of efficacy may be attributed to patients' diminishing enthusiasm or motivation with the passage of intervention, especially when most of the studies included in

this meta-analysis involve a regular or daily transmission of glucose data. Second, a lack of user-friendliness could also result in attenuated effect. As the mean age of recruited participants exceeds 40 years, new forms of technology may not be familiar and friendly for them. However, this result does not encourage the shortening of intervention periods, but rather it indicates that relative measures should be taken to increase patients' adherence and maintain their motivation, with the aim to prevent attenuated intervention effect over time. Furthermore, it is promising to develop a form of internet-based intervention that is both effective and friendly to use.

Types of Internet-Based Interventions

In terms of the types of internet-based interventions, there is not a consensus in previous meta-analyses on which form is the most effective. In this paper, combined website and mobile interventions have the largest WMD of HbA$_{1c}$, followed by Web-based only interventions. Mobile-based only interventions have the lowest mean difference, which indicates that mobile technologies may not be the optimal options in diabetes management. This result coincides with some findings in Toma's meta-analysis [23]. Toma discovered that the effect of mobile-only interventions in HbA$_{1c}$ outcomes was the poorest and not statistically important (WMD –0.20%, 95% CI –0.43 to 0.03; P=.09), whereas combined mobile and websites interventions yielded the best performance (WMD –0.54%, 95% CI –0.72 to –0.37; P<.001). Results of websites-only interventions were similar to those in this meta-analysis (WMD –0.51%, 95% CI –0.68 to –0.34). However, in another 2 meta-analyses exploring the effect of mobile-based interventions on diabetes management, the WMD in HbA$_{1c}$ outcomes was –0.40% (95% CI –0.69 to –0.11; P=.007) and –0.50% (95% CI –0.7 to –0.3) [22,55]. A possible explanation may be that in most studies included (7 out of 8) in this meta-analysis, a new mobile phone with functions related to diabetes management is provided to participants to replace their own cellular phones, rather than installing a diabetic application in their old ones [14,25,31,42,45,48,52]. Therefore, it is likely that such diabetic phones are not user-friendly and require some time to get familiar with. As a result, patients' satisfaction toward study and frequency of data transmission may be compromised, which further leads to the compromise of HbA$_{1c}$ performance ultimately. Furthermore, the fact that Pal's meta-analysis on mobile interventions only included 3 studies and Cui's meta-analysis also involved some studies that combined the use of mobile phone and websites may also contribute to the contradictory situation [22,55]. However, this result does not imply that we should altogether abandon mobile-based interventions, as it is drawn from the currently available evidence and therefore may lag in time. Due to the high penetration and convenience offered by mobile phones, they have great potentials in diabetes management, and future studies need to explore the effectiveness of apps that are easy to use and directly installed on patients' own cellular phones.

Forms of Feedback Provision and Function of Internet-Based Educational Contents

Automated feedbacks generated by algorithms yield similar effect to feedbacks provided manually. As automated algorithms could save manual labor and also provide real-time feedback as soon as patients enter data, it may have great potentials to alleviate the social and economic burden brought by various chronic diseases.

Furthermore, although various clinical guidelines have acknowledged the importance and effectiveness of diabetes education, only 40% of studies included in this meta-analysis embody this feature [4,5]. This result coincides with a previous review where among mobile diabetes apps on the market, only 20% of them had an educational module [3].

Limitations of This Meta-Analysis

This meta-analysis also has some limitations. First of all, because of limited time, we only searched 3 databases. Although we tried to identify studies from the reference lists of other reviews and include those published in other languages, it is possible that some studies in other databases and gray literature are overlooked. Second, although the studies included in this meta-analysis are all RCTs and the quality of evidence is relatively higher, only a few reported blinding of the outcome assessment process. Therefore, it is possible that performance bias is introduced. Third, in some studies, CGs also use some forms of intervention rather than purely usual care, such as paper-based education, note-keeping of glucose levels, and telephone follow-ups. Furthermore, contents of usual care in different locations may also vary. Therefore, this could also bias the genuine effects of internet-based interventions. Fourth, in some subgroup analyses, results are drawn from a small number of studies and heterogeneity is high. Finally, although the trim-and-fill method did not change the general effect of internet-based interventions, suggesting that the effect was not an artifact of unpublished negative studies, the possibility was not fully ruled out.

Implications for Future Studies and Clinical Practices

First of all, although blinding of participants and health care providers seems impossible because of the nature of the intervention, only a few studies explicitly reported that the outcome assessment process was blinded. In future studies, more importance should be attached to the blinding of outcome assessment because this is a process that should be achieved regardless of the type interventions. Only in this way, could performance bias be reduced to a minimum and a reliable effect be revealed.

Second, subgroup analysis indicates that automated feedback provision has a similar effect to manual feedback provision; therefore, it is promising to take full advantage of automated algorithms, as it could not only provide real-time feedbacks but also save manual labors. However, the number of studies in this aspect is limited and few have explored its cost-effectiveness and patients' satisfaction. As a result, future studies need not only examine the efficacy of automated feedback provision but also evaluate the feasibility of promotion on a large scale.

Third, future studies need to integrate diabetes-related functions to existing technologies, such as developing diabetic apps, which could be directly installed into patients' own mobile phones, rather than developing new types of phones. Only in this way, the true effect of mobile interventions can be revealed.

Finally, results in this meta-analysis also shed some light on clinical practices. If health care practitioners would like to use the internet as a way to help patients manage their glycemic levels, a combined use of websites and mobile technologies may achieve a better outcome than a standalone method. Furthermore, if there are currently available mobile apps on the market, practitioners may as well recommend patients to download such apps into their own mobile phones, rather than developing a new one for them. Moreover, as internet-based health education for diabetes management is proven to be effective, if due to various reasons health care providers could not provide a systematic and comprehensive intervention for patients, they could recommend some reliable and informative websites or information resources with diabetes-related contents that could be easily understood by patients.

Conclusions

In conclusion, the WMD of endpoint HbA_{1c} between internet-based interventions and usual care was -0.426%. Optimal outcomes appear in intervention duration of ≤3 months. In terms of the types of internet-based interventions, according to currently available evidence, combined interventions of websites and mobile technologies yield better results, and more studies are needed to explore the potentials of mobile technologies. Results also indicate that automated feedback has similar effects to feedbacks provided manually, and interventions with internet-based educational contents have better performance. As a result, interventions integrating the above elements may achieve more satisfactory results and help patients manage their glycemic levels effectively.

Authors' Contributions

XYS and YS provided the original idea for the meta-analysis. YS and FBW were responsible for the study design, study selection, and data extraction, with input from XYS, QDS, and XZ. YS, XZ, and XRZ performed statistical analysis. QDS contributed greatly to manuscript revision. XRZ, XYS, QDS, and EBF accomplished result interpretation. All authors wrote the manuscript and revised the paper for important intellectual content. All authors approved this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Details of search syntax.

[PDF File (Adobe PDF File), 122KB - jmir_v20i5e172_app1.pdf]

Multimedia Appendix 2

Characteristics of eligible studies included in meta-analysis.

[PDF File (Adobe PDF File), 182KB - jmir_v20i5e172_app2.pdf]

Multimedia Appendix 3

Characteristics of intervention group (IG) and control group (CG).

[PDF File (Adobe PDF File), 67KB - jmir_v20i5e172_app3.pdf]

Multimedia Appendix 4

Publication bias.

[PDF File (Adobe PDF File), 363KB - jmir_v20i5e172_app4.pdf]

Multimedia Appendix 5

Risk of bias.

[PDF File (Adobe PDF File), 392KB - jmir_v20i5e172_app5.pdf]

XSL•FO
RenderX

Multimedia Appendix 6

Sensitivity analysis.

[[PDF File (Adobe PDF File), 88KB - jmir_v20i5e172_app6.pdf](#)]

Multimedia Appendix 7

HbA$_{1c}$ value and related data of intervention group (IG) and control group (CG) at baseline and different intervention periods.

[[PDF File (Adobe PDF File), 55KB - jmir_v20i5e172_app7.pdf](#)]

References

1. Singh G, O'Donoghue J, Soon CK. Telemedicine: issues and implications. Technol Health Care 2002;10(1):1-10. [Medline: 11847443]
2. Smith JC, Schatz BR. Feasibility of mobile phone-based management of chronic illness. AMIA Annu Symp Proc 2010 Nov 13;2010:757-761 [FREE Full text] [Medline: 21347080]
3. Chomutare T, Fernandez-Luque L, Arsand E, Hartvigsen G. Features of mobile diabetes applications: review of the literature and analysis of current applications compared against evidence-based guidelines. J Med Internet Res 2011;13(3):e65 [FREE Full text] [doi: 10.2196/jmir.1874] [Medline: 21979293]
4. Sibal L, Home PD. Management of type 2 diabetes: NICE guidelines. Clin Med (Lond) 2009 Aug;9(4):353-357. [Medline: 19728510]
5. Funnell MM. Standards of care for diabetes: what's new? Nursing 2010 Oct;40(10):54-56. [doi: 10.1097/01.NURSE.0000388314.83393.bd] [Medline: 20844399]
6. Ramadas A, Quek KF, Chan CK, Oldenburg B. Web-based interventions for the management of type 2 diabetes mellitus: a systematic review of recent evidence. Int J Med Inform 2011 Jun;80(6):389-405. [doi: 10.1016/j.ijmedinf.2011.02.002] [Medline: 21481632]
7. Cho JH, Chang SA, Kwon HS, Choi YH, Ko SH, Moon SD, et al. Long-term effect of the Internet-based glucose monitoring system on HbA1c reduction and glucose stability: a 30-month follow-up study for diabetes management with a ubiquitous medical care system. Diabetes Care 2006 Dec;29(12):2625-2631. [doi: 10.2337/dc05-2371] [Medline: 17130195]
8. Noh JH, Cho YJ, Nam HW, Kim JH, Kim DJ, Yoo HS, et al. Web-based comprehensive information system for self-management of diabetes mellitus. Diabetes Technol Ther 2010 May;12(5):333-337. [doi: 10.1089/dia.2009.0122] [Medline: 20388042]
9. Tildesley HD, Mazanderani AB, Ross SA. Effect of Internet therapeutic intervention on A1C levels in patients with type 2 diabetes treated with insulin. Diabetes Care 2010 Aug;33(8):1738-1740 [FREE Full text] [doi: 10.2337/dc09-2256] [Medline: 20668152]
10. Avdal EU, Kizilci S, Demirel N. The effects of web-based diabetes education on diabetes care results: a randomized control study. Comput Inform Nurs 2011 Feb;29(2):101-106. [doi: 10.1097/NCN.0b013e3181fcbdc6] [Medline: 21099675]
11. Hsu WC, Lau KH, Huang R, Ghiloni S, Le H, Gilroy S, et al. Utilization of a cloud-based diabetes management program for insulin initiation and titration enables collaborative decision making between healthcare providers and patients. Diabetes Technol Ther 2016 Feb;18(2):59-67 [FREE Full text] [doi: 10.1089/dia.2015.0160] [Medline: 26645932]
12. Glasgow RE, Kurz D, King D, Dickman JM, Faber AJ, Halterman E, et al. Outcomes of minimal and moderate support versions of an internet-based diabetes self-management support program. J Gen Intern Med 2010 Dec;25(12):1315-1322 [FREE Full text] [doi: 10.1007/s11606-010-1480-0] [Medline: 20714820]
13. Holmen H, Torbjørnsen A, Wahl AK, Jenum AK, Småstuen MC, Arsand E, et al. A mobile health intervention for self-management and lifestyle change for persons with type 2 diabetes, part 2: one-year results from the Norwegian randomized controlled trial RENEWING HEALTH. JMIR Mhealth Uhealth 2014;2(4):e57 [FREE Full text] [doi: 10.2196/mhealth.3882] [Medline: 25499872]
14. Torbjørnsen A, Jenum AK, Småstuen MC, Arsand E, Holmen H, Wahl AK, et al. A low-intensity mobile health intervention with and without health counseling for persons with type 2 diabetes, part 1: baseline and short-term results from a randomized controlled trial in the Norwegian part of RENEWING HEALTH. JMIR Mhealth Uhealth 2014;2(4):e52 [FREE Full text] [doi: 10.2196/mhealth.3535] [Medline: 25499592]
15. Karhula T, Vuorinen A, Rääpysjärvi K, Pakanen M, Itkonen P, Tepponen M, et al. Telemonitoring and mobile phone-based health coaching among Finnish diabetic and heart disease patients: randomized controlled trial. J Med Internet Res 2015;17(6):e153 [FREE Full text] [doi: 10.2196/jmir.4059] [Medline: 26084979]
16. Dario C, Toffanin R, Calcaterra F, Saccavini C, Stafylas P, Mancin S, et al. Telemonitoring of type 2 diabetes mellitus in Italy. Telemed J E Health 2017;23:143-152. [doi: 10.1089/tmj.2015.0224] [Medline: 27379995]
17. Angeles RN, Howard MI, Dolovich L. The effectiveness of web-based tools for improving blood glucose control in patients with diabetes mellitus: a meta-analysis. Can J Diabetes 2011 Jan;35(4):344-352. [doi: 10.1016/S1499-2671(11)54011-0]
18. Azar M, Gabbay R. Web-based management of diabetes through glucose uploads: has the time come for telemedicine? Diabetes Res Clin Pract 2009 Jan;83(1):9-17. [doi: 10.1016/j.diabres.2008.09.055] [Medline: 19056140]

XSL•FO

RenderX

19. Hou C, Carter B, Hewitt J, Francisa T, Mayor S. Do mobile phone applications improve glycemic control (HbA1c) in the self-management of diabetes? A systematic review, meta-analysis, and grade of 14 randomized trials. Diabetes Care 2016 Dec;39(11):2089-2095. [doi: 10.2337/dc16-0346] [Medline: 27926892]

20. Mushcab H, Kernohan WG, Wallace J, Martin S. Web-based remote monitoring systems for self-managing type 2 diabetes: a systematic review. Diabetes Technol Ther 2015 Jul;17(7):498-509. [doi: 10.1089/dia.2014.0296] [Medline: 25830528]

21. Pereira K, Phillips B, Johnson C, Vorderstrasse A. Internet delivered diabetes self-management education: a review. Diabetes Technol Ther 2015 Jan;17(1):55-63. [doi: 10.1089/dia.2014.0155] [Medline: 25238257]

22. Cui M, Wu X, Mao J, Wang X, Nie M. T2DM self-management via smartphone applications: a systematic review and meta-analysis. PLoS One 2016;11(11):e0166718 [FREE Full text] [doi: 10.1371/journal.pone.0166718] [Medline: 27861583]

23. Toma T, Athanasiou T, Harling L, Darzi A, Ashrafian H. Online social networking services in the management of patients with diabetes mellitus: systematic review and meta-analysis of randomised controlled trials. Diabetes Res Clin Pract 2014 Nov;106(2):200-211. [doi: 10.1016/j.diabres.2014.06.008] [Medline: 25043399]

24. Zhang X, Yang S, Sun K, Fisher EB, Sun X. How to achieve better effect of peer support among adults with type 2 diabetes: a meta-analysis of randomized clinical trials. Patient Educ Couns 2016 Feb;99(2):186-197. [doi: 10.1016/j.pec.2015.09.006] [Medline: 26463156]

25. Orsama A, Lähteenmäki J, Harno K, Kulju M, Wintergerst E, Schachner H, et al. Active assistance technology reduces glycosylated hemoglobin and weight in individuals with type 2 diabetes: results of a theory-based randomized trial. Diabetes Technol Ther 2013 Aug;15(8):662-669. [doi: 10.1089/dia.2013.0056] [Medline: 23844570]

26. Forjuoh SN, Bolin JN, Huber Jr JC, Vuong AM, Adepoju OE, Helduser JW, et al. Behavioral and technological interventions targeting glycemic control in a racially/ethnically diverse population: a randomized controlled trial. BMC Public Health 2014;14:71 [FREE Full text] [doi: 10.1186/1471-2458-14-71] [Medline: 24450992]

27. Pressman AR, Kinoshita L, Kirk S, Barbosa GM, Chou C, Minkoff J. A novel telemonitoring device for improving diabetes control: protocol and results from a randomized clinical trial. Telemed J E Health 2014 Feb;20(2):109-114. [doi: 10.1089/tmj.2013.0157] [Medline: 24404816]

28. Kwon HS, Cho JH, Kim HS, Song BR, Ko SH, Lee JM, et al. Establishment of blood glucose monitoring system using the internet. Diabetes Care 2004 Feb;27(2):478-483. [Medline: 14747232]

29. Weinstock RS, Teresi JA, Goland R, Izquierdo R, Palmas W, Eimicke JP, et al. Glycemic control and health disparities in older ethnically diverse underserved adults with diabetes: five-year results from the Informatics for Diabetes Education and Telemedicine (IDEATel) study. Diabetes Care 2011 Feb;34(2):274-279 [FREE Full text] [doi: 10.2337/dc10-1346] [Medline: 21270184]

30. Wakefield BJ, Koopman RJ, Keplinger LE, Bomar M, Bernt B, Johanning JL, et al. Effect of home telemonitoring on glycemic and blood pressure control in primary care clinic patients with diabetes. Telemed J E Health 2014 Mar;20(3):199-205 [FREE Full text] [doi: 10.1089/tmj.2013.0151] [Medline: 24404819]

31. Rodríguez-Idígoras MI, Sepúlveda-Muñoz J, Sánchez-Garrido-Escudero R, Martínez-González JL, Escolar-Castelló JL, Paniagua-Gómez IM, et al. Telemedicine influence on the follow-up of type 2 diabetes patients. Diabetes Technol Ther 2009 Jul;11(7):431-437. [doi: 10.1089/dia.2008.0114] [Medline: 19580356]

32. Greenwood DA, Blozis SA, Young HM, Nesbitt TS, Quinn CC. Overcoming clinical inertia: a randomized clinical trial of a telehealth remote monitoring intervention using paired glucose testing in adults with type 2 diabetes. J Med Internet Res 2015;17(7):e178 [FREE Full text] [doi: 10.2196/jmir.4112] [Medline: 26199142]

33. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000 Jun;56(2):455-463. [Medline: 10877304]

34. Willi C, Bodenmann P, Ghali WA, Faris PD, Cornuz J. Active smoking and the risk of type 2 diabetes: a systematic review and meta-analysis. J Am Med Assoc 2007 Dec 12;298(22):2654-2664. [doi: 10.1001/jama.298.22.2654] [Medline: 18073361]

35. Kim HS, Jeong HS. A nurse short message service by cellular phone in type-2 diabetic patients for six months. J Clin Nurs 2007 Jun;16(6):1082-1087. [doi: 10.1111/j.1365-2702.2007.01698.x] [Medline: 17518883]

36. Kim CS, Park SY, Kang JG, Lee SJ, Ihm SH, Choi MG, et al. Insulin dose titration system in diabetes patients using a short messaging service automatically produced by a knowledge matrix. Diabetes Technol Ther 2010 Aug;12(8):663-669. [doi: 10.1089/dia.2010.0031] [Medline: 20615108]

37. Stone RA, Rao RH, Sevick MA, Cheng C, Hough LJ, Macpherson DS, et al. Active care management supported by home telemonitoring in veterans with type 2 diabetes: the DiaTel randomized controlled trial. Diabetes Care 2010 Mar;33(3):478-484 [FREE Full text] [doi: 10.2337/dc09-1012] [Medline: 20009091]

38. Bujnowska-Fedak MM, Puchała E, Steciwko A. The impact of telehome care on health status and quality of life among patients with diabetes in a primary care setting in Poland. Telemed J E Health 2011 Apr;17(3):153-163. [doi: 10.1089/tmj.2010.0113] [Medline: 21375410]

39. Quinn CC, Shardell MD, Terrin ML, Barr EA, Ballew SH, Gruber-Baldini AL. Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control. Diabetes Care 2011 Sep;34(9):1934-1942 [FREE Full text] [doi: 10.2337/dc11-0366] [Medline: 21788632]

40. Tang PC, Overhage JM, Chan AS, Brown NL, Aghighi B, Entwistle MP, et al. Online disease management of diabetes: engaging and motivating patients online with enhanced resources-diabetes (EMPOWER-D), a randomized controlled trial.

XSL•FO

RenderX

J Am Med Inform Assoc 2013 May 1;20(3):526-534 [FREE Full text] [doi: 10.1136/amiajnl-2012-001263] [Medline: 23171659]

41. Steventon A, Bardsley M, Doll H, Tuckey E, Newman SP. Effect of telehealth on glycaemic control: analysis of patients with type 2 diabetes in the Whole Systems Demonstrator cluster randomised trial. BMC Health Serv Res 2014;14:334 [FREE Full text] [doi: 10.1186/1472-6963-14-334] [Medline: 25100190]

42. Waki K, Fujita H, Uchimura Y, Omae K, Aramaki E, Kato S, et al. DialBetics: a novel smartphone-based self-management support system for type 2 diabetes patients. J Diabetes Sci Technol 2014 Mar 13;8(2):209-215 [FREE Full text] [doi: 10.1177/1932296814526495] [Medline: 24876569]

43. Zhou P, Xu L, Liu X, Huang J, Xu W, Chen W. Web-based telemedicine for management of type 2 diabetes through glucose uploads: a randomized controlled trial. Int J Clin Exp Pathol 2014;7(12):8848-8854 [FREE Full text] [Medline: 25674254]

44. Nicolucci A, Cercone S, Chiriatti A, Muscas F, Gensini G. A randomized trial on home telemonitoring for the management of metabolic and cardiovascular risk in patients with type 2 diabetes. Diabetes Technol Ther 2015 Aug;17(8):563-570. [doi: 10.1089/dia.2014.0355] [Medline: 26154338]

45. Kardas P, Lewandowski K, Bromuri S. Type 2 diabetes patients benefit from the COMODITY12 mHealth system: results of a randomised trial. J Med Syst 2016 Dec;40(12):259. [doi: 10.1007/s10916-016-0619-x] [Medline: 27722974]

46. Lim S, Kang SM, Kim KM, Moon JH, Choi SH, Hwang H, et al. Multifactorial intervention in diabetes care using real-time monitoring and tailored feedback in type 2 diabetes. Acta Diabetol 2016;53:189-198. [doi: 10.1007/s00592-015-0754-8] [Medline: 25936739]

47. Liu CT, Yeh YY, Lee TI, Li YC. Observations on online services for diabetes management. Diabetes Care 2005 Nov;28(11):2807-2808. [Medline: 16249564]

48. Faridi Z, Liberti L, Shuval K, Northrup V, Ali A, Katz DL. Evaluating the impact of mobile telephone technology on type 2 diabetic patients' self-management: the NICHE pilot study. J Eval Clin Pract 2008 Jun;14(3):465-469. [doi: 10.1111/j.1365-2753.2007.00881.x] [Medline: 18373577]

49. Kim H, Song MS. Technological intervention for obese patients with type 2 diabetes. Appl Nurs Res 2008 May;21(2):84-89. [doi: 10.1016/j.apnr.2007.01.007] [Medline: 18457747]

50. McMahon GT, Gomes HE, Hickson Hohne S, Hu TM, Levine BA, Conlin PR. Web-based care management in patients with poorly controlled diabetes. Diabetes Care 2005 Jul;28(7):1624-1629 [FREE Full text] [Medline: 15983311]

51. Takenga C, Berndt R, Musongya O, Kitero J, Katoke R, Molo K, et al. An ICT-based diabetes management system tested for health care delivery in the African context. Int J Telemed Appl 2014;2014:437307 [FREE Full text] [doi: 10.1155/2014/437307] [Medline: 25136358]

52. Yoo HJ, Park MS, Kim TN, Yang SJ, Cho GJ, Hwang TG, et al. A ubiquitous chronic disease care system using cellular phones and the internet. Diabet Med 2009 Jun;26(6):628-635. [doi: 10.1111/j.1464-5491.2009.02732.x] [Medline: 19538239]

53. Yoon KH, Kim HS. A short message service by cellular phone in type 2 diabetic patients for 12 months. Diabetes Res Clin Pract 2008 Feb;79(2):256-261. [doi: 10.1016/j.diabres.2007.09.007] [Medline: 17988756]

54. Stratton IM, Adler AI, Neil HA, Matthews DR, Manley SE, Cull CA, et al. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. Br Med J 2000 Aug 12;321(7258):405-412 [FREE Full text] [Medline: 10938048]

55. Pal K, Eastwood SV, Michie S, Farmer A, Barnard ML, Peacock R, et al. Computer-based interventions to improve self-management in adults with type 2 diabetes: a systematic review and meta-analysis. Diabetes Care 2014 Jun;37(6):1759-1766. [doi: 10.2337/dc13-1386] [Medline: 24855158]

Abbreviations

BMI: body mass index
CG: control group
HbA$_{1c}$: glycated hemoglobin
IG: intervention group
RCT: randomized controlled trial
T2DM: type 2 diabetes mellitus
WMD: weighted mean difference

Original Paper

Electronic Health Physical Activity Behavior Change Intervention to Self-Manage Cardiovascular Disease: Qualitative Exploration of Patient and Health Professional Requirements

Deirdre MJ Walsh[1,2], BA, MSc, PhD; Kieran Moran[1,3], PhD; Véronique Cornelissen[4], PhD; Roselien Buys[5], PhD; Nils Cornelis[4], BSc (Hons); Catherine Woods[6], PhD

[1]Insight Centre for Data Analytics, Dublin City University, Dublin, Ireland

[2]School of Health and Human Performance, Dublin City University, Dublin, Ireland

[3]MedEx Wellness, School of Health and Human Performance, Dublin City Unviersity, Dublin, Ireland

[4]Department of Rehabilitation Sciences, KU Leuven, Leuven, Belgium

[5]Department of Cardiovascular Sciences, KU Leuven, Leuven, Belgium

[6]Department of Physical Education and Sport Sciences, Faculty of Education and Health Sciences, University of Limerick, Limerick, Ireland

Corresponding Author:
Kieran Moran, PhD
MedEx Wellness
School of Health and Human Performance
Dublin City Unviersity
Room A246
Albert College
Dublin, Dublin 9
Ireland
Phone: 353 1 7008011
Email: kieran.moran@dcu.ie

Abstract

Background: Cardiovascular diseases are a leading cause of premature death worldwide. International guidelines recommend routine delivery of all phases of cardiac rehabilitation. Uptake of traditional cardiac rehabilitation remains suboptimal, as attendance at formal hospital-based cardiac rehabilitation programs is low, with community-based cardiac rehabilitation rates and individual long-term exercise maintenance even lower. Home-based cardiac rehabilitation programs have been shown to be equally effective in clinical and health-related quality of life outcomes and yet are not readily available.

Objective: Given the potential that home-based cardiac rehabilitation programs have, it is important to explore how to appropriately design any such intervention in conjunction with key stakeholders. The aim of this study was to engage with individuals with cardiovascular disease and other professionals within the health ecosystem to (1) understand the personal, social, and physical factors that inhibit or promote their capacity to engage with physical activity and (2) explore their technology competencies, needs, and wants in relation to an eHealth intervention.

Methods: Fifty-four semistructured interviews were conducted across two countries. Interviews were audiotaped, transcribed verbatim, and analyzed using thematic analysis. Barriers to the implementation of PATHway were also explored specifically in relation to physical capability and safety as well as technology readiness and further mapped onto the COM-B model for future intervention design.

Results: Key recommendations included collection of patient data and use of measurements, harnessing hospital based social connections, and advice to utilize a patient-centered approach with personalization and tailoring to facilitate optimal engagement.

Conclusions: In summary, a multifaceted, personalizable intervention with an inclusively designed interface was deemed desirable for use among cardiovascular disease patients both by end users and key stakeholders. In-depth understanding of core needs of the population can aid intervention development and acceptability.

(J Med Internet Res 2018;20(5):e163)   doi:10.2196/jmir.9181

XSL•FO
RenderX

Introduction

Background

Cardiovascular diseases (CVDs) are a leading cause of early death and disability within Europe and an economic burden worldwide [1]. Importantly, from a behavioral science perspective, approximately 80% of cases are precipitated by lack of self-management of key modifiable risk factors, including physical activity (PA), smoking, diet, alcohol consumption, stress management, and medication adherence [2]. Cardiac rehabilitation (CR) is an essential part of the gold standard management of CVD [3,4] and typically involves risk factor education, supervised exercise training, and psychological support. However, even though CR improves mortality and morbidity rates, uptake of CR remains suboptimal. There are several reasons for the low adherence rates, including travel distance, low self-efficacy, perceived body image, and lack of time [5]. Interestingly, home-based CR programs have been shown to be equally effective in clinical and health-related quality of life outcomes and overcome many of the stakeholder identified barriers to CR participation [6]; however, few CR programs offer a remote solution [7]. This gap in health care provision highlights the need to focus on the needs of individuals living with chronic illness and contemporary CVD self-management.

New Approaches to Self-Manage Cardiovascular Disease

As current CR delivery approaches do not suit everyone, new innovative ways are needed to match patient preferences and needs to improve uptake and completion of CR. PATHway (Physical Activity Towards Health way) aims to be such an innovative approach. PATHway proposes an Internet-enabled, sensor-based home exercise platform that allows remote participation in CR exercise programs at any time, either by oneself or by a small number of patients, from the comfort of their own living room. The home-based PATHway system will provide an individualized exercise prescription and program leveraging existing technology to facilitate participants to better self-manage their CVD. The proposed technology at this current development phase includes a portable personal computer, including PATHway software complimented by wearable sensors (eg, Microsoft Band 2 heart rate monitor to tailor the PATHway experience.) These sensors will facilitate the participant to engage in tailored exercise classes and games led by an avatar instructor who demonstrates each exercise to be conducted throughout the exercise session.

Saner highlighted how there are many obstacles to implementation, including how development of electronic health (eHealth) interventions is skewed toward technology development rather than user needs and expectations [8]. Therefore, it is important that qualitative development work be recognized as an integral aspect of creating a context specific, fit-for-purpose, user-centered intervention.

It has been identified that incremental stepped approaches to developing and evaluating behavior change interventions are most appropriate as per the Medical Research Council and Behavior Change Wheel frameworks [9,10]. A key tenet of both frameworks is to identify patients' key personal, social, and physical factors that inhibit or promote their capacity to engage in PA and also to identify their experience, needs, and wants from a technology-based intervention.

Furthermore, to ensure successful implementation of an eHealth behavioral change intervention to self-manage CVD, it is necessary to ascertain the views of key stakeholders from the health care ecosystem and across Europe. It is important to include stakeholders from across the spectrum (as captured by the Social Ecological Model, SEM; [11]), eg, from cardiologists to health policy makers. The SEM incorporates a wide range of individuals involved at various points of the CVD illness journey through its various levels (eg, individual, interpersonal, organizational, community, and policy). This includes those who refer patients to existing CR services and those who deliver those services. Interestingly, this study also seeks to explore the context of the implementation of the PATHway intervention, taking into account the potentially differing health care systems involved in the two sites (ie, Ireland and Belgium). These sites were selected as part of a consortium from European Union's Horizon 2020 Framework Programme for Research and Innovation Action under Grant Agreement no. 643491.

In an Irish context, CR services have been in development since the 1970's. CR services in Ireland grew at a fast pace following the implementation of the Building Healthier Hearts program [12]. Both health care systems aim to operate within a multidisciplinary context supported by cardiologists, physiotherapists, nurses, occupational therapists, dietitians, pharmacists, psychologists, and social workers.

In Belgium, patients receive reimbursement for involvement in CR (ie, maximum of 30 in-hospital sessions, 45 outpatient sessions), provided the patient meets certain criteria [13]. In Ireland, free standard CR services are generally offered (without reimbursement) to individuals. Unfortunately, particularly in Ireland, some CR units have suffered cutbacks in recent years.

The purpose of this study was to explore opinions and preferences to optimize PATHway intervention development before piloting the intervention (Trial registration number NCT02717806). This includes exploring the most appropriate content and viable approaches for PATHway intervention delivery. This was achieved through a comprehensive needs identification and analysis with PATHway patients and stakeholders using the capability, opportunity, motivation, and behavior (COM-B) model framework in two health care systems: Ireland and Belgium.

Methods

Interview Script Development

In-depth, semistructured, individual interviews were conducted in two trial locations (Dublin, Ireland and Leuven, Belgium). Interview scripts for both patient and stakeholders (see Multimedia Appendices 1 and 2) were developed using the COM-B model of health behavior change [10]. By adopting the COM-B framework, different aspects of behavior were investigated to design and implement the PATHway intervention appropriately. Author DW conducted all interviews in Ireland, whereas authors NC, VC, and RB conducted interviews in Belgium. All Irish interviews were transcribed by a selected transcription service, whereas in Belgium, all interviews were transcribed by the PATHway research team.

Patient Recruitment

To target individual and interpersonal levels, 33 CVD patients were interviewed in a combined total across both sites (Dublin, Ireland n=20; Leuven, Belgium n=13). These two sites were selected as they were the locations of the PATHway study team clinical partners, and importantly, these sites were the locations of the phase 2 and phase 3 CR programs with access to patient and health professional cohorts. These specific locations facilitated an in-depth view of the varied CR practice across different European countries. No other sites participated in this discrete phase of development with patients.

Patients were recruited by the PATHway study team in conjunction with hospital and community partners. Participants from phase 2 (hospital-based) and phase 3 (community-based) CR programs with different levels of engagement in CR were approached. That is, four groups were identified: (1) those attending phase 2 (hospital-based CR), (2) those who enrolled but did not complete phase 2, (3) those who were engaged with phase 3 (community-based CR), and (4) those who dropped out of phase 3. For patients still engaged with CR programs in the hospital or in the community, announcements and information sheets were made available at the beginning and end of CR classes to explain the study, and contact details were given. However, for patients who were no longer attending CR programs, CR staff contacted them, and if interested in participating, potential participants were given contact details of the PATHway study team. Before participating in an interview, all patients were asked to complete the technology usage questionnaire [14] to obtain further information on the general and technological background of the patients. This questionnaire and related findings were published by the PATHway study team (see [14]). Patient recruitment and interviews conducted were determined by whether new information or concepts emerged from the interviews [15].

Stakeholder Recruitment

The SEM was used to ensure inclusion of the full health ecosystem [11]. All stakeholders were approached by the PATHway study team following a brainstorming session with CR coordinators about potential participants to approach at each level of the SEM and each profession. All potential participants were then approached and given information sheets regarding the study and invited to participate. As such, interviews were conducted with a total of 21 stakeholders from each of the public policy, the community, and the organizational levels in Dublin (Ireland) and Leuven (Belgium): representatives from public policy, specifically individuals from the Department of Health (n=2) and from the Health Services Executive (n=1); representatives from the community, specifically general practitioners who refer patients to CR (n=3); public health nurses (n=1); local patient organization (n=1); national patient organization (n=1); representatives from the hospital, specifically the CR cardiologists (n=2); hypertension specialist (n=1); specialized cardiology nurses (n=3); physiotherapists or exercise physiologists involved in CR phase 2 and 3 (n=4); psychologist involved in CR (n=1); and technologists with experience of health care devices in CR (n=1). This sample endeavored to reflect the various levels and multidisciplinary nature of CR services insofar as possible at different specified levels of the SEM. This study aimed to capture the breadth of experience of multiple stakeholders across the CVD journey; however, it is acknowledged that potentially sufficient depth of these experiences may not be fully captured within this sample. Further sampling of each of the allied health professionals would be beneficial for future research as it was not possible within the scope of the current project to recruit further participants.

Patient and Stakeholder Interview Analysis

All interviews were transcribed verbatim by the PATHway team. Data were subject to a thematic analysis, guided by Braun and Clarke's [16] five-step framework and mapped to the COM-B framework. The five-step framework is listed as follows:

1. Step one includes familiarizing yourself with data through multiple readings.
2. Step two generates an initial list of ideas about what is in the data and what is interesting about them and involves the production of initial codes from the data.
3. Step three, themes begin to emerge, and this refocuses the analysis at the broader level of themes.
4. Step four involves reviewing themes whereby a set of candidate themes are explored and refined, including similarities and differences between interviews. This is an important step given the multisite approach in PATHway, which may offer conflicting findings.
5. Step five involves defining and naming themes.

Audit trails were used throughout to ensure transparency from raw interview transcripts to themes to use case formulations in both trial sites. Each sited coded all site-specific data separately. Preliminary coding was shared at each step of analysis as listed above, and group discussions were held. Authors from the Belgian site translated all emerging codes, and numerous representative quotes were translated from Dutch to English for the full team to discuss. All data were then combined to facilitate data synthesis and integration of qualitative data from both sites. This synthesis process was done in English and findings corroborated by authors from the Belgian site to ensure that codes and representative quotes were integrated appropriately. This synthesis of data was especially important given the need to reflect both the similarities but also the differences between

the two health systems represented within this data. This analysis was done inductively and then mapped across to the COM-B framework to allow transparent translation to PATHway intervention requirements. Analysis was done separately for patients and stakeholders.

Results

Results of Cardiovascular Disease Patient Interviews

A total of 33 patients took part in individual interviews across the two sites. Thirteen patients (39%, 13/33) were from Leuven and 60% (20/33) from Ireland (mean age=60 years; female=21%, 7/33) and various levels of education (eg, second level education or lower=36%, 12/33; third level education [including undergraduate and postgraduate programs]= 64%, 21/33). There were 55% (18/33) from phase 2 and 45% (15/33) from phase 3 with various reasons for attending CR listed. There were varying levels of technology use with 58% (19/33) reporting high, 24% (8/33) moderate, and 18% (6/33) low technology use. An overview of the patient sample is shown in Table 1.

Textbox 1 represents the main themes and subthemes that emerged from the individual patient interviews.

Theme 1: Capability

Physical Capability

Capability was first explored with CVD patients to establish whether individuals felt they were "physically capable" of engaging with PA. Patients were unsure of their physical fitness levels to engage fully with a home-based program. This lack of confidence in their own physical ability highlights the need to introduce exercises at a suitable level and carefully monitor participant satisfaction following each session.

Some patients raised the issue of how they were concerned whether the exercises would be "age-appropriate," including any potential negative impact on comorbidities that some patients felt was an issue in their capability of initiating and maintaining PATHway use:

> *That circuit in the gym, I do have difficulties when I run...My hip starts to ache.* [Participant 11, male, 65 years, low technology use, Leuven]

Psychological Capability

Psychological Readiness

Apprehensions were raised about potentially not having the "psychological capability" or knowledge of CVD to know what physiological reactions are acceptable to experience during remote exercise participation. Patients suggested that information should be provided on what to expect while exercising, especially in relation to any symptoms that may suggest an adverse event. This was suggested to help understand their "new normal." One patient stated:

> *I didn't really know anything about physical activity. Now thanks to this program, I learned that it takes a lot more than you would think at first.* [Participant 10, male, 42 years, high technology use, Leuven]

Patients were interested to understand the different readings from devices and knowing safe and optimal ranges for exercise:

> *I'll be anxious maybe initially to make sure that I'd be doing it within the confines that I'm supposed to.* [Participant 1, male, 62 years, moderate technology use, Dublin]

Further educational training (ie, information on optimal target heart rate zones, recommended daily step counts, and recommended weekly minutes of PA) was suggested as something to enhance capability so that patients could understand different physical measurements and what they mean for their CVD risk profile. Interestingly, although many patients had attended and engaged with both phase 2 and phase 3, many felt a lack of knowledge and capability surrounding the type of exercises to be conducted at home. Exercise examples and information regarding specific exercises, muscle groups, training types, and their respective benefits were listed as areas of interest that PATHway could provide. One patient stated:

> *I'm not sure, we're piecing it together what we do myself and my wife. I wouldn't be totally confident now that it's a perfect training programme.* [Participant 26, male, 60 years, high technology use, Dublin]

Technology Readiness

Although capability for using technology to self-manage CVD among the target population was quite high, users were less confident in their ability to "set up" the PATHway system for the first time. This suggests that there may be an important role of mentoring within PATHway. One patient stated the following:

> *I wouldn't have a clue but my son or daughter would. I wouldn't have a clue how to set it up.* [Participant 17, female, 36 years, high technology use, Dublin]

Familiarization with the system was seen as crucial as patients wanted a face-to-face demo with the PATHway team initially. Information technology support and phone support were suggested as the most acceptable, with further suggestions of email support and YouTube videos for technically capable users. Some patients suggested a manual with diagrams and pictures with a quick guide to enable ease of use:

> *I think once you've had one session with people and they show you how to use it, it would probably be ok and maybe then if there was a place that you could go...if you weren't able to manage it.* [Participant 14, female, 74 years, moderate technology use, Dublin]

Theme 2: Opportunity

Physical Opportunity

Overall, patients expressed three common obstacles to physical opportunity. This was with particular reference to resources such as time, equipment (including high speed Wi-Fi), and space in the home. Patients showed various preferences in relation to having a set space for PATHway use at home vs a moveable system to facilitate varying family and home life needs. Flexibility of class start times within PATHway was seen as a positive feature for those who were still working; however,

XSL•FO

RenderX

many felt that with flexibility came the trade-off of procrastination and no set time therefore leading to less accountability:

We all have the best intentions. So I think to have facilities at home...you want to be motivated, unless there was something you had signed up to. [Participant 16, male, 67 years, high technology use, Dublin]

Table 1. Demographics of patients in qualitative interviews (Leuven and Dublin sites).

ID	Site	Phase	Technology use	Sex	Age (years)	Education	Reason for attending CR[a]
P1	Leuven	Adult congenital heart disease (phase 3)	Moderate	Male	25	PhD-student	Congenital heart disease
P2	Leuven	Phase 3	Moderate	Male	63	Master	Percutaneous coronary intervention (PCI)
P3	Leuven	Phase 2	High	Male	59	Master	PCI
P4	Leuven	Phase 2	Moderate	Male	50	Bachelor	Coronary artery bypass graft (CABG)
P5	Leuven	Phase 2	Low	Male	69	Master	PCI
P6	Leuven	Phase 3	Low	Male	64	Bachelor	PCI
P7	Leuven	Phase 3	High	Male	78	PhD (professor emeritus)	CABG
P8	Leuven	Phase 3	Moderate	Male	74	Master	CABG, ICD
P9	Leuven	Phase 2	Low	Male	48	Leaving certificate	CABG
P10	Leuven	Phase 2	High	Male	42	Bachelor	Heart failure
P11	Leuven	Phase 2	Low	Male	65	PhD	Heart failure
P12	Leuven	Phase 2 (dropout)	High	Male	55	Leaving certificate	CABG
P13	Leuven	Phase 3	High	Male	69	Bachelor	PCI
P14	Dublin	Phase 3 (dropout)	Moderate	Female	74	Bachelor	Heart attack or stenting
P15	Dublin	Phase 2	Moderate	Male	62	Intermediate or junior or group certificate	Heart attack or stenting
P16	Dublin	Phase 2	High	Male	67	Leaving certificate	Heart attack or stenting
P17	Dublin	Phase 2	High	Female	36	Intermediate or junior or group certificate	Heart attack or stenting
P18	Dublin	Phase 2 (dropout)	High	Male	61	Intermediate or junior or group certificate	Stenting
P19	Dublin	Phase 2	High	Male	70	Primary level	Heart attack or stenting
P20	Dublin	Phase 2 (dropout)	High	Male	60	Diploma or certificate	Stenting or pacemaker implantable cardioverter d*efibrillator*
P21	Dublin	Phase 2 (dropout)	High	Female	56	Intermediate or junior or group certificate	Heart attack or stenting
P22	Dublin	Phase 2	Moderate	Male	60	Bachelor	Heart attack or stenting
P23	Dublin	Phase 2	High	Female	46	Diploma or certificate	Heart attack or stenting
P24	Dublin	Phase 2 (dropout)	High	Male	49	Leaving certificate	Heart attack or stenting
P25	Dublin	Phase 3 (dropout)	High	Male	53	Intermediate or junior or group certificate	Prediabetes
P26	Dublin	Phase 3(dropout)	High	Male	60	Postgraduate or higher degree	Heart attack or stenting
P27	Dublin	Phase 2	High	Male	62	Leaving certificate	Bypass grafting
P28	Dublin	Phase 3	Low	Female	82	Diploma or certificate	Stenting
P29	Dublin	Phase 3	Moderate	Male	72	Master	Heart attack or stenting
P30	Dublin	Phase 3	High	Male	62	Leaving certificate	Heart attack or stenting
P31	Dublin	Phase 3	High	Female	—	Bachelor	—
P32	Dublin	Pilot phase 3	High	Male	62	Bachelor	Valve surgery
P33	Dublin	Phase 3	Low	Female	77	Diploma or certificate	Stenting

[a]CR: cardiac rehabilitation.

Textbox 1. Patient themes and subthemes.

Capability

- physical capability

- psychological capability, including psychological readiness and technological readiness

Opportunity

- physical opportunity

- social opportunity

Motivation

- goal setting

- social interaction

- perceptions

- structured approach to exercise

- personalization

- present and future health and well-being

Social Opportunity

Patients felt a high level of support from their social environments, including family and friend contexts. However, a number of patients felt that exercise had to be self-motivated to ensure long-term maintenance. Several patients suggested that PATHway could be introduced as a family-wide intervention. Leveraging family connections to promote intergenerational lifestyle could enable PA, while not necessarily prioritizing it above other family commitments, as illustrated in the following statement:

> *I don't want to become obsessed and say "no kids I'm not talking to you tonight, forget about the homework I've got my run." I did that for long enough you know and now I'm trying to match everything up, not to be too greedy with my own time.* [Participant 26, male, 60 years, high technology use, Dublin]

Patients felt that they would like a bridge between hospital-based CR and community-based CR to maintain social connections forged within the hospital-based CR program:

> *We're with people we all know, we know how each other feel. We're with each other three times a week...If we could stay together, we continue together.* [Participant 17, female, 36 years, high technology use, Dublin]

This included connections with health care professionals (HCPs). PATHway was seen as a potential way of on-going professional support; however, an agreed structured approach to monitoring was seen as desirable:

> *Would I use it? YES! At least, if I'm still allowed to come on a consultation from time to time.* [Participant 9, male, 48 years, low technology use, Leuven]

Home exercise was seen as a positive alternative to gym use for long term maintenance. Patients shared concerns regarding dislike of gyms, embarrassment, and feeling out of place in the gym context:

> *You go to a gym, you're looking at the person next to you, you're wondering are they looking at you. For somebody who is overweight they don't want to go to a gym and have anybody look at them. They feel insecure. If you're at home you don't have that.* [Participant 23, female, 46 years, high technology use, Dublin]

Theme 3: Motivation for Cardiovascular Disease Patients

Various factors were explored in relation to motivation. Central subthemes were as follows: (1) goal setting, (2) social interaction, (3) perceptions, (4) structured approach to exercise, (5) monitoring, (6) personalization, and (7) present and future health and well-being.

Goal Setting

Goal setting was a key motivator for most of the patients; most identified that they felt less obligated to exercise at home. Patients felt that tracking PA and creating concrete action plans were key to success. Some patients wanted to set these goals with HCPs or face-to-face to create a sense of accountability, such as a contract agreement. Patients highlighted the need for personal fitness goals and functional improvement:

> *I'd be checking to make sure I'd be getting near my goal and if I wasn't getting near then I'd make the extra effort to do it.* [Participant 21, female, 56 years, high technology use, Dublin]

Patients wanted progress and indicators of performance to be tracked to enable visual feedback, as well as physical testing at intermittent periods. Risk profiles were deemed important information by patients as many expressed not fully understanding the underlying cause of their CVD. Prompts were also seen as an important and acceptable part of engaging in maintenance of lifestyle change.

XSL•FO

RenderX

Social Interaction

Social interaction was the motivating factor for individuals currently partaking in CR and general PA. Many patients were interested in PATHway, allowing them access to a further social network or an avenue to augment their existing network:

> *You meet the other people who are in the same boat you'll all have similar stories. So that alone means that you're not alone.* [Participant 32, male, 62 years, high technology use, Dublin]

> *I like to talk, I like to have people around me and I have always been together with people. Being on my own is nothing for me.* [Participant 3, male, 59 years, high technology use, Leuven]

The group was also cited as creating a positive social pressure; however, some patients were concerned about unwise comparisons with fellow patients:

> *There's no sense in comparing me with a 20 stone man. It's inaccurate you know.* [Participant 15, male, 62 years, moderate technology use, Dublin]

Illness Perceptions

Participant's personal illness perceptions (ie, their current health or CVD status and its subsequent impact) largely influenced participant views of their personal exercise ability and subsequent participation. Many patients did not consider CVD to be a chronic condition that needed to be managed and had difficulty understanding their personal risk profiles, with a majority of patients assigning their CVD risk factor to family history.

Certain exercises and sports were associated with specific age groups. Many patients were not active before the CVD incident and felt that fitness was purely functional, with many experiencing low self-efficacy in relation to exercise. This had an interesting impact on how different potential components of PATHway were received by participants. The gaming module of PATHway was often not considered as a viable option for rigorous targeted exercise as it did not fit with many of the participant's traditional perception of what exercise "should be." Some felt that implementing a "game" as a health care self-management solution was not appropriate:

> *It wasn't used as an exercise thing; it was just something to do.* [Participant 19, male, 70 years, high technology use, Dublin]

Conversely, many users were interested in a low-technology solution such as engaging with PATHway through the Active Lifestyle component. Key motivators were enjoyment of the outdoors, engaging in exercise with a purpose, and getting out of the house.

Structured Approach to Exercise

A key point that certain individuals felt was necessary was a structured option to create a sense of obligation to engage:

> *Rather than going off and doing something which would not be of any benefit for me but would nearly be detrimental I think having a kind of a laid down programme, you'd use different bits of it but it's the*

> *programme that suits your physical state, would be a very good idea.* [Participant 32, male, 62 years, high technology use, Dublin]

Monitoring

Monitoring, feedback, and HCP's recommendation were key aspects of motivation to engage in PA. Patients' felt that feedback could be delivered in a variety of ways including motivational messages, recordings, videos, immediate audio and visual feedback from the avatar and wrist-worn sensors, summaries, and graphs. However, patients noted that they did not want to become obsessive with the monitoring:

> *I think I'd become obsessed. I'd like to be able to go and say check my own when I need to but not to be in my face all the time.* [Participant 23, female, 46 years, high technology use, Dublin]

Patients valued health information and feedback could be utilized by an HCP but wanted their information to be private. Interestingly, patients referred to the PATHway system and remote monitoring as a "big brother" type of device:

> *It's a bit like big brother as well...I personally think it would work.* [Participant 21, female, 56 years, high technology use, Dublin]

Personalization

Many patients felt that this individualization and monitoring would give them a greater insight into their personal health. Personal data was seen as a way to further their own awareness and self-reflection and aid the management of CVD. A further aspect for patients was the ability to control the exercise intensity, as illustrated in the following statement:

> *So long as it's encouraging and it recognises people not as a number as such, as an individual...that to me now would be a big; I feel everybody kind of is inclined to be a number nowadays. That would motivate me as well.* [Participant 23, female, 46 years, high technology use, Dublin]

Present and Future Health and Well-Being

Prevention of further CVD incidences was a core motivator for many patients. Patients noted that there was an immense sense of achievement following exercise:

> *I know people who have had similar problems to me and they just haven't bothered but for me it was the right thing. I still think it's a great idea; it's one way of making sure that I'm staying fitter and I'm doing the right things for the heart muscle to keep me going.* [Participant 32, male, 62 years, high technology use, Dublin]

PA was seen as a core component of stress management by patients who exercised regularly:

> *If you're really focused on something...all the normal day to day stresses and everything are gone...and I like that.* [Participant 31, Female, age not given, high technology use, Dublin]

Results of Stakeholder Interviews

Stakeholder interviews provided information from the stakeholders' perspective of the key barriers preventing and the motivators increasing the likelihood of stakeholders in general to use PATHway as a tool to facilitate long-term maintenance of health behavior change among community-based CR patients. Stakeholders also reflected on elements that would act as barriers and motivators to engage. Stakeholders reflected on both professional and patient elements within this model.

Stakeholders considered aspects of patient engagement within community-based CR programs (eg, the psychological capability of some of the cohort to engage with new technology). However, there is also professional reflection in terms of barriers (eg, whole team buy-in). These themes are presented in Textbox 2 using the COM-B model framework.

Theme 1: Capability

Capability was discussed in light of both psychological and physical capability for both patients and stakeholders.

Psychological Capability

Mismatch Between Current Cardiovascular Disease Patients and Future Cardiovascular Disease Patients

Many stakeholders suggested that future CVD patients in the coming years will become increasingly knowledgeable of health care technology and its use in self-management; however, some were unsure whether current typical profile CVD patients would be familiar enough with technology and willing to use it as part of their long-term CR:

> *There's going to be new generation of patients going into the future that are more tech savvy than my generation.* [Stakeholder 14, specialized cardiac nurse, Dublin]

Other stakeholders recommended that data protection was an important issue, and all information recorded and retained by the system would need to be explained in full to the patient with details of why such information is necessary or beneficial:

> *This is like Big Brother...they would need to be reassured…it is people watching what's going on but for a very good reason...there's an extra element there with people who haven't grown up with this as part of normal experience.* [Stakeholder 15, health service executive, Dublin]

Textbox 2. Stakeholder themes mapped to capability, opportunity, motivation, and behavior (COM-B) model.

Capability

- psychological capability

  - mismatch between current cardiovascular disease (CVD) patients and future CVD patients

  - ease of use

- physical capability

  - alarm and emergency protocols

Opportunity

- social opportunity

  - peer-to peer social connections

  - health care connections

  - whole team buy-in

  - general social support

  - patient-led participation

- physical opportunity

  - finite resources

Motivation

- potential for real-time assessment and feedback

- technology augmented care

- positive patient reinforcement

- personalization

- feedback

Ease of Use

Stakeholders emphasized that the system above all else needed to be easy-to-use for both clinicians and patients. Overly complicated systems were seen as a barrier to use. Stakeholders also suggested that the avatar should be able to issue teaching points as well as visual cues to the patient to aid interaction with the system, including an explanation of how to wear and use the sensors within PATHway. One stakeholder stated:

> We're dealing with patients putting on monitors every day and it can be very difficult for some of them, technically very difficult. [Stakeholder 14, specialized cardiac nurse, Dublin]

Stakeholders felt that information technology support should be made available for patients:

> I think people don't mind being talked through something over the phone but certainly if you've to go online and try and chat to someone I think there's a barrier. [Stakeholder 13, local patient organization, Dublin]

This ease of use theme also included a clinical and patient interface. Stakeholders suggested that summary information would have to be presented in a simple easy-to-use way, including graphics of patient progress and the information and progress in relation to important risk factors. These summaries were recommended to be no longer than 1 page long and easily read during a consultation with information that did not just provide a duplicate of what is already available:

> [People are] terribly impatient with stuff, you know? It has to be the cat sat on the mat [Stakeholder 1, cardiac rehabilitation cardiologist, Dublin]

Physical Capability

Alarm and Emergency Protocol

Stakeholder emphasized how clearly the system needed to alert the patients to adverse events. Stakeholders also highlighted the need to explain the potential for false alarms to patients before interaction with the system:

> You need a big red stop...Stop what you're doing, sit down, have a glass of water, in big writing and it has to be red, because that's alarm. [Stakeholder 14, specialized cardiac nurse, Dublin]

Theme 2: Opportunity

In terms of opportunity, stakeholders reflected on aspects of both social and physical opportunity. This included how patients engage but also how HCPs may also engage (see subthemes "whole team buy-in" and "patient-led participation").

Social Opportunity

New Social Peer-to-Peer and Health Care Connections

To ensure patient engagement with PATHway, stakeholders suggested that PATHway should aim to harness initial hospital-based social connections for optimal results in facilitating recruitment and retention among current CVD patients within the hospital. By using hospital-based CR as a platform to launch and familiarize PATHway with patients, preexisting groups and social interactions can be augmented, whereby patients have the option to motivate one another, provide peer mentorship, and strengthen habit-forming and routine created during attendance at the hospital-based CR. One stakeholder stated:

> There's also the sense of belonging to a group, which is very attractive for some patients. [Stakeholder 2, public policy, Dublin]

Stakeholders felt that PATHway could be seen as a type of remote follow-up for patients to provide additional support and as a way to further educate and inform patients of best practice and current guidelines. This was deemed, by the stakeholders, to be integral to on-going patient self-management:

> There is that instant ready assessment of how they're doing with a straight access to whoever the key players are including the patient and their health professional. [Stakeholder 13, local patient organization, Dublin]

A "patient-centered approach" was deemed very important, particularly in terms of engagement and long-term use of the system:

> It is not only about what you think is important but also what the patient thinks is important, it always the combination. [Stakeholder 1, cardiac rehabilitation cardiologist, Dublin]

General Social Support

Some stakeholders suggested the participation of family and peers is a practical way to provide implement a "patient-centered approach." Peer mentors who have experienced CVD are in a unique position to guide, advise, and motivate fellow CVD recoverees to engage with CR and PATHway in particular. Family members can be active within the intervention as support in PA maintenance, potentially playing a monitoring role to create accountability for the CVD patient, as illustrated in the following statement:

> The participation of other members of the family, I think that's really positive and I think children, younger people love to take on and support a parent particularly in getting better you know. So I think there's huge potential there...we know it from other settings, the peer and the group initiative. [Stakeholder 13, local patient organization, Dublin]

Whole Team "Buy-In"

"Buy-in" from the cardiology unit was seen as crucial to the implementation of such an intervention, including consultants and senior staff members. The "buy-in" of the HCP was seen as an advantage with regards to patient recruitment and retention, and whole team "buy-in" was ideally envisaged as a multidisciplinary effort to optimize patient experience.

Stakeholders within the Belgian system felt that further information on past patients would be valuable and rewarding, whereas within the Irish system, this extra information and follow-up seemed to create a sense of further responsibility, as illustrated in the following statements:

I would feel satisfied to objectively see that the patients keep exercising. I consider it very important and I've been missing this aspect throughout my years at the cardiac rehab. What is the long-term results of all our investments, ie, time and effort we put in the patients to clarify to him what is important the long term. [Stakeholder 2, public policy, Dublin]

That's a whole new role though...I wouldn't feel comfortable doing that in my daily work because I have my own work to do and that's a responsible job you know? You've people's lives at risk at home...they put a lot of onus on their own health, their responsibility on to your shoulders and that can be difficult. [Stakeholder 14, specialized cardiac nurse, Dublin]

Patient-Led Participation

Several stakeholders identified that engagement with PATHway needs to be patient-led. Stakeholders identified that if a patient requests information from the PATHway system to be a part of their consultation then it "forces" them—the stakeholders—into action and engagement with the system and can generate a good foundation for further focused follow-up appointments, as illustrated in the following quote:

Basically, they're patient held, they're patient produced and it's part of the consultation. If you're going to start to ask primary health care to provide it at another time access a portal of some sort to view patient data I think there's challenges around the uptake of that. [Stakeholder 21, GP, Dublin]

Physical Opportunity

Stakeholders discussed barriers to PATHway implementation that were mainly focused around resource issues.

Finite Resources

Time and money were seen as finite resources within the clinical setting. A key concern regarding implementation of PATHway within a clinical setting is whether PATHway would deliver better patient care and useful clinical information while maintaining or minimizing current workload:

These are serious time invests...If you really want to do an intensive follow up it surely would be a serious time-investment compared to just receiving a weekly summary where you can see how frequent and at what intensities the patient has been training. [Stakeholder 2, public policy, Dublin]

Further concerns were raised regarding the follow-up of patients and whether this task fit into existing CR roles and whether the obtained information would be used in existing practices. This highlighted the need for an examination of intended financial and staffing resources:

We're giving them information to go on to do [community-based CR] but we don't bring them back. [Stakeholder 14, specialized cardiac nurse, Dublin]

Theme 3: Motivation

Age of Measurement

Stakeholders identified that technology as an integral part of future health care, calling it the "age of measurement" where patients and HCPs alike expect to have health status measurements and feedback. Giving feedback and providing various health measurements was identified as increasing the acceptability and adding value to the use of PATHway among patients:

I think it can or could offer added value for certain topics, especially in terms of objectivizing a number of parameters and general items that we are often still evaluating in an approximate or very subjective way. [Stakeholder 2, public policy, Dublin]

HCPs felt that these measurements could increase efficacy within patient consultations and act a good starting point for follow-up appointments.

Mode of Feedback

Feedback was suggested in a variety of forms particularly in relation to information available through PATHway and also push notifications and prompts:

Technology can sometimes be a reminder, a stimulus in order to motivate your patients to practice, so that they can rehabilitate and make some progress. [Stakeholder 2, public policy, Dublin]

It was advised that patients should give feedback to aid tailoring and program improvement. This was suggested to be easily captured via brief feedback from the participant on their satisfaction and enjoyment levels following their exercise class.

Many stakeholders mentioned the issue of patients becoming too dependent and obsessed with measurement and numbers (ie, with wrist worn sensors, etc). This was a cause for concern, as ideally patients should feel safe to exercise limits using the Rate of Perceived Exertion scale, and this would be a better gauge of progress, as illustrated in the following statement:

They become obsessed with the numbers and if there's any fluctuation they start panicking so that's why I don't think it's, personally, it's a good method. And we have always used the RPE scale here and I find it works very well if you explain it properly to the patient. [Stakeholder 14, specialized cardiac nurse, Dublin]

In terms of feedback for stakeholders, key outcomes of interest were as follows: body mass index, height, weight, cholesterol, alcohol consumption, smoking cessation, medication adherence, blood pressure, steps, PA (Frequency, Intensity, Type, and Time), nutrition, well-being, general mental health, quality of life, satisfaction using PATHway, self-efficacy, engagement with pathway, social support, and social interaction.

Technology Augmented Care

Some stakeholders expressed that patients should be educated in a holistic way, including exercise confidence and psychological well-being. This was seen as important outcome

to facilitate a positive view of health and self-management, potentially having an impact on illness and exercise perceptions:

> *I think it is also important to validate the psychological well-being of the patients...does the patient feel better, are patients more confident while exercising and is he more motivated to exercise.* [Stakeholder 12, psychologist involved in CR, Dublin]

Stakeholders clearly identified that although technology could augment care and many were very supportive of the role of technology in health care whereby a system such as PATHway would be able to complement existing care practices and lead a more efficient and satisfactory follow-on service:

> *It would make it possible to interact with the patients on a regular basis besides the more formal contacts we have during the follow-up consults.* [Stakeholder 6, exercise physiologist, Leuven]

However, several stakeholders clearly raised the issue that the therapeutic relationship should not be replaced totally:

> *You...would be very wary of programmes...taking over...the irreplaceable role of the therapeutic relationship...I only think of it as augmentative.* [Stakeholder 12, psychologist involved in CR]

Technology was seen something that could provide comfort and safety to CVD patients who lacked confidence while exercising:

> *The greatest challenge for me is to translate the feeling of safety to the patient, that they really get the feeling "I can exercise here safely." I think that's the most important.* [Stakeholder 10, exercise physiologist, Leuven]

Positive Patient Reinforcement

Positive reinforcement in the absence of negative reinforcement was seen as important in terms of motivation and engagement. Many stakeholders suggested a multidisciplinary approach whereby patients set goals in tandem with key health professionals:

> *You should do it in cooperation with all healthcare providers involved and set up certain goals together with your patients on beforehand.* [Stakeholder 4, exercise physiologist, Leuven]

Positive reinforcement was a key point that was mentioned in terms of feedback to the user. Nonconfrontational gradual approaches were seen as a productive way to move forward with health and lifestyle information and intervention in a primary care setting:

> *Setting goals together with the patients and evaluating with which means you'll try to achieve those goals. If you impose a certain technology to your patients, in whom they don't fully believe themselves, you'll fail. It must be integrated within some kind of motivational strategy.* [Stakeholder 4, exercise physiologist, Leuven]

Personalization

Personalization was cited to be key for both users and clinicians. Provision of personalized exercise programs was deemed important in terms of patient capability and enjoyment:

> *There has to be provided something whereby the patient himself could personalize his therapy and be personally responsible together with the health care provider and also helps shaping his own programme.* [Stakeholder 4, exercise physiologist, Leuven]

Discussion

Principal Findings

This study addresses both patient and stakeholder requirements for a CVD eHealth intervention. This development work highlights the need for certain aspects of capability, opportunity, and motivation to be addressed for the implementation of an eHealth self-management system among CVD patients and stakeholders. This study highlights the core combined requirements of such a system for both patients and other key stakeholders. In summary, Textbox 3 broadly highlights high-level needs such as (1) personalized feedback delivered through easy-to-use technology, (2) facilitating social connection and support (both peer and professional), (3) the importance of the availability of resources, and (4) augmentation of existing relationships through technology.

Initially, before patients began PATHway use, physical and psychological capabilities were felt to be important considerations in terms of use of the system. Participants felt that the PATHway system should be as integrated into daily living as possible. Passive data collection or "sensed data," including heart rate, steps, calories, and minutes of PA will allow a wide range of information to be available for the users to get up-to-date feedback facilitating engagement with behavioral change components of PATHway. It has been shown that personalized interventions have superior efficacy over time compared with those that base their tailoring on single or infrequent assessments (eg, baseline) [17,18]. This also creates a balance of user engagement but also allows flexibility around potential participant difficulties or reluctance to enter information manually.

Once issues of capability had been discussed, it was apparent that patients valued information and measurement of their health status. However, despite the desire for improved measurement of core outcomes and increased reporting, stakeholder and patient enthusiasm for "the age of measurement" must be tempered with caution, as both patient and stakeholders had reservations in relation to patient technology competencies. This concern was reflected in the patient theme "psychological capability" and subthemes "psychological readiness" and "technological readiness," whereas stakeholders flagged a "mismatch between current CVD patients and future CVD patients" technology competencies. Digital health literacy is an important consideration within eHealth interventions, and participants need to be supported appropriately within intervention standard operating procedures [19]. Potential solutions for PATHway include a user manual, a familiarization

phase within CR classes, how-to videos, and a support line. These options will be further explored within the PATHway trial and further qualitative work in PATHway debrief interviews. These will be guided by the Health information technology Usability Evaluation Model [20] and will be used to evaluate the experience with selected SEM stakeholders.

Despite these concerns regarding digital health literacy, findings from previous studies [14] suggest that the current target CVD population *are* familiar with technology and, importantly, have regular access to smartphones and the Internet, with the majority of patients reporting the concept of PATHway as appealing. These findings support previous formative work from a mobile health (mHealth) CR exercise intervention in New Zealand [21]. However, it is important to embed this formative work within the context of recent evidence synthesis highlighting key behavior change techniques to be used in eHealth interventions

with CVD patients [22,23]. eHealth offers a tangible opportunity to provide pervasive connections between CVD patients and HCPs, thus "harnessing phase 2 connections" (as recommended by key stakeholders) for optimal phase 3 CR adherence and engagement.

Findings revealed that "the instructor" was an essential component of the theme "motivation" for patient CR adherence. Previous research has cited the use of an avatar within mHealth interventions as a potentially "nonthreatening conversational agent" [24]. Previous research has shown that often patients perceive avatars positively, assigning high levels of empathy and alliance to the avatar [25]. Significantly, this also has implications for patients with low levels of health literacy. From the qualitative interviews, it was clear that many patients did not engage with the health information and guidance provided for them by traditional HCPs and stakeholders in phase 2.

Textbox 3. Combined patient and stakeholder capability, opportunity, motivation, and behavior (COM-B). Italics indicate COM-B subcategories.

Capability

- psychological capability
  - mismatch between current cardiovascular disease (CVD) patients and future CVD patients
  - ease of use
  - psychological readiness
  - technological readiness

- physical capability
  - alarm and emergency protocols

Opportunity

- social opportunity
  - new social peer-to-peer and health care connections
  - whole team buy-in
  - general social support

- Physical opportunity
  - finite resources
  - clinical and patient interface

Motivation

- patient-led participation
- mode of feedback
- potential for real-time assessment and feedback
- positive patient reinforcement
- personalization
- goal setting
- social interaction
- perceptions
- structured approach to exercise
- present and future health and well-being

Perhaps the use of an avatar or "conversational agent" within PATHway can initiate engagement with certain at-risk individuals who would otherwise not engage with regular CR routes of care.

The theme "opportunity" relates to both "physical opportunity" and "social opportunity." Within the subtheme "social opportunity," a "community-of-practice" was called for with the implementation of a family-wide intervention and workplace intervention to enable greater adoption of CR guidelines. This suggestion from patients and stakeholders to use the CVD incident as an opportunity or "intervention window" whereby patients and their families use the incident of CVD to promote healthy behavior change within their home. This finding is important for future CR interventions to optimize engagement initially within the intervention but also within eHealth trials for the purposes of recruitment and retention. This use of a "teachable moment" has been used in previous PA interventions to leverage a greater impact than stand-alone individual intervention components [26].

Finite resources were an issue for stakeholders. A key goal for eHealth interventions is to provide an effective, evidence-based, low-cost alternative to traditional health care routes. eHealth can reduce the burden of patient management through enabling remote participation in a phase 3 CR program. However, key stakeholders have identified issues with role delineation and remain concerned regarding the burden of remote patient monitoring in conjunction with their existing duties. This highlights the importance of the stakeholder identified theme "whole team buy-in" and consideration of the health ecosystem. This is particularly relevant given the different contexts in which the PATHway system may be implemented within. The Irish Association of Cardiac Rehabilitation has stated that CR services in Ireland have received cutbacks during a time of economic recession in Ireland. It is therefore noteworthy that Irish stakeholders found the prospect of further patient information and potential related duties and responsibilities difficult to envisage within their current context, whereas the Belgian stakeholders found these proposals predominantly positive and feasible to implement.

Goal setting and monitoring were integral elements of the patient theme "motivation." Implementing goal setting is a key behavior change technique to be used within PATHway and one that was the most cited by patients as desirable. Previous research has examined implementing different levels of goal options to promote participant self-efficacy in relation to exercise [27]. This is very important for remote eHealth interventions given participant concerns regarding "physical capability" and "exercise perceptions." Goal feedback is important to buffer against patient disengagement and integral to participant retention with the PATHway system. This relates back to the stakeholder recommendation of using mainly positive reinforcement. King and colleagues [27] employed positive reinforcement statements that were delivered when participants either met their weekly goal or exceeded their weekly goal. Future PATHway development phases will be able to explore how best to present such goal-setting and monitoring functions to foster motivation for participants. This finding is particularly

relevant for studies that will implement exercise prescriptions for participants and adapt goals based on monitored progress.

Personalization was also expected in PATHway feedback, prompts, and notifications. This expectation is interesting as it has been previously demonstrated that tailored health messages are more engaging and effective in terms of health behavior change than untailored, generic messages [28]. As participant engagement and retention are critical factors in successful behavior change research, it is important for further exploration as considerable resources are needed to deliver tailored content.

Social interaction with others is a core driver of motivation and is closely linked to the "social opportunity" subtheme within the opportunity category of the COM-B model. Social opportunity can be defined as an opportunity created by cultural context that shapes how we think about things [10]. In essence, it can be seen that the social context (ie, social opportunity) facilitates initial engagement with an eHealth system such as PATHway (eg, your CR coordinator recommending its use). It has been previously evidenced that people are inclined to expose themselves to innovations that not only provide a solution to their needs but that also appear to be consistent with and reinforce their attitudes or value systems [29]. Further to this "social opportunity," social interaction as a motivation focuses on the social connection that PATHway could potentially facilitate. This subtheme highlighted again the importance of the stakeholder call for harnessing phase 2 connections both in relation to the peer-to-peer social connections made but also the important health care connections. The importance of social interaction as a motivation for engagement with the PATHway system tied in with the stakeholder recommendation of a "patient-centered approach," employing key strategies such as that of structured social support and peer mentoring for patients. Augmenting existing social connections can be effective in increasing PA, along with other healthy lifestyle behaviors. The harnessing of these existing connections may be more effective for patient engagement and satisfaction than trying to create a whole new online community.

Strengths and Limitations

The use of qualitative interviews in this study allows in-depth understandings of PATHway patients and stakeholder views that may not be reached using quantitative methods alone, while the use of COM-B driven interview scripts facilitates evidence-based intervention development and highlights areas to address for future intervention implementation across different European health care systems.

The recruitment of a broad spectrum of users was an important task for PATHway and a key strength given the multisite nature of the project. Recruitment was balanced for age, gender, and socioeconomic status insofar as possible. However, in terms of limitations, "dropouts" from phase 2 and phase 3 were difficult to recruit because of their disengagement with services.

Future Research and Conclusions

This is the first phase of development for the PATHway system. Further extensive user testing will be conducted, and a process evaluation in terms of the feasibility, acceptability, satisfaction, and usability of the PATHway system from the participants'

perspective will be assessed before a randomized trial. This future trial will also involve a health economics evaluation of the cost-effectiveness of PATHway. Furthermore, following this trial, debrief interviews will be conducted with participants and stakeholders (ie, as defined by the SEM used previously). This is crucial for future feasible implementation of the PATHway system within other European health care systems.

In conclusion, several broad learnings emerged from the in-depth qualitative work with individuals with CVD and HCPs. It is clear that a multifaceted, personalizable intervention is desirable. Key learnings include the need for maximal patient tailoring, simplicity within the platform, technology-augmented care, enabling or increasing individual self-management through eHealth, and capitalizing on an appropriate time to intervene in the CR journey.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1

Patient interview script.

[PDF File (Adobe PDF File), 53KB - jmir_v20i5e163_app1.pdf]

Multimedia Appendix 2

Stakeholder interview script.

[PDF File (Adobe PDF File), 47KB - jmir_v20i5e163_app2.pdf]

References

1. World Health Organization. 2011. Essential medicines and health products information portal URL: http://apps.who.int/medicinedocs/en/d/Js18804en/ [WebCite Cache ID 6uBVhEMyS]
2. World Health Organization. 2015. Fact sheet N317: cardiovascular diseases (CVDs) URL: http://www.who.int/mediacentre/factsheets/fs317/en/ [WebCite Cache ID 6uBVaclZV]
3. Balady G, Williams M, Ades P, Bittner V, Comoss P, Foody J. Ahajournals. Core components of cardiac rehabilitation/secondary prevention programs: 2007 update URL: http://circ.ahajournals.org/content/115/20/2675 [accessed 2018-03-28] [WebCite Cache ID 6yG8Sr5P2]
4. Piepoli M, Corrà U, Adamopoulos S, Benzer W, Bjarnason-Wehrens B, Cupples M. Secondary prevention in the clinical management of patients with cardiovascular diseases. Eur J Prev Cardiol 2014;21(6):664-681 [FREE Full text] [Medline: 10.1177/2047487312449597]
5. Dunlay S, Witt B, Allison T. Barriers to participation in cardiac rehabilitation. Am Heart J 2009;158(5):852-859 [FREE Full text] [doi: 10.1016/j.ahj.2009.08.010]
6. Taylor R, Kirby B. The evidence base for the cost effectiveness of cardiac rehabilitation. Heart 1997 Jul;78(1):5-6 [FREE Full text] [Medline: 9290390]
7. Dalal HM, Zawada A, Jolly K, Moxham T, Taylor RS. Home based versus centre based cardiac rehabilitation: cochrane systematic review and meta-analysis. BMJ 2010 Jan 19;340(jan19 4):b5631. [doi: 10.1136/bmj.b5631]
8. Saner H. e-Cardiology and e-Health: from industry-driven technical progress to clinical application. Eur J Prev Cardiol 2014 Nov;21(2 Suppl):2-3. [doi: 10.1177/2047487314553160] [Medline: 25354947]
9. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new medical research council guidance. BMJ 2008 Sep 29;337(sep29 1):a1655. [doi: 10.1136/bmj.a1655]
10. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions. London: Silverback Publishing; 2014.
11. Centre for Disease Control and Prevention. Social ecological model URL: https://www.cdc.gov/cancer/crccp/sem.htm [accessed 2018-02-02] [WebCite Cache ID 6wvpHQacM]
12. Department of Health and Children. Health.gov. 1999. Building healthier hearts: introduction to the report of the cardiovascular health strategy group URL: http://health.gov.ie/wp-content/uploads/2014/03/Building-Healthier-Hearts-Report.pdf [accessed 2018-02-02] [WebCite Cache ID 6wvp7SJ50]

13. Belgian Health Care Knowledge Centre. 2010. Cardiac rehabilitation: clinical effectiveness and utilisation in Belgium URL: https://kce.fgov.be/sites/default/files/atoms/files/kce_140c_cardiac_rehabilitation_0.pdf [WebCite Cache ID 6wvpw0nzeOriginal]

14. Buys R, Claes J, Walsh D, Cornelis N, Moran K, Budts W, et al. Cardiac patients show high interest in technology enabled cardiovascular rehabilitation. BMC Med Inform Decis Mak 2016 Dec 19;16:95 [FREE Full text] [doi: 10.1186/s12911-016-0329-9] [Medline: 27431419]

15. Charmaz K. Constructing grounded theory. London: Sage Publications Ltd; 2006.

16. Braun V, Clarke V. Using thematic analysis in psychology. Qual res psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

17. Adams MA, Sallis JF, Norman GJ, Hovell MF, Hekler EB, Perata E. An adaptive physical activity intervention for overweight adults: a randomized controlled trial. PLoS One 2013 Dec;8(12):e82901 [FREE Full text] [doi: 10.1371/journal.pone.0082901] [Medline: 24349392]

18. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010 Sep;51(3-4):214-221 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.004] [Medline: 20558196]

19. van der Vaart R, Drossaert C. Development of the digital health literacy instrument: measuring a broad spectrum of health 1.0 and health 2.0 skills. J Med Internet Res 2017 Jan 24;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]

20. Brown W, Yen P, Rojas M, Schnall R. Assessment of the health IT usability evaluation model (Health-ITUEM) for evaluating mobile health (mHealth) technology. J Biomed Inform 2013 Dec;46(6):1080-1087 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.001] [Medline: 23973872]

21. Pfaeffli L, Maddison R, Whittaker R, Stewart R, Kerr A, Jiang Y, et al. A mHealth cardiac rehabilitation exercise intervention: findings from content development studies. BMC Cardiovasc Disord 2012 May 30;12:36 [FREE Full text] [doi: 10.1186/1471-2261-12-36] [Medline: 22646848]

22. Duff OM, Walsh DM, Furlong BA, O'Connor NE, Moran KA, Woods CB. Behavior change techniques in physical activity ehealth interventions for people with cardiovascular disease: systematic review. J Med Internet Res 2017 Aug 02;19(8):e281 [FREE Full text] [doi: 10.2196/jmir.7782] [Medline: 28768610]

23. Goodwin L, Ostuzzi G, Khan N, Hotopf MH, Moss-Morris R. Can we identify the active ingredients of behaviour change interventions for coronary heart disease patients? a systematic review and meta-analysis. PLoS One 2016 Apr;11(4):e0153271 [FREE Full text] [doi: 10.1371/journal.pone.0153271] [Medline: 27105435]

24. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013 Jul;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]

25. Bickmore T, Gruber A, Picard R. Establishing the computer–patient working alliance in automated health behavior change interventions. Patient Educ Couns 2005;59(1):21-30 [FREE Full text] [doi: 10.1016/j.pec.2004.09.008]

26. Demark-Wahnefried W, Jones LW, Snyder DC, Sloane RJ, Kimmick GG, Hughes DC, et al. Daughters and mothers against breast cancer (DAMES): main outcomes of a randomized controlled trial of weight loss in overweight mothers with breast cancer and their overweight daughters. Cancer 2014 Aug 15;120(16):2522-2534 [FREE Full text] [doi: 10.1002/cncr.28761] [Medline: 24804802]

27. King AC, Hekler EB, Grieco LA, Winter SJ, Sheats JL, Buman MP, et al. Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults. PLoS One 2013 Apr;8(4):e62613 [FREE Full text] [doi: 10.1371/journal.pone.0062613] [Medline: 23638127]

28. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009 Feb;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]

29. Hassinger E. Stages in the adoption process. Rural Sociol 1959;24:52-53.

Abbreviations

COM-B: capability, opportunity, motivation, and behavior
CR: cardiac rehabilitation
CVD: cardiovascular disease
eHealth: electronic health
HCP: health care professional
mHealth: mobile health
PA: physical activity
PATHway: Physical Activity Towards Health way
SEM: Social Ecological Model

XSL·FO

RenderX

Original Paper

Effects of Treatment Length and Chat-Based Counseling in a Web-Based Intervention for Cannabis Users: Randomized Factorial Trial

Benjamin Jonas[1], Dipl-Psych; Marc-Dennan Tensil[1], Dipl-Psych; Peter Tossmann[1], PhD, Dipl-Psych; Evelin Strüber[2], Dipl-Soz

[1]Delphi Gesellschaft für Forschung, Beratung und Projektentwicklung, Berlin, Germany
[2]Federal Centre for Health Education, Cologne, Germany

Corresponding Author:
Benjamin Jonas, Dipl-Psych
Delphi Gesellschaft für Forschung, Beratung und Projektentwicklung
Kaiserdamm 8
Berlin, 14057
Germany
Phone: 49 30 39 40 97 82
Fax: 49 30 39 40 97 89
Email: jonas@delphi.de

Abstract

Background: Digital interventions show promise in reducing problematic cannabis use. However, little is known about the effect of moderators in such interventions. The therapist-guided internet intervention Quit the Shit provides 50 days of chat-based (synchronous) and time-lagged (asynchronous) counseling.

Objective: In the study, we examined whether the effectiveness of Quit the Shit is reduced by shortening the program or by removing the chat-based counseling option.

Methods: We conducted a purely Web-based randomized experimental trial using a two-factorial design (factor 1: real-time-counseling via text-chat: yes vs no; factor 2: intervention duration: 50 days vs 28 days). Participants were recruited on the Quit the Shit website. Follow-ups were conducted 3, 6, and 12 months after randomization. Primary outcome was cannabis-use days during the past 30 days using a Timeline Followback procedure. Secondary outcomes were cannabis quantity, cannabis-use events, cannabis dependency (Severity of Dependence Scale), treatment satisfaction (Client Satisfaction Questionnaire), and working alliance (Working Alliance Inventory-short revised).

Results: In total, 534 participants were included in the trial. Follow-up rates were 47.2% (252/534) after 3 months, 38.2% (204/534) after 6 months, and 25.3% (135/534) after 12 months. Provision of real-time counseling (factor 1) was not significantly associated with any cannabis-related outcome but with higher treatment satisfaction (P=.001, d=0.34) and stronger working alliance (P=.008, d=0.22). In factor 2, no significant differences were found in any outcome. The reduction of cannabis use among all study participants was strong (P<.001, d≥1.13).

Conclusions: The reduction of program length and the waiver of synchronous communication have no meaningful impact on the effectiveness of Quit the Shit. It therefore seems tenable to abbreviate the program and to offer a self-guided start into Quit the Shit. Due to its positive impact on treatment satisfaction and working alliance, chat-based counseling nevertheless should be provided in Quit the Shit.

Trial Registration: International Standard Randomized Controlled Trial Number ISRCTN99818059; http://www.isrctn.com/ISRCTN99818059 (Archived by WebCite at http://www.webcitation.org/6uVDeJjfD)

Introduction

Cannabis is the most widely used illegal drug in Europe with a last-year prevalence of 14% among young adults and around 1% daily or almost daily users [1]. In 2015, cannabis use of 550,000 adults in Germany was clinically relevant [2].

In the past years, several online interventions targeting individuals with cannabis use disorder (CUD) were made available to the public in Europe [3]. The guided program "Quit the Shit" (QTS) is the only intervention that has been tested in a randomized controlled trial (RCT) [4-7]. Besides the Australian intervention "Reduce your use," QTS is currently the only evidence-based internet intervention targeting cannabis users that is freely available for the public [8].

As part of the drug prevention website drugcom, QTS is operated by the German Federal Centre for Health Education (BZgA) since 2004 and is one of the components of the prevention strategy for CUD of the BZgA [4,9]. With 50 days of individual counseling by trained therapists, QTS offers more intense support than most other evaluated interventions related to cannabis use [5,6]. Free and anonymous QTS provides direct (synchronous) and time-lagged (asynchronous) counseling. Synchronous counseling via live chat is mainly offered during admission and termination of the intervention. Asynchronous support is delivered by weekly feedbacks on participants' input in the cannabis use diary and exercises of QTS. The intervention is described in detail in the Methods section.

To account for the increasing demand of QTS and to make the program less dependent on prescheduled chat counseling, our interest was to test whether shortening the intervention and whether eliminating chat-based counseling has negative impact on the effectiveness of QTS. A shorter version of QTS would presumably reduce the counselor's effort per client and thereby allow increasing the number of participants. Moreover, a program without synchronous communication would be more flexible because participants would not be dependent on prescheduled chats to start the intervention. However, those changes should not significantly reduce the effectiveness of QTS or decrease the user satisfaction or working relation between the counselor and client.

Although internet interventions targeting substance-related disorders have shown to be effective, little is known about moderators of their effectiveness [6,10]. Despite some indication in favor of longer alcohol-related treatments, evidence is still mixed [10]. For internet interventions targeting CUD, there is no such evidence in either direction [6]. Although internet interventions based on synchronous interaction have shown to be generally effective [11], their effects compared with asynchronous-only counseling were not investigated systematically, pointing to a need of further research in this area [12]. Evidence from an earlier meta-analysis indicated no better performance of either synchronous or asynchronous mode of interaction [13].

To secure the planned changes to QTS with empirical evidence, we therefore examined whether shortening the intervention or eliminating chat-based counseling has negative impact on the program. The main outcome was cannabis use frequency during the past 30 days. Secondary outcomes were cannabis quantity, cannabis use events, cannabis dependency, treatment satisfaction, and working alliance.

Methods

Study Design

We conducted a purely Web-based pragmatic randomized experimental trial using a two-factorial design. In the study, participants were allocated to 1 of the 4 versions of QTS (Table 1). In the first experimental factor, we compared intervention versions containing chat-based counseling with versions which only consisted of counseling via asynchronous communication channels. In the second factor, program versions with the regular length of 50 days were compared with versions with the reduced duration of 28 days.

The study was conducted on the website of the intervention. Trial participants were directly recruited from all individuals who were interested in signing up for QTS. At the beginning of the program registration, potential participants were informed about the study. A PDF file containing all relevant study details was offered for download and in the confirmation email for study participants. The PDF file is included in the Multimedia Appendix 1. Individuals who were willing to participate were asked to register and provide their informed consent by checking an "I agree to participate" checkbox. The study outcomes were included in the regular registration questionnaire of QTS. Users of the intervention who opted not to participate in the study or who did not meet the eligibility criteria had full access to the regular version of the intervention and were not included in any follow-up surveys.

After registration, study participants were to choose an appointment from a schedule provided by QTS. By logging into the program at this appointment, participants were randomized automatically to 1 of the 4 intervention versions and directly forwarded into the program. Neither the researchers nor the counselors could influence or predict the randomization result. Participants were blind to the results of the randomization because they only received detailed information about the program version they were allocated to. None of the intervention versions underwent any changes during the evaluation process.

Follow-up surveys were conducted in online questionnaires 3, 6, and 12 months after randomization. Each follow up participation was compensated for with a gift voucher for a major internet-based retailer worth 10 euros.

The study was approved by the ethics committee of the Department of Applied Human Sciences at the University of Magdeburg-Stendal (Ref 4973-35) and was registered with ISRCTN (ISRCTN99818059). Note: In the registry, the study design was erroneously declared as RCT. The right indication, however, should have been randomized factorial trial.

Table 1. Experimental design.

Experimental design	Factor 2: Intervention length	
	50 days	28 days
Factor 1: Chat-based counseling		
Yes	Version 1	Version 2
No	Version 3	Version 4

Measures

Cannabis use was measured using the Timeline Followback method (TLFB) [14]. In the TLFB form, participants had to indicate the number of joints, bongs, and other cannabis use for each day in the past 30 days in a calendar. They were encouraged to use anchor events such as birthdays, appointments, or holidays to get a better orientation. The number of cannabis use days and the number of use events (ie, the sum of joints, bongs, and other cannabis use) were derived from the input from the TLFB form. To measure the cannabis quantity in grams, participants were asked to estimate this sum over the past 30 days.

Cannabis dependence was measured by the German version of the Severity of Dependence Scale (SDS) [15,16], changing the reference period from 12 to 3 months. Participants with a SDS score of at least 4 points were categorized as cannabis dependent [17]. Effects on treatment satisfaction were measured using the German version of the Client Satisfaction Questionnaire (CSQ-8) [18,19]. Data on the working alliance between counselor and participant were collected using the German version of the Working Alliance Inventory-short revised (WAI-sr) [20,21]. The CSQ-8 and WAI-sr were not listed in the study registry, as they were introduced after registering the study.

As a measure of the program usage, we tracked the number of logins and the duration of chat-based counseling for each participant.

Study Criteria

To be eligible for the study, individuals had to be at least 18 years old and to be first-time users of the intervention. Exclusion criteria were alcohol use disorder operationalized by a score of at least 3 in the CAGE [22], diagnosed psychotic or bipolar disorder based on self-report, current use of other illicit drugs on more than 4 days during the last 30 days, and suicidal thoughts as measured by selected items of Becks Scale for Suicide Ideation [23]. Individuals who displayed suicidal thoughts were given detailed information on suitable psychosocial support options such as telephone helplines or local institutions.

Interventions

Version 1 (Regular Quit the Shit)

The first intervention version of the trial was identical to the regular QTS program and therefore was 50 days long. QTS in general is based on the principles of self-regulation and self-control [24]. The weekly feedbacks are based on the solution-focused approach [25] and motivational interviewing

[26]. Therefore, the participants' responsibility to achieve personal change is accentuated, current personal developments and achievements are reviewed, and clear recommendations for further steps concerning the solution of problems are given. After the registration, the intervention comprises 3 consecutive phases:

First, admission takes place during a prescheduled appointment in a one-to-one chat with a counselor of QTS. The chat takes 50 min and is mandatory to enter QTS. It aims at clarifying the situation of the client and helps determine individual consumption goals and coping strategies. Usually, each user stays with the same counselor throughout the whole program.

Second, after the admission chat, the login area of QTS is activated. It contains a diary where participants are required to write down all relevant aspects of their cannabis use over the whole duration of 50 days. Moreover, the login area includes coping exercises, for example, aiming to develop control strategies, enhancing quality of life, or balancing the pros and cons of using cannabis. Once a week, participants receive detailed feedback by their counselor on their entries in diary and exercises. Depending on the involvement of each participant, up to 7 feedbacks are given. They discuss the current cannabis use, the psychosocial situation, and the counseling process as such.

Third, at the end of the program, clients are invited to a concluding chat, where the initial consumption goals and the applied control strategies are reflected. If necessary, participants are referred to local addiction counseling or therapy.

Version 2

Version 2 is an abbreviated variant of the original intervention and thus is only usable for 28 days. As the counselor feedback is timed weekly, version 2 therefore includes only 4 feedbacks as opposed to 7 in version 1. Apart from that, there are no differences.

Version 3

In contrast to versions 1 and 2, this variant of QTS does not offer synchronous communication between the counselor and client. Instead of starting the intervention via live chat, users of version 3 are to describe their current situation and their program goals in a self-guided tour. In open text boxes, participants are requested to outline their cannabis use, their personal pros and cons for using cannabis, and their ideas of what or who might help them to reduce their use. After concluding the self-guided tour, the login area, as described above, is activated automatically. The first counselor feedback 1 week later refers to the first-week entries in diary and to the input in the self-guided tour.

Like version 1, this variant of QTS is 50 days long and therefore provides up to 7 weekly feedbacks. Instead of a live chat at the end of the program, the last feedback is used to summarize the progress made during participation, to reconsider the working strategies, and to determine whether further professional help is required.

Version 4

Identical to version 3, version 4 does not include chat-based interaction and instead consists of the same self-guided admission procedure as described above. The only difference in comparison with version 3 is the shortened duration of 28 days. For a comparison of the intervention versions, see Table 2. A screenshot of the QTS home page is shown in Figure 1. Furthermore, a screenshot of the program diary is provided in Multimedia Appendix 2.

Statistical Analysis

Generalized estimating equations were used to examine the effects of the experimental factors on all cannabis-related study outcomes (ie, cannabis use and cannabis dependence). Treatment-related satisfaction and working alliance only were collected at the first follow-up. To measure the effects on these outcomes, we therefore used generalized linear models. In a first step of data analysis, we included both factors (Table 1) as main effects, the interaction of each factor with time, the interaction of both factors with each other, and the 3-way interaction of both factors and time in each model. Moreover, we tested whether group differences at baseline and group differences in follow-up participation moderated the factorial effects on each study outcome. If statistically significant, the respective term and its interaction with each factor were included in the models. Otherwise, it was not considered in the effectiveness testing. We assumed factorial effects on each

outcome if the interaction between the respective factor and the time variable was statistically significant. To measure the overall development for each outcome, the main effect of time was examined.

The study was powered to detect a difference between factor levels in the reduction of cannabis use frequency (ie, consumption days in the past 30 days) of at least 20%. We utilized means and SDs of an earlier trial [4] to compute the associated effect size [27]. In that trial, participants of the regular QTS intervention reduced their cannabis use by approximately 14 days (SD=12.0). We, therefore, aimed to detect a difference between 14 days and 11.2 days [14 days × (1−0.2)] reduction. To detect the corresponding effect size of $f=0.12$, a total sample of n=552 is required (n=138 for each cell of the factorial design; two-sided alpha=.05; power=0.80).

We conducted intention-to-treat analyses, including all randomized participants according to their group allocation. Missing data were estimated by multiple imputations. We performed 50 imputations. The effectiveness results of the imputed datasets were compared with the results of the nonimputed dataset (completer-only analyses).

Logistic regression analyses were conducted to compare study participants with study nonparticipants (ie, regular users of QTS) at baseline, to compare factor levels at baseline, and to determine whether baseline measures were predicting follow-up participation.

All analyses were conducted with R 3.4.1 (R Foundation for Statistical Computing, Vienna, Austria) [28], utilizing the following commands: glm [28], geeglm [29], and des [30]. Multiple imputations were estimated with R's mice package [31].

Table 2. Comparison of the interventions.

Characteristics	Version 1[a]	Version 2	Version 3	Version 4
Duration	50 days	28 days	50 days	28 days
Chat-based counseling	Yes	Yes	No	No
Registration (baseline)	Registration questionnaire	Registration questionnaire	Registration questionnaire	Registration questionnaire
Admission to intervention	Live chat (~50 min)	Live chat (~50 min)	Self-guided	Self-guided
Cannabis use diary and exercises	Up to 50 days	Up to 28 days	Up to 50 days	Up to 28 days
Weekly feedback	Up to 7	Up to 4	Up to 7	Up to 4
Conclusion of intervention	Live chat (~30 min)	Live chat (~30 min)	The last weekly feedback	The last weekly feedback

[a]Version 1 is identical to the original Quit the Shit intervention.

Figure 1. Home page of "Quit the Shit" (QTS).

Results

Flow of Participants

During the study, 876 individuals accessed the baseline questionnaire of QTS and therefore were assessed for eligibility (Figure 2). In total, 339 individuals did not take part in the study, mainly because they refused to participate (n=239). In total, 100 persons did not meet all study criteria, primarily due to problematic alcohol use (n=46), suicidal thoughts (n=34), or because they had used QTS before (n=21). The randomization of the 534 participants resulted in similar-sized study groups. In total, 252 individuals provided data at the first follow-up, 204 at the second, and 135 participants filled out the last follow-up survey 12 months after randomization, resulting in follow-up rates of 47.2%, 38.2%, and 25.3%, respectively.

Follow-up participation was significantly predicted by a higher age (OR 1.03, 95% CI 1.00-1.06, P=.02, d=0.19), lower cannabis quantity (OR 0.99, 95%-CI: 0.98-1.00, P=.004, d=0.25), higher school education (OR 1.76, 95% CI 1.35-2.30, P<.001), and higher number of logins during program participation (OR 1.05, 95% CI 1.04-1.06, P<.001, d=0.88). The allocation to either factor level, however, was no significant predictor for follow-up participation (Factor 1: OR 0.43, 95% CI 0.18-0.99, P=.05; Factor 2: OR 1.05, 95% CI 0.71-1.54, P=.82).

Sample Description

Baseline characteristics and program usage of the study participants are shown in Table 3. The majority of participants were male (65.7%) and had a high educational level with 64.7% attending or having successfully finished the highest German secondary school type ("Gymnasium") [32]. As expected, cannabis use was high with only few abstinent days during the last month.

Individuals who used the regular QTS intervention without taking part in the study (n=339, see Figure 2) had comparable values in most baseline variables. The only exception was found in the cannabis use days, which was slightly higher among individuals who were excluded from the study (OR 1.02, 95% CI 1.01-1.04, P=.01, d=0.19).

Except for a small age-related difference in factor 2 (OR 0.97, 95% CI 0.95-0.996, P=.03, d=0.20), randomization resulted in similar groups. As expected, the duration of chat-based counseling differs within factor 1 (OR 1.08, 95% CI 1.07-1.09, P<.001, d=2.72) with 105.3 min of counseling chats among participants who were in the chat-based versions of QTS. In factor 2, individuals who used the longer versions logged in significantly more often than participants allocated to the shorter versions (OR 1.01, 95% CI 1.01-1.02, P=.001, d=0.32). However, with 31.6 as opposed to 23.2 logins, that number is disproportionally low regarding the difference of 50 versus 28 days program length.

Comparison of Effectiveness

The effectiveness results reveal no significant differences between factor levels in any of the cannabis-related outcomes both in the imputed and in the nonimputed dataset (Tables 4 and 5; Multimedia Appendix 3). In working alliance and treatment satisfaction, however, significant differences with small effect sizes were found favoring chat-based communication.

None of the 3-way interactions (factor 1×factor 2×time) on the cannabis-related outcomes were significant, suggesting no relevant effectiveness differences between particular program versions (eg, versions 1 and 4). In the imputed dataset, this also applies to the two-way interactions (factor 1×factor 2) on working alliance and treatment satisfaction, which were only measured during the first follow-up (see Table 5). However, in the nonimputed dataset, we found significant two-way interactions on working alliance and treatment satisfaction (WAI: beta=−.55, CI −1.00 to −0.10, P=.02; CSQ: beta=−.41, CI −0.72 to −0.10, P=.01; Multimedia Appendix 3). The highest WAI and CSQ ratings were identified in the short version including chat-based counseling and the lowest ratings were found in the short version without chat-based counseling (see Multimedia Appendix 3).

Significant and strong time effects indicate a great overall reduction of cannabis use and use-related symptoms. The strongest reduction in the imputed dataset is found in the cannabis use days (beta=−.34, CI −0.45 to −0.23, P<.001, d=2.05) followed by the number of use events (beta=−.51, CI −0.68 to −0.34, P<.001, d=1.21; see Multimedia Appendix 3). The proportion of participants with cannabis dependence dropped from 98.5% during baseline to 78.4% (3 months), 67.0% (6 months), and 62.6% (12 months).

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of participants.
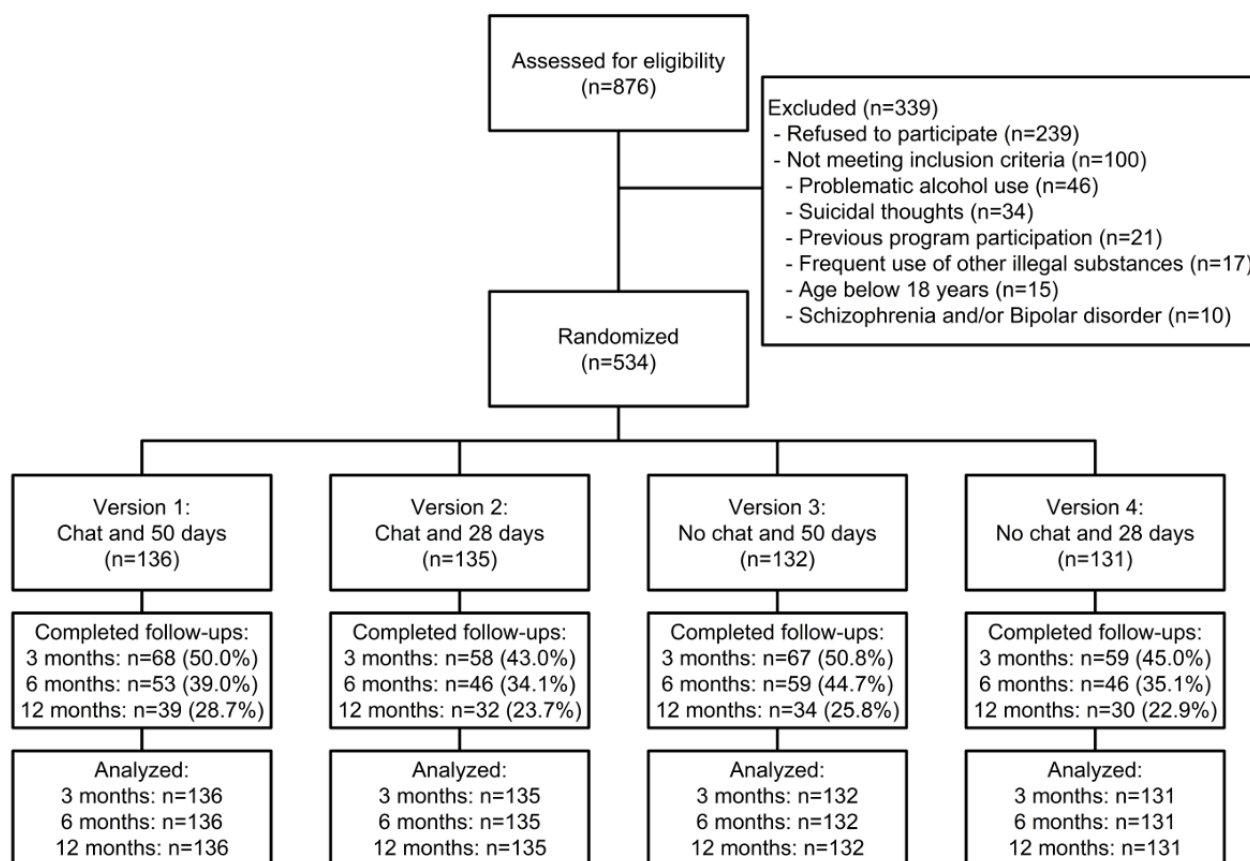
Table 3. Participant characteristics at baseline and usage of the intervention.

Characteristics	Factor 1: Chat-based communication		Factor 2: Length		All participants (n=534)
	No (n=263)	Yes (n=271)	28 days (n=266)	50 days (n=268)	
Intervention versions	3, 4	1, 2	2, 4	1, 3	1, 2, 3, 4
Gender					
Female, n (%)	85 (32.3)	98 (36.2)	91 (34.2)	92 (34.3)	183 (34.3)
Male, n (%)	178 (67.7)	173 (63.8)	175 (65.8)	176 (65.7)	351 (65.7)
Age, mean (SD)	27.5 (7.3)	27.6 (6.7)	28.2 (7.1)	26.8 (6.8)	27.5 (7.0)
Educational level, n (%)					
Basic school (Hauptschule)	25 (9.5)	29 (10.7)	30 (11.3)	24 (9.0)	54 (10.1)
Middle school (Realschule)	64 (24.3)	57 (21.0)	61 (22.9)	60 (22.4)	121 (22.7)
High school (Gymnasium)	165 (62.7)	181 (66.8)	167 (62.8)	179 (66.8)	346 (64.8)
Other school	9 (3.4)	4 (1.5)	8 (3.0)	5 (1.9)	13 (2.4)
Cannabis					
Use days[a], mean (SD)	24.7 (7.3)	25.1 (6.5)	24.9 (7.0)	24.9 (6.8)	24.9 (6.9)
Use events[a], mean (SD)	122.5 (111.6)	120.1 (104.1)	123.7 (108.9)	118.9 (106.8)	121.2 (107.7)
Quantity (grams)[a], mean (SD)	23.2 (18.8)	21.3 (18.6)	23.2 (19.6)	21.3 (17.8)	22.2 (18.7)
SDS[b,c], mean (SD)	9.9 (2.8)	10.0 (2.7)	10.1 (2.5)	9.8 (2.9)	10.0 (2.7)
SDS[b,c] cannabis dependence, n (%)	260 (98.9)	266 (98.2)	265 (99.6)	261 (97.4)	526 (98.5)
Currently no professional help, n (%)	196 (74.4)	217 (80.1)	206 (77.4)	207 (77.2)	413 (77.3)
Usage of the intervention, mean (SD)					
Number of logins	25.1 (26.4)	29.7 (27.2)	23.2 (20.8)	31.6 (31.2)	27.4 (26.9)
Chat-based counseling (min)	3.2 (30.1)	105.3 (43.8)	53.8 (59.9)	56.8 (66.9)	55.3 (63.5)

[a]During the past 30 days.

[b]SDS: Severity of Dependence Scale.

[c]Cutoff of ≥4 for cannabis dependence [17].

Table 4. Outcome scores at all 4 study points.

Characteristics	Factor 1: Chat-based communication[a]		Factor 2: Length[a]	
	No (n=263)	Yes (n=271)	28 days (n=266)	50 days (n=268)
Cannabis use[b] (days), mean (SD)				
Baseline	24.7 (7.3)	25.1 (6.5)	24.9 (7.0)	24.9 (6.8)
3 months	8.1 (9.4)	7.6 (9.6)	8.6 (9.9)	7.1 (9.1)
6 months	8.5 (9.7)	8.5 (9.9)	9.4 (10.2)	7.6 (9.3)
12 months	9.6 (10.3)	9.1 (10.3)	10.3 (10.6)	8.4 (9.9)
Cannabis use[b] (number of events), mean (SD)				
Baseline	122.5 (111.6)	120.1 (104.1)	123.7 (108.9)	118.9 (106.8)
3 months	25.2 (43.5)	20.3 (35.7)	25.3 (44.3)	20.1 (35.0)
6 months	29.1 (48.9)	25.5 (45.0)	30.4 (49.8)	24.2 (44.1)
12 months	33.9 (55.2)	29.3 (51.1)	35.2 (56.4)	27.9 (49.5)
Cannabis use[b] (grams), mean (SD)				
Baseline	23.2 (18.8)	21.3 (18.6)	23.2 (19.6)	21.3 (17.8)
3 months	6.4 (9.8)	5.0 (8.2)	6.3 (9.7)	5.2 (8.2)
6 months	6.5 (9.5)	5.5 (8.7)	6.9 (10.1)	5.1 (8.0)
12 months	7.3 (10.6)	6.8 (10.4)	7.6 (10.8)	6.5 (10.1)
Cannabis dependence (SDS[c]), mean (SD)				
Baseline	9.9 (2.8)	10.0 (2.7)	10.1 (2.5)	9.8 (2.9)
3 months	7.2 (3.5)	6.8 (3.6)	7.0 (3.5)	6.9 (3.6)
6 months	5.4 (3.5)	5.1 (3.8)	5.4 (3.6)	5.1 (3.7)
12 months	5.5 (3.6)	5.4 (3.8)	5.7 (3.6)	5.2 (3.8)
Working alliance (WAI-sr[d]), mean (SD)				
3 months	3.3 (1.0)	3.5 (0.9)	3.5 (0.9)	3.3 (1.0)
Treatment satisfaction (CSQ-8[e]), mean (SD)				
3 months	1.8 (0.7)	2.1 (0.7)	2.0 (0.6)	1.9 (0.7)

[a]Intention-to-treat analyses following multiple imputation. Results of the nonimputed datasets can be found in Multimedia Appendix 3.

[b]During the past 30 days.

[c]SDS: Severity of Dependence Scale.

[d]WAI-sr: Working Alliance Inventory-short revised.

[e]CSQ: Client Satisfaction Questionnaire.

Table 5. Group comparisons and interactions between both factors.

Characteristics	Group difference: chat no versus yes[a,b]			Group difference: 28 versus 50 days[a,b]			Interaction factor 1×factor 2×time[a,c]	
	Beta (95% CI)	P value	Effect size d (95% CI)	Beta (95% CI)	P value	Effect size d (95% CI)	Beta (95% CI)	P value
Cannabis use[d] (days), mean (SD)								
Baseline	−.02 (−0.14 to 0.10)	.74	N/A[e]	−.09 (−0.23 to 0.05)	.20	N/A	−.01 (−0.18 to 0.16)	.91
3 months	N/A	N/A	0.10 (−0.07 to 0.27)	N/A	N/A	0.15 (−0.02 to 0.32)	N/A	N/A
6 months	N/A	N/A	0.07 (−0.10 to 0.24)	N/A	N/A	0.18 (0.01 to 0.35)	N/A	N/A
12 months	N/A	N/A	0.10 (−0.07 to 0.27)	N/A	N/A	0.17 (0.00 to 0.35)	N/A	N/A
Cannabis use[d] (number of events), mean (SD)								
Baseline	−.05 (−0.25 to 0.16)	.64	N/A	−.09 (−0.31 to 0.13)	.43	N/A	−.03 (−0.30 to 0.24)	.82
3 months	N/A	N/A	0.10 (−0.07 to 0.27)	N/A	N/A	0.09 (−0.08 to 0.26)	N/A	N/A
6 months	N/A	N/A	0.05 (−0.12 to 0.22)	N/A	N/A	0.09 (−0.08 to 0.26)	N/A	N/A
12 months	N/A	N/A	0.06 (−0.11 to 0.23)	N/A	N/A	0.09 (−0.08 to 0.26)	N/A	N/A
Cannabis use[d] (grams), mean (SD)								
Baseline	.01 (−0.17 to 0.19)	.88	N/A	−.04 (−0.23 to 0.15)	.70	N/A	−.05 (−0.27 to 0.18)	.69
3 months	N/A	N/A	0.06 (−0.11 to 0.23)	N/A	N/A	0.02 (−0.15 to 0.19)	N/A	N/A
6 months	N/A	N/A	0.00 (−0.17 to 0.17)	N/A	N/A	0.09 (−0.08 to 0.26)	N/A	N/A
12 months	N/A	N/A	−0.06 (−0.23 to 0.11)	N/A	N/A	0.01 (−0.16 to 0.18)	N/A	N/A
Cannabis dependence (SDS[f]), mean (SD)								
Baseline	−.02 (−0.09 to 0.05)	.63	N/A	−.02 (−0.0 to 0.05)	.57	N/A	.01 (−0.09 to 0.10)	.91
3 months	N/A	N/A	0.14 (−0.03 to 0.31)	N/A	N/A	−0.10 (−0.27 to 0.07)	N/A	N/A
6 months	N/A	N/A	0.11 (−0.06 to 0.28)	N/A	N/A	−0.04 (−0.21 to 0.13)	N/A	N/A
12 months	N/A	N/A	0.06 (−0.11 to 0.23)	N/A	N/A	0.00 (−0.17 to 0.17)	N/A	N/A
Working alliance (WAI-sr[g]), mean (SD)								
3 months	.36 (0.09 to 0.63)	.008	0.22 (0.05 to 0.39)	−.05 (−0.32 to 0.22)	.71	−0.21 (−0.38 to −0.04)	−.27 (−0.64 to 0.09)	.15
Treatment satisfaction (CSQ-8>[h]), mean (SD)								
3 months	.33 (0.14 to 0.52)	.001	0.34 (0.17 to 0.51)	−.03 (−0.21 to 0.16)	.76	−0.17 (−0.34 to 0.00)	−.16 (−0.42 to 0.10)	.22

[a]Intention-to-treat analyses following multiple imputation. Results of the nonimputed datasets can be found in Multimedia Appendix 3.

[b]Between-group comparisons were conducted with the interaction of each factor with time, except for the effects on WAI-sr and CSQ-8 which were analyzed with the main effect of each factor.

XSL•FO
RenderX

[c] Effects on WAI-sr and CSQ-8 were analyzed with the interaction between factor 1 and factor 2.

[d] During the past 30 days.

[e] N/A: not applicable.

[f] SDS: Severity of Dependence Scale.

[g] WAI-sr: Working Alliance Inventory-short revised.

[h] CSQ-8: Client Satisfaction Questionnaire.

Discussion

Principal Findings

This study examined whether the effectiveness of the internet intervention QTS is reduced by removing chat-based counseling or by shortening the program. To our knowledge, it is the first trial designed to systematically analyze these key characteristics of guided internet interventions. According to the results, neither of these changes has meaningful impact on the effectiveness of QTS. As study participants were comparable to nonparticipants, we assume results can be generalized to regular users of QTS.

The results correspond to outcomes of meta-analyses, which also found no effects of changes in program duration on the outcomes of internet interventions for substance users [6,33,34]. In QTS, similar results of the shorter and longer program versions might be explained by a relatively fast onset of effects in the first weeks of participation and by a stabilization of these effects afterwards. Similar developments of use-related outcomes were found in another trial about an internet intervention for cannabis users [35].

The similar effects of the longer versions might also be related to a disproportionally low user engagement in these versions, a pervasive phenomenon coined as the law of attrition [36]. Therefore, the increment of received support was probably too small to significantly enhance the effects as compared with the shorter program versions.

The similar performance of the nonchat-based program versions corresponds with results of an earlier meta-analysis, which found no effectiveness difference between synchronous and asynchronous communication [13]. Our results suggest that the removal of chat-based counseling can be compensated by other elements of QTS, like the self-guided tour at the beginning or by enriching the weekly feedbacks with more information. However, extended feedbacks might not be sufficient to compensate a lack of an effective intervention element, as results from another study in this field of research suggest [37]. In contrast to our results, Schaub et al [35] found additional effects of chat counseling in a Web-based intervention for cannabis users. One key reason for the differences between their and our findings may be due to the reference conditions: the nonchat-group (active control) in the Swiss study received an automated self-help program, whereas the nonchat-conditions in our study included therapist guidance. Therefore, in our study, the chat and nonchat groups were probably more similar in terms of received support than the corresponding group in the study of Schaub and colleagues.

In contrast to the nonsignificant cannabis-related outcomes, we found a stronger working alliance and higher satisfaction ratings among users of the chat-based versions. These results are in line with outcomes of an earlier comparison study [38], supporting the assumption that direct interaction leads to a closer cooperation between the client and counselor and thus to better satisfaction ratings. These findings, therefore, should be taken into account in future modifications of QTS.

With a within-group effect size of $d=2.05$ between study baseline and the 3-month follow-up, study participants strongly reduced the frequency of their cannabis use. This effect surpasses the reductions found among QTS participants in our earlier trial (within-group $d=1.47$ for use frequency) [4,39] and also goes beyond the effects found among users of the Web-based intervention with chat counseling studied by Schaub et al (within-group $d=0.75$ for use frequency) [35] and the reductions found in the self-guided treatment examined by Rooke and colleagues (within-group $d=1.08$ for use frequency) [40]. Although within-group changes should always be interpreted carefully, the strong overall reductions in this study presumably reflect the high level of support provided by QTS.

Strengths and Limitations

We took several measures to ensure validity of the study results. We strictly adhered to the CONSORT rules, implemented a randomized factorial study design, tested and controlled potential confounders in the analyses, and compared results from the main analyses with those of completer-only analyses. Furthermore, the original intervention was already successfully tested in a randomized study [4] and is conducted by qualified staff with several years' experiences in online counseling.

As in other studies in this field of research [4,35,40,41], a major weakness is the low follow-up rate. Although we applied multiple imputations on the dataset, validity of the longer-term results is probably reduced. However, results are coherent across follow ups and, except for the significant interaction on the WAI and CSQ score, across datasets. Nevertheless, future studies in this field of research should look for ways of decreasing participant attrition. Compared with our earlier trial on the effectiveness of QTS [4], we were able to increase follow-up rates significantly. We suppose this mainly goes back to improvements in the follow up-recruitment, like offering shopping vouchers for each follow-up, emphasizing the short duration of each questionnaire, repeatedly inviting every participant for each follow-up, and addressing each nonresponder personally.

As all purely Web-based RCTs, we relied on self-reported data. This poses a threat to validity, especially for the cannabis-related outcomes, as cannabis use still is illegal in Germany and therefore might be understated by the participants. However, we deemed a biological validation of these outcomes unfeasible, as the collection of biological markers is associated with high costs, a low measurement precision, and a narrow selection of participants to those who are willing to provide these data. It

must also be noted that the outcomes "quantity of cannabis" and "number of use events" did not account for the type of cannabis product (eg, hashish, marijuana), its administration (eg, joint, bong), or its THC content. Therefore, the measurement precision of these outcomes is probably reduced. Other studies aim to improve the estimation by a standardization formula [37,40] or by gaining more detailed information on the cannabis product typically used [35]. Despite their apparent advantages, we chose not to use either of these procedures as they are not validated for application in the German setting.

Conclusions

The reduction of program length and the waiver of synchronous communication have no meaningful impact on the effectiveness of QTS. It therefore seems tenable to abbreviate the program and to offer a self-guided start into QTS. As chat-based counseling shows higher user ratings, it should be provided for those users who prefer to be supported that way.

Authors' Contributions

All authors conceived the study and participated in its design. BJ coordinated the study, the data collection, conducted the analysis, and drafted the manuscript. MT and ES carried out the modifications of QTS necessary for the study. All the authors approved the final version of the manuscript submitted for publication.

Conflicts of Interest

BJ, MT, and PT are researchers at Delphi Gesellschaft, which developed "Quit the Shit" on behalf of the BZgA. ES is the project manager for drugcom website and "Quit the Shit" at the BZgA.

Multimedia Appendix 1

Information for trial participants.

[PDF File (Adobe PDF File), 92KB - jmir_v20i5e166_app1.pdf]

Multimedia Appendix 2

Screenshot of the program diary.

[PNG File, 198KB - jmir_v20i5e166_app2.png]

Multimedia Appendix 3

Additional Tables.

[PDF File (Adobe PDF File), 55KB - jmir_v20i5e166_app3.pdf]

Multimedia Appendix 4

CONSORT‐EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 9MB - jmir_v20i5e166_app4.pdf]

References

1. European Monitoring Centre for Drugs and Drug Addiction. European Drug Report: Trends and Developments. Luxembourg: Publications Office of the European Union; 2017.
2. de Matos EG, Atzendorf J, Kraus L, Piontek D. Substanzkonsum in der Allgemeinbevölkerung in Deutschland. SUCHT 2016 Oct;62(5):271-281. [doi: 10.1024/0939-5911/a000445]
3. European Monitoring Centre for Drugs and Drug Addiction. 2014. Internet-based drug treatment (Perspectives on drugs) URL: http://www.emcdda.europa.eu/publications/pods/internet-based-drug-treatment [WebCite Cache ID 6uRrcOURO]
4. Tossmann P, Jonas B, Tensil M, Lang P, Strüber E. A controlled trial of an internet-based intervention program for cannabis users. Cyberpsychol Behav Soc Netw 2011 Nov;14(11):673-679. [doi: 10.1089/cyber.2010.0506] [Medline: 21651419]
5. Hoch E, Preuss UW, Ferri M, Simon R. Digital interventions for problematic cannabis users in non-clinical settings: findings from a systematic review and meta-analysis. Eur Addict Res 2016;22(5):233-242 [FREE Full text] [doi: 10.1159/000445716] [Medline: 27160333]

XSL•FO

RenderX

6. Tait RJ, Spijkerman R, Riper H. Internet and computer based interventions for cannabis use: a meta-analysis. Drug Alcohol Depend 2013 Dec 1;133(2):295-304. [doi: 10.1016/j.drugalcdep.2013.05.012] [Medline: 23747236]

7. European Monitoring Centre for Drugs and Drug Addiction. Treatment of cannabis-related disorders in Europe. Luxembourg: Publications Office of the European Union; 2015.

8. Rogers MA, Lemmen K, Kramer R, Mann J, Chopra V. Internet-delivered health interventions that work: systematic review of meta-analyses and evaluation of website availability. J Med Internet Res 2017 Mar 24;19(3):e90 [FREE Full text] [doi: 10.2196/jmir.7111] [Medline: 28341617]

9. Die Drogenbeauftragte der Bundesregierung. 2017. Drogen- und Suchtbericht [Drug and Addiction Report] URL: http://www.drogenbeauftragte.de/fileadmin/dateien-dba/Drogenbeauftragte/4_Presse/1_Pressemitteilungen/2017/2017_III_Quartal/Drogen-_und_Suchtbericht_2017_V2.pdf [WebCite Cache ID 6uRtk5RAj]

10. Sundström C, Blankers M, Khadjesari Z. Computer-based interventions for problematic alcohol use: a review of systematic reviews. Int J Behav Med 2017 Oct;24(5):646-658 [FREE Full text] [doi: 10.1007/s12529-016-9601-8] [Medline: 27757844]

11. Hoermann S, McCabe KL, Milne DN, Calvo RA. Application of synchronous text-based dialogue systems in mental health interventions: systematic review. J Med Internet Res 2017 Jul 21;19(8):e267 [FREE Full text] [doi: 10.2196/jmir.7023] [Medline: 28784594]

12. Dowling M, Rickwood D. Online counseling and therapy for mental health problems: a systematic review of individual synchronous interventions using chat. J Technol Hum Serv 2013 Jan;31(1):1-21. [doi: 10.1080/15228835.2012.728508]

13. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. J Technol Hum Serv 2008 Jul 03;26(2-4):109-160. [doi: 10.1080/15228830802094429]

14. Litten RZ, Allen JP, editors. Timeline follow-back: A technique for assessing self-reported alcohol consumption. In: Measuring alcohol consumption: Psychosocial and biochemical methods. Totowa, NJ, US: Humana Press; 1992.

15. 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. [Medline: 7795497]

16. Steiner S, Baumeister SE, Kraus L. Severity of Dependence Scale: establishing a cut-off point for cannabis dependence in the German adult population. SUCHT 2008;54:57-63. [doi: 10.1024/2008.07.07]

17. Martin G, Copeland J, Gates P, Gilmour S. The Severity of Dependence Scale (SDS) in an adolescent population of cannabis users: reliability, validity and diagnostic cut-off. Drug Alcohol Depend 2006 Jun 9;83(1):90-93. [doi: 10.1016/j.drugalcdep.2005.10.014] [Medline: 16310973]

18. Larsen DL, Attkisson CC, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfaction: development of a general scale. Eval Program Plann 1979 Jan;2(3):197-207. [doi: 10.1016/0149-7189(79)90094-6]

19. Schmidt J, Nübling R, Lamprecht F, Wittmann W. Patientenzufriedenheit am Ende psychosomatischer Reha-Behandlungen. Zusammenhänge mit Behandlungs- und Ergebnisvariablen und prognostische Bedeutung. Salutogenese. Ein neues Konzept in der Psychosomatik? Kongreßband der 40. Jahrestagung des Deutschen Kollegiums für Psychosomatische Medizin e-book. Frankfurt a. M. Frankfurt a. M: VAS Verlag für Akademische Schriften; 1994:271-283.

20. Horvath AO, Greenberg LS. Development and validation of the Working Alliance Inventory. J Couns Psychol 1989;36(2):223-233. [doi: 10.1037/0022-0167.36.2.223]

21. Wilmers F, Munder T, Linster H. Die deutschsprachige Version des Working Alliance Inventory - short revised (WAI-sr) - Ein schulenübergreifendes, ökonomisches und empirisch validiertes Instrument zur Erfassung der therapeutischen Allianz. Klinische Diagnostik Und Evaluation serial online 2008;1(3):343-358. [doi: 10.1002/cpp.658]

22. Ewing JA. Detecting alcoholism: The CAGE questionnaire. J Am Med Assoc 1984;252(14):1905-1907. [doi: 10.1001/jama.1984.03350140051025]

23. Beck AT, Rush AJ, Shaw BF, Emery G. Kognitive Therapie der Depression (eBook). Munich: Urban & Schwarzenberg; 1981.

24. Miller WR, Heather N, editors. Implications of a self-regulation model of therapy for treatment of addictive behaviors. In: Treating Addictive Behaviors: Processes of Change. New York: Plenum Press; 1986.

25. De Shazer S, Dolan Y, Korman H, McCollum E, Trepper T, Berg I. More Than Miracles: The State Of The Art Of Solution-Focused Brief Therapy. New York, NY, US: Haworth Press; 2007.

26. Miller WR, Rollnick S. Motivational Interviewing: Preparing People To Change Addictive Behavior. New York: The Guilford Press; 1991.

27. Cohen J. Statistical Power Analysis for the Behavioral Sciences (2nd Edition). Abingdon, United Kingdom: Routledge; 1988.

28. R-project. The R Project for Statistical Computing URL: https://www.r-project.org/ [accessed 2018-04-06] [WebCite Cache ID 6yTnyIyj7]

29. Højsgaard S, Højsgaard S, Yan J. The R package geepack for generalized estimating equations. J Stat Softw 2006;15(2):-. [doi: 10.18637/jss.v015.i02]

30. AC Del Re. 2013. compute.es: Compute Effect Sizes URL: http://cran.r-project.org/web/packages/compute.es/ [WebCite Cache ID 6yToAzKIi]

31. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. J Stat Softw 2011;45(3):1-67 [FREE Full text] [doi: 10.18637/jss.v045.i03]

32. Statistisches Bundesamt. Absolventen/Abgänger nach Abschlussart [Graduates/graduates after graduation] URL: https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/BildungForschungKultur/Schulen/Tabellen/AbsolventenAbgaenger_Abschlussart.html [WebCite Cache ID 6uRzxBVE1]

33. Rooke S, Thorsteinsson E, Karpin A, Copeland J, Allsop D. Computer-delivered interventions for alcohol and tobacco use: a meta-analysis. Addiction 2010 Aug;105(8):1381-1390. [doi: 10.1111/j.1360-0443.2010.02975.x] [Medline: 20528806]

34. Riper H, Blankers M, Hadiwijaya H, Cunningham J, Clarke S, Wiers R, et al. Effectiveness of guided and unguided low-intensity internet interventions for adult alcohol misuse: a meta-analysis. PLoS One 2014 Jun 17;9(6):e99912 [FREE Full text] [doi: 10.1371/journal.pone.0099912] [Medline: 24937483]

35. Schaub MP, Wenger A, Berg O, Beck T, Stark L, Buehler E, et al. A web-based self-help intervention with and without chat counseling to reduce cannabis use in problematic cannabis users: three-arm randomized controlled trial. J Med Internet Res 2015;17(10):e232 [FREE Full text] [doi: 10.2196/jmir.4860] [Medline: 26462848]

36. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]

37. Copeland J, Rooke S, Rodriquez D, Norberg MM, Gibson L. Comparison of brief versus extended personalised feedback in an online intervention for cannabis users: short-term findings of a randomised trial. J Subst Abuse Treat 2017;76:43-48. [doi: 10.1016/j.jsat.2017.01.009] [Medline: 28162849]

38. Cook JE, Doyle C. Working alliance in online therapy as compared to face-to-face therapy: preliminary results. Cyberpsychol Behav 2002;5(2):95-105. [doi: 10.1089/109493102753770480]

39. Hedman E, Ljótsson B, Lindefors N. Cognitive behavior therapy via the Internet: a systematic review of applications, clinical efficacy and cost-effectiveness. Expert Rev Pharmacoecon Outcomes Res 2012 Dec;12(6):745-764. [doi: 10.1586/erp.12.67] [Medline: 23252357]

40. Rooke S, Copeland J, Norberg M, Hine D, McCambridge J. Effectiveness of a self-guided web-based cannabis treatment program: randomized controlled trial. J Med Internet Res 2013 Feb 15;15(2):e26 [FREE Full text] [doi: 10.2196/jmir.2256] [Medline: 23470329]

41. Arnaud N, Baldus C, Elgán TH, De Paepe N, Tønnesen H, Csémy L, et al. Effectiveness of a web-based screening and fully automated brief motivational intervention for adolescent substance use: a randomized controlled trial. J Med Internet Res 2016;18(5):e103 [FREE Full text] [doi: 10.2196/jmir.4643] [Medline: 27220276]

Abbreviations

BZgA: Bundeszentrale für gesundheitliche Aufklärung, Federal Centre for Health Education
CUD: cannabis use disorder
CSQ-8: Client Satisfaction Questionnaire
RCT: randomized controlled trial
QTS: Quit the Shit
SDS: Severity of Dependence Scale
TLFB: Timeline Followback
WAI-sr: Working Alliance Inventory-short revised

XSL•FO
RenderX

XSL•FO

RenderX

Original Paper

Web-Based Physical Activity Intervention for Latina Adolescents: Feasibility, Acceptability, and Potential Efficacy of the Niñas Saludables Study

Britta Larsen[1], PhD; Tanya Benitez[1], MSW, PhD; Mayra Cano[1], BS; Shira S Dunsiger[2], PhD; Bess H Marcus[1,3], PhD; Andrea Mendoza-Vasconez[1], MPH; James F Sallis[1], PhD; Michelle Zive[4], RD, PhD

[1]Department of Family Medicine & Public Health, University of California, San Diego, San Diego, CA, United States

[2]Centers for Behavioral and Preventive Medicine, Department of Psychiatry and Human Behavior, Miriam Hospital, Providence, RI, United States

[3]Department of Behavioral and Social Sciences, Brown University School of Public Health, Providence, RI, United States

[4]Center for Community Health, Department of Pediatrics, University of California, San Diego, San Diego, CA, United States

Corresponding Author:
Britta Larsen, PhD
Department of Family Medicine & Public Health
University of California, San Diego
9500 Gilman Drive
San Diego, CA, 92093 0725
United States
Phone: 1 8585348429
Email: blarsen@ucsd.edu

Abstract

Background: Physical activity is markedly low in Latina adolescents, yet few physical activity interventions have been attempted in this population. Web-based interventions can incorporate theory-based components, be appealing to adolescents, and have potential for low-cost dissemination.

Objective: This study aimed to assess the feasibility, acceptability, and potential efficacy of a Web-based physical activity intervention for Latina adolescents in a single-arm pilot trial.

Methods: A total of 21 Latina adolescents (aged 12-18 years) who could read and write in English and were underactive (<90 min/week) participated in a 12-week, theory-informed Web-based physical activity intervention. The intervention website was modified from a previous Web-based intervention for Latina adults. Web content was individually tailored based on the responses to monthly questionnaires. Feasibility was measured by recruitment, retention, and adherence/engagement, and acceptability was measured by satisfaction surveys. Physical activity was measured at baseline and follow-up (12 weeks) using the 7-day physical activity recall (PAR) interview and accelerometers.

Results: Baseline activity as measured by the 7-day PAR and accelerometers was 24.7 (SD 26.11) and 24.8 (SD 38.3) min/week, respectively. At 12 weeks, 19 participants (90%, 19/21) returned. Adherence and engagement with materials were low, but 72% (15/21) of the participants indicated that they were satisfied with the intervention. Activity at 12 weeks increased by 58.8 (SD 11.33) min/week measured by the 7-day PAR ($P<.001$). Accelerometer-measured activity did not increase. Activities reported at follow-up were more varied than at baseline, including some activities measured poorly by accelerometers (eg, biking and swimming). Participants suggested simplifying the website and incorporating other technologies.

Conclusions: Good retention and increases in self-reported activity suggest a promising approach to delivering a physical activity intervention to Latina adolescents. Incorporating other technologies, such as smartphone apps, could make the intervention more engaging, acceptable, and effective.

KEYWORDS

exercise; health behavior; internet; eHealth; telemedicine

XSL•FO
RenderX

Introduction

The health benefits of physical activity for children and adolescents are extensive and include improved cardiovascular and metabolic health, decreased rates of obesity, and improved mental health [1]. To realize these health benefits, physical activity guidelines recommend that school-aged children engage in at least 60 min of moderate to vigorous physical activity (MVPA) on at least 5 days per week [2]. Rates of meeting guidelines in children in the United States, however, are quite low, particularly among adolescents. Although 42% of children aged 6 to 11 years meet guidelines based on objective measures, this drops sharply to only 8% in adolescents [2]. As health habits developed during adolescence predict health behavior and status later in adulthood [3-5], low rates of MVPA during this period could translate to insufficient MVPA for many years.

Although participation in MVPA is low during adolescence, it is especially low for girls, particularly racial/ethnic minority females. Compared with 17.9% of Mexican American boys, only 2.9% of adolescent Mexican American girls meet activity guidelines [6]. Paralleling this, Latina adults report less MVPA than non-Latino white and non-Latino black women [7] and are at higher risk of chronic diseases related to inactivity including overweight/obesity [8] and diabetes [9]. Developing effective interventions to increase MVPA in Latina adolescents is thus essential to promote health throughout the life course and reduce growing disparities.

Despite the low rates of activity and high risk of chronic disease in young Latinas, few MVPA interventions have targeted this group. Evaluations of past interventions showed that simply providing more information or more opportunities to be active was not effective in changing MVPA habits in Latino children [10,11], emphasizing a need for new theory-based and tailored interventions. Interventions grounded in psychosocial theories may be especially appropriate for increasing MVPA in youth, as self-efficacy and social support consistently emerge as the strongest correlates of MVPA in adolescents [12,13], and are consistently lower in adolescent girls than in boys [14,15]. Translating theory-based interventions into a Web-based format could be both appealing to adolescents and have potential for broad dissemination. Recent data showed that 92% of adolescents reported accessing the Internet daily, and 82% of Latino adolescents had access to a computer in their home [16].

We previously developed and tested a Web-based MVPA intervention for Latina adults (*Pasos Hacia La Salud*), which effectively increased MVPA over 6 months [17] and maintained increases 6 months later [18]. This intervention was adapted for Latinas and grounded in Social Cognitive Theory and the Transtheoretical Model, and could thus be an appropriate intervention approach for Latina adolescents. In this study, we modified this Web-based intervention based on formative research to ensure appropriateness for Latina adolescents and then tested the modified intervention in a 12-week single-arm demonstration trial. The aim of this study was to report the feasibility, acceptability, and potential efficacy of this theory-based, Web-delivered intervention for Latina adolescents.

Methods

Participants

The study sample comprised 21 Latina girls aged 12-18 years. Participants were eligible to enroll in the study if they (1) identified as Latina; (2) could read, write, and speak English fluently; (3) were physically inactive (≤90 min/week MVPA); and (4) had regular access to the Internet. Participants were ineligible for the study if they reported a health condition that would make unsupervised physical activity unsafe (according to the Physical Activity Readiness Questionnaire [19]). The study received human subjects approval from the University of California, San Diego's Institutional Review Board. All participants gave written informed assent, and parents/guardians provided consent.

Recruitment

The primary mode of recruiting participants was through health-focused community events, including health fairs, onsite school presentations, and church youth group meetings in San Diego. We also contacted adult Latinas who participated in the *Pasos Hacia La Salud* study to inquire if they had family or friends who could be eligible for the study. Other methods of recruitment included participant referrals and advertisements posted in churches, gyms, grocery stores, high schools, and parks.

Protocol Overview

After a screening interview in person or via telephone to determine eligibility, participants came to an orientation session with their mother (or other primary caregiver) to learn about the study and complete the informed consent process. Participants were given an ActiGraph GT3X+ accelerometer (ActiGraph, Pensacola, FL), with instructions to wear it the following week. One week after the orientation visit, participants returned with the accelerometer and a completed packet of psychosocial questionnaires. Certified staff then administered the 7-day physical activity recall (PAR) interview to assess self-reported physical activity. Participants then received the individually tailored intervention, including a goal setting session and access to the personalized intervention website for 12 weeks. Check-in calls were made at 1 week and 1 month. Follow-up visits were completed approximately 12 weeks after study initiation.

Tailored Intervention

The *Niñas Saludables* Web-based intervention was adapted from the *Pasos Hacia la Salud* intervention for adult Latinas. Semistructured interviews with 11 Latina girls, aged 12-19 years, were conducted to modify the content of the website. Similar to adult Latinas, the main theme across all age groups was finding time to exercise, followed by wanting support and not feeling motivated. As girls emphasized the importance of support, we included information on support throughout the website and challenged them to identify sources of support for that week's goals. Girls also interacted with and commented on the *Pasos Hacia la Salud* website to guide intervention modification and make it more appealing for this age group. Main themes in the website feedback included needing more

pictures throughout the website, particularly pictures that featured girls their age. Participants also reported wanting a more colorful template and reducing the amount of writing to provide a less cluttered look. The website was then modified to include more pictures of girls being active and a simplified homepage with highlighted shortcuts to goal setting and activity reporting. All tip sheets were redesigned to be more visual and reduce the amount of writing. We also included links to exercise videos for activities they expressed interest in, including Zumba and hip-hop dance. The website was then formatted to be mobile phone friendly, as many of the girls reported that they would only be accessing the website on their cell phones.

The first part of the intervention was a one-on-one goal setting session based on principles of motivational interviewing. A trained interventionist taught participants to set specific physical activity goals and performed guided problem-solving. Participants were given a pedometer to wear daily and were encouraged to track their steps and minutes of activity on a logging calendar on the website. All participants then received access to the *Niñas Saludables* study website for 12 weeks. The intervention was based on Social Cognitive Theory and the Transtheoretical Model and emphasized behavioral strategies for increasing activity levels (eg, goal setting, self-monitoring, and increasing social support). Intervention components included Internet-delivered activity manuals that were matched to participants' current level of motivational readiness, computer-expert system tailored reports, activity tip sheets, and a guide of local activity resources. The tailored reports and activity tip sheets were both updated to reflect activities and common barriers that girls had reported in the formative interviews, such as dancing and how to resolve time conflicts.

Participants completed monthly questionnaires on the website (see the section Measures) that generated individually tailored content: (1) stage-matched webpages provided information about MVPA that was matched to the participant's level of readiness, or stage of change, for becoming more physically active, according to the Transtheoretical Model and Social Cognitive Theory (see Figure 1), and (2) personalized computer expert system reports. The expert system draws from a bank of 330 messages addressing psychosocial and environmental factors influencing MVPA and automatically generates personalized reports on (1) their current stage of motivational readiness for physical activity; (2) increasing self-efficacy for physical activity; (3) cognitive and behavioral strategies associated with physical activity behavior change (processes of change); (4) how the participant's answers compared with their prior responses (progress feedback); (5) how the participant's responses compared with other adolescents who are physically active (normative feedback); and (6) self-monitoring of physical activity behavior (using online activity logging calendars). This expert system has been used in multiple intervention studies, and targeted theoretical constructs were shown to mediate changes in physical activity [20,21].

Figure 1. Screenshot of stage-matched moderate to vigorous physical activity information page.

In addition, participants received tip sheets addressing barriers identified by Latina adolescents, such as support from family and friends, beating boredom, and finding time to exercise. New tip sheets were made available on a weekly basis during the first month, then biweekly during months 2 and 3. Per feedback from the formative interviews, participants were also sent email reminders when a new tip sheet was available. The local activity resource guide on the website included information and links to free and low-cost activity resources in the area, such as recreation center fitness schedules and mapped walking and hiking routes.

We also developed intervention materials for mothers to help them support their daughters. On a monthly basis, mothers were sent tip sheets on topics identified in formative interviews and literature to be important for supporting MVPA behavior change in their daughters, including positive support strategies, helping daughters choose activities, and tips for getting the whole family active. Tip sheets were available in English and Spanish.

Measures

Adolescent's weight, height, and resting blood pressure were collected at baseline. Parents provided basic demographic information (income, marital status, and family size) and filled out the Brief Acculturation Scale, which assesses language use across 4 life contexts [22]. All adolescent participants were fluent in English and were all relatively highly acculturated; thus, we did not assess acculturation in adolescents.

Psychosocial and environmental access variables were used for both evaluation and intervention-tailoring purposes, and all were assessed at baseline and post-test. The PACE+Adolescent Psychosocial measure assessed motivational stage for MVPA change (3 items), MVPA change strategies (15), self-efficacy (6), decisional balance (10), family influence and support (4), peer influence and support (6), enjoyment for activity (2), and environmental access (4). Items in each domain included questions that participants rated 1-5, with options differing by domain. For example, for self-efficacy, participants responded to questions such as, "do you feel you can do physical activity even when you feel sad or stressed?" (1: I'm sure I can't, 5: I'm sure I can), or for family support, "during a typical week, how often has a member of your household done a physical activity or played sports with you?" (1: never, 5: every day). This measure has been shown to be reliable in an ethnically diverse sample of children as young as 11 years [23]. The Stages of Change for Physical Activity, Processes of Change, and Self-efficacy for Physical Activity were administered monthly on the website to generate stage-matched manuals and expert system reports. These 3 measures have been used extensively in physical activity research [24,25].

A consumer satisfaction questionnaire used in our past trials [18,26] was adapted specifically to assess the feasibility and acceptability of the current intervention.

Physical Activity Measurement

Physical activity was measured at baseline and 12 weeks using the 7-day PAR, a semistructured interview to assess the frequency, duration, and intensity of MVPA, which has been validated against objective measures and shown to be sensitive to change over time [27-29]. At baseline and follow-up, this was preceded by a 10-min treadmill walk or walk outdoors to demonstrate what MVPA should feel like. Activity was also measured objectively using ActiGraph GT3X+accelerometers. Accelerometer wear was considered sufficient if participants wore it for ≥10 hours/day on ≥5 days. After processing accelerometer files using the Choi 2008 wear time validation algorithm [30], wear time was further visually examined to determine whether the accelerometer was in fact worn during short (<3 hours) and long (≥16 hours) periods of continuous wear. Once examined and unnecessary data categorized as nonwear, the data were scored using procedures identified by Treuth et al [31], which have been validated specifically for adolescent females.

Data Analysis

Feasibility was determined by considering recruitment, retention, and adherence to the intervention at 12 weeks. For recruitment, we considered the number of people who needed to be contacted and screened to enroll 20 participants and the proportion of interested individuals who were deemed ineligible. The study was considered feasible if at least 80% of participants were retained, defined as attending the 12-week assessment. Adherence was determined by the percentage of monthly questionnaires that were completed and the percentage of materials participants reported reading. Acceptability was assessed by the satisfaction with the intervention at 12 weeks. The intervention was considered acceptable if at least 75% of participants indicated they were satisfied/very satisfied with the program.

The primary aim was to assess feasibility and acceptability; thus, we did not calculate power to detect efficacy. However, we examined changes in MVPA to explore potential efficacy. Unadjusted within-subject changes in MVPA were examined using t tests, and generalized linear models were used to assess changes in MVPA from baseline to 12 weeks (for both reported and objectively measured MVPA). Models use a likelihood-based approach to estimation and thus make use of all available data without directly imputing missing outcomes. All accelerometer data were adjusted for wear time. We tested the sensitivity of the findings to completers only and found no significant differences ($P>.05$ for sensitivity parameter). Our results present the intent-to-treat analyses with alpha level set at .05.

Results

Recruitment

A total of 50 individuals expressed interest in the study. The majority of these (n=39) came from community events, such as health fairs or school presentations. The next most successful recruitment strategy was contacting women who had recently completed the *Pasos Hacia la Salud* study (n=8). Passive recruitment, through advertisements posted in community locations, yielded only 2 calls. Another participant referred the final individual.

Of the 50 individuals who expressed interest, 30 were screened, 26 were deemed eligible, and 21 were enrolled in the study (see

Figure 2). Participant yield was lowest from health fairs, where parents often signed up on behalf of their daughters, and we were unable to contact daughters to complete screenings. Enrollment was highest from presentations at schools and other sites where adolescents were directly targeted. Reasons for ineligibility included too much activity (n=2) and medical conditions (n=2).

Baseline

Baseline characteristics are displayed in Table 1. The sample comprised 21 adolescent Latinas, aged 12-18 years (mean age=14.7, SD=2.1). Body mass index (BMI) varied greatly, ranging from underweight (BMI=17.4) to obese (BMI=33.5), with a mean BMI of 25.3. Families were generally low income, with 38% (8/21) reporting an annual family income below US $20,000. Overall, mothers had low acculturation levels, with over two-thirds reporting being the first generation in the United States, and the majority (76%, 16/21) speaking only Spanish or more Spanish than English.

Physical activity at baseline was relatively low, with a mean of 24.7 min/week of MVPA measured by the 7-day PAR (Table 1), and 24% (5/21) of girls reporting no activity at all. One outlier (4+ SDs above mean) was excluded from analysis. A large percentage of participants (28%, 6/21) had insufficient accelerometer wear time at baseline. Mean objectively measured MVPA at baseline was 24.8 (SD 38.3) in the entire sample and 21.1 (30.1) min/week among the subsample with sufficient wear time. Baseline stage of change was evenly split between contemplation (those thinking about becoming physically active) and preparation (those doing some activity, but not regularly). Of those reporting any activity at baseline, the majority (71%, 10/14) reported walking as their primary activity.

Girls reported relatively high levels of enjoyment for physical activity (mean=3.47; range 1-5), but low levels of family and friend support (2.06 and 1.84, respectively). Psychosocial variables were modestly correlated with baseline physical activity (see Table 2). The strongest correlates of baseline activity were self-efficacy (r=.34) and number of physically active friends (r=.32).

After 1 month, 81% (17/21) completed a check-in call, of which 59% (10/17) reported meeting their initial goals. The most commonly cited barrier for MVPA by far was time (70%, 12/17), followed by low motivation/energy (18%, 3/17), finding activities to do (6%, 1/17), and weather (6%, 1/17). None mentioned environmental barriers to activity. The most common source of support identified was their mothers (43% 6/14), followed by friends (29%, 4/14), other family members (21%, 3/14), and teachers (7%, 1/14).

Twelve-Week Follow-Up

Feasibility

Of the 21 participants who began the study, 19 (90%) returned for the 12-week assessment. One was placed on an indefinite medical hold due to serious illness, and one could not be contacted again (Figure 2). This exceeded the predefined threshold of 80%, suggesting good feasibility. Adherence to the intervention, however, was moderate. Participants logged on to the website on a mean of 4.29 days (SD=3.24) and visited a mean of 32.2 different pages on the website (SD=15.9). Approximately two-thirds (63%, 40/63) of monthly questionnaires were completed. Only 42% (8/19) reported reading most/all of the individually tailored reports, and 42% (8/19) reported reading at least half of the additional emailed tip sheets.

Figure 2. Recruitment and retention of participants.
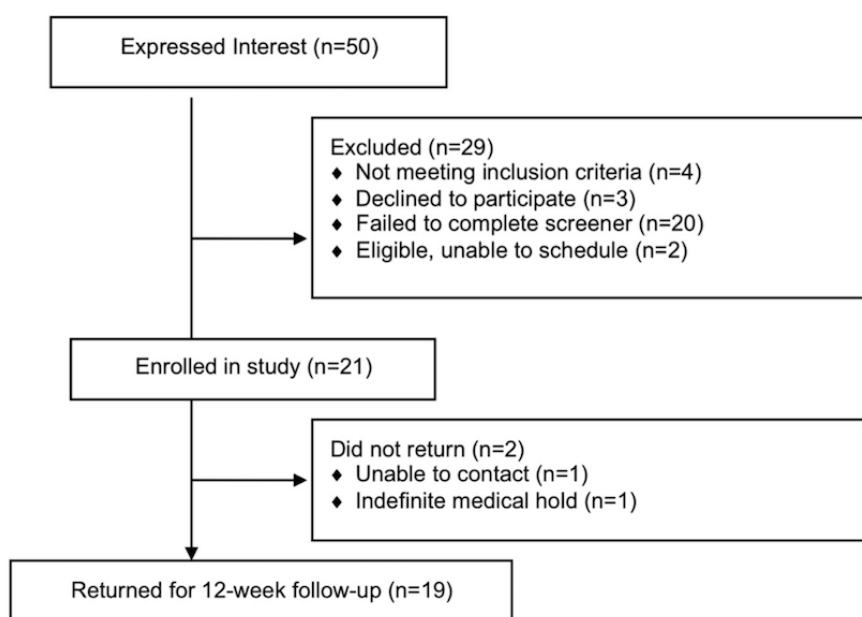
Table 1. Baseline characteristics of study sample.

Baseline characteristic	Statistics
Study sample, N	21
Age, mean (SD)	14.7 (2.1)
Body mass index, mean (SD)	25.3 (4.2)
Annual family income, n (%)	
<US $20,000	8 (38)
US $20,000-$40,000	8 (38)
>US $40,000	5 (24)
Parent's marital status, n (%)	
Married/living with partner	10 (47)
Never married	4 (20)
Separated/divorced	7 (33)
Parent's acculturation level	
Generation status, n (%)	
First generation	14 (67)
Second or third	7 (33)
Language, n (%)	
Mostly/only Spanish	16 (76)
Spanish and English	3 (14)
Mostly English	2 (10)
Number of other children living in the home, n (%)	
0	0 (0)
1	8 (38)
2	5 (24)
3+	8 (38)
Minutes/week of MVPA[a], mean (SD)	
7-day physical activity recall	24.7 (26.1)
ActiGraph GT3X+ accelerometer	24.8 (38.3)
Baseline stage of change, n (%)	
Contemplation	10 (47)
Preparation	11 (53)
Self-efficacy (1-5), mean (SD)	2.82 (0.6)
Physical activity enjoyment (1-5), mean (SD)	3.47 (0.7)
Environmental access (1-5), mean (SD)	3.39 (0.7)
Family support (1-5), mean (SD)	2.06 (0.8)
Friend support (1-5), mean (SD)	1.84 (0.9)
Number of physically active friends, n (%)	
0	0 (0)
1	6 (28)
2	6 (28)
3+	9 (44)

[a]MVPA: moderate to vigorous physical activity.

Table 2. Predictors of baseline weekly minutes of activity (7-day physical activity recall).

Baseline psychosocial measures	r
Self-efficacy	.34
Family support	.28
Friend support	.29
Number of active friends	.32
Environmental access	.01
Physical activity enjoyment	−.01

Acceptability

A total of 18 participants completed final consumer satisfaction surveys. Overall satisfaction was moderate, with 72% (13/18) of participants saying they were satisfied/very satisfied with the program, 28% (5/18) indicating they were dissatisfied, and none saying they were very dissatisfied. Similarly, 72% (13/18) said they were likely/very likely to recommend the program to family or friends. This was just below the predefined threshold of 75%. Participants rated the most useful features of the website (in order) to be (1) the goal setting calendar, (2) the activity logging calendar, and (3) the information on local activity resources. The message board and Ask the Expert forum were rated the least helpful.

The most common complaint about the program was having to log in to use the website, particularly when accessing the website through a smartphone. Participants expressed enthusiasm for expanding to other media channels, particularly a smartphone app and/or texting, and making the website more "youthful" and decreasing the amount of writing. Another common theme was to make the program more social; suggestions included enrolling friends together, incorporating teams or clubs, offering it at a school, and expanding the program to all girls, not just Latinas.

Physical Activity

Measured by the 7-day PAR, participants increased weekly minutes of MVPA from a mean 24.7 (SD 26.1) at baseline (range: 0-85) to 79.4 (SD 46.8) at follow-up (range: 14-177), with a mean increase of 58.8 (SD 46.3) min/person ($P<.001$; range: 4-155). Likelihood-based estimates showed significant increases baseline to 12 weeks ($P<.001$). All participants who completed follow-up visits reported doing some physical activity, ranging from 14 to 177 min/week. Activity types reported were more varied than at baseline and included running, walking, soccer, Zumba, swimming, and cycling. Accelerometer-measured MVPA decreased from 24.8 min/week at baseline to 10.4 (SD 30.2) at 12 weeks. Some girls who reported high activity at follow-up participated in activities in which the accelerometer could not be worn (eg, swimming or contact sports) or that were not well measured by the accelerometer (eg, cycling). Figure 3 shows total weekly minutes reported across the sample for different types of activities at baseline and 12-week follow-up.

Figure 3. Total weekly minutes reported in different activities at baseline and follow-up. PE: physical education.
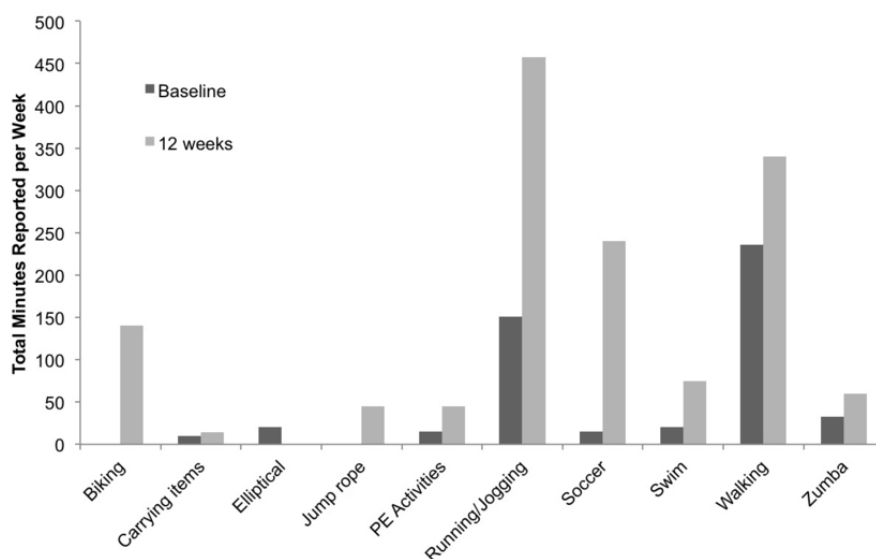
Discussion

Principal Findings

These data suggest good feasibility of the MVPA intervention for Latina adolescents, as demonstrated by the study recruitment and retention rates. Moreover, self-reported MVPA results suggest the potential efficacy of the *Niñas Saludables* intervention in increasing MVPA among Latina adolescents. These results are consistent with previous studies that have successfully used adaptations of this intervention to promote MVPA among other populations, including adult Latina women [32-34].

Although the theory-based intervention was modified based on feedback from the target population, qualitative interviews and consumer satisfaction questionnaires conducted on intervention completion showed that *Niñas Saludables* was only somewhat acceptable for Latina adolescents, and additional modifications may be necessary to increase acceptability. A recurring theme from postintervention interviews was the need to further tailor the reports and stage-matched manuals to be more age-specific and concise. Girls expressed a preference for shorter reports more often, as opposed to longer reports once per month. This was also captured by the low adherence rates, with participants reading a relatively small amount of the information provided in response to their monthly assessments. These findings are consistent with the literature: overall, Internet-based interventions for the promotion of physical activity have been found to be effective [35], yet adherence remains an obstacle that may account for small effect sizes [36]. Previous research has found that harnessing the persuasive capability of technology (using tools such as tailoring, rewards, and competition) improves adherence to Web-based interventions.

Web-based interventions will need to further capitalize on the use of technological tools, which have become increasingly popular with the rise of mobile health (mHealth) [37]. For example, use of multimedia elements like videos and images has been shown to increase understanding and improve performance in a variety of behaviors and populations [38-40], and could address the common complaint in this study of having too much writing on the website. *Niñas Saludables* participants endorsed the possibility of using newer technologies, including smartphone apps and texting, which could circumvent the need to log in to a website. Interactive features such as goal setting and activity logging were identified as the most useful features, along with the personalized reports and goal-setting session. Given our increasing capability to incorporate technology and interactive features in promoting physical activity, future research should aim to include more of these features in interventions to appeal to this specific population, while maintaining theory-based strategies to promote behavior change. Future qualitative research may be necessary to understand how to incorporate these elements and how to develop effective physical activity interventions that address the needs of Latina adolescents, and to further investigate the role of environmental access and enjoyment, which were surprisingly unrelated to MVPA in this study. To gain greater insight into the feasibility of technology-based MVPA interventions among young Latinas,

consumer satisfaction measures in future studies can also include an evaluation of social acceptability as described in the study by Poder et al [41].

Results from the 7-day PAR showed that changes in activity at 12 weeks were seen not just in quantity but also in type. Increases in self-reported MVPA in our study were comparable to those found in *Pasos Hacia la Salud* at 6 months (ie, mean increases in min/week MVPA 58.8 vs 50, respectively); however, changes in accelerometer-measured MVPA differed substantially between the 2 studies. A potential explanation for the decrease in objectively measured MVPA in *Niñas Saludables* is that the broad range of activity types that participants reported engaging in at 12 weeks may not be accurately measured using accelerometers (eg, swimming, cycling). Although activity at baseline was mostly restricted to walking, activities reported at follow-up were more varied. This could be in response to intervention materials that specifically encouraged trying new activities, including tip sheets for keeping activity interesting, tip sheets for mothers to help daughters find new activities, and the activity resource guide, which provided information on sports and activity classes in the community. This could be an important feature of interventions, as previous literature suggests individuals who enjoy activity are more likely to increase and maintain their MVPA. These findings are contrary to the results from previous studies among adult Latina women, who have mostly increased walking as a result of similar interventions [33,42], although walking was still a popular activity in our study. Our findings are thus encouraging and merit further research to understand the different results obtained among adolescent Latinas compared with adult Latinas, and whether engagement in activities other than walking may result in increased enjoyment and maintenance of MVPA. Our findings also differ from previous studies in that Latina adolescents did not endorse incorporation of Latino culture into intervention components and expressed preference for a program their non-Latina friends could also participate in. An intervention targeted to all adolescent girls could be more generalizable and prevent Latinas from feeling singled out.

Our findings also revealed challenges associated with the measurement of physical activity among Latina adolescents. Compliance with accelerometer wear was low compared with our previous studies [17]. Accelerometers may also have underestimated activity, as participants could not wear accelerometers during certain activities such as swimming or contact sports, and activities such as biking may have been inadequately captured by accelerometers. Given the high proportion of participants who reported engaging in these types of activities, accelerometry provided an incomplete measure of MVPA. Moreover, there are a wide range of suggested MVPA accelerometer cut points for adolescents, and there is no consensus on which is most appropriate [43-45]. Although the cut points used here were validated specifically for adolescent girls, the validation study found that cycling and step aerobics had poor calibration with accelerometers and found a wide range of metabolic equivalents in individuals for each level of accelerometer counts. These difficulties emphasize the importance of continuing to use validated self-report measures, both to measure quantity and type of activity, and to improve

tools of objective measurement of MVPA among this and similar populations. Girls in our study also reported a good amount of activity in 10-min bouts, which, if it were slightly less than 10 min, would not have been counted by accelerometers. Few girls participated in the treadmill walk at follow-up, which may have also increased reporting of activities that were not moderate intensity. Social desirability could have also influenced self-reported physical activity. Options for objective assessment of physical activity are expanding and include integration of geographic positioning system, accelerometer, and geographic information system data to reveal where people are active; use of heart rate monitors along with accelerometers; and devices such as ActivPAL to collect complementary data on sedentary behavior [46]. However, all these options require wearing additional devices, which increases participant burden.

Strengths and Limitations

Limitations of this study included the small sample size and single-arm design, which were appropriate for the current feasibility study but prevented formal efficacy analyses or more detailed explorations of data, including potential mediators and moderators. Given the preliminary nature and small sample size

of this pilot study, results cannot be generalized to a broader population of Latina adolescents or to locations outside the research setting. Results need to be replicated among larger samples in a randomized trial with a control group to determine whether this intervention is efficacious in helping Latina adolescents increase their MVPA. Strengths of this study included the novel, high-risk population; Web-based intervention channel; formative research; and individually tailored, theory-based intervention content.

Conclusions

The findings detailed here are promising, particularly given the paucity of research regarding the promotion of MVPA among Latina girls and adolescents and the importance of MVPA to lifelong health. These data suggest that delivering an individually tailored Web-based physical activity intervention to Latina adolescents is feasible and has potential to be efficacious, particularly when the intervention is made more acceptable by incorporating more mHealth strategies and technologies that are becoming increasingly available on a broad scale in this and other populations.

Conflicts of Interest

None declared.

References

1. Janssen I, Leblanc AG. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. Int J Behav Nutr Phys Act 2010 May 11;7:40 [FREE Full text] [doi: 10.1186/1479-5868-7-40] [Medline: 20459784]
2. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
3. Hancox RJ, Milne BJ, Poulton R. Association between child and adolescent television viewing and adult health: a longitudinal birth cohort study. Lancet 2004 Jul;364(9430):257-262. [doi: 10.1016/S0140-6736(04)16675-0] [Medline: 15262103]
4. Barnekow-Bergkvist M, Hedberg G, Janlert U, Jansson E. Adolescent determinants of cardiovascular risk factors in adult men and women. Scand J Public Health 2001 Sep;29(3):208-217. [Medline: 11680773]
5. Sacker A, Cable N. Do adolescent leisure-time physical activities foster health and well-being in adulthood? Evidence from two British birth cohorts. Eur J Public Health 2006 Jun;16(3):332-336. [doi: 10.1093/eurpub/cki189] [Medline: 16157610]
6. Whitt-Glover MC, Taylor WC, Floyd MF, Yore MM, Yancey AK, Matthews CE. Disparities in physical activity and sedentary behaviors among US children and adolescents: prevalence, correlates, and intervention implications. J Public Health Policy 2009;30(Suppl 1):S309-S334. [doi: 10.1057/jphp.2008.46] [Medline: 19190581]
7. Blackwell DL, Lucas JW, Clarke TC. Summary health statistics for U.S. adults: national health interview survey, 2012. Vital Health Stat 10 2014 Feb(260):1-161 [FREE Full text] [Medline: 24819891]
8. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of obesity among adults: United States, 2011-2012. NCHS Data Brief 2013 Oct(131):1-8 [FREE Full text] [Medline: 24152742]
9. Menke A, Casagrande S, Geiss L, Cowie CC. Prevalence of and trends in diabetes among adults in the United States, 1988-2012. JAMA 2015 Sep 08;314(10):1021-1029. [doi: 10.1001/jama.2015.10029] [Medline: 26348752]
10. Huhman M, Potter LD, Wong FL, Banspach SW, Duke JC, Heitzler CD. Effects of a mass media campaign to increase physical activity among children: year-1 results of the VERB campaign. Pediatrics 2005 Aug;116(2):e277-e284. [doi: 10.1542/peds.2005-0043] [Medline: 16061581]
11. Alhassan S, Sirard JR, Robinson TN. The effects of increasing outdoor play time on physical activity in Latino preschool children. Int J Pediatr Obes 2007;2(3):153-158. [doi: 10.1080/17477160701520108] [Medline: 17852547]

12. Sallis JF, Prochaska JJ, Taylor WC. A review of correlates of physical activity of children and adolescents. Med Sci Sports Exerc 2000 May;32(5):963-975. [Medline: 10795788]

13. Craggs C, Corder K, van Sluijs EM, Griffin SJ. Determinants of change in physical activity in children and adolescents: a systematic review. Am J Prev Med 2011 Jun;40(6):645-658. [doi: 10.1016/j.amepre.2011.02.025] [Medline: 21565658]

14. Cardon G, Philippaerts R, Lefevre J, Matton L, Wijndaele K, Balduck AL, et al. Physical activity levels in 10- to 11-year-olds: clustering of psychosocial correlates. Public Health Nutr 2005 Oct;8(7):896-903. [Medline: 16277806]

15. Biddle S, Goudas M. Analysis of children's physical activity and its association with adult encouragement and social cognitive variables. J Sch Health 1996 Feb;66(2):75-78. [Medline: 8930014]

16. Lenhart A. PewInternet. 2015 Apr 09. Teens, Social Media & Technology Overview 2015 URL: http://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/ [accessed 2018-03-29] [WebCite Cache ID 6yHxCWmOj]

17. Marcus BH, Hartman SJ, Larsen BA, Pekmezi D, Dunsiger SI, Linke S, et al. Pasos Hacia La Salud: a randomized controlled trial of an internet-delivered physical activity intervention for Latinas. Int J Behav Nutr Phys Act 2016 May 28;13:62 [FREE Full text] [doi: 10.1186/s12966-016-0385-7] [Medline: 27234302]

18. Hartman SJ, Dunsiger SI, Bock BC, Larsen BA, Linke S, Pekmezi D, et al. Physical activity maintenance among Spanish-speaking Latinas in a randomized controlled trial of an Internet-based intervention. J Behav Med 2017 Jun;40(3):392-402. [doi: 10.1007/s10865-016-9800-4] [Medline: 27752866]

19. Thomas S, Reading J, Shephard RJ. Revision of the physical activity readiness questionnaire (PAR-Q). Can J Sport Sci 1992 Dec;17(4):338-345. [Medline: 1330274]

20. Marcus BH, Dunsiger SI, Pekmezi DW, Larsen BA, Bock BC, Gans KM, et al. The Seamos Saludables study: a randomized controlled physical activity trial of Latinas. Am J Prev Med 2013 Nov;45(5):598-605 [FREE Full text] [doi: 10.1016/j.amepre.2013.07.006] [Medline: 24139773]

21. Marcus BH, Napolitano MA, King AC, Lewis BA, Whiteley JA, Albrecht A, et al. Telephone versus print delivery of an individualized motivationally tailored physical activity intervention: Project STRIDE. Health Psychol 2007 Jul;26(4):401-409. [doi: 10.1037/0278-6133.26.4.401] [Medline: 17605559]

22. Norris AE, Phillips RE, Kang HP. Evaluation of a generic brief acculturation scale in a sample of male enlisted naval personnel. Clin Nurse Spec 2008;22(4):172-178. [doi: 10.1097/01.NUR.0000311705.38404.14] [Medline: 18596485]

23. Norman GJ, Sallis JF, Gaskins R. Comparability and reliability of paper- and computer-based measures of psychosocial constructs for adolescent physical activity and sedentary behaviors. Res Q Exerc Sport 2005 Sep;76(3):315-323. [doi: 10.1080/02701367.2005.10599302] [Medline: 16270708]

24. Marcus BH, Bock BC, Pinto BM, Forsyth LH, Roberts MB, Traficante RM. Efficacy of an individualized, motivationally-tailored physical activity intervention. Ann Behav Med 1998;20(3):174-180. [doi: 10.1007/BF02884958] [Medline: 9989324]

25. Marcus BH, Owen N. Motivational readiness, self-efficacy, and decision making for exercise. J Applied Soc Psychol 1992;22:3-16.

26. Marcus BH, Dunsiger SI, Pekmezi D, Larsen BA, Marquez B, Bock BC, et al. Twelve-month physical activity outcomes in Latinas in the Seamos Saludables trial. Am J Prev Med 2015 Feb;48(2):179-182 [FREE Full text] [doi: 10.1016/j.amepre.2014.08.032] [Medline: 25442225]

27. Blair SN, Haskell WL, Ho P, Paffenbarger Jr RS, Vranizan KM, Farquhar JW, et al. Assessment of habitual physical activity by a seven-day recall in a community survey and controlled experiments. Am J Epidemiol 1985 Nov;122(5):794-804. [Medline: 3876763]

28. Rauh MJ, Hovell MF, Hofstetter CR, Sallis JF, Gleghorn A. Reliability and validity of self-reported physical activity in Latinos. Int J Epidemiol 1992 Oct;21(5):966-971. [Medline: 1468861]

29. Leenders NY, Sherman WM, Nagaraja HN, Kien CL. Evaluation of methods to assess physical activity in free-living conditions. Med Sci Sports Exerc 2001 Jul;33(7):1233-1240. [Medline: 11445774]

30. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. Med Sci Sports Exerc 2011 Feb;43(2):357-364 [FREE Full text] [doi: 10.1249/MSS.0b013e3181ed61a3] [Medline: 20581716]

31. Treuth MS, Schmitz K, Catellier DJ, McMurray RG, Murray DM, Almeida MJ, et al. Defining accelerometer thresholds for activity intensities in adolescent girls. Med Sci Sports Exerc 2004 Jul;36(7):1259-1266 [FREE Full text] [Medline: 15235335]

32. Carr LJ, Dunsiger SI, Lewis B, Ciccolo JT, Hartman S, Bock B, et al. Randomized controlled trial testing an internet physical activity intervention for sedentary adults. Health Psychol 2013 Mar;32(3):328-336 [FREE Full text] [doi: 10.1037/a0028962] [Medline: 22823069]

33. Marcus BH, Dunsiger SI, Pekmezi DW, Larsen BA, Bock BC, Gans KM, et al. The Seamos Saludables study: A randomized controlled physical activity trial of Latinas. Am J Prev Med 2013 Nov;45(5):598-605 [FREE Full text] [doi: 10.1016/j.amepre.2013.07.006] [Medline: 24139773]

34. Marcus BH, Hartman SJ, Larsen BA, Pekmezi D, Dunsiger SI, Linke S, et al. Pasos Hacia La Salud: a randomized controlled trial of an internet-delivered physical activity intervention for Latinas. Int J Behav Nutr Phys Act 2016 May 28;13:62 [FREE Full text] [doi: 10.1186/s12966-016-0385-7] [Medline: 27234302]

XSL•FO

RenderX

35.   Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012 Apr 30;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]

36.   Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]

37.   mHealth: New horizons for health through mobile technologies. Switzerland: WHO Press; 2011.

38.   Campbell MK, Honess-Morreale L, Farrell D, Carbone E, Brasure M. A tailored multimedia nutrition education pilot program for low-income women receiving food assistance. Health Educ Res 1999 Apr;14(2):257-267. [Medline: 10387505]

39.   Armstrong AW, Idriss NZ, Kim RH. Effects of video-based, online education on behavioral and knowledge outcomes in sunscreen use: a randomized controlled trial. Patient Educ Couns 2011 May;83(2):273-277. [doi: 10.1016/j.pec.2010.04.033] [Medline: 20570081]

40.   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]

41.   Poder TG, Bellemare CA, Bédard SK, Lemieux R. Social acceptance and population confidence in telehealth in Quebec. BMC Health Serv Res 2015 Feb 21;15:72 [FREE Full text] [doi: 10.1186/s12913-015-0727-1] [Medline: 25889230]

42.   Marcus BH, Hartman SJ, Larsen BA, Pekmezi D, Dunsiger SI, Linke S, et al. Pasos Hacia La Salud: a randomized controlled trial of an internet-delivered physical activity intervention for Latinas. Int J Behav Nutr Phys Act 2016 May 28;13:62 [FREE Full text] [doi: 10.1186/s12966-016-0385-7] [Medline: 27234302]

43.   Cain KL, Sallis JF, Conway TL, Van Dyck D, Calhoon L. Using accelerometers in youth physical activity studies: a review of methods. J Phys Act Health 2013 Mar;10(3):437-450. [Medline: 23620392]

44.   Atkin AJ, Ekelund U, Møller NC, Froberg K, Sardinha LB, Andersen LB, et al. Sedentary time in children: influence of accelerometer processing on health relations. Med Sci Sports Exerc 2013 Jun;45(6):1097-1104. [doi: 10.1249/MSS.0b013e318282190e] [Medline: 23274612]

45.   Logan GRM, Duncan S, Harris NK, Hinckson EA, Schofield G. Adolescent physical activity levels: discrepancies with accelerometer data analysis. J Sports Sci 2016 Nov;34(21):2047-2053. [doi: 10.1080/02640414.2016.1150599] [Medline: 26890884]

46.   Ainsworth B, Cahalin L, Buman M, Ross R. The current state of physical activity assessment tools. Prog Cardiovasc Dis 2015;57(4):387-395. [doi: 10.1016/j.pcad.2014.10.005] [Medline: 25446555]

Abbreviations

BMI: body mass index
mHealth: mobile health
MVPA: moderate to vigorous physical activity
PAR: physical activity recall

XSL•FO
RenderX

Original Paper

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

Belinda J Lawford[1], BHSc (Hons), BMSc; Rana S Hinman[1], BPhysio (Hons), PhD; Jessica Kasza[2], PhD; Rachel Nelligan[1], BPhysio; Francis Keefe[3], MS, PhD; Christine Rini[4], PhD; Kim L Bennell[1], BAppSci (Physio), PhD

[1]Centre for Health, Exercise and Sports Medicine, Department of Physiotherapy, School of Health Sciences, The University of Melbourne, Melbourne, Australia

[2]Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Australia

[3]Duke Pain Prevention and Treatment Research Program, Durham, NC, United States

[4]John Theurer Cancer Center, Department of Biomedical Research, Hackensack University Medical Center, Hackensack, NJ, United States

Corresponding Author:
Kim L Bennell, BAppSci (Physio), PhD
Centre for Health, Exercise and Sports Medicine
Department of Physiotherapy, School of Health Sciences
The University of Melbourne
161 Barry Street
Melbourne, 3010
Australia
Phone: 61 3 8344 4135
Email: k.bennell@unimelb.edu.au

Abstract

Background: Internet-delivered exercise, education, and pain coping skills training is effective for people with knee osteoarthritis, yet it is not clear whether this treatment is better suited to particular subgroups of patients.

Objective: The aim was to explore demographic and clinical moderators of the effect of an internet-delivered intervention on changes in pain and physical function in people with knee osteoarthritis.

Methods: Exploratory analysis of data from 148 people with knee osteoarthritis who participated in a randomized controlled trial comparing internet-delivered exercise, education, and pain coping skills training to internet-delivered education alone. Primary outcomes were changes in knee pain while walking (11-point Numerical Rating Scale) and physical function (Western Ontario and McMaster Universities Osteoarthritis Index function subscale) at 3 and 9 months. Separate regression models were fit with moderator variables (age, gender, expectations of outcomes, self-efficacy [pain], education, employment status, pain catastrophizing, body mass index) and study group as covariates, including an interaction between the two.

Results: Participants in the intervention group who were currently employed had significantly greater reductions in pain at 3 months than similar participants in the control group (between-group difference: mean 2.38, 95% CI 1.52-3.23 Numerical Rating Scale units; interaction *P*=.02). Additionally, within the intervention group, pain at 3 months reduced by mean 0.53 (95% CI 0.28-0.78) Numerical Rating Scale units per unit increase in baseline self-efficacy for managing pain compared to mean 0.11 Numerical Rating Scale units (95% CI –0.13 to 0.35; interaction *P*=.02) for the control group.

Conclusions: People who were employed and had higher self-efficacy at baseline were more likely to experience greater improvements in pain at 3 months after an internet-delivered exercise, education, and pain coping skills training program. There was no evidence of a difference in the effect across gender, educational level, expectation of treatment outcome, or across age, body mass index, or tendency to catastrophize pain. Findings support the effectiveness of internet-delivered care for a wide range of people with knee osteoarthritis, but future confirmatory research is needed.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12614000243617; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=365812&isReview=true (Archived by WebCite at http://www.webcitation.org/6z466oTPs)

XSL•FO
RenderX

Introduction

Knee osteoarthritis (OA) is prevalent, affecting approximately one-quarter of adults [1]. People with knee OA often experience persistent pain, impaired function, reduced quality of life, and psychological disability [2]. Education and exercise are key nonsurgical, nondrug strategies recommended by all clinical guidelines for the management of knee OA [3]. In addition, because low self-efficacy, increased pain catastrophizing, and maladaptive pain coping strategies are common among people with OA [4-6], interventions that target these psychological factors may also provide benefits. Growing evidence supports the use of pain coping skills training (PCST) to improve pain and physical and psychological functioning in chronic pain conditions [7-9]. However, many people with OA may have difficulties accessing health professionals skilled in the prescription of exercise or psychological interventions [10,11]. As such, more accessible models of health service delivery are needed.

Providing care remotely via the internet is one way to potentially improve access to treatments for people with OA. For example, programs such as Skype or FaceTime are freely available, offering potentially accessible ways in which people can consult via video with health professionals such as physical therapists. Our recent randomized controlled trial (RCT) found that an internet-delivered intervention combining online educational material, an automated PCST program, and physical therapist-prescribed exercise via Skype led to greater improvements in pain and function in people with knee OA compared to online educational material alone [12]. These improvements were also maintained at long-term follow-up (9 months). These findings are consistent with evidence in other populations, where care delivered via videoconferencing for people who have undergone knee joint replacement surgery led to improvements in physical activity and function that were similar to those after face-to-face consultations [13]. Collectively, these studies provide support for the effectiveness of telerehabilitation as an alternative model of service delivery.

When evaluating novel models of service delivery, it is important to consider whether there are subgroups of patients that respond more or less favorably. Moderation analysis of RCT data can be used to determine if the effect of a treatment relative to the control differs for distinct subgroups of patients. In these analyses, baseline characteristics that interact with the treatment to influence clinical outcomes are identified as moderators [14]. As such, the aim of moderation analysis is to determine whether groups of people with different characteristics respond differently to the treatment. Identifying moderators of the effect of a treatment on an outcome can help match patients to the most effective and appropriate treatment thereby improving outcomes, in turn helping to control treatment costs and promoting efficient use of resources. In addition, moderation analyses can help guide future refinements to an intervention

so that it may be more effective for subgroups who do not respond as well.

To our knowledge, few previous studies have conducted moderator analyses on outcomes following an intervention involving exercise or PCST for people with OA. There is some evidence that pain coping style, expectation of treatment response, radiographic disease severity, age, and educational level may moderate the effects of a face-to-face PCST intervention on treatment outcomes among people with hip and knee OA [15]. Another study found that body mass index (BMI) moderated the effects of a supervised aquatic exercise intervention on changes in quality of life among people with OA [16]. Although a number of studies have identified predictors of outcomes following exercise for people with OA (eg, self-efficacy for managing pain [17], gender [18], and age [19]), outcome predictors are not necessarily the same as moderators of the effect of a treatment. For example, participants with a particular characteristic may improve over time no matter what treatment they receive and, although that characteristic might predict improvement, it may not necessarily moderate the effect of a treatment. Thus, predictors of outcome do not allow identification of subgroups that respond, or do not respond, to a given intervention. Therefore, the aim of this study was to explore potential demographic and clinical moderators of the effect of a combined internet-delivered exercise, education, and PCST program on changes in pain and physical function for people with knee OA. This will provide future directions for confirmatory studies.

Methods

This study involved exploratory moderation analyses using data from a parallel, two-group pragmatic RCT (Australian New Zealand Clinical Trials Registry: ACTRN12614000243617) aiming to evaluate the effectiveness of internet-delivered physical therapist-prescribed home exercise, education, and PCST compared to internet-delivered education alone (IMPACT trial). Study procedures were approved from the University of Melbourne Human Research Ethics Committee and all participants provided written informed consent. The trial protocol [20] and outcomes [12] have been published.

Study Population

A total of 148 people with chronic knee pain were recruited Australia-wide to participate in the RCT. Briefly, inclusion criteria included age 50 years or older, knee pain for more than 3 months and on most days of the previous month, knee pain during walking in the previous week (≥4 on an 11-point Numerical Rating Scale [NRS]), mild to moderate physical dysfunction (>20 out of 68 on the physical function subscale of the Western Ontario and McMaster Universities Osteoarthritis Index [WOMAC]), and having an active email account and computer with internet access.

Intervention

Participants in the intervention arm of the trial received three internet-delivered treatments: (1) educational material about exercise and physical activity, pain management, emotions, healthy eating, complementary therapies, and medications (freely available on the Arthritis Australia website [21]); (2) an online interactive automated PCST program [9,22], involving completion of one 35 to 45 minute training module per week for 8 weeks; and (3) seven physical therapist consultations via Skype over 12 weeks, with each consultation lasting 30 to 45 minutes. Participants were randomly allocated to one of eight different physical therapists, who delivered all subsequent Skype sessions. Physical therapists performed a brief assessment and prescribed a lower-limb strengthening home exercise program to be completed by the participant three times per week. Participants were also encouraged to increase their physical activity levels and were given the opportunity to use a pedometer for motivation if desired, and were also encouraged to practice pain coping skills daily.

The control group received access to the same educational material as the intervention group (ie, material about exercise and physical activity, pain management, emotions, healthy eating, complementary therapies, and medications through the Arthritis Australia website), but did not have access to the PCST program.

Dependent Variables

Participants completed questionnaires online at baseline, 3 months, and 9 months. Primary outcomes were valid and reliable measures of pain and function [23]. Pain during walking over the last week was measured using an 11-point NRS ranging from 0 ("no pain") to 10 ("worst pain possible"). Function was measured using the WOMAC physical function subscale [24] with scores ranging from 0 (no dysfunction) to 68 (maximum dysfunction). Thus, we calculated the 3-month change in pain and function as baseline minus 3-month values, and 9-month change in pain and function as baseline minus 9-month values.

Selected Moderators

Based on previous research [15-19] and/or theoretical plausibility (Multimedia Appendix 1), we investigated whether the effect of the internet-delivered treatment relative to the control treatment was moderated by each of the following baseline variables: gender, age, level of education (dichotomized as no tertiary training or some tertiary training), employment situation (dichotomized as employed or not employed), pain self-efficacy using the pain subscale of the Arthritis Self-Efficacy Scale [25] (scores ranging from 1 to 10, with higher scores indicating greater self-efficacy), pain catastrophizing using the Pain Catastrophizing Scale [26] (scores ranging from 0 to 52, with higher scores indicating greater catastrophizing), expectation of treatment effects using a five-point Likert scale ranging from "no effect" to "complete recovery" (dichotomized into "no effect to moderate improvement" or "large improvement to complete recovery"), and BMI based on self-reported height and weight.

Analysis

To determine whether the effect of the treatment relative to the control on change in each of the primary outcomes was moderated by the selected baseline variables, separate linear regression models were fit with the potential moderator variable and study group as covariates, and an interaction between the two. This analysis provides an answer to the question of whether the effect of the treatment, relative to the control condition, differs for different subgroups of participants. For binary moderators, the estimated effect of treatment and a 95% confidence interval was determined for each of the moderator levels. For continuous moderators, results were calculated as the effect of a one-unit increase of that moderator in each of the control and intervention groups. The linear regression assumptions of linearity, heteroscedasticity, and normality were assessed using standard diagnostic plots. Scatterplots of the outcome against each continuous moderator by treatment group were examined to determine if more complex models (including nonlinear terms for moderators) were warranted. All statistical analyses were performed using Stata version 14.1 (StataCorp LLC, College Station, TX, USA).

Results

A total of 148 people with knee OA were enrolled in the study, with just over half female (56.1%, 83/148) and currently employed either full-time or part-time (57.4%, 85/148; Table 1). At baseline, participants in the intervention group had higher educational levels than those in the control group. At 9 months, nine people in the intervention group and seven people in the control group had been lost to follow-up (unable to contact, family issues/illness, deceased).

Moderators of the Effect of the Intervention on Change in Walking Pain

There was no evidence for moderation of the effect of the intervention on change in walking pain at 3 or 9 months by most binary variables, including gender, level of education, and treatment expectations (Table 2). There was some evidence for moderation of the treatment effect by employment status for change in walking pain at 3 months (interaction P=.02). Among those who were currently employed, participants assigned to the intervention group had greater reductions in pain than those in the control group, with an estimated difference between groups of mean 2.38 (95% CI 1.52-3.23) NRS units. Among unemployed participants, the estimated difference in reduction in pain at 3 months between the intervention and control groups was mean 0.86 (–0.13 to 1.85) NRS units.

Table 1. Baseline descriptive characteristics (N=148).

Characteristic	Intervention (n=74)	Control (n=74)
Age (years), mean (SD)	60.8 (6.5)	61.5 (7.6)
Body mass index (kg/m^2), mean (SD)	32.0 (13.9)	30.1 (10.2)
Gender (female), n (%)	43 (58)	40 (54)
Level of education, n (%)		
No tertiary training	16 (22)	24 (32)
Some tertiary training	58 (78)	50 (68)
Employment status, n (%)		
Employed	40 (54)	45 (61)
Unemployed	34 (46)	29 (39)
Expectation of treatment outcomes, n (%)		
No effect to moderate improvement	11 (15)	21 (29)
Large improvement to complete recovery	63 (85)	52 (71)
Self-efficacy (pain)[a]	6.1 (1.8)	5.9 (1.8)
Pain catastrophizing[b]	8.8 (9.2)	10.1 (9.6)

[a]ASES: Arthritis Self-Efficacy Scale (range 1-10; higher scores indicate greater self-efficacy).

[b]PCS: Pain Catastrophizing Scale (range 0-52; higher scores indicate greater catastrophizing).

Scatterplots of changes in pain against each continuous moderator (Multimedia Appendices 2 and 3) indicated that exploration of linear terms for each moderator was sufficient, and more complex models were not warranted. Results of linear models for each of the continuous moderators are presented in Figures 1 and 2. There was no evidence that age, pain catastrophizing, and BMI interacted with treatment group in the model for changes in walking pain at 3 or 9 months (Table 3). However, there was evidence that self-efficacy for managing pain interacted with treatment group at 3 months (P=.02). Within the intervention group, for each additional unit of self-efficacy (pain) at baseline, pain at 3 months reduced by an additional mean 0.53 (95% CI 0.28-0.78) NRS units, whereas in the control group, the association between self-efficacy (pain) and change

in pain was estimated as mean 0.11 (95% CI –0.13 to 0.35) NRS units. That is, people with higher self-efficacy for managing pain at baseline demonstrated greater reductions in walking pain with the intervention compared to control.

Moderators of the Effect of the Intervention on Change in Physical Function

Scatterplots of changes in function against each continuous moderator (Multimedia Appendices 2 and 3) indicated that exploration of linear terms for each moderator was sufficient. Results of linear models are presented in Figures 1 and 2. None of the selected binary (Table 2) or continuous (Table 3) moderators had a significant interaction with the treatment in the model for changes in physical function at 3 or 9 months.

Table 2. Results of the moderation analysis for binary moderators for change in walking pain and physical function.

Moderator	3 months		9 months	
	Intervention-control difference (95% CI)[a]	Interaction (P value)	Intervention-control difference (95% CI)[b]	Interaction (P value)
Change in walking pain				
Gender		.64		.45
Male	1.88 (0.85, 2.90)		0.74 (–0.47, 1.94)	
Female	1.56 (0.69, 2.42)		1.33 (0.31, 2.36)	
Level of education		.22		.58
No tertiary training	0.96 (–0.29, 2.21)		0.57 (–0.97, 2.12)	
Some tertiary training	1.87 (1.10, 2.64)		1.08 (0.18, 1.97)	
Treatment expectation		.70		.85
No effect-moderate improvement	1.76 (1.01, 2.51)		1.08 (0.22, 1.94)	
Large improvement/complete recovery	1.45 (0.06, 2.85)		1.26 (-0.41, 2.93)	
Employment		.02		.86
Not employed	0.86 (–0.13, 1.85)		1.06 (–0.13, 2.25)	
Employed	2.38 (1.52, 3.23)		1.20 (0.17, 2.22)	
Change in physical function				
Gender		.43		.88
Male	7.80 (2.46, 13.14)		6.87 (1.63, 12.12)	
Female	10.63 (6.11, 15.15)		7.42 (2.94, 11.90)	
Level of education		.22		.25
No tertiary training	5.66 (–0.88, 12.20)		3.25 (–3.46, 9.96)	
Some tertiary training	10.44 (6.41, 14.48)		7.75 (3.86, 11.64)	
Treatment expectation		.73		.99
No effect-moderate improvement	9.77 (5.87, 13.69)		7.25 (3.46, 11.04)	
Large improvement/complete recovery	8.35 (1.05, 15.65)		7.27 (–0.12, 14.65)	
Employment		.14		.81
Not employed	6.88 (1.74, 12.01)		6.72 (1.50, 11.93)	
Employed	11.94 (7.48, 16.41)		7.57 (3.05, 12.08)	

[a]Baseline-3 months.

[b]Baseline-9 months.

Figure 1. Difference in change in NRS walking pain and WOMAC function (baseline-3 months) between treatment groups for each potential continuous moderator. Negative values favor the intervention group. Solid line indicates the difference between the control and intervention arms. Shaded areas indicate 95% confidence intervals. ASES: Arthritis Self-Efficacy Scale (range 1-10; higher scores indicate greater self-efficacy); NRS: Numerical Rating Scale (range 0-10; lower scores indicate less pain); PCS: Pain Catastrophizing Scale (range 0 to 52; higher scores indicate greater catastrophizing); WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index; ranges from 0 to 68, where lower scores indicate better function.
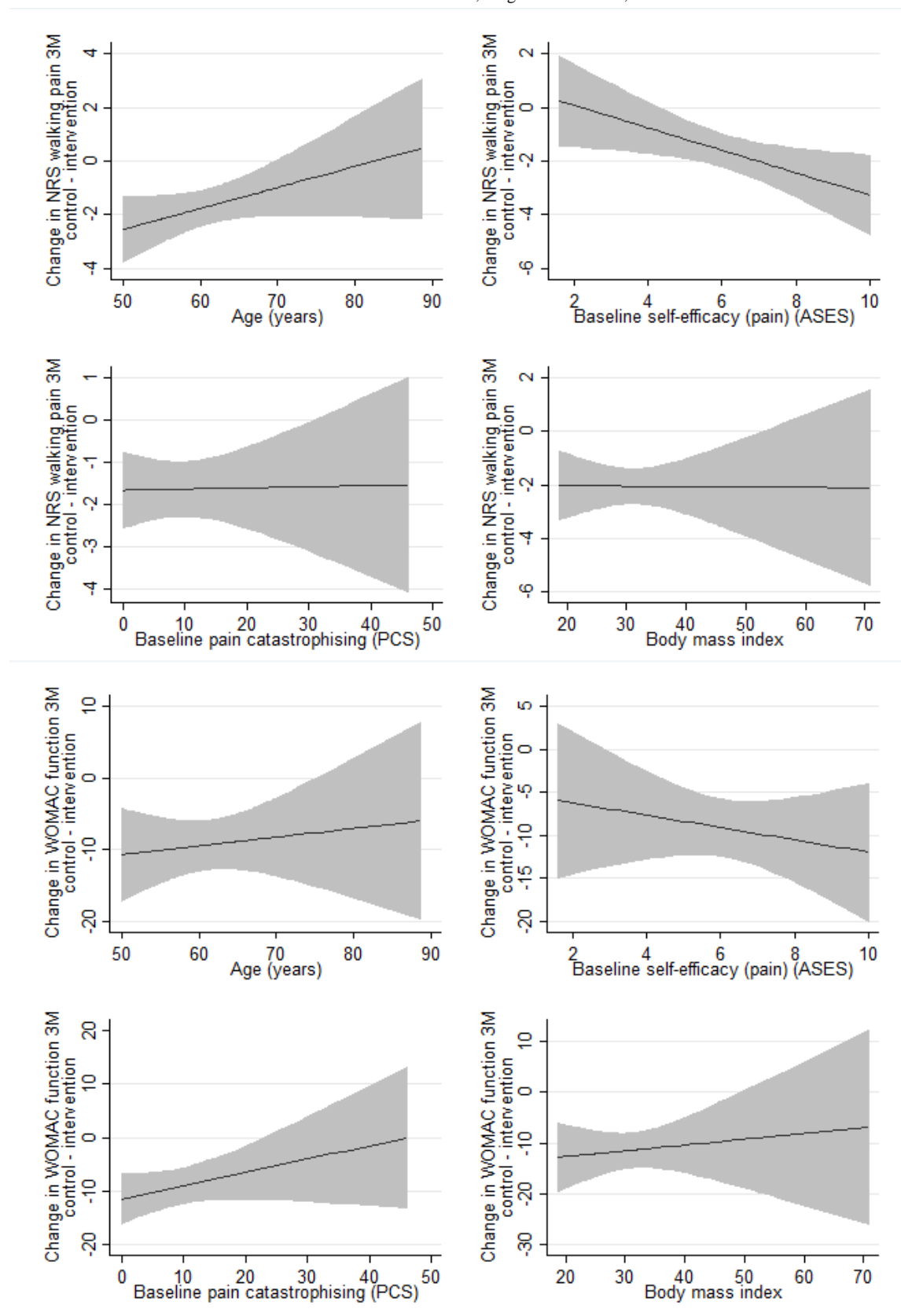
Figure 2. Difference in change in NRS walking pain and WOMAC function (baseline–9 months) between treatment groups for each potential continuous moderator. Negative values favor the intervention group. Solid line indicates the difference between the control and intervention arms. Shaded areas indicate 95% confidence intervals. ASES: Arthritis Self-Efficacy Scale (range 1-10; higher scores indicate greater self-efficacy); PCS: Pain Catastrophizing Scale (range 0-52; higher scores indicate greater catastrophizing); WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index (range 0-68, lower scores indicate better function).
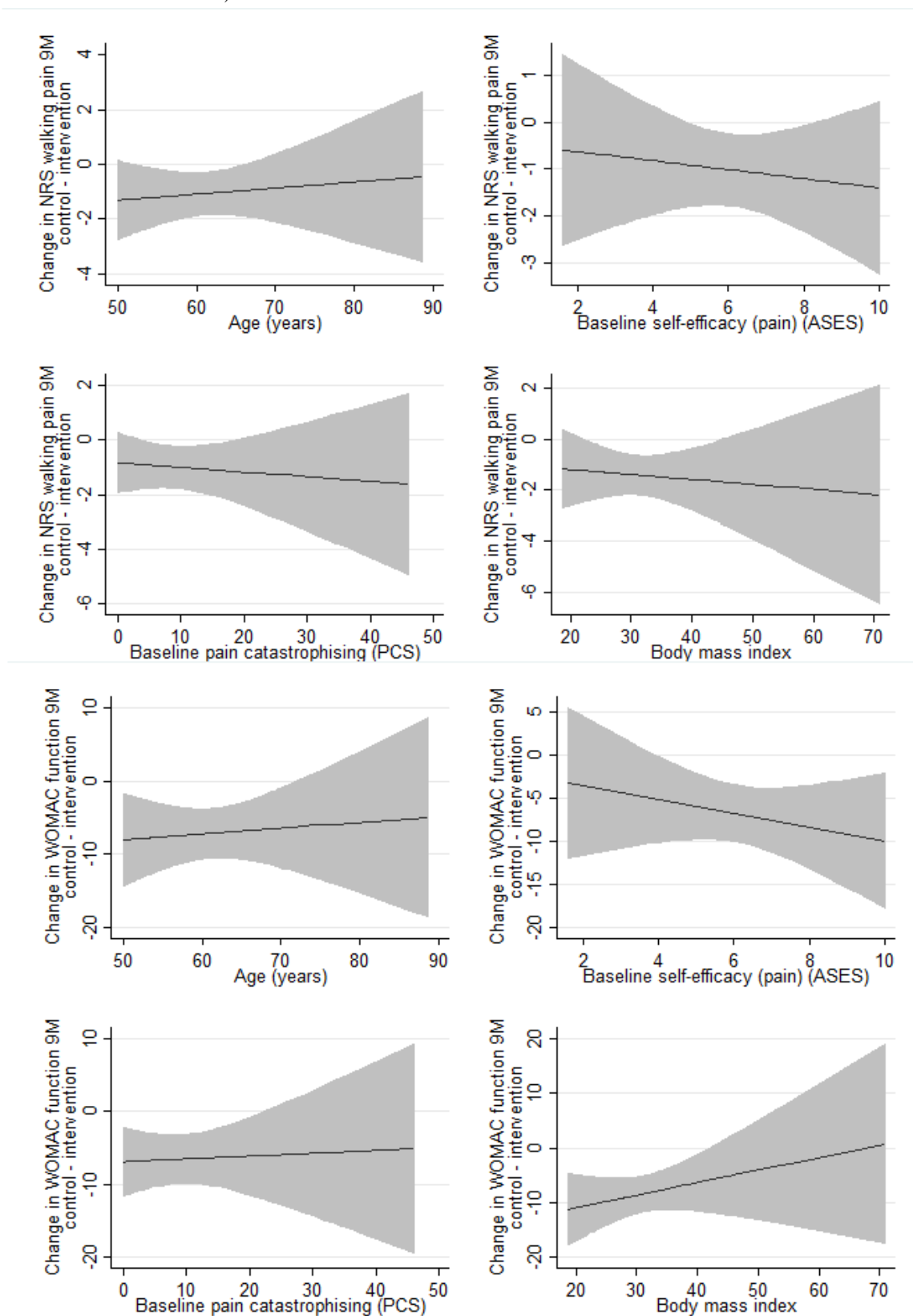
Table 3. Results of the moderation analysis for continuous moderators for change in walking pain and physical function.

Moderator	3 months			9 months		
	Estimated moderator coefficient (95% CI)		Interaction (P value)	Estimated moderator coefficient (95% CI)		Interaction (P value)
	Control group	Intervention group		Control group	Intervention group	
Change in walking pain						
Age (years)	0.02 (–0.04, 0.08)	–0.06 (–0.13, 0.01)	.10	–0.01 (–0.08, 0.06)	–0.03 (–0.12, 0.05)	.69
Self-efficacy (pain)[a]	0.11 (–0.13, 0.35)	0.53 (0.28, 0.78)	.02	0.21 (–0.08, 0.51)	0.31 (0.00, 0.61)	.66
Pain catastrophizing[b]	–0.04 (–0.09, 0.01)	–0.04 (–0.09, 0.01)	.94	0.01 (–0.01, 0.03)	0.02 (–0.00, 0.04)	.49
Body mass index	–0.02 (–0.09, 0.04)	–0.02 (–0.09, 0.04)	.98	–0.07 (–0.15, 0.01)	–0.05 (–0.12, 0.02)	.72
Change in physical function						
Age (years)	–0.08 (–0.40, 0.23)	–0.21 (–0.58, 0.17)	.62	–0.01 (–0.32, 0.31)	–0.09 (–0.46, 0.29)	.74
Self-efficacy (pain)[a]	0.96 (–0.35, 2.26)	1.67 (0.32, 3.03)	.45	0.88 (–0.40, 2.17)	1.68 (0.36, 3.00)	.39
Pain catastrophizing[b]	–0.14 (–0.39, 0.11)	–0.39 (–0.64,–0.13)	.17	–0.03 (–0.13, 0.07)	0.04 (–0.05, 0.14)	.28
Body mass index	–0.08 (–0.44, 0.28)	–0.19 (–0.52, 0.14)	.64	–0.04 (–0.38, 0.30)	–0.27 (–0.59, 0.05)	.32

[a]ASES: Arthritis Self-Efficacy Scale (range 1-10; higher scores indicate greater self-efficacy).

[b]PCS: Pain Catastrophizing Scale (range 0-52; higher scores indicate greater catastrophizing).

Discussion

Principal Findings

The aim of this study was to explore potential demographic and clinical moderators of the effect of an internet-delivered exercise, education, and PCST program for people with knee OA on changes in pain and physical function. We found that the effect of the intervention at 3 months differed by employment status and self-efficacy, such that the effect of the intervention on reductions in pain was greater among those who were employed and had higher self-efficacy for managing their pain at baseline. There was no evidence that any of the selected baseline variables moderated the effect of the intervention on changes in physical function at 3 months, and no evidence that any moderated the effect of the intervention on changes in pain or function at 9 months.

We found some evidence that employment status moderated the effect of the intervention on changes in walking pain at 3 months. Those who were currently employed had greater improvements in walking pain with the treatment compared to the control, whereas among those who were not employed there was no difference in changes on walking pain at 3 months between the intervention and control groups. Wright and colleagues [19] found that employment status did not significantly predict changes in pain or function after an exercise and/or manual therapy intervention for people with hip OA. However, they looked at predictors, rather than moderators, and used seven different categories to define employment status (ranging from "homemaker" to "full-time or part-time employment") with a small number of participants in each group, which may have limited the study's power. Being employed exposes people to a range of psychosocial environments and experiences (eg, structured and meaningful use of time, opportunities to use new and existing skills, variety in tasks, and social contact outside of the home) that have been linked to decreases in emotional distress within chronic pain populations, independent of pain severity [27-29]. People who are employed may also be more motivated to improve so that they can continue working; furthermore, being employed has been associated with higher self-efficacy among people with chronic musculoskeletal pain [30,31], which may have contributed to greater reduction in pain with treatment at 3 months. However, there was no evidence that employment status moderated treatment effects on changes in pain at 9 months. The reasons for this are not clear, but support the longer-term effectiveness of this intervention for people who are employed or unemployed.

To our knowledge, no previous research has investigated self-efficacy as a potential moderator of an exercise or PCST intervention on outcomes in people with chronic pain conditions. Skou and colleagues [17] found that self-efficacy for managing pain at 3 months predicted pain and quality-of-life outcomes at 1 year among people with knee or hip OA who received an education and exercise intervention. However, they did not investigate whether baseline self-efficacy also predicted outcomes and only looked at predictors of outcomes, rather than moderators of the effect of treatment. According to Social Cognitive Theory [32], self-efficacy influences a person's choice of effort and persistence in the face of adversity (eg, knee pain) [33]. As such, our findings make intuitive sense in that people who are more confident in their ability to succeed despite the presence of pain are more likely to experience greater improvements with a treatment program that emphasizes self-management in the form of home exercise and utilizing pain coping skills. However, there was no evidence that self-efficacy for managing pain moderated intervention effects on changes in pain at 9 months. Although the reasons for this are not clear, the results support the longer-term effectiveness of the intervention for people with either high or low self-efficacy at baseline.

In contrast to previous studies exploring moderators of PCST [15] and exercise [16], we found no evidence that age, BMI, expectation of treatment outcome, pain coping style, and education moderated the effect of the intervention relative to the control on changes in pain and function. In the literature, there appears to be some heterogeneity in moderators of treatment effects following exercise or PCST interventions for people with chronic pain conditions [34], which might be because of differences in intervention design or delivery across studies. Our intervention was unique in that it was remotely delivered and combined both physical and psychological interventions. Other studies involved face-to-face supervised exercise programs [16,35] or 10 sessions of PCST delivered face-to-face by nurse practitioners [15]. The exercise interventions of other studies also differed to ours, in which they investigated moderators of aquatic exercise for people with OA [16] or a combined intervention of dietary weight loss and exercise (combining aerobic and resistance training) [35]. In addition, our sample comprised fewer females who were also, on average, younger than those of previous studies [15,16,35]. These differences in delivery mode and frequency/type of exercise/PCST, as well as heterogeneity in sample characteristics and measures of outcomes or moderator variables, might explain the differences between our findings and the limited existing literature.

To our knowledge, this is the first study to explore moderators of the effect of internet-delivered care on outcomes for people with a chronic musculoskeletal condition. We found no evidence that age or level of education moderated the effect of internet-delivered care on changes in pain or function at 3 or 9 months. Similarly, previous studies found that age had no association with outcomes following an intervention involving behavioral treatment delivered via Skype for veterans with posttraumatic stress disorder [36,37]. Our recent survey also found that, among people with knee and/or hip OA, there was no evidence that age and level of education influenced interest in receiving exercise from a physical therapist via video [38]. These findings counter commonly held misconceptions of telehealth, namely that it may be most suitable for people who are younger and/or highly educated, and are therefore "technology savvy" [39,40]. Our findings suggest that people of varying ages, and those with high or low levels of education, benefited just as much with internet-delivered care. This study thus provides further support for an increased use of telerehabilitation as an alternative model of service delivery for people with OA.

Our findings have a number of practical implications. Internet-delivered education, exercise, and PCST appears to be effective for a range of people with OA, including both men and women with varying educational backgrounds, of different ages, as well as those who are obese or normal weight. This suggests that the intervention does not need to be targeted at specific subgroups of patients. For greater improvements in pain at 3 months, or if services or resources are limited, service providers or physical therapists could consider using measures to screen for people with low self-efficacy for managing pain as a means of identifying potentially vulnerable patients who might need some additional support before receiving this kind of intervention or an alternative such as face-to-face care.

Strengths and Limitations

This study has a number of strengths, including the robust study design (moderation/subgroup analysis using RCT data), the range of potential moderators included, the use of meaningful patient-reported clinical outcomes of pain and physical function, and the fact that we recruited participants from regional/remote and metropolitan areas across Australia. This study also has some limitations. Because the analyses in our study were exploratory in nature, our results should be interpreted with caution. For example, it is possible that some of the interactions in our analysis were nonsignificant due to lack of power rather than the absence of a significant effect. As such, further research is required to confirm our findings. Our intervention included multiple components (ie, Skype-delivered exercise in addition to PCST and online educational materials), so it is not possible to determine which treatment component interacts with each of the moderator variables. In addition, more than 75% of the people in the intervention arm of the study had completed some tertiary education and, therefore, our results may not be generalizable to people who are less educated.

Conclusions

In summary, people who were employed and had higher self-efficacy at baseline were more likely to experience greater improvements in pain at 3 months after an internet-delivered exercise, education, and PCST program. This may be because people who are more confident in their ability to succeed despite the presence of pain are more likely to experience greater improvements with a treatment program that emphasizes self-management. There was no evidence of a difference in the effect across gender, educational level, expectation of treatment outcome, or across age, BMI, or tendency to catastrophize pain. This study provides further support for telerehabilitation as an alternative model of service delivery that is suitable for a broad range of people with OA. Further research is required to confirm our findings and identify moderators of the effect of exercise and PCST on long-term changes in pain and function, and also identify moderators of the effects of other modes of service delivery (eg, telephone-delivered care).

Authors' Contributions

The authors declare the following contributions to the preparation of the manuscript: study conception and design (KLB, RSH, FK, CR), inclusion and data collection (RN), data analysis (JK) and interpretation of data (all authors), drafting of the manuscript (BJL), and critical revision of the manuscript (all authors). All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Overview of selected demographic and clinical moderators.

[PDF File (Adobe PDF File), 28KB - jmir_v20i5e10021_app1.pdf]

Multimedia Appendix 2

Change in NRS walking pain and WOMAC function (baseline-3 months) against each continuous potential moderator, by treatment group.

[PNG File, 47KB - jmir_v20i5e10021_app2.png]

Multimedia Appendix 3

Change in NRS walking pain and WOMAC function (baseline-9 months) against each continuous potential moderator, by treatment group.

[PNG File, 46KB - jmir_v20i5e10021_app3.png]

References

1. 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]

2. Murphy LB, Sacks JJ, Brady TJ, Hootman JM, Chapman DP. Anxiety and depression among US adults with arthritis: prevalence and correlates. Arthritis Care Res (Hoboken) 2012 Jul;64(7):968-976 [FREE Full text] [doi: 10.1002/acr.21685] [Medline: 22550055]

3. Nelson AE, Allen KD, Golightly YM, Goode AP, Jordan JM. A systematic review of recommendations and guidelines for the management of osteoarthritis: The chronic osteoarthritis management initiative of the U.S. bone and joint initiative. Semin Arthritis Rheum 2014 Jun;43(6):701-712. [doi: 10.1016/j.semarthrit.2013.11.012] [Medline: 24387819]

4. Scopaz KA, Piva SR, Wisniewski S, Fitzgerald GK. Relationships of fear, anxiety, and depression with physical function in patients with knee osteoarthritis. Arch Phys Med Rehabil 2009 Nov;90(11):1866-1873 [FREE Full text] [doi: 10.1016/j.apmr.2009.06.012] [Medline: 19887210]

5. Sale JE, Gignac M, Hawker G. The relationship between disease symptoms, life events, coping and treatment, and depression among older adults with osteoarthritis. J Rheumatol 2008 Feb;35(2):335-342. [Medline: 18203312]

6. Keefe FJ, Bonk V. Psychosocial assessment of pain in patients having rheumatic diseases. Rheum Dis Clin North Am 1999 Feb;25(1):81-103. [Medline: 10083960]

7. Dixon K, Keefe FJ, Scipio CD, Perri LM, Abernethy AP. Psychological interventions for arthritis pain management in adults: a meta-analysis. Health Psychol 2007 May;26(3):241-250. [doi: 10.1037/0278-6133.26.3.241] [Medline: 17500610]

8. Bennell K, Ahamed Y, Jull G, Bryant C, Hunt MA, Forbes AB, et al. Physical therapist-delivered pain coping skills training and exercise for knee osteoarthritis: randomized controlled trial. Arthritis Care Res (Hoboken) 2016 May;68(5):590-602 [FREE Full text] [doi: 10.1002/acr.22744] [Medline: 26417720]

9. Rini C, Porter LS, Somers TJ, McKee DC, DeVellis RF, Smith M, et al. Automated Internet-based pain coping skills training to manage osteoarthritis pain: a randomized controlled trial. Pain 2015 May;156(5):837-848 [FREE Full text] [doi: 10.1097/j.pain.0000000000000121] [Medline: 25734997]

10. Papandony M, Chou L, Seneviwickrama M, Cicuttini FM, Lasserre K, Teichtahl AJ, et al. Patients' perceived health service needs for osteoarthritis (OA) care: a scoping systematic review. Osteoarthritis Cartilage 2017 Jul;25(7):1010-1025. [doi: 10.1016/j.joca.2017.02.799] [Medline: 28232144]

11. Dobson F, Bennell K, French SD, Nicolson PJ, 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]

12. Bennell K, Nelligan R, Dobson F, Rini C, Keefe F, Kasza J, et al. Effectiveness of an Internet-delivered exercise and pain-coping skills training intervention for persons with chronic knee pain: a randomized trial. Ann Intern Med 2017 Apr 04;166(7):453-462. [doi: 10.7326/M16-1714] [Medline: 28241215]

13. Shukla H, Nair SR, Thakker D. Role of telerehabilitation in patients following total knee arthroplasty: evidence from a systematic literature review and meta-analysis. J Telemed Telecare 2017 Feb;23(2):339-346. [doi: 10.1177/1357633X16628996] [Medline: 26843466]

14. 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. [Medline: 12365874]

15. Broderick J, Keefe FJ, Schneider S, Junghaenel DU, Bruckenthal P, Schwartz JE, et al. Cognitive behavioral therapy for chronic pain is effective, but for whom? Pain 2016 Dec;157(9):2115-2123. [doi: 10.1097/j.pain.0000000000000626] [Medline: 27227692]

16. Cadmus L, Patrick MB, Maciejewski ML, Topolski T, Belza B, Patrick DL. Community-based aquatic exercise and quality of life in persons with osteoarthritis. Med Sci Sports Exerc 2010 Jan;42(1):8-15. [doi: 10.1249/MSS.0b013e3181ae96a9] [Medline: 20010135]

17. Skou ST, Simonsen ME, Odgaard A, Roos EM. Predictors of long-term effect from education and exercise in patients with knee and hip pain. Dan Med J 2014 Jul;61(7):A4867. [Medline: 25123117]

18. French H, Galvin R, Cusack T, McCarthy GM. Predictors of short-term outcome to exercise and manual therapy for people with hip osteoarthritis. Phys Ther 2014 Jan;94(1):31-39. [doi: 10.2522/ptj.20130173] [Medline: 23929827]

19. Wright A, Cook CE, Flynn TW, Baxter GD, Abbott JH. Predictors of response to physical therapy intervention in patients with primary hip osteoarthritis. Phys Ther 2011 Apr;91(4):510-524. [doi: 10.2522/ptj.20100171] [Medline: 21310898]

20. Dobson F, Hinman RS, French S, Rini C, Keefe F, Nelligan R, et al. Internet-mediated physiotherapy and pain coping skills training for people with persistent knee pain (IMPACT - knee pain): a randomised controlled trial protocol. BMC Musculoskelet Disord 2014 Aug 13;15:279 [FREE Full text] [doi: 10.1186/1471-2474-15-279] [Medline: 25125068]

21. Arthritis Australia. 2018. URL: http://www.arthritisaustralia.com.au/ [accessed 2018-04-15] [WebCite Cache ID 6yhxpo9ju]

22. Rini C, Porter LS, Somers TJ, McKee DC, Keefe FJ. Retaining critical therapeutic elements of behavioral interventions translated for delivery via the Internet: recommendations and an example using pain coping skills training. J Med Internet Res 2014 Dec 19;16(12):e245 [FREE Full text] [doi: 10.2196/jmir.3374] [Medline: 25532216]

23. McAlindon T, 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]

24. 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]

25. 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. [Medline: 2912463]

26. Osman A, Barrios FX, Gutierrez PM, Kopper BA, Merrifield T, Grittmann L. The Pain Catastrophizing Scale: further psychometric evaluation with adult samples. J Behav Med 2000 Aug;23(4):351-365. [Medline: 10984864]

27. Jackson T, Iezzi A, Lafreniere K, Narduzzi K. Relations of employment status to emotional distress among chronic pain patients: a path analysis. Clin J Pain 1998 Mar;14(1):55-60. [Medline: 9535314]

28. Jackson T, Iezzi A, Lafreniere K. The impact of psychosocial features of employment status on emotional distress in chronic pain and healthy comparison samples. J Behav Med 1997 Jun;20(3):241-256. [Medline: 9212379]

29. Jackson T, Wang Y, Fan H. Associations between pain appraisals and pain outcomes: meta-analyses of laboratory pain and chronic pain literatures. J Pain 2014 Jun;15(6):586-601. [doi: 10.1016/j.jpain.2014.01.499] [Medline: 24524845]

30. Rahman A, Ambler G, Underwood MR, Shipley ME. Important determinants of self-efficacy in patients with chronic musculoskeletal pain. J Rheumatol 2004 Jun;31(6):1187-1192. [Medline: 15170934]

31. Jackson T, Wang Y, Wang Y, Fan H. Self-efficacy and chronic pain outcomes: a meta-analytic review. J Pain 2014 Aug;15(8):800-814. [doi: 10.1016/j.jpain.2014.05.002] [Medline: 24878675]

32. Bandura A. Human agency in social cognitive theory. Am Psychol 1989 Sep;44(9):1175-1184. [Medline: 2782727]

33. Rejeski W, Ettinger WH, Martin K, Morgan T. Treating disability in knee osteoarthritis with exercise therapy: a central role for self-efficacy and pain. Arthritis Care Res 1998 Apr;11(2):94-101. [Medline: 9668732]

34. McCracken LM, Turk DC. Behavioral and cognitive-behavioral treatment for chronic pain: outcome, predictors of outcome, and treatment process. Spine (Phila Pa 1976) 2002 Nov 15;27(22):2564-2573. [doi: 10.1097/01.BRS.0000032130.45175.66] [Medline: 12435995]

35. Rejeski WJ, Focht BC, Messier SP, Morgan T, Pahor M, Penninx B. Obese, older adults with knee osteoarthritis: weight loss, exercise, and quality of life. Health Psychol 2002 Sep;21(5):419-426. [Medline: 12211508]

XSL•FO

RenderX

36. Price M, Gros DF. Examination of prior experience with telehealth and comfort with telehealth technology as a moderator of treatment response for PTSD and depression in veterans. Int J Psychiatry Med 2014;48(1):57-67. [doi: 10.2190/PM.48.1.e] [Medline: 25354926]

37. Gros DF, Price M, Yuen EK, Acierno R. Predictors of completion of exposure therapy in OEF/OIF veterans with posttraumatic stress disorder. Depress Anxiety 2013 Nov;30(11):1107-1113. [doi: 10.1002/da.22207] [Medline: 24151141]

38. Lawford B, Bennell KL, Hinman RS. Consumer perceptions of and willingness to use remotely delivered service models for exercise management of knee and hip osteoarthritis: a cross-sectional survey. Arthritis Care Res (Hoboken) 2017 May;69(5):667-676. [doi: 10.1002/acr.23122] [Medline: 27813360]

39. Kayyali R, Hesso I, Mahdi A, Hamzat O, Adu A, Nabhani GS. Telehealth: misconceptions and experiences of healthcare professionals in England. Int J Pharm Pract 2017 Jun;25(3):203-209. [doi: 10.1111/ijpp.12340] [Medline: 28261891]

40. Bradford N, Caffery LJ, Smith AC. Awareness, experiences and perceptions of telehealth in a rural Queensland community. BMC Health Serv Res 2015 Sep 28;15:427 [FREE Full text] [doi: 10.1186/s12913-015-1094-7] [Medline: 26416464]

41. Somers T, Keefe FJ, Pells JJ, Dixon KE, Waters SJ, Riordan PA, et al. Pain catastrophizing and pain-related fear in osteoarthritis patients: relationships to pain and disability. J Pain Symptom Manage 2009 May;37(5):863-872 [FREE Full text] [doi: 10.1016/j.jpainsymman.2008.05.009] [Medline: 19041218]

42. Edwards RR, Bingham CO, Bathon J, Haythornthwaite JA. Catastrophizing and pain in arthritis, fibromyalgia, and other rheumatic diseases. Arthritis Rheum 2006 Apr 15;55(2):325-332 [FREE Full text] [doi: 10.1002/art.21865] [Medline: 16583384]

Abbreviations

ASES: Arthritis Self-Efficacy Scale
BMI: body mass index
NRS: Numerical Rating Scale
OA: osteoarthritis
PCST: pain coping skills training
RCT: randomized controlled trial
WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index

XSL•FO
RenderX

Original Paper

Increasing Physical Activity in Mothers Using Video Exercise Groups and Exercise Mobile Apps: Randomized Controlled Trial

Maya Nina Mascarenhas[1,2], MPH, PhD; June Maylin Chan[2,3], ScD; Eric Vittinghoff[2], PhD; Erin Lynn Van Blarigan[2,3], ScD; Frederick Hecht[1], MD

[1]Osher Center for Integrative Medicine, University of California, San Francisco, San Francisco, CA, United States

[2]Department of Epidemiology & Biostatistics, University of California, San Francisco, San Francisco, CA, United States

[3]Department of Urology, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Maya Nina Mascarenhas, MPH, PhD
Osher Center for Integrative Medicine
University of California, San Francisco
1545 Divisadero Street, 4th Floor
San Francisco, CA,
United States
Phone: 1 415 890 2878
Email: Maya.Mascarenhas@ucsf.edu

Abstract

Background: Women significantly decrease their activity levels in the transition to motherhood. Digital health technologies are low cost, scalable, and can provide an effective delivery mechanism for behavior change. This is the first study that examines the use of videoconferencing and mobile apps to create exercise groups for mothers.

Objective: The aim of the study was to test the feasibility, acceptability, and effectiveness of an individually adaptive and socially supportive physical activity intervention incorporating videoconferencing and mobile apps for mothers.

Methods: The Moms Online Video Exercise Study was an 8-week, 2-armed, Web-based randomized trial comparing the effectiveness of a group exercise intervention with a waitlist control. Healthy mothers with at least 1 child under the age of 12 years were recruited through Facebook and email listservs. Intervention participants joined exercise groups using videoconferencing (Google Hangouts) every morning on weekdays and exercised together in real time, guided by exercise mobile apps (eg, Nike+, Sworkit) of their choice. Waitlist control participants had access to recommended mobile apps and an invitation to join an exercise group after the 8-week study period. Main outcomes assessed included changes in self-reported moderate, vigorous, and moderate to vigorous physical activity (MVPA) minutes per week in aggregate and stratified by whether women met Centers for Disease Control and Prevention guidelines for sufficient aerobic activity at baseline. Outcomes were measured through self-assessed Web-based questionnaires at baseline and 8 weeks.

Results: The intervention was effective at increasing exercise for inactive women and proved to be feasible and acceptable to all participants. A total of 64 women were randomized, 30 to intervention and 34 to control. Women attended 2.8 sessions per week. There was a strong, but not statistically significant, trend toward increasing moderate, vigorous, and MVPA minutes for all women. As hypothesized, in the prespecified stratum of women who were inactive at baseline (n=51), intervention participants significantly increased their activity by an average of 50 (95% CI 4.0-95.9, P=.03) MVPA minutes per week more than control participants. They had a corresponding statistically significant net increase of 19 (95% CI 3.2-34.8, P=.02) minutes of vigorous activity. Inactive women in the intervention arm also experienced promising reductions in depression, reporting a statistically significant net decrease in their depression score (−3.8, 95% CI −7.0 to −0.6; P=.02).

Conclusions: We found that a group exercise intervention using videoconferencing and mobile apps was a feasible and acceptable way to deliver a physical activity intervention to mothers. The intervention increased physical activity in inactive mothers. Further studies are needed to better establish how long these changes in physical activity can be maintained and whether these findings can be reproduced in a more diverse population.

Trial Registration: ClinicalTrials.gov NCT02805140; https://clinicaltrials.gov/ct2/show/NCT02805140 (Archived by WebCite at http://www.webcitation.org/6yYZwRveg)

XSL•FO
RenderX

Introduction

Background

Despite strong evidence of the health benefits of physical activity and decades of efforts to increase activity levels, almost half of the United States adult population fails to meet Centers for Disease Control and Prevention (CDC) exercise guidelines of 150 minutes of moderate or 75 minutes of vigorous exercise per week, and 70% of the population fails to meet the biweekly muscle strengthening guidelines [1-6]. One group with unique challenges to being sufficiently active is women with young children. Women significantly reduce their activity levels in the transition to motherhood [7-9]. Mothers are less likely to be active than fathers, women of the same age who do not have children, and compared with their own activity levels before having children [7]. The proportion of hours per week that mothers with young children are physically active has decreased by 14 hours per week in the past 45 years, whereas sedentary activities such as watching television and driving have increased by 6 hours per week [10]. This decrease in physical activity is not only a concern for the health of mothers but also for their potential impact on their children. Active mothers have a positive influence on the activity levels of their children [11-13]. In addition, when mothers exercise, they report being able to better manage the demands of raising children [14-17]. Due to mothers' unique needs and risks, it is important that we design appropriate interventions to help mothers be more physically active.

Mothers experience a wide range of barriers to exercising including isolation, a lack of leisure time, lack of social support, lack of child care, lack of spousal support, and the need to put family obligations ahead of themselves [16-19]. Reviews suggest that 2 elements of effective physical activity interventions that can help overcome such barriers include (1) adapting to individual needs and (2) incorporating community-based social support [20,21]. Individually adaptive interventions are able to tailor to individuals' needs, preferences, and contexts. Social support interventions often draw upon support and accountability that individuals within a network can provide one another. These elements have each been tested successfully in physical activity interventions with mothers, but their combined impact is not known [7]. Unfortunately, individually adaptive and group physical activity interventions can be costly and complicated to deliver, and in-person groups can be particularly inconvenient for mothers.

Digital technology interventions represent a convenient, cost-effective, and scalable delivery mechanism for providing socially supportive and individually adaptive interventions [22]. In the United States, 77% of the adult population owns a mobile phone, and this proportion continues to increase rapidly [23]. More than half of downloaded apps are in the health and fitness domain, yet few exercise apps incorporate evidence-based content [24-28]. Mothers, in particular, are heavy users of technology, and thus represent an important group to test evidence-based technology interventions [29]. Technology interventions have a growing evidence base for being effective at increasing activity, though this research is in its early stages [30-33]. Additionally, videoconferencing tools such as Google Hangouts and Skype have been tested for exercise video coaching but not as a way to bring participants, and mothers specifically, together for real-time exercise video groups [34].

Objectives

In this study, we assessed the feasibility and acceptability and estimated the effectiveness of a group physical activity intervention that incorporated videoconferencing and exercise mobile apps. This intervention relied on providing evidence-based elements of social support and individualization to increase physical activity in mothers.

Methods

Study Design

The Moms Online Video Exercise (MOVE) Study was an 8-week, 2-armed, parallel, Web-based randomized trial comparing the effectiveness of an intervention arm consisting of exercise groups that used videoconferencing and mobile apps with a waitlist control arm. We detailed our methods below and in an eHealth checklist (Multimedia Appendix 1).

Recruitment

We recruited participants using advertisements that included a link to our study website in parent-specific Facebook groups and email listservs. Participants were recruited from all over the country, though the recruiting efforts and time zones available were targeted to the West Coast. In addition to email and Facebook advertisements, all recruited participants were asked to share the advertisement with any relevant email listservs or Facebook groups and any individuals they thought might be interested. Once on the study website, women were able to sign up for an introductory phone call in which study staff reviewed study procedures and consent forms using DocuSign (DocuSign, California, USA) before enrollment began. Recruitment efforts took place between July 2016 and November 2016. Before recruitment, we received approval from the University of California, San Francisco Institutional Review Board (14-15344), and registered our trial with the Clinical Trials Registry (NCT02805140).

Participants

Our eligibility criteria stipulated that women needed to be between the ages of 18 and 60 years, speak and understand English, be able to give consent, and have at least 1 child under the age of 12 years. Enrolled women could not be pregnant or plan on being pregnant during the study period. Participants

XSL•FO
RenderX

were also required to have access and understand how to operate 2 devices, one with videoconferencing capacity and one with mobile app capacity. These devices could include cell phones, computers, and smart tablets. Participants had to be capable of exercising safely, which was assessed using the validated Physical Activity Readiness Questionnaire [35]. We targeted women who were inactive (did not meet CDC physical activity guidelines), but hypothesized that even physically active women with young children might benefit from the intervention, and hence included them, but planned a priori to analyze them separately.

Protocol

Women who were eligible for the study were asked to complete one introductory phone call, a baseline survey, and a practice group exercise video session to be randomized. Informed consent was obtained by study staff in the introductory phone call, and consent forms were electronically signed during or after the call. In practice group video sessions, participants signed into Google hangouts, introduced themselves, and then opened up a mobile app to complete a short workout using the Johnson & Johnson mobile app 7-minute workout routine [36]. Participants who confirmed their continued interest in participating in the study and who completed a baseline survey were randomized to the intervention or waitlist control. All participants were provided with access to a list of recommended mobile exercise apps. Women randomized to the intervention were additionally assigned to a video exercise group at a time of their preference and provided an exercise prescription. The exercise prescription over 8 weeks for intervention participants consisted of 5 weekday video exercise sessions lasting between 5 and 30 minutes and varying in type (interval training, dance, yoga, etc) and intensity (low to high) depending on the participant's choice of mobile app and associated routine for each session. Adherence to this prescription was monitored via self-report, and staff support was provided if needed via email. After the 8-week study period, all participants were asked to fill out end of study surveys. Women in the intervention arm were then given the option of continuing for an additional 8 weeks, whereas those in the waitlist control were invited to join an exercise group for 8 weeks. The main analysis included only data from the 8-week study period, during which the waitlist control participants were not participating in video sessions.

Before randomization, women were asked to pick a morning time slot that they could attend every weekday for 8 weeks. We offered exercise group time slots on the half hour from 6:00 AM to 9:30 AM Pacific Standard Time Zone (PST). Exercise group sizes ranged from 2 to 5 participants. We enrolled participants over a period of 5 months. Groups grew over time as new participants enrolled and those from the waitlist arm joined groups after their 8-week waiting period. Group sessions lasted no more than a total of 30 minutes, beginning with a check-in lasting up to 5 minutes. Participants had an individualized website that contained a link to their respective Google Hangouts videoconferencing group calls and a tracking form that they filled out before each session (Multimedia Appendix 2). Women usually did their workouts while remaining on video to provide accountability (passive monitoring) and support (solidarity from working out

simultaneously), much like having a "gym buddy." We recommended freely available mobile apps and YouTube exercise videos routines that were updated on our study website as the study progressed (Multimedia Appendix 3). Participants were also encouraged to find exercise mobile apps and videos that were not on the list. Women were encouraged to individualize their choice of workout during a session, so participants in groups were often performing a wide range of workouts simultaneously. One of the study goals was to assess whether the hypothesized benefits in accountability and support would still be obtained while providing individual choice of workout routines. In addition to group exercise sessions, participants were connected to their group members via email primarily so they could communicate about planned or unplanned absences. Study staff monitored attendance via tracking forms that were filled out at each session and reached out to participants who had missed more than a week of workout sessions to check in over email.

Randomization

Participants were randomized using parallel arms, equal allocation (1:1), and block randomization (random block sizes of 2 and 4 participants). The randomization was stratified on the participant's morning time slot of choice and the participant's baseline activity status, a binary variable of whether they met CDC guidelines of 150+ moderate to vigorous physical activity (MVPA) minutes per week. We stratified on these 2 factors to address potential confounding by baseline activity status and to ensure evenly sized intervention and control arms within time strata. Our statistician generated a stratified block random sequence using Stata 14 (StataCorp, Texas, USA) and stored it in Research Electronic Data Capture (REDCap), a secure, Web-based database application hosted at the University of California, San Francisco [37]. The sequence was concealed from the primary investigator who used REDCap to reveal the computer-assigned randomization once participants were enrolled. The assignment was not blinded to investigators or participants.

Measures

Study Measurement Procedure

During the recruitment phase, participants filled out a screening survey to establish eligibility. Once eligibility was confirmed by study staff, participants were asked to complete a self-assessed baseline survey that included primary and secondary outcomes. At 8 weeks, all trial participants were asked to fill out self-assessed questionnaires with the same outcomes. Mothers who were randomized to the intervention arm were asked additional evaluation questions. All surveys were Web-based and completed online using Qualtrics (Qualtrics, Utah, USA) software.

Physical Activity

We assessed our primary outcome of physical activity using a self-assessed validated questionnaire, the Active Australia Survey [38,39]. Participants reported the frequency and duration of the past 7 days of activity in the following categories: walking (for at least half a mile), moderate activity (makes you breathe harder than normal), and vigorous activity (makes you sweat,

out of breath). MVPA minutes per week were calculated by the sum of vigorous minutes multiplied by a factor of 2 plus the number of moderate minutes. The Active Australia Survey has good reliability and good validity compared with accelerometry and was found to be responsive to change in clinical trials [40,41]. Furthermore, it has been used in a number of physical activity trials with mothers [42,43].

Secondary Measures and Study Evaluation

We collected a self-report of weight. We assessed psychosocial measures specific to physical activity, which included social support for physical activity and physical activity self-efficacy [42,44,45]. We used Patient-Reported Outcome Measurement Information System (PROMIS) short form measures for anxiety, sleep disturbance, depression, and fatigue, and converted summary scores into standardized T-scores [46]. PROMIS T-scores are a standardized score based on a mean of 50 and a standard deviation of 10 using the reference population of a sample of the 2000 General Census [47-50]. We assessed participant adherence by monitoring their session attendance per week throughout their 8-week participation. Adherence took into account holiday weeks; the rate for the week excluding the holiday was applied to the whole holiday week. Acceptability was assessed through survey evaluation questions administered to participants in the intervention arm at the end of the study.

Statistical Analysis

We used an intention-to-treat analysis. We analyzed all women who completed baseline and 8-week surveys (complete cases) according to their randomization status. On the basis of our a priori hypothesis that inactive women would benefit most from the study, we analyzed results for all women who completed 8-week surveys, followed by an analysis stratified by whether women met CDC aerobic guidelines (150+ minutes of MVPA per week) at baseline. We used linear regression to compare changes in minutes per week of physical activity from baseline to 8 weeks across randomized arms for the following categories: MVPA, moderate, and vigorous minutes per week. We included the following additional covariates in our model: baseline value of the outcome and the timeslot at which women chose to join their sessions. Time was included as an 8-part variable (time slots from 6:00 AM to 9:30 AM PST) and included in the model using dummy variables. We did not include time as a covariate in models for women who met activity guidelines due to inadequate sample size. We used these same linear regression models and covariates to analyze secondary outcomes of changes in weight and psychosocial measures. We assessed recruitment and retention rates, adherence (measured by attendance of video sessions in the intervention arm), and acceptability (through questionnaire feedback from intervention participants). We carried out 4 sensitivity analyses for inactive and all mothers for the physical activity outcome measures of MVPA, moderate, and vigorous minutes per week. They included omitting time from the model, adjusting for total number of children which was imbalanced at baseline, replacing missing values assuming no change from baseline, and finally a "worst case scenario" where we replaced missing values with the respective randomization arm mean plus a standard deviation for control participants and minus a standard deviation for

intervention participants. Our sample size was estimated based on informal pilot data where we found an average increase of 30 minutes per week (standard deviation of 15 minutes per week) in 5 adherent participants over 8 weeks using a single intervention arm. We calculated that we needed at least 32 participants to have 80% power (with alpha=0.05) to detect a 20-minute difference in MVPA minutes per week between randomization arms if attrition was less than 10%, and we assumed an increase of 10 minutes per week in the control arm. As we found that it was feasible to recruit more participants during the planned recruitment period, we exceeded the minimum number of participants we aimed to enroll based on these sample size calculations.

Results

Participant Characteristics

We randomized 64 participants who were recruited over 5 months (July 2016-November 2016), 30 were allocated to the intervention and 34 to the control arm (Figure 1). All participants completed baseline surveys in which the majority reported not meeting CDC activity guidelines (54/64, 84%). At the 8-week follow-up time (October 2016-January 2017), 3 out of 64 participants were unable to be contacted, 2 from the control arm and 1 from the intervention arm, resulting in an overall loss of 5% for all participants (in analyses stratified by baseline activity, this equated to a 6% (3/54) loss for the inactive group, to which all 3 missing participants belonged). Participants with complete data (n=61) were included in analyses of primary, secondary, adherence, and feasibility outcomes.

The mean age of all women enrolled in the trial (n=64) was 37 years, and on average, they had less than 2 children (Table 1). Participants were predominately married, white, and had a high level of education, the majority with a post-graduate degree. Most women worked full-time or part-time jobs.

Physical Activity

Mothers in the intervention arm (n=29) increased their mean number of MVPA minutes per week by 42.2 (95% CI –11.3 to 95.7, P=.12) more minutes than mothers in the control arm (n=32), adjusted for baseline MVPA minutes per week and group time slot (Table 2). The intervention arm mothers increased moderate activity by 13.8 (95% CI –4.8 to 32.4, P=.14) more minutes per week and vigorous activity by 13.2 (95% CI –7.3 to 33.8, P=.20) more minutes per week.

Mothers who were inactive at baseline (n=51) increased their MVPA minutes per week by 50 (95% CI 4.0- 95.9, P=.03) more minutes in the intervention arm (n=23) compared with the control arm (n=28; see Table 2 and Figure 2). Inactive mothers at baseline assigned to the intervention arm increased their vigorous minutes per week by a net 19 (95% CI 3.2- 34.8, P=.02) minutes compared with the controls and increased their moderate activity minutes by 9.7 (95% CI –11.3 to 30.7, P=.36) min. In contrast, we found no statistically significant differences in physical activity outcomes across arms in mothers who were active at baseline (n=10).

Post Hoc Sensitivity Analysis

Our sensitivity analyses indicated that our physical activity models for all and inactive mothers were neither sensitive to omitting exercise time slot as a covariate in the model nor to various strategies for imputing the values of missing data (n=3, see Methods for a description of the imputation approaches). However, the estimated effect of treatment on MVPA in inactive women was attenuated and no longer statistically significant after adjustment for how many children women had, a baseline variable which was imbalanced across randomization arms. Women had more children in the control arm, and having additional children was independently associated with lower activity levels. However, the significant treatment effect on vigorous activity in inactive women was not sensitive to adjustment for this imbalance in number of children across arms (Multimedia Appendix 4).

Secondary Outcomes

We examined changes in several secondary outcomes: weight, social support for physical activity, physical activity self-efficacy, and 4 health-related quality of life measures (Table 3). Among women who were inactive at baseline, the intervention arm lost 2 (95% CI −4.2 to 0.2) more kilograms

than the control arm ($P=.07$). Social support for physical activity increased more for women in the intervention arm than in the control arm ($P=.04$). The difference in physical activity self-efficacy changes across trial arms was not statistically significant ($P=.80$).

Women in the intervention arm compared with the control arm did not experience statistically significant changes in the health-related quality of life measures of sleep, anxiety, fatigue, or depression across the 8 weeks. In stratified analyses, inactive women had a greater decrease in their depression score, a statistically significant difference of −3.8 (95% CI −7.0 to −0.6, $P=.02$).

Adherence and Acceptability

Women in the intervention arm (n=30) attended 2.8 group video sessions per week on average for over 8 weeks. The attendance had a standard deviation of 1.17 and a skewed distribution with a median of 3.5. Participants attended 3.3 sessions per week in the first half of the study and 2.4 sessions per week in the second half. Moreover, 5 participants from the intervention arm did not complete the entire 8 weeks, though all, except 1, completed end-of-study assessments. Women reported multiple reasons for noncompletion, including work, ill health, and lack of sleep.

Figure 1. Participant flow diagram. One participant declined to participate after randomization to the intervention arm, and that same participant did not complete end-of-study assessments. For our main analysis, we included complete cases (n=61); we also completed 2 sensitivity analyses to assess the possible effects of missing data that included all randomized participants (n=64).
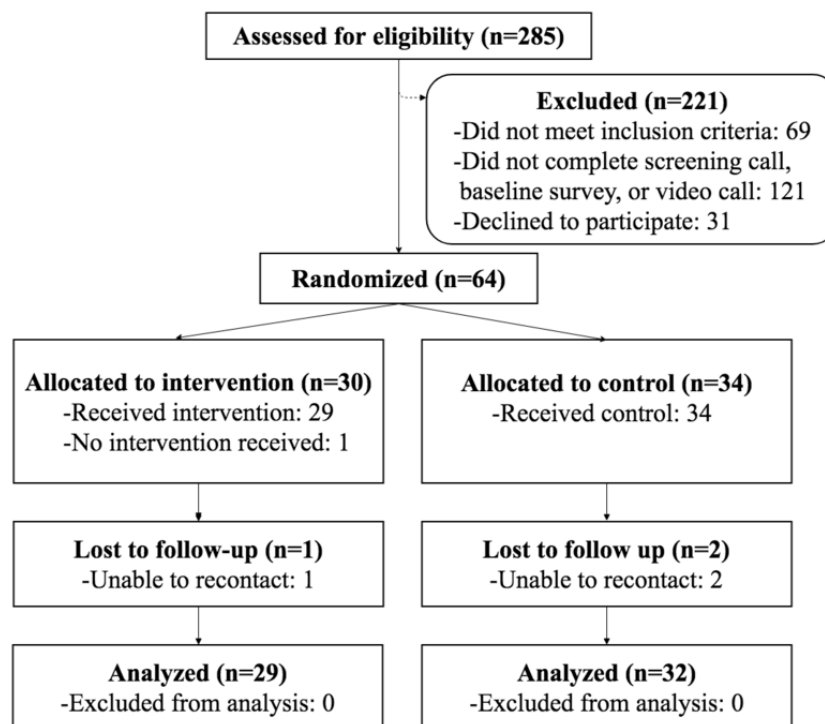
Table 1. Baseline characteristics.

Characteristic	Control (n=34)[a]	Intervention (n=30)[a]
Mother's age in years, mean (SD)	36.8 (6.5)	37.3 (4.0)
Children's age in years, mean (SD)	2.5 (1.9)	2.9 (2.1)
Number of children, mean (SD)	1.8 (0.8)	1.4 (0.5)
Marital status[b], n (%)		
Married or living as married	29 (88)	28 (93)
Never married	3 (9)	1 (3.3)
Separated or divorced	1 (3)	1 (3)
Race ethnicity, n (%)		
African American	1 (3)	1 (3)
Asian	4 (12)	3 (10)
Latina	0 (0)	1 (3)
Middle Eastern	2 (6)	0 (0)
Two or more races[c]	5 (15)	4 (13)
White	22 (65)	21 (70)
Employment, n (%)		
Full time	17 (50)	18 (60)
Not employed	9 (27)	4 (13)
Part time	7 (21)	6 (20)
Student	1 (3)	2 (7)
Education level, n (%)		
Some college	1 (3)	1 (3)
Bachelor degree	9 (27)	10 (33)
Post college degree	24 (71)	19 (63)
Currently breastfeeding, n (%)	15 (44)	12 (40)
Physical activity in minutes per week, mean (SD)		
Moderate to vigorous	59.1 (80.1)	89.5 (112.5)
Vigorous	13.5 (29.6)	24 (44.8)
Moderate	32.1 (38.2)	41.5 (50.3)
BMI[d] (kg/cm^2), mean SD)	24.1 (3.3)	25.6 (4.6)
Physical activity measures in score, mean (SD)		
Physical activity self-efficacy	3.5 (0.7)	3.6 (0.7)
Physical activity social support	2.0 (0.6)	2.1 (0.8)
PROMIS[e] measures in T-score[f], mean (SD)		
Depression	48.2 (6.8)	48.7 (7.6)
Sleep	57.7 (7.6)	57 (6.9)
Fatigue	60.0 (8.7)	59.4 (6.0)
Anxiety	50.4 (9.2)	51.9 (7.9)

[a]We used stratified randomization (time and baseline activity status), which resulted in intervention and control arms of unequal sizes.

[b]One person (from control arm) chose not to answer.

[c]Race/ethnicity—Two or more races category includes the following (n): Latina/white (1), Latina/Middle Eastern (1), Middle Eastern/white (3), Asian/white (2), American Indian/white (1), Mixed race-not specified (1).

[d]BMI: body mass index.

XSL•FO

RenderX

[e]PROMIS: Patient-Reported Outcome Measurement Information System.

[f]See Methods for scoring of PROMIS measures

Table 2. Changes in minutes of physical activity over 8 weeks by randomization arm and differences in changes across randomization arms.

Outcome measures	8-week change in physical activity minutes per week[a] (95% CI)		Difference across arms in 8-week changes[a,b] (95% CI)	P value
	Control arm	Intervention arm		
All mothers (n=61)				
MVPA[c]	−7.3 (−43.7 to 29.2)	34.9 (−3.4 to 73.2)	42.2 (−11.3 to 95.7)	.12
Vigorous	0.9 (−13.0 to 14.9)	14.2 (−0.5 to 28.8)	13.2 (−7.3 to 33.8)	.20
Moderate	−8.2 (−21.0 to 4.5)	5.6 (−7.8 to 19)	13.8 (−4.8 to 32.4)	.14
Inactive mothers (n=51)				
MVPA[c]	1.5 (−29.3 to 32.2)	51.4 (17.5-85.4)	50.0 (4.0-95.9)	.03
Vigorous	−0.1 (−10.7 to 10.4)	18.9 (7.2-30.5)	19.0 (3.2-34.8)	.02
Moderate	2.8 (−11.2 to 16.8)	12.5 (−3 to 27.9)	9.7 (−11.3 to 30.7)	.36
Active mothers (n=10)				
MVPA[c]	−9.1 (−204 to 185.8)	−68.1 (−225.5 to 89.3)	−59.0 (−315.9 to 197.9)	.60
Vigorous	10.8 (−81.8 to 103.3)	−5.5 (−81.0 to 70.0)	−16.3 (−135.9 to 103.3)	.76
Moderate	−65.3 (−110.6 to −20.0)	−34.0 (−70.5 to 2.5)	31.3 (−28.7 to 91.3)	.26

[a]Adjusted for baseline value of outcome and exercise time slot of choice for all mothers and inactive mothers. Adjusted for baseline value of outcome for active mothers.

[b]Difference of the within-group change for intervention versus control arm.

[c]MVPA: moderate to vigorous physical activity.

Figure 2. Moderate to vigorous physical activity (MVPA) minutes per week at baseline and 8 weeks for inactive (n=51) and active (n=10) women. MVPA minutes are calculated as follows: moderate minutes+(2×vigorous minutes). Unadjusted minutes of MVPA per week at baseline are in dark gray and at 8 weeks in light gray.
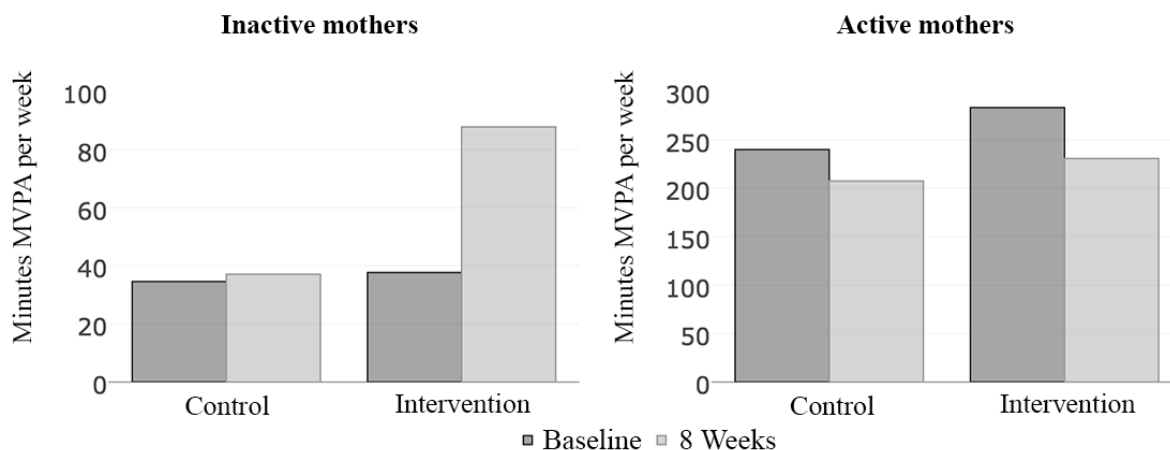
Table 3. Differences across randomization arms in changes in secondary outcomes over 8 weeks.

Outcome measures	All mothers (n=61)		Inactive mothers (n=51)		Active mothers (n=10)	
	Treatment effect[a] (95% CI)	P value	Treatment effect[a] (95% CI)	P value	Treatment effect[b] (95% CI)	P value
Weight loss (kg)	−1.6 (−3.7 to 0.6)	.15	−2.0 (−4.2 to 0.2)	.07	−2.6 (−18.2 to 12.9)	.69
Physical activity social support[c]	0.3 (0.0-0.6)	.04	0.2 (−0.1 to 0.5)	.15	0.9 (0.1-1.7)	.04
Physical activity self efficacy[c]	0.0 (−0.3 to 0.2)	.80	−0.1 (−0.3 to 0.2)	.71	0.0 (−0.6 to 0.6)	.87
PROMIS[d] measures						
Depression[e]	−3.0 (−5.9 to 0.0)	.05	−3.8 (−7.0 to −0.6)	.02	2.5 (−7.2 to 12.3)	.56
Sleep disturbance[e]	−1.3 (−3.8 to 1.2)	.32	−1.2 (−4.0 to 1.6)	.40	−0.3 (−8.1 to 7.6)	.94
Fatigue[e]	−1.1 (−5.1 to 2.8)	.57	−1.6 (−5.6 to 2.5)	.44	0.5 (−13.5 to 14.5)	.93
Anxiety[e]	−1.1 (−4.5 to 2.2)	.50	−1.3 (−5.2 to 2.7)	.52	1.6 (−6.5 to 9.7)	.66

[a]Adjusted for baseline value of outcome and exercise time slot of choice.

[b]Adjusted for baseline value of outcome for active mothers.

[c]Higher scores indicate a more optimal outcome.

[d]PROMIS: Patient-Reported Outcome Measurement Information System.

[e]Lower scores indicate a more optimal outcome.

Textbox 1. Satisfaction with the Moms Online Video Exercise (MOVE) intervention—participants' qualitative survey assessments.

Things we liked best

- "That it got me doing SOMETHING physical which I really, really needed."
- "I liked having the time set out for me to do the workout and having other people 'keeping me company.' That was a HUGE motivator."
- "Creating a structured time for myself and following through."
- "Loved the group motivation."
- "Working out from home, having accountability, the "come as you are" mentality, the other gals were great!"
- "I discovered that 15 minutes of morning exercise made my body feel better immediately and often for the rest of the day."
- "The workouts. You really can notice results with 15 minutes per day."
- "The 'live' nature of the sessions."
- "Knowing that there were other moms in the same boat as me."
- "Having a program to participate in created more support from [my] partner around exercise."

Things we would change

- "Some way to help push yourself to increasingly challenging programs in a measured way."
- "More workout options."
- "Offer more flexibility in the time."
- "Better introductions when a new person starts."
- "It would be nice to be able to join a later group if we can't make our regularly scheduled group."

The majority of mothers (86%) expressed satisfaction (extremely or somewhat satisfied) with the intervention. All mothers said they would recommend it to a friend, either certainly (96%) or maybe (4%). Mothers reported that the most significant impact from their participation was increasing their fitness levels (36%), being a good role model for their kids (14%), improving mood (11%), and feeling better about their body (7%). The most frequently (sometimes and often) used apps and YouTube videos included Sworkit, Yoga YouTube videos, Johnson and Johnson, and Nike+. All women reported feeling a benefit after sessions, for example, "energized," "great!," and "proud." A little less than half of the women in the intervention arm (42.9%) reported increasing their activity levels outside of the study and described these increases as: "The kids wanted to start doing more yoga (Cosmic Kids on YouTube) and dance parties as a family" and "I had more energy to do other activities throughout the day." Most women reported that their biggest barriers to attendance were lack of sleep, family commitments, and work

commitments. Most women (78%) reported in the survey that their commitment to the group and the expectation that others would be there and rely on them being present were the main motivators to attending sessions. In open survey responses to why participants liked the study, most listed social support, accountability, and convenience as their favorite features, as well as ones they would like further strengthened in future iterations of the program (Textbox 1).

Discussion

Summary of Results

The MOVE trial assessed an exercise group intervention using videoconferencing and mobile apps for mothers over an 8-week period through a randomized controlled design. The intervention was feasible and acceptable to participants. There was a trend toward increasing MVPA, moderate, and vigorous minutes of physical activity per week for all women, although this did not reach statistical significance. As hypothesized, women in the prespecified stratum who were inactive at baseline statistically significantly increased their MVPA minutes by an average of 50 minutes per week more in the intervention arm. A corresponding statistically significant increase of 19 minutes of vigorous activity drove the increase in total MVPA minutes per week for inactive women.

Feasibility and Acceptability

Digital tools were the driving force behind the feasibility and acceptability of this intervention. Recruitment, enrollment, data collection, and intervention delivery were all conducted remotely using technology, which was convenient for participants and study staff. Programs that can adapt to the individual context of their participants and ones that provide strong social support have proved effective at increasing physical activity [20,21]. The digital tools we used helped us address individual needs of participants while creating a socially supportive exercise space. Mobile apps allowed participants to choose short, and often, vigorous workouts, which could be customized to individual abilities and interests. Using mobile exercise apps provided participants with a way of efficiently exercising without having to make major changes to their existing routines. Women exercised from the convenience of their home at the time of their choosing, usually alongside their children. The videoconferencing tools helped create a supportive social group, which enabled women to check in face-to-face regularly, and facilitated accountability. Simultaneous exercise sessions also provided a sense of solidarity, even when the individual exercise routines were not coordinated within groups.

The participants' enthusiasm for the program was important in the early recruitment efforts, where participants shared study advertisements with multiple types of mother support group networks, and in the retention of participants who almost uniformly filled out end of study surveys, even if they no longer were able to participate in sessions. Many physical activity trials for mothers require fairly high time commitments from participants, primarily through coaching and education in person [43,51,52], remotely via telephone and texts [53,54] or both [55]. Participants' time in this study went almost entirely toward exercising. Participants reported a strong appreciation for the

convenience and flexibility of the intervention, which are particularly important features for mothers of young children who report feeling overwhelmed and unable to prioritize their own self-care [15,17-19,56]. Accordingly, our retention rates of 94% to 95% were higher than rates for 2 comparable technology trials on physical activity with mothers at equivalent time points of 86% at 1 month [53], 87% at 13 weeks [55], and among the highest of physical activity trials with mothers [43,51,52,54]. The high feasibility and acceptability of this trial has implications for future Internet physical activity trials targeting mothers.

Effectiveness

Randomized trials of physical activity with mothers have mixed results. Some trials have found statistically significant increases in physical activity [43,51-53,55], whereas others report nonstatistically significant changes [54]. There is great heterogeneity in the types of interventions delivered and even inconsistency in the definition of MVPA. Some studies use a simple equation [moderate + vigorous = MVPA], whereas others use a vigorous enhanced equation [moderate + vigorous × 2 = MVPA] as used in these analyses. Moreover, 2 comparable randomized technology trials of physical activity with mothers that incorporated technology found statistically significant increases in MVPA minutes in the range of the increases we found in inactive women [53,55]. One trial that utilized a physical activity website with resources, pedometers, and telephone counselors to provide motivational interviewing to help mothers incrementally work up to a goal of 150 MVPA found an increase of 92 MVPA minutes per week for mothers of babies 3+ months comparable with our difference of 50 MVPA minutes per week using the vigorous enhanced equation; a second trial that had mothers set their own exercise goals and provided support in meeting these goals over 13 weeks via individually tailored text messages found an increase of 49 MVPA minutes per week comparable with our difference of 30 MVPA minutes per week using the simple equation. These 2 studies were larger and longer and they differed from this study in that they had a large coaching component, did not include any group social support, and did not use apps or videoconferencing tools. Changes in vigorous minutes were not disaggregated from MVPA minutes per week in either of these studies.

Secondary Findings

In addition to the increases in physical activity, we observed improvements in several secondary measures. Social support specific to physical activity increased for mothers in this trial. Mothers have a uniquely challenging set of barriers to physical activity. Our participants reported that they were motivated to show up for one another (social support), and the presence of other mothers re-enforced their own capacity to exercise consistently (self-efficacy). We observed a statistically significant decrease in depression among inactive women in the intervention arm across the trial period. The increases in physical activity and social support that we observed could both contribute to decreased depression [4,57]. These are mechanisms that could be tested individually and synergistically in future trials.

XSL•FO

RenderX

Limitations

Our digital tools helped create an efficient recruiting process; however, our recruitment and enrollment strategies and inclusion criteria resulted in a sample that was not representative of the United States population. Our requirements of completing a phone call, survey, and a practice video session before randomization could have produced a sample that was more adherent than the general population. Furthermore, recruitment strategies that targeted mothers' groups through Facebook and email as well as the inclusion criteria that required an ability to own and use 2 devices while potentially caring for a child could have contributed to our sample consisting primarily of women who were highly educated, married, white, of an older age at first child, and typically lived in large cities on the West Coast. Future trials are needed to test whether this type of intervention could be replicated in a more diverse population.

We relied on a self-report measure of physical activity, which though validated and widely used, could have introduced bias. Participants and investigators were not blinded to their randomization status, which could have also introduced bias. Our sample size limited our ability to fully explore the differences in outcomes by baseline activity status. In particular, the group of mothers who were physically active at baseline was quite small (n=10). Although our results clearly suggest that this type of intervention is most likely to benefit mothers who are inactive, it would be premature to conclude, based on our data, that this approach does not benefit all mothers.

In post hoc sensitivity analyses, we found that our model was neither sensitive to the removal of exercise time slot nor to 2 imputation strategies to address missing data. However, we did find that our results for MVPA among inactive women were sensitive to adjustment for number of children, which was imbalanced at baseline. In contrast, results for vigorous activity retained statistical significance, suggesting that our intervention may be most effective at increasing higher intensity activity. Although the imbalance resulted by chance in this small sample, stratification of the randomization by number of children could be worth exploring in future studies in this population.

We were unable to test a longer intervention due to limited resources, and we could not assess whether the intervention effect could be maintained over a longer time due to our waitlist design. In addition, we were unable to fully disentangle the impact of videoconferencing separate from the impact of mobile apps. The waitlist group participated in a required initial videoconferencing session using a mobile app and was given access to the recommended list of apps, which suggests that the videoconferencing drove the difference between randomization arms; however, the intervention arm additionally had a specific prescription of exercise as well as support and monitoring by study staff to meet the prescription, which limits our ability to attribute any differences solely to videoconferencing.

Conclusions

This study suggests that technology can be used to create an individualized physical activity intervention with social support using a scalable and cost-effective delivery mechanism for mothers. There is great excitement in the use of new technology to solve old problems; however, often new technology alone cannot overcome the barriers to behavior change. We utilized technology to deliver evidence-based components of individualization and social support in a physical activity program that was convenient and compelling for our busy participants. To our knowledge, this is the first study that examines the use of videoconferencing paired with exercise mobile apps to create exercise groups. We found that using videoconferencing and mobile apps was a feasible and acceptable way to deliver a physical activity group intervention for mothers. Furthermore, we showed our intervention increased physical activity in inactive mothers. Further studies are needed to better establish how long these changes in physical activity can be maintained and whether these findings can be reproduced in a more diverse population.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT‐EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 101KB - jmir_v20i5e179_app1.pdf]

Multimedia Appendix 2

Participant personal website.

[PNG File, 408KB - jmir_v20i5e179_app2.png]

XSL·FO

RenderX

Multimedia Appendix 3

Recommended exercise apps and websites.

[PDF File (Adobe PDF File), 28KB - jmir_v20i5e179_app3.pdf]

Multimedia Appendix 4

Post hoc sensitivity analyses—differences across randomization arms in physical activity changes over 8 weeks for inactive mothers and all mothers.

[PDF File (Adobe PDF File), 73KB - jmir_v20i5e179_app4.pdf]

References

1. Centers for Disease Control and Prevention (CDC). Adult participation in aerobic and muscle-strengthening physical activities--United States, 2011. MMWR Morb Mortal Wkly Rep 2013 May 03;62(17):326-330 [FREE Full text] [Medline: 23636025]

2. Orozco LJ, Buchleitner AM, Gimenez-Perez G, Roqué I Figuls M, Richter B, Mauricio D. Exercise or exercise and diet for preventing type 2 diabetes mellitus. Cochrane Database Syst Rev 2008 Jul 16(3):CD003054. [doi: 10.1002/14651858.CD003054.pub3] [Medline: 18646086]

3. Bauman AE. Updating the evidence that physical activity is good for health: an epidemiological review 2000-2003. J Sci Med Sport 2004 Apr;7(1 Suppl):6-19. [Medline: 15214597]

4. Fox KR. The influence of physical activity on mental well-being. Public Health Nutr 1999 Sep;2(3A):411-418. [Medline: 10610081]

5. Callaghan P. Exercise: a neglected intervention in mental health care? J Psychiatr Ment Health Nurs 2004 Aug;11(4):476-483. [doi: 10.1111/j.1365-2850.2004.00751.x] [Medline: 15255923]

6. US Department of Health and Human Services. 2008 Physical Activity Guidelines for Americans: Be Active, Healthy, and Happy! URL: http://www.health.gov/paguidelines [WebCite Cache ID 6yes6rr4j]

7. Bellows-Riecken KH, Rhodes RE. A birth of inactivity? A review of physical activity and parenthood. Prev Med 2008 Feb;46(2):99-110. [doi: 10.1016/j.ypmed.2007.08.003] [Medline: 17919713]

8. Engberg E, Alen M, Kukkonen-Harjula K, Peltonen JE, Tikkanen HO, Pekkarinen H. Life events and change in leisure time physical activity: a systematic review. Sports Med 2012 May 01;42(5):433-447. [doi: 10.2165/11597610-000000000-00000] [Medline: 22512413]

9. Allender S, Hutchinson L, Foster C. Life-change events and participation in physical activity: a systematic review. Health Promot Int 2008 Jun;23(2):160-172 [FREE Full text] [doi: 10.1093/heapro/dan012] [Medline: 18364364]

10. Archer E, Lavie CJ, McDonald SM, Thomas DM, Hébert JR, Taverno Ross SE, et al. Maternal inactivity: 45-year trends in mothers' use of time. Mayo Clin Proc 2013 Dec;88(12):1368-1377. [doi: 10.1016/j.mayocp.2013.09.009] [Medline: 24290110]

11. Hesketh KR, Goodfellow L, Ekelund U, McMinn AM, Godfrey KM, Inskip HM, et al. Activity levels in mothers and their preschool children. Pediatrics 2014 Apr;133(4):e973-e980 [FREE Full text] [doi: 10.1542/peds.2013-3153] [Medline: 24664097]

12. Trost SG, Sallis JF, Pate RR, Freedson PS, Taylor WC, Dowda M. Evaluating a model of parental influence on youth physical activity. Am J Prev Med 2003 Nov;25(4):277-282. [Medline: 14580627]

13. Verloigne M, Van Lippevelde W, Maes L, Brug J, De Bourdeaudhuij I. Family- and school-based correlates of energy balance-related behaviours in 10-12-year-old children: a systematic review within the ENERGY (EuropeaN Energy balance Research to prevent excessive weight Gain among Youth) project. Public Health Nutr 2012 Aug;15(8):1380-1395. [doi: 10.1017/S1368980011003168] [Medline: 22269173]

14. Norman E, Sherburn M, Osborne RH, Galea MP. An exercise and education program improves well-being of new mothers: a randomized controlled trial. Phys Ther 2010 Mar;90(3):348-355. [doi: 10.2522/ptj.20090139] [Medline: 20056720]

15. Currie J. Motherhood, stress and the exercise experience: freedom or constraint? Leisure Studies 2004 Jul;23(3):225-242. [doi: 10.1080/0261436042000251987]

16. Lewis B, Ridge D. Mothers reframing physical activity: family oriented politicism, transgression and contested expertise in Australia. Soc Sci Med 2005 May;60(10):2295-2306. [doi: 10.1016/j.socscimed.2004.10.011] [Medline: 15748677]

17. Hamilton K, White KM. Understanding parental physical activity: meanings, habits, and social role influence. Psychol Sport Exerc 2010 Jul;11(4):275-285. [doi: 10.1016/j.psychsport.2010.02.006]

18. Barkin JL, Wisner KL. The role of maternal self-care in new motherhood. Midwifery 2013 Sep;29(9):1050-1055. [doi: 10.1016/j.midw.2012.10.001] [Medline: 23415369]

19. Verhoef MJ, Love EJ. Women and exercise participation: the mixed blessings of motherhood. Health Care Women Int 1994;15(4):297-306. [doi: 10.1080/07399339409516122] [Medline: 8056646]

XSL•FO

RenderX

20. Bauman AE, Reis RS, Sallis JF, Wells JC, Loos RJ, Martin BW, Lancet Physical Activity Series Working Group. Correlates of physical activity: why are some people physically active and others not? Lancet 2012 Jul 21;380(9838):258-271. [doi: 10.1016/S0140-6736(12)60735-1] [Medline: 22818938]

21. Kahn EB, Ramsey LT, Brownson RC, Heath GW, Howze EH, Powell KE, et al. The effectiveness of interventions to increase physical activity. A systematic review. Am J Prev Med 2002 May;22(4 Suppl):73-107. [Medline: 11985936]

22. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006 Jun 23;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]

23. Pew Research Center. Washington, DC; 2015 Feb 05. Mobile fact sheet URL: http://www.pewinternet.org/fact-sheet/mobile/ [WebCite Cache ID 6yesZ5t48]

24. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]

25. Conroy DE, Yang CH, Maher JP. Behavior change techniques in top-ranked mobile apps for physical activity. Am J Prev Med 2014 Jun;46(6):649-652. [doi: 10.1016/j.amepre.2014.01.010] [Medline: 24842742]

26. Cowan LT, Van Wagenen SA, Brown BA, Hedin RJ, Seino-Stephan Y, Hall PC, et al. Apps of steel: are exercise apps providing consumers with realistic expectations?: a content analysis of exercise apps for presence of behavior change theory. Health Educ Behav 2013 Apr;40(2):133-139. [doi: 10.1177/1090198112452126] [Medline: 22991048]

27. Modave F, Bian J, Leavitt T, Bromwell J, Harris I, Vincent H. Low quality of free coaching apps with respect to the American College of Sports Medicine Guidelines: a review of current mobile apps. JMIR Mhealth Uhealth 2015;3(3):e77 [FREE Full text] [doi: 10.2196/mhealth.4669] [Medline: 26209109]

28. Pagoto S, Bennett GG. How behavioral science can advance digital health. Transl Behav Med 2013 Sep;3(3):271-276 [FREE Full text] [doi: 10.1007/s13142-013-0234-z] [Medline: 24073178]

29. Duggan M, Lenhart A, Lampe C, Ellison NB. Pew Research Center. 2015 Jul. Parents and social media URL: http://www.pewinternet.org/2015/07/16/parents-and-social-media/ [WebCite Cache ID 6yetTdCtD]

30. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]

31. Foster C, Richards J, Thorogood M, Hillsdon M. Remote and web 2.0 interventions for promoting physical activity. Cochrane Database Syst Rev 2013 Sep 30;9:CD010395 [FREE Full text] [doi: 10.1002/14651858.CD010395.pub2] [Medline: 24085594]

32. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]

33. Muntaner A, Vidal-Conti J, Palou P. Increasing physical activity through mobile device interventions: a systematic review. Health Informatics J 2016 Sep;22(3):451-469. [doi: 10.1177/1460458214567004] [Medline: 25649783]

34. Alley S, Jennings C, Plotnikoff RC, Vandelanotte C. Web-based video-coaching to assist an automated computer-tailored physical activity intervention for inactive adults: a randomized controlled trial. J Med Internet Res 2016 Aug 12;18(8):e223 [FREE Full text] [doi: 10.2196/jmir.5664] [Medline: 27520283]

35. Adams R. Revised Physical Activity Readiness Questionnaire. Can Fam Physician 1999 Apr;45:992, 995, 1004-992, 995, 1005 [FREE Full text] [Medline: 10216799]

36. Johnson & Johnson Health & Wellness Solutions. 7 Minute Workout URL: https://7minuteworkout.jnj.com/ [WebCite Cache ID 6rRZlSLhP]

37. 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]

38. Australian Institute of Health and Welfare. The Active Australia Survey: A Guide and Manual for Implementation, Analysis and Reporting. Canberra: Australian Institute of Health and Welfare; 2003. URL: https://www.aihw.gov.au/reports/physical-activity/active-australia-survey/contents/table-of-contents

39. Armstrong T, Bauman A, Davies J. Australian Institute of Health and Welfare. Canberra, Australia; 2000. Physical activity patterns of Australian adults: results of the 1999 National Physical Activity Survey URL: https://www.aihw.gov.au/reports/physical-activity/physical-activity-patterns-of-australian-adults/contents/table-of-contents [WebCite Cache ID 6ym7G6yP0]

40. Brown WJ, Burton NW, Marshall AL, Miller YD. Reliability and validity of a modified self-administered version of the Active Australia physical activity survey in a sample of mid-age women. Aust N Z J Public Health 2008 Dec;32(6):535-541. [doi: 10.1111/j.1753-6405.2008.00305.x] [Medline: 19076744]

41. Reeves MM, Marshall AL, Owen N, Winkler EA, Eakin EG. Measuring physical activity change in broad-reach intervention trials. J Phys Act Health 2010 Mar;7(2):194-202. [Medline: 20484758]

42. Albright CL, Steffen AD, Novotny R, Nigg CR, Wilkens LR, Saiki K, et al. Baseline results from Hawaii's Nā Mikimiki Project: a physical activity intervention tailored to multiethnic postpartum women. Women Health 2012;52(3):265-291 [FREE Full text] [doi: 10.1080/03630242.2012.662935] [Medline: 22533900]

43. Miller YD, Trost SG, Brown WJ. Mediators of physical activity behavior change among women with young children. Am J Prev Med 2002 Aug;23(2 Suppl):98-103. [Medline: 12133744]

44. Sallis JF, Pinski RB, Grossman RM, Patterson TL, Nader PR. The development of self-efficacy scales for health-related diet and exercise behaviors. Health Educ Res 1988;3(3):283-292. [doi: 10.1093/her/3.3.283]

45. Sallis JF, Grossman RM, Pinski RB, Patterson TL, Nader PR. The development of scales to measure social support for diet and exercise behaviors. Prev Med 1987 Nov;16(6):825-836. [Medline: 3432232]

46. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol 2010 Nov;63(11):1179-1194 [FREE Full text] [doi: 10.1016/j.jclinepi.2010.04.011] [Medline: 20685078]

47. HealthMeasures. 2016. Anxiety - A brief guide to the PROMIS Anxiety instruments URL: http://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Anxiety_Scoring_Manual_09202016.pdf [accessed 2018-04-16] [WebCite Cache ID 6ym7i079D]

48. HealthMeasures. 2015. Sleep Disturbance - A brief guide to the PROMIS Sleep Disturbance instruments URL: http://www.healthmeasures.net/images/promis/manuals/PROMIS_Sleep_Disturbance_Scoring_Manual.pdf [accessed 2018-04-17] [WebCite Cache ID 6ym8KKfir]

49. HealthMeasures. 2017. Depression - A brief guide to the PROMIS Depression instruments URL: http://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Depression_Scoring_Manual_02222017.pdf [accessed 2018-04-17] [WebCite Cache ID 6ym8OBSaz]

50. HealthMeasures. 2017. Fatigue - A brief guide to the PROMIS Fatigue instruments URL: http://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Fatigue_Scoring_Manual.pdf [accessed 2018-04-18] [WebCite Cache ID 6ym8WVaPo]

51. Cramp AG, Brawley LR. Moms in motion: a group-mediated cognitive-behavioral physical activity intervention. Int J Behav Nutr Phys Act 2006 Aug 22;3:23 [FREE Full text] [doi: 10.1186/1479-5868-3-23] [Medline: 16925809]

52. Cody R, Lee C. Development and evaluation of a pilot program to promote exercise among mothers of preschool children. Int J Behav Med 1999;6(1):13-29. [doi: 10.1207/s15327558ijbm0601_2] [Medline: 16250689]

53. Albright CL, Steffen AD, Wilkens LR, White KK, Novotny R, Nigg CR, et al. Effectiveness of a 12-month randomized clinical trial to increase physical activity in multiethnic postpartum women: results from Hawaii's Nā Mikimiki Project.Effectiveness of a 12-month randomized clinical trial to increase physical activity in multiethnic postpartum women: results from Hawaii's Nā Mikimiki Project. Prev Med 2014 Dec;69:214-223 [FREE Full text] [doi: 10.1016/j.ypmed.2014.09.019] [Medline: 25285751]

54. Chang MW, Nitzke S, Brown R. Design and outcomes of a Mothers In Motion behavioral intervention pilot study. J Nutr Educ Behav 2010;42(3 Suppl):S11-S21 [FREE Full text] [doi: 10.1016/j.jneb.2010.01.010] [Medline: 20399404]

55. Fjeldsoe BS, Miller YD, Graves N, Barnett AG, Marshall AL. Randomized controlled trial of an improved version of MobileMums, an intervention for increasing physical activity in women with young children. Ann Behav Med 2015 Aug;49(4):487-499. [doi: 10.1007/s12160-014-9675-y] [Medline: 25582987]

56. Harrington M, Dawson D, Bolla P. Objective and subjective constraints on women's enjoyment of leisure. Loisir Société/Society Leisure 1992 Jan;15(1):203-221. [doi: 10.1080/07053436.1992.10715414]

57. Gariépy G, Honkaniemi H, Quesnel-Vallée A. Social support and protection from depression: systematic review of current findings in Western countries. Br J Psychiatry 2016 Dec;209(4):284-293. [doi: 10.1192/bjp.bp.115.169094] [Medline: 27445355]

Abbreviations

CDC: Centers for Disease Control and Prevention
MOVE: Moms Online Video Exercise
MVPA: moderate to vigorous physical activity

XSL•FO

RenderX

Review

The Effectiveness of Smartphone Apps for Lifestyle Improvement in Noncommunicable Diseases: Systematic Review and Meta-Analyses

Pernille Lunde[1], MSc; Birgitta Blakstad Nilsson[1,2], PhD; Astrid Bergland[1], PhD; Kari Jorunn Kværner[3,4], PhD; Asta Bye[5,6], PhD

[1]Department of Physiotherapy, Faculty of Health Sciences, OsloMet—Oslo Metropolitan University, Oslo, Norway

[2]Section for Physiotherapy, Division of Medicine, Oslo University Hospital, Oslo, Norway

[3]Center for Connected Care, Oslo University Hospital, Oslo, Norway

[4]Department of Strategy and Entrepreneurship, BI Norwegian School of Business, Oslo, Norway

[5]Department of Nursing and Health Promotion, Faculty of Health Sciences, OsloMet—Oslo Metropolitan University, Oslo, Norway

[6]European Palliative Care Research Centre, Department of Oncology, Oslo University Hospital and Institute of Clinical Medicine, University of Oslo, Oslo, Norway

Corresponding Author:
Pernille Lunde, MSc
Department of Physiotherapy
Faculty of Health Sciences
OsloMet—Oslo Metropolitan University
1st Floor
Pilestredet 50
Oslo, 0130
Norway
Phone: 47 48063537
Email: plunde@oslomet.no

Abstract

Background: Noncommunicable diseases (NCDs) account for 70% of all deaths in a year globally. The four main NCDs are cardiovascular diseases, cancers, chronic pulmonary diseases, and diabetes mellitus. Fifty percent of persons with NCD do not adhere to prescribed treatment; in fact, adherence to lifestyle interventions is especially considered as a major challenge. Smartphone apps permit structured monitoring of health parameters, as well as the opportunity to receive feedback.

Objective: The aim of this study was to review and assess the effectiveness of app-based interventions, lasting at least 3 months, to promote lifestyle changes in patients with NCDs.

Methods: In February 2017, a literature search in five databases (EMBASE, MEDLINE, CINAHL, Academic Research Premier, and Cochrane Reviews and Trials) was conducted. Inclusion criteria was quantitative study designs including randomized and nonrandomized controlled trials that included patients aged 18 years and older diagnosed with any of the four main NCDs. Lifestyle outcomes were physical activity, physical fitness, modification of dietary habits, and quality of life. All included studies were assessed for risk of bias using the Cochrane Collaboration`s risk of bias tool. Meta-analyses were conducted for one of the outcomes (glycated hemoglobin, HbA_{1c}) by using the estimate of effect of mean post treatment with SD or CI. Heterogeneity was tested using the I^2 test. All studies included in the meta-analyses were graded.

Results: Of the 1588 records examined, 9 met the predefined criteria. Seven studies included diabetes patients only, one study included heart patients only, and another study included both diabetes and heart patients. Statistical significant effect was shown in HbA_{1c} in 5 of 8 studies, as well in body weight in one of 5 studies and in waist circumference in one of 3 studies evaluating these outcomes. Seven of the included studies were included in the meta-analyses and demonstrated significantly overall effect on HbA_{1c} on a short term (3-6 months; P=.02) with low heterogeneity (I^2=41%). In the long term (10-12 months), the overall effect on HbA_{1c} was statistical significant (P=.009) and without heterogeneity (I^2=0%). The quality of evidence according to Grading of Recommendations Assessment, Development and Evaluation was low for short term and moderate for long term.

XSL•FO
RenderX

Conclusions: Our review demonstrated limited research of the use of smartphone apps for NCDs other than diabetes with a follow-up of at least 3 months. For diabetes, the use of apps seems to improve lifestyle factors, especially to decrease HbA_{1c}. More research with long-term follow-up should be performed to assess the effect of smartphone apps for NCDs other than diabetes.

Introduction

Noncommunicable diseases (NCDs) account for as much as 70% of all deaths globally [1]. The four main NCDs are cardiovascular diseases (CVDs), cancers, chronic pulmonary diseases, and diabetes mellitus (DM), which all share the same behavioral risk factors: physical inactivity, unhealthy diet, tobacco use, and harmful use of alcohol [1]. Lifestyle changes toward a more healthy behavior are of great importance in both prevention and treatment of these NCDs [2-5].

Adherence to treatment is the most important modifiable factor that compromises treatment outcome. Traditionally, adherence has focused on medication, which also is reflected in the World Health Organization's (WHO) definition of adherence; "the extent to which the patient follows medical instructions" [6]. However, adherence also encompasses numerous health-related behaviors such as smoking cessation and changes in physical activity (PA), exercise, or diet, which are considered as a major challenge in treatment of NCDs [6]. Usually the interventions designed to promote healthy behavior are conducted as face-to-face modes of delivery, and their mainly short-term effectiveness has been extensively documented in a number of systematic reviews [7-10]. One reason of the inconclusive long-term results are probably lack of systematic follow-up and monitoring, which are crucial elements of all effective health behavior change [11].

Feedback seems to be essential for success in behavioral change [12]. Modern technology such as electronic devices permits structured monitoring of important health parameters and follow-up of patients with NCD [13]. A meta-analyses (n=43,200) documented that mixed mode of delivery interventions where traditional behavioral change techniques (BCTs) were used together with dedicated digital tools were more effective than traditional techniques for behavioral change alone [12]. Another meta-analyses (n=20,000) supports this and concludes that tailored Web-based interventions was significantly more effective in improving health outcomes compared with nontailored Web-based interventions [14]. Although several interventions such as Web portals, SMS text messaging (short message service, SMS), and phone calls to improve health for patients with NCDs are promising [15-19], smartphone technology has been emphasized because of its possibility to monitor and follow-up patients' health from anywhere at any time [20].

Thus, the purpose of this systematic review was to examine the effectiveness of interventions with smartphone apps, lasting at least 3 months, to promote lifestyle changes such as PA, physical fitness, modification of dietary habits, and quality of life (QoL) in patients with NCDs.

Methods

Reporting Standards

This systematic review and its procedures were planned, conducted, and reported in accordance to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidance. The review protocol was registered in the International Prospective Register of Systematic Reviews, registration number CRD42017057796.

Inclusion Criteria

Randomized and nonrandomized clinical trials with a minimum of 3-months follow-up that evaluated the effect of interventions with apps aiming to monitor PA and/or dietary habits were considered for inclusion. Patients had to be aged 18 years or above and diagnosed with CVD, cancer, chronic pulmonary disease, or DM. Case series with 10 or less participants were not included. If change of lifestyle was not the main goal of the intervention, studies were excluded. Due to limited resources for translation, the review was restricted to publications in Norwegian and English.

Outcome Measures

Primary outcomes of interest were PA, physical fitness, modification of dietary habits, and/or QoL. Regarding PA and physical fitness, the following measures were considered relevant: steps, self-reported minutes in activity, self-reported minutes of exercise, maximal oxygen consumption, 6-min walk test, shuttle walk tests, and submaximal physical fitness tests. Regarding effect on modification of dietary habits, measures included body weight, body mass index (BMI), waist circumference, and glycated hemoglobin (HbA_{1c}). Both generic and disease-specific QoL questionnaires were evaluated.

Search Strategy

Five databases (EMBASE, MEDLINE, CINAHL, Academic Research Premier, and Cochrane Reviews and Trials) were systematically searched for relevant studies with help from a research librarian. Boolean operators were used to expand, exclude, or join keywords in the search using the terms "AND" and "OR." Articles published before February 23, 2017 in English were included in this systematic review. The search strategy of each database is listed in Multimedia Appendix 1.

Selection of Studies

Figure 1 shows the PRISMA flowchart of reviewed and included studies.

The first author conducted the database search assisted by a research librarian. After conducting the search, duplicates were removed, and 2 authors independently reviewed title and abstract of all studies. We kept relevant reviews to hand screen the reference lists in case some articles got lost in the initial search. Disagreements between the two authors conducting the title and abstract review were discussed until a consensus was reached. All the studies that met the inclusion criteria went through a full-text screening process by two reading pairs. The first author reviewed all the studies. In the full-text screening phase, we hand screened the reference lists of all reviews, and we also screened the study characteristics of the included studies in the reviews. Additional studies were identified for inclusion to full-text screening. In case of disagreement in this phase, the other reading pair contributes to achieve consensus. In case of uncertainty related to the intervention used in some of the studies, we contacted authors. In addition, phone developers' own description was used if there was any uncertainty whether the phones were smartphones or not.

The first author extracted data from the studies. In studies with more than one intervention arm, data from the most intervening arm were extracted [21,22]. This was done to make the interventions in the different studies as homogeneous as possible. Data extracted from the studies included authors, year, country, study design, patient group (sample size and disease), inclusion criteria, details of the interventions, outcomes, and results.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram of reviewed and included studies.

Figure 2. Risk of bias within studies.

Quality Assessment

In total, five reviewers (two reading pairs and first author reading all papers) independently assessed each included study for risk of bias (high, low, or unclear) using the Cochrane Collaboration`s risk of bias tool [23]. Again, the other reading pair assisted to reach consensus if necessary. Regarding scoring the different studies with respect to "selective reporting," we followed the judging criteria given by Cochrane Collaboration and read all protocols available in clinical trials or in journals

if published. The results of the risk of bias assessment were then exported to the software RevMan, version 5.3 [24] to create visual representation of the publication (Figure 2). Difficulty in scoring some of the studies was handled by reading the protocol if published, either in paper or in Clinical Trials and/or by contacting study authors.

Data Synthesis and Statistical Analyses

After the first author had extracted data from the studies, all authors evaluated the preliminary results of the review. Meta-analyses were performed based on sufficient homogeneity across most of the included studies with regard to disease (DM) and primary outcome (HbA_{1c}). One meta-analysis for short-term effect (3-6 months) and one for long-term effect (10-12 months) were performed. In the end, all studies included in the meta-analyses were evaluated using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) [25].

Meta-analyses were conducted by using estimate of effect of mean posttreatment HbA_{1c} values for both intervention and control group with SD. In studies where mean change were the only presented result [26,27], we used this in addition to SD for both groups. If only CIs were presented, we calculated SD. In one study [27], both adjusted and nonadjusted estimate of effect were presented. We chose to use the adjusted estimate as authors reported this as results. To carry out the meta-analyses, we pooled studies based on length of the interventions. Heterogeneity was tested using the I^2 test. Forest plot were constructed to visualize the results. All analyses were performed with RevMan version 5.3 software, with double entry of the estimate of effect.

Results

Study Selection

Our search results are summarized in the PRISMA flow diagram (Figure 1). A total of 1588 studies were identified. Duplicates were removed, leaving 1228 studies to screen. After screening title and abstracts, 1181 papers were excluded for not meeting the inclusion criteria, leaving 47 studies for full-text review. On the basis of the full-text review, 6 studies were included in this review. From a hand search of review paper references and study characteristics in the reviews, an additional 5 studies were identified as potentially eligible, of which 3 studies were included. In total, 9 studies were included in the systematic review and 8 were eligible for meta-analysis [21,22,26-31]. One study [28] was not included in the meta-analysis because of poor outcome reporting and lack of response on email.

Study Characteristics

The characteristics of the included studies are presented in Table 1. Six of the included studies included patients with DM type 2 only [21,22,27-30]. One study included both type 1 and type 2 DM [31], 1 study included patients with DM type 2 or CVD (ischemic heart disease and/or heart failure) [26], and 1 study included patients with myocardial infarction [32]. Four studies were carried out in Europe [21,26,27,32], 3 in North- and East

America [22,28,30], and 2 in Asia [29,31]. Study duration ranged from 3 months to 1 year of follow up; 3 months (n=3) [28,29,31], 6 months (n=2) [30,32], 10 months (n=1) [27], and 1 year (n=3) [21,22,26]. All the included studies had a control group, where 6 studies involved "usual care" or "standard medical care" as control. Two of the included studies gave the control group part of the intervention, whereas one of them received a simplified smartphone app with drug adherence e-diary [32], and the other received health coaching [30]. One study did not describe what the control group received [28].

Intervention Characteristics

An overview of the characteristics of the interventions used in included studies is presented in Table 2. All the included studies [21,22,26-32] used apps where it was possible to register blood glucose data. All studies had registration of lifestyle factors, whereas 6 studies registered exercise and/or PA in the app [21,26,27,29,30,32] and/or registration of dietary habits [21,22,28-31]. Patients in all studies monitored themselves on lifestyle factors and clinical measurements. In 7 of the studies [22,26-31], health personal and/or researchers monitored them as well. Patients received feedback based on what they had registered in the app, whereas 4 of the studies had automatic feedback [21,22,28,32], 3 studies had individualized feedback [26,30,31], 1 study had automatic feedback and individualized if warranted [27], and 1 study had both automatic and individualized [29]. In 6 of the included studies, they had additional support to the app [21,22,26-28,30]; however, the app was the main part of the intervention.

Risk of Bias Within Studies

There was high risk of bias in all the included studies (Figure 2).

Six of the studies were registered in clinical trials [21,22,26,30-32]. All of them reported on more outcomes than registered. The additional outcomes were not exclusively positive. The reason might be not updating the study protocol rather than selective outcome reporting. Therefore, they got "low risk of bias" on "selective reporting" score. The three other studies were neither registered in clinical trials nor published elsewhere [27-29]. However, the published reports included all expected outcomes, and therefore, they were all given "low risk of bias."

Effects of Smartphone App

An overview of effects of apps on lifestyle factors including physical fitness, PA, modification of dietary habits, and QoL is presented Table 1. Five of 8 studies evaluating HbA_{1c} reported statistical significant differences between groups in favor of the intervention groups [22,27-29,31]. One of 3 studies evaluating waist circumference reported a statistical significant effect between groups in favor of the intervention group [26]. Additionally, 1 study reported a statistical significant within group change for the intervention group [30]. One of 5 studies evaluating body weight reported statistical significant differences between groups in favor of the intervention group [27], and 2 studies reported a statistical significant change in body weight within the intervention groups [26,30].

Table 1. Study characteristics.

Reference (year), country	Study design and study duration	Sample size; disease	Intervention group (IG) or control group (CG)	Outcomes of interest[a]	Results[b]
Holmen et al (2014), Norway [21]	3-arm randomized controlled trial (RCT), multicenter, 12 months	N=151; Diabetes mellitus (DM) type 2	IG 1: app to increase self-management ; IG 2: IG 1 + five health counseling sessions by a diabetes nurse; CG: usual care	*Glycated hemoglobin (HbA_{1c})*, Weight, Health-related quality of life (HRQoL; 36-item short form survey, SF-36), Lifestyle change (dietary and physical activity)	No statistical differences between groups (NS) in outcomes of interest
Johnston et al (2016), Sweden [32]	2-arm RCT, multicenter, 6 months	N=174; Myocardial infarction	IG: app to register information about drug adherence, exercise, weight, smoking, blood pressure, low-density lipoprotein cholesterol, and blood glucose; CG: simplified app with drug adherence e-diary	Cardiovascular risk (body mass index, physical activity), QoL (EuroQoL-5D)	NS in outcomes of interest
Karhula et al (2015), Finland [26]	2-arm RCT, 12 months	N=519; Heart disease patients (ischemic and/or heart failure) or DM type 2	IG: app with health coaching and self-monitoring of health parameters; CG: usual care	*HRQoL (SF-36), HbA_{1c} (in DM patients)*, Body weight, Waist circumference	Diabetics: Change in waist circumference, P=.01; IG: −2.03, 95% CI (−2.76 to −1.29), CG: −0.29, 95% CI (−1.47 to 0.9); NS in other outcomes of interest; Heart patients: NS in all outcomes of interest
Orsama et al (2013), Finland [27]	2-arm RCT, 10 months	N=53; DM type 2	IG: app for monitoring and remote reporting of diabetes health-related parameters; CG: usual care	*HbA_{1c}*, Body weight	Change in HbA$_{1c}$, P=.02, IG: −0.4, 95% CI (−0.67 to −0.14), CG: 0.004, 95% CI (−0.35 to 0.36)-Change in body weight, P=.02, IG: −2.1 kg, 95% CI (−3.6 to −0.6), CG: 0.4 kg, 95% CI (−1.1 to 1.9)
Quinn et al (2008), Maryland, United States [28]	2-arm RCT, multicenter, 3 months	N=30; DM type 2	IG: app with monitoring of health parameters; CG: not mentioned	*HbA_{1c}*	Change in HbA$_{1c}$, P=.04, IG: −2.03%, CG: −0.68%
Quinn et al (2011), Maryland, United States [22]	4-arm cluster RCT, 12 months	N=163; DM type 2	IG 1: app allowing patients to enter diabetes self-care data. Web portal that augmented the app. Health providers had access to analyzed patient data; IG 2: as IG 1, but in the Web portal, health providers had access to unanalyzed patient data; IG 3: as IG 2, but the health providers had only access to patient data if the patients chose to share it; CG: usual care	*HbA_{1c}*	Change in HbA$_{1c}$, P=.001, 95% CI change in IG: −2.3 to −1.5, CG: −1.1 to −0.3
Waki et al (2014), Japan [29]	2-arm RCT, 3 months	N=54; DM type 2	IG: app aiming to increase self-management; CG: usual care, continue their self-care regimen	*HbA_{1c}*, Body mass index (BMI)	Change in HbA$_{1c}$, P=.015, IG: −0.4%, CG: 0.1%; NS in other outcomes of interest
Wayne et al (2015), Canada [30]	2-arm RCT, multicenter, 6 months	N=131; DM type 2	IG: app monitoring health parameters; CG: usual care and health coaching	*HbA_{1c}*, Body weight, BMI, Waist circumference, QoL (SF-12)	NS in outcomes of interest.

Reference (year), country	Study design and study duration	Sample size; disease	Intervention group (IG) or control group (CG)	Outcomes of interest[a]	Results[b]
Zhou et al (2016), China [31]	2-arm RCT, 3 months	N=100; DM type 1 and type 2	IG: app monitoring health parameters; CG: usual care	HbA_{1c}, Body weight, BMI, Waist circumference	Change in HbA_{1c}, $P<.01$, IG: −1.95%, CG: −0.79%; NS in other outcomes of interest.

[a]Outcome in italics indicate primary outcome in the study.

[b]Results are reported as difference between groups (*P* value) and as mean change in each group in accordance what is used by the authors.

Table 2. Intervention characteristics.

Smartphone app First author (year)	Logging lifestyle factors	Clinical measurements logging	Monitoring personnel	Education or information	Feedback	Additional support[a]
Holmen et al (2014) [21]	✓[b]	Blood glucose (BG)	Patient	✓	Automatic	✓ (1,3)
Johnston et al (2016) [32]	✓	Blood pressure (BP), BG, Low-density lipoprotein cholesterol, Weight	Patient	✓	Automatic	
Karhula et al (2015) [26]	✓	BP, Weight, BG (diabetics)	Patient, Health-coach	✓	Individualized via telephone every 4-6 weeks	✓ (2,3)
Orsama et al (2013) [27]	✓	BP, Weight, BG	Patient, Study nurses	✓	Automatic, Individualized if warranted	✓ (2)
Quinn et al (2008) [28]	✓	BG	Research team, Patient, Physician	✓	Automatic	✓ (2,4)
Quinn et al (2011) [22]	✓	BG	Patient, Health care provider	✓	Automatic	✓ (2,3)
Waki et al (2014) [29]	✓	BG, BP, Weight	Patient, Research team, Dietitian		Automatic, Individualized	
Wayne et al (2015) [30]	✓	BG, Mood	Patient, Health coach		Individualized	✓ (1,3)
Zhou et al (2016) [31]	✓	BG, BP	Patient, Research team	✓	Individualized	

[a]1: Exercise advice; 2: Patient Web portal; 3: Telephone contact or coaching; 4: Email.

[b]Check mark denotes characteristic is present.

Effect of Smartphone App for Patients With Diabetes With Regard to Glycated Hemoglobin

Seven studies were included in the quantitative synthesis; 3 studies evaluated the effect of apps on short term [29-31], and 4 studies on long term [21,22,26,27]. The overall effect on short term was statistically significant (*P*=.02; Figure 3). The heterogeneity was acceptable with I^2 at 41%. The overall effect on long term was statistically significant (*P*=.009) with no heterogeneity (I^2=0%; Figure 4).

The quality of evidence (GRADE) is presented in Table 3. The quality of evidence in the included studies in short- and long-term effect analysis was scored as low and moderate, respectively. In the short-term effect analysis, the quality was downgraded because of risk of bias and imprecision [29-31]. In the long-term effect analysis, the quality was downgraded to moderate because of imprecision in the estimate of effect [21,22,26,27].

Figure 3. Forest plot: short-term effect on glycated hemoglobin (HbA1c).
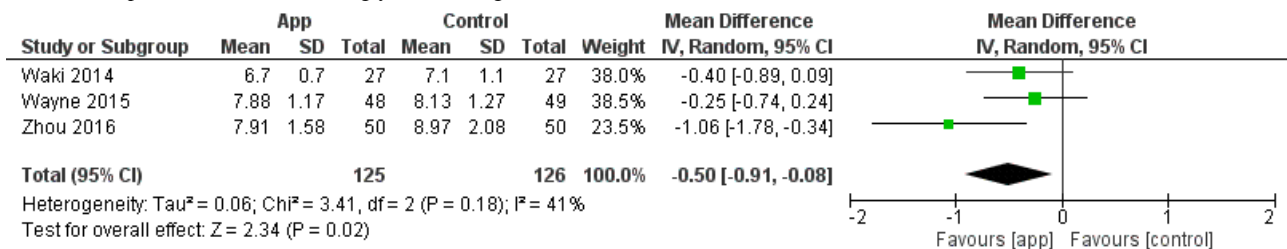
Figure 4. Forest plot: long-term effect on glycated hemoglobin (HbA_{1c}).
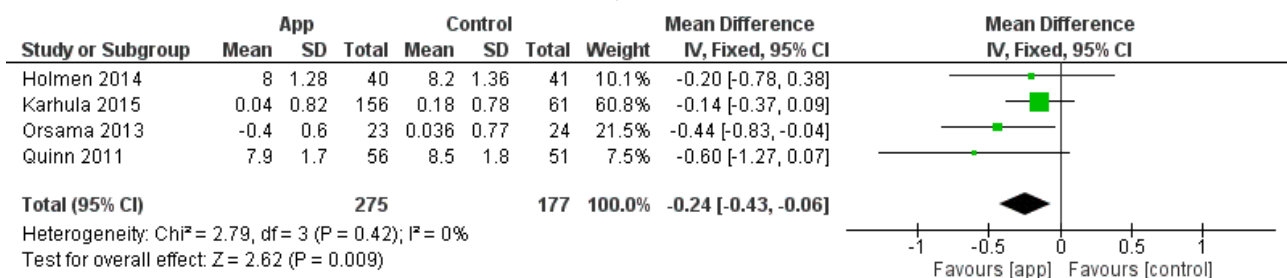
Table 3. Quality of evidence of glycated hemoglobin (HbA_{1c}).

Outcome	Number of participants (number of studies)	Standardized mean differences (95% CI)	Quality of evidence (GRADE[a])
HbA_{1c} short term	251 (3)	−0.50 (−0.91 to −0.08)	Low[b,c]
HbA_{1c} long term	452 (4)	−0.24 (−0.43 to −0.06)	Moderate[d]

[a]GRADE: Grading of Recommendations Assessment, Development, and Evaluation.

[b]Downgraded because of risks of biases (such as attrition bias, blinding, and other bias).

[c]Downgraded because of imprecision (few participants, less than 300).

[d]Downgraded because of imprecision (variation in the estimate of effect).

Discussion

This is, to our knowledge, the first systematic review that examines the effectiveness, for at least 3 months, of apps to promote lifestyle changes for patients with NCD. Eight of 9 studies in this review were performed in persons with DM. In this group, the app showed better effectiveness to improve lifestyle factors than traditional ways to intervene and/or no intervention, especially regarding decrease of HbA_{1c}. Only 2 studies had included persons with CVD, and no differences were found in variables reflecting lifestyle.

A major strength of this review are the authors' attempt to identify all relevant studies by using a comprehensive search strategy in multiple databases led by a research Liberian, as well as well documented methodological strictness performing the systematic review and meta-analyses. In total, five authors participated in this process, which also included hand searching of review paper references to identify additional studies that may have been lost in the initial search. All authors also independently screened the studies for risk of bias. As the search results turned out to be relatively homogeneous, it also was possible to pool the results of one common outcome into two meta-analyses and grade them. However, despite the existence of hundreds of studies involving apps used by CVD, cancer, chronic pulmonary, and/or DM patients, there is a lack of

rigorous trials regarding specific lifestyle outcomes such as PA, physical fitness, modification of dietary habits, and QoL.

Statistical significant improvements between groups on lifestyle factors were reported in 6 of 9 studies (67%). To our knowledge, only one systematic review has previously evaluated the impact of mobile health (mHealth), which WHO has defined as medical and public health practice supported by mobile devices such as mobile phones patient monitoring devices and other wireless devices [33], in more than one chronic disease. They reported significant improvements between groups on disease-specific outcomes in 39% of the 41 included studies [34]. The different results may be explained by different study aims. Although the aim of this review was to study the effectiveness of apps to promote lifestyle changes, the former review aimed to assess the usability, feasibility, and acceptability of mHealth interventions. It is therefore likely that the included studies [34] also were designed to assess usability, feasibility, and acceptability and not necessarily to improve lifestyle and disease-specific outcomes.

To our knowledge, only one systematic review and meta-analyses on the effect of apps to improve HbA_{1c} has previously been conducted [35]. This review included several studies also included in this review [21,22,27-29], but they did not have any exclusion criteria based on follow-up, and the results were pooled into meta-analyses based on methodological

quality. They reported a mean reduction in HbA$_{1c}$ in participants using an app compared with control of 0.49% (95% CI 0.3-0.68; I^2=10%). Studies with fair or good quality showed lower effect compared with studies with poor quality [35]. In the current meta-analyses, the mean reduction in HbA$_{1c}$ in participants using an app compared with controls were 0.50% (95% CI 0.08-0.91; I^2=41%) and 0.24% (95% CI 0.06-0.43; I^2=0%) for short term and long term, respectively.

Despite the fact that the majority of the included studies showed significant efficacy, 3 of the included studies [21,30,32] did not show any effect on outcomes of interest, and significant effect were not found in health-related QoL [21,26,30,32]. One explanation for this might be the fact that the studies did not have enough power to detect such differences, as HbA$_{1c}$ was the primary outcome, and statistical power and the intervention design were based on this. In addition, we should not ignore the fact that it might be with apps similar to other lifestyle interventions, it is hard to actually get a change that lasts over time [6]. This may be what we see as a tendency in our meta-analyses regarding HbA$_{1c}$ as well, where short-term effect is superior to long-term effect.

A recent systematic meta-review evaluated telehealth interventions, which are also regarded as mHealth, to support self-management of long-term conditions [36]. It revealed that most of the research in the field of technology-based interventions is currently conducted in patients with DM, and their results support our findings. Monitoring of blood glucose and feedback improved glycemic control in patients with DM [36]. Meta-analysis on the effects of mHealth in patients with DM have reported a significant reduction in HbA$_{1c}$ of 0.33% [37]. Such interventions may also have a potential to improve well-being in patients with DM type 2, although the results did not reach statistical significance in favor of the intervention [38], which is in line with our results.

The use of mobile technologies and their innovative apps to address lifestyle change in patients with NCD seems to be in its early days, which can explain our limited findings in other NCDs than DM. However, mHealth interventions have been demonstrated as effective to reduce CVD outcomes, body weight, and BMI and to increase adherence to medical therapy, as well as adherence to nonpharmacologic therapy for patients with CVD [18,39]. Telehealth interventions have been demonstrated as potentially effective interventions to improve outcomes in cancer patients [40]. Apps to support self-management in patients with asthma have been pointed as potential effective [41]. Although apps for lifestyle improvement in patients with DM seems to be ahead compared with the other NCDs, we believe that in a few years more studies will exist for CVD, cancer, and chronic pulmonary diseases as well. We screened many studies evaluating apps for CVD, cancer, and chronic pulmonary diseases in the screening phase of this

systematic review; however, most of them were excluded because of follow-up time and outcomes. The reason why apps for DM is major and ahead compared with the other NCDs may be because of difficultness in developing apps that are feasible and with high utility for the more complex NCDs.

App as an intervention can be defined as a complex intervention defined as interventions containing several interacting components [42]. All studies included in his review used apps with several and different components as the main part of the intervention. Most of the studies also had additional support (see Table 2). It can be difficult to understand the cause of any effects of a complex intervention, and therefore, it is crucial to have an idea of the underlying theory of the intervention [42]. In this review, 5 studies [21,22,26,27,30] showed some underlying theory of their intervention. However, it was just one of the included studies that explicitly mentioned their predefined theoretical framework for the intervention [21]. For Internet interventions, it is shown that if a theoretical framework based on several BCTs is incorporated, the interventions are more effective [12]. This may be because different techniques target different stages of a behavioral change process [43]. All studies included in this review used different kinds of feedback and monitoring as BCTs in the app (Table 2). BCTs that have been reported as effective and feasible, especially individualized feedback, have been pointed as being essential to behavioral change and improvement of lifestyle factors [12,44].

Self-management is an important part of the treatment in NCDs. PA, exercise, and a health-promoting diet are the keys to enable a good life while coping with the disease, as well as a possibility to reduce morbidity and mortality [45]. As these are all such important aspects, it is interesting that none of the studies included in this review objectively measured PA or physical fitness.

In conclusion, the results of this study demonstrate that there is limited research of the use of apps for other NCDs than DM with a follow-up of at least 3 months. For DM, the use of apps seems promising to improve lifestyle factors, especially to decrease HbA$_{1c}$. As self-management, including PA and healthy diet, is the key in treatment for all NCDs, it is plausible to believe that such an intervention may also be promising for other NCDs than DM. However, this systematic review clearly indicates a need of further research to evaluate the effect of apps for follow-up for NCDs before implications for practice can be concluded. Especially, there is a need of powered long-term (at least a year) studies for NCDs to be able to evaluate the real effect as NCD patients need to handle their diseases for the rest of their lives. Furthermore, this review reaffirms that future studies must ensure that complex interventions, such as apps, are based on a theoretical framework to bring out the desired behavior change and to understand the impact of the intervention. Finally, appropriate measurements based on the aim of the intervention are always warranted.

XSL•FO

RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy.

[PDF File (Adobe PDF File), 106KB - jmir_v20i5e162_app1.pdf]

References

1. World Health Organization. 2017. Noncommunicable Diseases Progress Monitor, 2017 URL: http://apps.who.int/iris/bitstream/handle/10665/258940/9789241513029-eng.pdf;jsessionid=3B08B426AB7B0F41995FDDDC083DDFE5?sequence=1 [WebCite Cache ID 6yNfMdZJ7]

2. Task Force on diabetes, pre-diabetes, and cardiovascular diseases of the European Society of Cardiology (ESC), European Association for the Study of Diabetes (EASD), Rydén L, Grant PJ, Anker SD, Berne C, et al. ESC guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD-summary. Diab Vasc Dis Res 2014;11(3):133-173. [doi: 10.1177/1479164114525548]

3. Doyle C, Kushi LH, Byers T, Courneya KS, Demark-Wahnefried W, Grant B, 2006 Nutrition, Physical Activity and Cancer Survivorship Advisory Committee, American Cancer Society. Nutrition and physical activity during and after cancer treatment: an American Cancer Society guide for informed choices. CA Cancer J Clin 2006;56(6):323-353. [Medline: 17135691]

4. Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med 2007 Sep 15;176(6):532-555. [doi: 10.1164/rccm.200703-456SO] [Medline: 17507545]

5. Arends J, Bachmann P, Baracos V, Barthelemy N, Bertz H, Bozzetti F, et al. ESPEN guidelines on nutrition in cancer patients. Clin Nutr 2017 Dec;36(1):11-48. [doi: 10.1016/j.clnu.2016.07.015] [Medline: 27637832]

6. Sabaté E. World Health Organization. Adherence to long-term therapies: evidence for action URL: http://www.who.int/chp/knowledge/publications/adherence_full_report.pdf?ua=1 [WebCite Cache ID 6yNi9WBdD]

7. van den Berg MH, Schoones JW, Vliet Vlieland TP. Internet-based physical activity interventions: a systematic review of the literature. J Med Internet Res 2007;9(3):e26. [Medline: 17942388]

8. Simons-Morton DG, Calfas KJ, Oldenburg B, Burton NW. Effects of interventions in health care settings on physical activity or cardiorespiratory fitness. Am J Prev Med 1998 Nov;15(4):413-430. [Medline: 9838981]

9. van der Bij AK, Laurant MG, Wensing M. Effectiveness of physical activity interventions for older adults: a review. Am J Prev Med 2002 Feb;22(2):120-133. [Medline: 11818183]

10. Kahn EB, Ramsay LT, Brownson RC, Heath GW, Howze EH, Powell KE, et al. The effectiveness of interventions to increase physical activity. A systematic review. Am J Prev Med 2002;22(4 Suppl):73-107. [Medline: 11985936]

11. Maher D, Ford N. Action on noncommunicable diseases: balancing priorities for prevention and care. Bull World Health Organ 2011 Aug 1;89(8):547-547A [FREE Full text] [doi: 10.2471/BLT.11.091967] [Medline: 21836749]

12. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4. [doi: 10.2196/jmir.1376] [Medline: 20164043]

13. Powell-Cope G, Nelson AL, Patterson ES. Patient Care Technology and Safety. In: Hughes RG, editor. Patient Safety and Quality: An Evidence-Based Handbook for Nurses. Rockville (MD): Agency for Healthcare Research and Quality; 2008.

14. 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]

15. Beatty AL, Fukuoka Y, Whooley MA. Using mobile technology for cardiac rehabilitation: a review and framework for development and evaluation. J Am Heart Assoc 2013;2(6):e000568. [doi: 10.1161/JAHA.113.000568]

16. Cajita MI, Gleason KT, Han HR. A systematic review of mHealth-based heart failure interventions. J Cardiovasc Nurs 2016;31(3):E10-E22 [FREE Full text] [doi: 10.1097/JCN.0000000000000305] [Medline: 26544175]

17. Cotter AP, Durant N, Agne AA, Cherrington AL. Internet interventions to support lifestyle modification for diabetes management: a systematic review of the evidence. J Diabetes Complications 2014;28(2):243-251. [doi: 10.1016/j.jdiacomp.2013.07.003] [Medline: 24332469]

18. Gandhi S, Chen S, Hong L, Sun K, Gong E, Li C, et al. Effect of mobile health interventions on the secondary prevention of cardiovascular disease: systematic review and meta-analysis. Can J Cardiol 2017 Feb;33(2):219-231. [doi: 10.1016/j.cjca.2016.08.017] [Medline: 27956043]

19. Goode AD, Lawler SP, Brakenridge CL, Reeves MM, Eakin EG. Telephone, print, and Web-based interventions for physical activity, diet, and weight control among cancer survivors: a systematic review. J Cancer Surviv 2015 Dec;9(4):660-682. [doi: 10.1007/s11764-015-0442-2]

20. Marzano L, Bardill A, Fields B, Herd K, Veale D, Grey N, et al. The application of mHealth to mental health: opportunities and challenges. Lancet Psychiatry 2015 Oct;2(10):942-948. [doi: 10.1016/S2215-0366(15)00268-0] [Medline: 26462228]

21. Holmen H, Torbjørnsen A, Wahl AK, Jenum AK, Småstuen MC, Årsand E, et al. A mobile health intervention for self-management and lifestyle change for persons with type 2 diabetes, Part 2: one-year results from the Norwegian randomized controlled trial RENEWING HEALTH. JMIR mHealth and uHealth 2014;2(4):e57. [doi: 10.2196/mhealth.3882] [Medline: 25499872]

22. Quinn CC, Shardell MD, Terrin ML, Barr EA, Ballew SH, Gruber-Baldini AL. Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control. Diabetes Care 2011;34(9):1934-1942. [doi: 10.2337/dc11-0366] [Medline: 21788632]

23. Higgins JPT, Altman DG, Sterne JAC. Assessing risk of bias in included studies. In: Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. Monash University, Australia: The Cochrane Collaboration, 2011; 2011.

24. Cochrane Community. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration; 2014. RevMan 5 URL: http://community.cochrane.org/tools/review-production-tools/revman-5 [WebCite Cache ID 6ye5kc25L]

25. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336(7650):924-926. [doi: 10.1136/bmj.39489.470347.AD] [Medline: 18436948]

26. Karhula T, Vuorinen AL, Rääpysjärvi K, Pakanen M, Itkonen P, Tepponen M, et al. Telemonitoring and mobile phone-based health coaching among Finnish diabetic and heart disease patients: randomized controlled trial. J Med Internet Res 2015 Jun 17;17(6):e153 [FREE Full text] [doi: 10.2196/jmir.4059] [Medline: 26084979]

27. Orsama AL, Lähteenmäki J, Harno K, Kulju M, Wintergerst E, Schachner H, et al. Active assistance technology reduces glycosylated hemoglobin and weight in individuals with type 2 diabetes: results of a theory-based randomized trial. Diabetes Technol Ther 2013 Aug;15(8):662-669. [doi: 10.1089/dia.2013.0056] [Medline: 23844570]

28. Quinn CC, Clough SS, Minor JM, Lender D, Okafor MC, Gruber-Baldini A. WellDoc mobile diabetes management randomized controlled trial: change in clinical and behavioral outcomes and patient and physician satisfaction. Diabetes Technol Ther 2008 Jun;10(3):160-168. [doi: 10.1089/dia.2008.0283] [Medline: 18473689]

29. Waki K, Fujita H, Uchimura Y, Omae K, Aramaki E, Kato S, et al. DialBetics: a novel smartphone-based self-management support system for type 2 diabetes patients. J Diabetes Sci Technol 2014 Mar;8(2):209-215 [FREE Full text] [doi: 10.1177/1932296814526495] [Medline: 24876569]

30. Wayne N, Perez DF, Kaplan DM, Ritvo P. Health coaching reduces HbA1c in type 2 diabetic patients from a lower-socioeconomic status community: a randomized controlled trial. J Med Internet Res 2015;17(10):e224. [doi: 10.2196/jmir.4871] [Medline: 26441467]

31. Zhou W, Chen M, Yuan J, Sun Y. Welltang - a smart phone-based diabetes management application - improves blood glucose control in Chinese people with diabetes. Diabetes Res Clin Pract 2016 Jun;116:105-110. [doi: 10.1016/j.diabres.2016.03.018] [Medline: 27321324]

32. Johnston N, Bodegard J, Jerstrom S, Akesson J, Brorsson H, Alfredsson J, et al. Effects of interactive patient smartphone support app on drug adherence and lifestyle changes in myocardial infarction patients: a randomized study. Am Heart J 2016 Aug;178:85-94. [doi: 10.1016/j.ahj.2016.05.005] [Medline: 27502855]

33. WHO Global Observatory for eHealth. In: Kay M, Santos J, Takane M, editors. mHealth: New horizons for health through mobile technologies. Switzerland: World Health Organization; 2011.

34. 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;17(2):e52. [doi: 10.2196/jmir.3951] [Medline: 25803266]

35. Can H, Carter B, Hewitt J, Francisa T, Mayor S. Do mobile phone applications improve glycemic control (HbA1c) in the self-management of diabetes? A systematic review, meta-analysis, and GRADE of 14 randomized trials. Diabetes Care 2016 Nov;39(11):2089-2095. [doi: 10.2337/dc16-0346] [Medline: 27926892]

36. 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. Send to J Med Internet Res 2017;19(5):e172. [doi: 10.2196/jmir.6688] [Medline: 28526671]

37. Alharbi NS, Alsubki N, Jones S, Khunti K, Munro N, de Lusignan S. Impact of information technology-based interventions for type 2 diabetes mellitus on glycemic control: a systematic review and meta-analysis. J Med Internet Res 2016;18(11):e310. [doi: 10.2196/jmir.5778] [Medline: 27888169]

38. Hadjiconstantinou M, Byrne J, Bodicoat DH, Robertson N, Eborall H, Khunti K, et al. Do web-based interventions improve well-being in type 2 diabetes? A systematic review and meta-analysis. J Med Internet Res 2016;18(10):e270. [doi: 10.2196/jmir.5991] [Medline: 27769955]

39.   Widmer RJ, Allison TG, Lerman LO, Lerman A. Digital Health Intervention as an Adjunct to Cardiac Rehabilitation Reduces Cardiovascular Risk Factors and Rehospitalizations. J Cardiovasc Transl Res 2015 Jul;8(5):283-292 [FREE Full text] [doi: 10.1007/s12265-015-9629-1] [Medline: 25946990]

40.   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;17(3):e65. [doi: 10.2196/jmir.4009] [Medline: 25793945]

41.   Marcano Belisario JS, Huckvale K, Greenfield G, Car J, Gunn LH. Smartphone and tablet self management apps for asthma. Cochrane Database Syst Rev 2013(11):CD010013. [doi: 10.1002/14651858.CD010013.pub2] [Medline: 24282112]

42.   Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. Int J Nurs Stud 2013 May;50(5):587-592. [doi: 10.1016/j.ijnurstu.2012.09.010] [Medline: 23159157]

43.   Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]

44.   Ferrier S, Blanchard CM, Vallis M, Giacomantonio N. Behavioural interventions to increase the physical activity of cardiac patients: a review. Eur J Cardiovasc Prev Rehabil 2011;18(1):15-32. [doi: 10.1097/HJR.0b013e32833ace0e] [Medline: 20502341]

45.   Global action plan for the prevention and control of noncommunicable diseases 2013-2020. Geneva, Switzerland: World Health Organization; 2013.

Abbreviations

BCT: behavioral change technique
BMI: body mass index
CVD: cardiovascular disease
DM: diabetes mellitus
GRADE: Grading of Recommendations Assessment, Development, and Evaluation
HbA$_{1c}$: glycated hemoglobin
mHealth: mobile health
NCD: noncommunicable disease
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QoL: quality of life
WHO: World Health Organization

Original Paper

Exploring Hospitals' Use of Facebook: Thematic Analysis

Nima Kordzadeh[1], PhD; Diana K Young[2], PhD

[1]Robert A Foisie School of Business, Worcester Polytechnic Institute, Worcester, MA, United States
[2]Department of Finance and Decision Science, Trinity University, San Antonio, TX, United States

Corresponding Author:
Nima Kordzadeh, PhD
Robert A Foisie School of Business
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA, 01609
United States
Phone: 1 5088315218
Fax: 1 5088315720
Email: nkordzadeh@wpi.edu

Abstract

Background: Although health care organizations such as hospitals and clinics have widely embraced social media as a means to educate the community on health topics and increase patient loyalty and satisfaction, little is known about the content these organizations actually share when using social media channels.

Objective: This study aimed to explore the types of content US hospitals post on their Facebook pages and how hospitals' Facebook activities differ with regard to content types.

Methods: We collected and thematically analyzed more than 1700 Facebook posts made over a 3-month period by 17 US hospitals. During the first phase, the 2 researchers coded a set of 159 posts and created an initial thematic web of content. During the second phase, the researchers coded the remaining posts and then revised, refined, and validated the initial web of content accordingly. Coding consensus was achieved on 1184 of the 1548 analyzable sampled posts (76.49%).

Results: We identified a list of 13 unique health social media post themes and classified those themes into 3 thematic groups that included *announcing*, *sharing*, and *recognizing* activities. The most frequently used theme was *sharing health information*, which appeared in 35.81% (424/1184) of the posts analyzed. Such posts sought to provide health tips and advice to community members. *Recognizing special days* and *recognizing employees* were the second and third most frequently used themes, respectively, with 14.95% (177/1184) and 11.82% (140/1184) of the posts containing those themes. The frequency of these themes was surprising as the content was geared more toward stakeholders internal to the organization, although most previous literature has focused on social media as a tool to connect with external stakeholders. In addition, we found many of the posts involved more than one theme, and selected sets of themes co-occurred frequently. For example, 25.4% (45/177) of the posts *recognizing special days* also included content to *share health information,* and approximately 38% (32/85) of the posts *announcing research activities* also included content to *share health information*. Finally, we found similarities and differences between the sampled hospitals in terms of the types of content they posted more frequently on their Facebook pages.

Conclusions: Hospitals use Facebook as an inexpensive way to educate people on health and wellness topics and to communicate different types of information and news to the public audience. Hospitals and clinics that are expanding their social media activities or are starting to embark on social media strategies can use the results of this study to better formulate their activities on Facebook.

(J Med Internet Res 2018;20(5):e190) doi:10.2196/jmir.9549

Introduction

Over the past decade, social media has infiltrated the health care industry [1]. Hospitals, clinics, and other health care providers use social media channels such as Facebook and Twitter to communicate with their patients and the surrounding community. The main goals of using social media platforms by health care providers such as hospitals and clinics include

disseminating health information, tips, and advice; promoting health care services and products; managing brand recognition and reputation; and strengthening ties with their audience [2-5]. Some of these goals such as brand management and recognition are common with organizations in other sectors such as travel and tourism, fashion, and restaurants [6,7]. Other goals such as educating the community about health issues, promoting monetary and organ donation, and sharing patient success stories are specific to the health care domain.

As of April 2018, over 1600 US-based hospitals maintained officially sponsored social media accounts [8]. For example, Johns Hopkins Hospital, a world-class institution based in Baltimore, Maryland, has more than 580,000 Facebook followers [9], whereas the organization's YouTube channel and Twitter page have more than 47,000 subscribers [10] and 500,000 followers [11], respectively. Use of social media by European-based health care organizations has also increased drastically. A mere 10% of 873 European hospitals sampled in 2010 reported having an institutionally sponsored Facebook account [12], although the following year, over 67% of those institutions had such accounts.

Health care institutions use social media platforms for different purposes. In a survey of 36 US hospitals and health systems conducted by Computer Sciences Corporation's (CSC) Global Institute for Emerging Healthcare Practices in 2012 [5], respondents were asked *for what purpose does your organization use social media* and *what are your organization's primary objectives in using social media*? The top responses to the first question were to promote wellness and healthy behaviors as well as marketing services or products. The top responses to the second question included to engage patients or consumers, build greater brand recognition, and attract new customers.

Despite the number of studies concerning how health care organizations use social media, little is known about the content of the posts hospitals make on their social media pages. Among the very few studies that have addressed these questions, Richter et al [13] analyzed hospitals' Facebook fan page posts to determine which of those institutions had postings related to patient education, staff discussion, staff awards, hospital awards, and consumer engagement. The results showed Facebook was primarily used for educational purposes (91%), followed by staff discussion (76%) and staff awards (63%). These categories used did not emerge from the data but were considered a priori knowledge. In addition, Richter et al [13] employed a binary coding schema to show whether or not there was at least one post related to each of the aforementioned categories. Therefore, their findings did not show what proportion of posts was related to each of the categories. For instance, having one post related to hospital awards was treated the same as having 100 posts related to the category. To fill those gaps and to provide more insights into the actual use of social media platforms in the health care domain, we conducted this study. Our main objectives are as follows: (1) understanding the main content themes of the posts made by major US hospitals on Facebook and (2) examining the differences between hospitals in terms of the types of content they post on their Facebook pages.

Methods

Overview

We employed a two-phased thematic analysis to identify the types of content health care institutions post on their institutionally sponsored Facebook pages. Thematic analysis is a qualitative method used to identify, analyze, and report on patterns found in text data [14-16]. Use of the method requires "careful reading and rereading of the data" [17] to identify explicit and implicit meaning embedded within the text [18]. Thematic analysis has been widely employed in health care and medical informatics research. Kneafsey et al [19] used the method to analyze health care professionals' compassion perceptions to reveal 4 overarching components of the concept. Holm and Severinsson [20] used thematic analysis to identify key patterns found in geriatric patients' narratives regarding their experiences of surviving with depression. Hickey et al [21] employed thematic analysis in their mixed-method, longitudinal study of factors that influence nurses' career choices and aspirations. Finally, Amann and Rubinelli [22] used thematic analysis to understand the views of community managers on knowledge cocreation in Web-based health communities for people with disabilities.

The thematic analysis process involves 6 steps with iteration allowed between steps [14,23,24]. The process begins with the research team becoming familiar with the data to generate an initial set of themes to code the data. Next, the researchers read and reread the text data to tag individual elements with theme codes. When using the method inductively, tagging can be a highly iterative activity with the researchers actively updating and revising the list of theme codes and data elements throughout the search process [18]. Once all data elements have been tagged, the final list of themes are reviewed and corresponding names and definitions developed. Finally, a graphical thematic web of content is developed to visually depict the relationship between the data and the final list of identified themes. In this study, we performed 2 phases of thematic analysis. In the first phase, we developed an initial thematic web of content using a small sample of Facebook posts. In the second phase, we used a larger sample to validate the thematic web developed in the first phase of the study.

Sampling and Data Collection

To identify the types of content shared by health care institutions, we first collected the Facebook posts made during a 3-month period (March 2014 to May 2014) by a set of highly ranked, US-based hospitals, selected from the US News Best Hospitals 2013-2014 list. We chose the 3-month time window because a shorter window could raise the possibility of month-specific patterns in the data, which would decrease the generalizability of the results. Conversely, a longer window would result in a much larger dataset, making the content analysis process significantly harder and more complex. It would also potentially increase the likelihood of errors and time to publication of the results.

From the US News Best Hospitals 2013-2014 list, we selected the top 10 institutions from each category. We then eliminated duplicate institutions and removed all institutions offering

services in a limited health care domain such as psychiatry. We felt this last step was necessary as we were most interested in how general care hospitals use social media and believed specialty institutions would be more specialty topic focused. This resulted in a final list of 54 general medicine hospitals.

Next, we searched Facebook to determine which of those institutions offered an institutionally sponsored Facebook page. A total of 28 hospitals met that criteria; however, there was wide diversity in the popularity of the institutions' Facebook main pages. Three institutions, including Cleveland Clinic, Mayo Clinic, and Johns Hopkins Hospital, had very popular main pages with 1,008,069; 517,519; and 205,115 main page likes, respectively. Seventeen institutions had between 5000 and 65,000 main page likes, and the remaining 8 institutions had fewer than 5000 main page likes.

To ensure we compared data from similar organizations, we chose to ignore the 3 institutions with highly popular main pages and the institutions with fewer than 5000 main page likes. We included, instead, the 17 organizations with between 5000 and 65,000 main page likes in our sample frame. We chose these thresholds because fewer than 5000 likes may indicate the organization is not following a systematic social media strategy

and hence may not be used as a reliable benchmark for other medical institutions. Moreover, Mount Sinai Medical Center with nearly 65,000 page likes was the most popular page after Mayo Clinic's main page with over 500,000 likes. Thus, we felt the difference between those two represented a natural cut-off point in the data. Table 1 presents information about the sampled hospitals and the data collected from each hospital's Facebook page.

Data Analysis

Phase 1

For the first phase of the study, our goal was to identify an initial list of themes that represented the intent or purpose of the posts made by hospitals on their institutionally sponsored Facebook pages. To do so, we randomly selected 5 of the 17 organizations included in our sample frame. The selected institutions included Mount Sinai Medical Center, Rush University Medical Center, Massachusetts General Hospital, Yale-New Haven Hospital, and National Jewish Health. These 5 hospitals made a total of 159 posts between March 1, 2014, and April 30, 2014, on their Facebook pages. Following the thematic analysis steps proposed by Braun and Clarke [14] and widely adopted in the extant literature (eg, [22-24]), we analyzed the 159 posts.

Table 1. Sample characteristics.

Hospital's name[a]	Number of page likes as of the data collection date	Number of posts			
		March 2014	April 2014	May 2014	Total
Mount Sinai Medical Center	64,932	21[b]	22[b]	54	97
Barnes-Jewish Hospital Washington University	62,936	27	43	37	107
New York-Presbyterian University Hospital of Columbia and Cornell	42,159	23	36	47	106
Rush University Medical Center	30,254	16[b]	17[b]	13	46
University of Pittsburgh Medical Center	27,337	39	53	58	150
Stanford Hospital and Clinics	26,316	14	15	14	43
Brigham and Women's Hospital	20,109	36	45	30	111
National Jewish Health, Denver-University of Colorado Hospital	19,074	5[b]	18[b]	54	77
Shepherd Center	15,153	31	35	36	102
Cedars-Sinai Medical Center	12,300	25	25	32	82
Yale-New Haven Hospital	10,540	11[b]	16[b]	28	55
Northwestern Memorial Hospital	9452	47	84	58	189
Thomas Jefferson University Hospital	8326	74	65	73	212
Massachusetts General Hospital	7574	8[b]	25[b]	20	53
NYU Langone Medical Center	7185	26	31	28	85
Florida Hospital	6977	40	34	27	101
Magee-Women's Hospital of UPMC	5622	36	37	33	106
Total	*376,246*	*479*	*601*	*642*	*1722*

[a]The hospitals are ordered in the table based on the number of their followers (ie, page likes) on Facebook.

[b]The numbers are associated with the posts analyzed during the first phase of the study.

XSL•FO
RenderX

Figure 1. Thematic web of content (phase 1).

We first familiarized ourselves with the sample data and independently generated initial lists of theme codes. We then met, compared our lists, discussed differences, and settled upon an initial set of codes to use when tagging the sampled data elements. In addition, we agreed each researcher could add codes to the list, and we would discuss those additions later. Next, we each independently reviewed each of the sample data elements and inspected any photos or videos included in the post to discern meaning. We then tagged it with up to 3 theme codes believed to be relevant to the post. We met on 2 separate occasions during the coding process to compare our updated lists of theme codes, resolve differences, and discuss the processes we each employed. During the early course of those discussions, we agreed to shift away from applying a strictly semantic lens (word meaning) when analyzing the data elements and toward a more interpretive lens (statement intent) [25]. This was necessary, as intent was not always clearly conveyed in just the post text. For example, the text of one post announced the date and time of an upcoming social event; however, the attached image contained a graphic indicating the event was sponsored by an organization concerned with organ donation. An interpretive lens was necessary to induce meaning from the photo, which ultimately resulted in coding the post as both *announcing a social event* and *promoting organ donation*.

We made several iterations through the entire sample. Once we reached consensus on the final set of codes and coding of all data elements, we developed names and definitions for each

theme code. We then spent time analyzing the individual theme codes to identify similarities in purpose and intent to group the codes into theme content categories and subcategories. Finally, we used that information to develop our thematic web of content (Figure 1).

Phase 2

According to Lincoln and Guba [26], the validity and soundness of qualitative research is determined based on the credibility, transferability, and confirmability of the results. In the context of thematic analysis, these criteria can be achieved by conducting a thematic content analysis on a second dataset to confirm the findings of the first round of analysis [26]. Thus, to enhance the overall robustness of the phase 1 thematic web of content and to validate the list of content themes identified, we conducted a second thematic analysis on the remaining posts made by the 17 hospitals between March 1, 2014, and May 30, 2014. As such, for the 12 institutions not included in the first phase, we analyzed all posts made during the 3-month sample period. For the 5 institutions included in the first phase of the study, we only analyzed the posts made in May 2014. As such, in the second phase, we analyzed 1563 Facebook posts.

Over the course of several months and using the theme categories and subcategories identified during the first phase of the study, we each independently reviewed and tagged each data element in the new sample with up to 3 theme codes deemed to be most relevant to the post intent. Again, an

interpretive approach was applied with each of us reading every post multiple times to induce the underlying purpose before assigning codes. In cases where URL links, photos, or videos were included, we accessed those links and reviewed the multimedia content to ensure any meaning included in those items was considered. As potential new codes were identified during the coding process, we tagged the corresponding elements with the *Other* code and a note specifying a possible code addition.

At 3 points during the phase 2 coding process, we met to discuss our progress, issues we had encountered, and any elements tagged with the *Other* code. During those discussions, we found that no elements were consistently tagged with the *Other* code by both researchers. Accordingly, no additional codes were added to the agreed upon list. Again, we applied a highly iterative process in which individual data elements were reviewed multiple times by each researcher.

Results

Phase 1

At the end of the first phase, we had identified 33 unique theme codes. Of the 159 posts, we did not code 6 posts as they were cover photo changes and did not contain any textual content. We tagged most posts (122) with a single theme code, whereas 29 posts were tagged with 2 theme codes and 2 posts were tagged with 3 theme codes. Review, comparison, and grouping of the individual theme codes resulted in a list of 11 theme categories: *donation information, events, research activities, organizational news, recognizing special days, recognizing employees, recruiting employees, recruiting volunteers, sharing patient success stories, sharing health career information,* and *sharing health information.* Moreover, 3 of the 11 categories could be further broken into multiple subcategories. We identified 2 subcategories within the donation information category, 5 subcategories within the event category, and 3 subcategories within the organizational news category. We further organized the 11 theme categories into 4 theme purpose groups, which included *announcing and reporting, sharing, recognizing,* and *recruiting activities.*

We found that 2 of the individual theme codes (*issue debate* and *nonmedical product promotion*) did not fit within any of the identified theme categories. A single post was coded as *issue debate* and 2 posts were coded as *nonmedical product promotions.* The *issue debate* post was related to the pediatric immunization issue. The *nonmedical product promotion* posts concerned t-shirt sales at a single hospital. As we did not believe either of those codes were of significant importance for our study, we did not include them in the thematic web of content but did retain the codes for use in phase 2 of the study. The *announcing and reporting* theme purpose group involved posts communicating information concerning *donations,* upcoming or past *events,* institution *research activities,* and *organizational news.* A total of 94 of the collected posts (59%) involved *announcing or reporting* activities. Table 2 lists the category and subcategory themes identified within the *announcing and reporting* purpose group and provides sample posts for each of the subcategories and categories identified within that group.

The *sharing* purpose group involved posts disseminating information concerning *health information, health careers information,* and *patient success stories.* A total of 44 posts (28%) involved *sharing information.* The *recognizing* purpose group involved posts *acknowledging employees* and *special days.* Furthermore, 42 posts (26%) involved *recognition* activities. Finally, the *recruiting* purpose group involved posts soliciting *new employees and volunteers.* Only 3 posts (2%) involved *recruiting* activities. Table 3 lists the theme categories identified within the *sharing, recognizing,* and *recruiting* purpose groups and provides sample posts for each category.

Figure 1 presents the thematic web of content we developed during phase 1 of the study. This model shows the relationships between the identified theme purpose groups, categories, and subcategories.

Phase 2

We began phase 2 with the 20 theme codes identified during the first phases of the study plus the *Other* code for a total of 21 theme codes. These theme codes represented all subcategories and single-dimensional categories included in the phase 1 thematic web, as well as the *nonmedical product promotion,* and the *issue debate* theme codes that were not included because of low representativeness. During the coding process, we found we were consistently assigning the *organizational news-achievements* and the *organizational news-new facilities* concurrently. This was because new facilities announcements were made in conjunction with new services that were promoted as institutional achievements. Accordingly, we decided to combine the 2 codes into a single theme code midway through the coding process.

Of the 1563 posts analyzed, 15 posts were not assigned any theme code, as they were cover photo changes and did not contain analyzable content. The first researcher tagged 988 posts with a single theme code, 487 posts with 2 theme codes, and 73 posts with 3 theme codes. The second researcher tagged 1063 posts with a single theme code, 415 posts with 2 theme codes, and 70 posts with 3 theme codes. We achieved coding consensus on 1184 of the sampled posts (76.49%). We calculated interrater reliability based on Cohen kappa statistic using a weighted average of each theme code's kappa value [27]. The resulting reliability was measured at 73.44%, (κ=.7344), which is considered a very good level of interrater reliability [28].

As our goal was to generate a validated set of content theme codes the 2 researchers achieved a consensus upon, we felt it was important to proceed with only the data elements where we both agree on their themes. Accordingly, we eliminated the 337 data elements to which we assigned different codes. This reduced the analyzable sample size to 1184 data elements. Next, we conducted frequency analysis on this reduced sample to determine how representative each of the 20 theme codes was. In line with our phase 1 results, the *sharing-health information* code was assigned most frequently, with 35.81% of the data elements representing this theme code. *Recognizing-special days* and *recognizing-employees* were the second and third most frequently assigned codes, representing 14.95% and 11.82% of the data elements, respectively. This too was in line with our phase 1 results where *recognizing-special days* was also the

second most frequently assigned code and *recognizing-employees* was the fifth most frequently assigned

code. Figure 2 shows the percent frequency of each of the content theme codes assigned during the phase 2 analysis.

Table 2. Announcing and reporting theme categories, subcategories, and example post excerpts.

Category	Count	Example post excerpt
Donation		
Monetary	5	*Thank you New York Rangers' Brad Richards for your generous gift of US $25,000 to support our Pediatric #Palliative Care Program!*
Organ	6	*What does it feel like to save a life? At our #OrganDonor Appreciation Ceremony, donors shared personal stories of why they gave the #GiftofLife.*
Events		
Charity	6	*Are you a Tough Mudder? Learn how you can participate in and support the Student Veteran Society from Columbia College's Paved in Mud campaign to benefit the Center for Veterans and Their Families at Rush.*
Health information	17	*Arthritis in your knee can make even simple daily tasks hard to complete. Join an orthopedic surgeon and rheumatologist at Rush on April 9 to learn about nonsurgical and surgical treatment options.*
Health screening	9	*Join our #HealthFair this Saturday from 10 AM -3 PM for FREE health screenings including #HepC, #Diabetes, #BloodPressure, and #Cholesterol.*
Social	16	*National Jewish Health and the LA Professional Services Black and White Ball presented by Debbie & Stu Steinberg and Paul Zaffaroni.*
Support groups	1	*CPAP Support Group May 10 at 10 AM in MDT National Jewish Health in Denver, Colorado.*
Research activities	9	*A research team is investigating whether the body's own immune system can be encouraged to mount a defense against #cancer before healthy tissue is damaged.*
Organizational news		
New facilities	7	*Thank you to everyone who celebrated the opening of our new #Maternal#Fetal Medicine Center and #Pediatric Specialty Center expansion at One Long Wharf in New Haven last night.*
Services	8	*Here is the Department of Maternal-Fetal Medicine's Photo of the Month! Our MFM specialists and registered sonographers use state-of-the-art 3D or 4D ultrasounds*
Achievements	10	*Mass General is proud to again be listed on DiversityInc's list of Top 10 Hospital Systems. DiversityInc, a publication about diversity and business, ranked Mass General 7th on its annual list of top employers, noting the hospital's commitment to mentoring minority physicians and nurses. For more information on the DiversityInc list, visit link*

Table 3. Sharing, recognizing, and recruiting theme categories and example post excerpts.

Theme purpose and category	Count	Example post excerpt
Sharing		
Health information	31	*How much fiber is on your plate? You need 25-35 g/day. Rethink your #salad*
Patient success and feel good stories	11	*Amazing story about a 3-year old who beat the odds following a failed kidney transplant.*
Health career information	2	*What is a Child Life specialist? Learn more: link*
Recognizing		
Employees	28	*YNHH's Friedlaender earns top leadership honor from American Academy of#Orthopaedic #Surgeons. http://ow.ly/uAGHV*
Special days	14	*Everyone loves to get a thank you note, including your doctor. You can express your gratitude by sending your doctor at Rush a thank you eCard and making a gift in his or her honor in celebration of National Doctor's Day, which is Sunday, March 30.*
Recruiting		
Volunteers	2	*National Jewish Health is currently seeking volunteers who are looking for a challenging opportunity to use their retail and/or professional skills in Nan and Dollie's Gift Shop.*
Employees	1	*Do you want to become a part of Mount Sinai and help transform the #healthcare landscape? Register for our #Research #OpenHouse in May*

XSL•FO

RenderX

Figure 2. Theme frequency distribution by percentage.
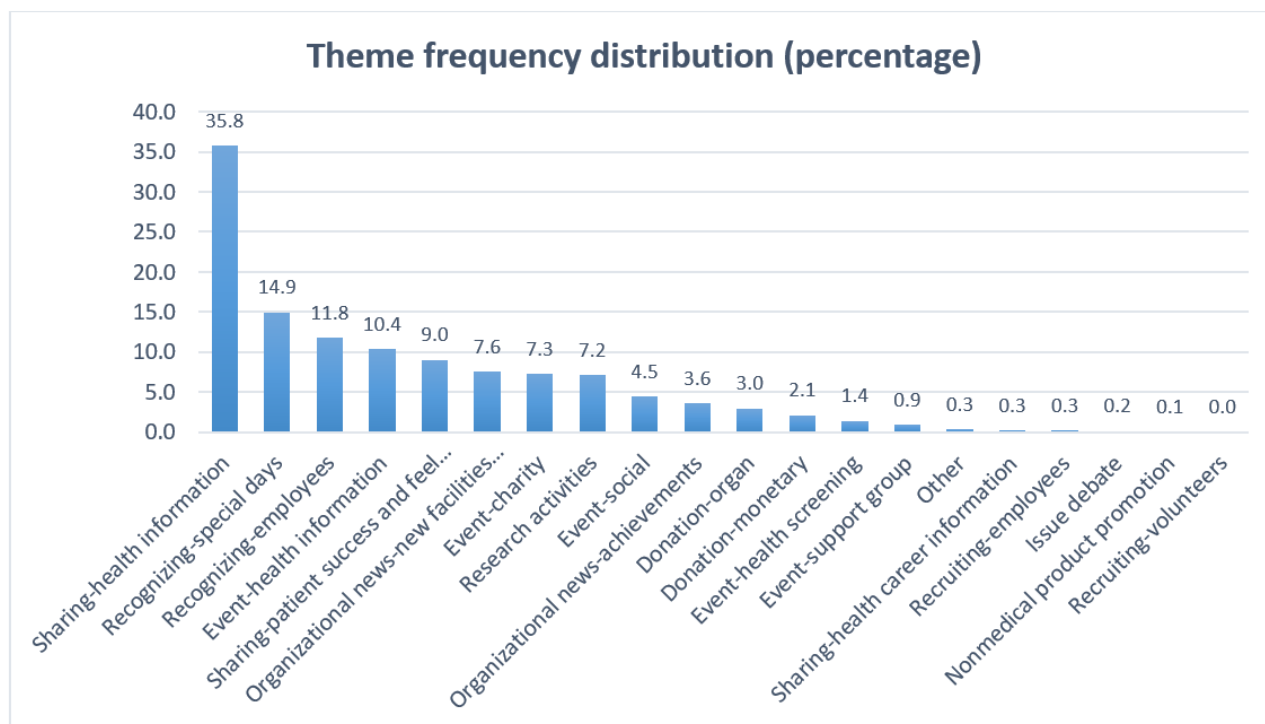
Figure 3. Thematic web of content (phase 2).

As 6 of the identified theme codes each represented less than 1% of the posts in the sample, we decided those codes were not representative of the entire sample and should not be included in our validated thematic web of content. Accordingly, we dropped the *announcing and reporting events-support groups, sharing-health career information, recruiting employees, issue debate, nonmedical product promotion,* and *recruiting volunteers* theme codes. In addition, we agreed to drop the *Other*

code as no distinct themes arose from the items tagged with that code. Finally, we updated the thematic web of content to reflect this validated set of theme code categories and subcategories. Figure 3 presents the revised model.

Next, we examined the frequency distribution of the post themes for each hospital and compared the top 3 themes (in terms of frequency of use) that each of the 17 hospitals included in their Facebook posts. The results (Table 4) demonstrated *sharing-health information* was consistently among the top 3 most frequently used themes by the hospitals. The only exception was New York-Presbyterian Hospital that had used the theme in only 6 of the 55 posts; thus, it was not among that hospital's top 3 themes.

Table 4. Most frequently used themes by each hospital.

Hospital[a]	Total number of posts analyzed in phase 2	Theme and n (%) of posts containing the theme		
		First most frequent	Second most frequent	Third most frequent
Mount Sinai Medical Center	38[b]	Sharing health information, 9 (24)	Recognizing employees, 9 (24)	Announcing or reporting event-health information, 6 (16)
Barnes-Jewish Hospital Washington University	68	Sharing health information, 22 (32)	Recognizing special days, 19 (28)	Sharing patient success stories, 16 (24)
New York-Presbyterian University Hospital of Columbia and Cornell	55	Announcing or reporting events-health information, 10 (18)	Recognizing special days, 9 (16)	Sharing patient success stories, 7 (13)
Rush University Medical Center	10[b]	Recognizing special days, 5 (50)	Sharing health information, 4 (40)	Recognizing employees, 3 (30)
University of Pittsburgh Medical Center	111	Sharing health information, 52 (46.9)	Recognizing special days, 20 (18.0)	Announcing or reporting organ donation and recognizing employees, 9 (8.1)
Stanford Hospital and Clinics	35	Sharing health information, 12 (34)	Recognizing special days, 10 (29)	Sharing patient success stories, 3 (9)
Brigham and Women's Hospital	92	Sharing health information, 38 (41)	Announcing or reporting events-charity, 11 (12)	Announcing or reporting research activities, 11 (12)
National Jewish Health, Denver-University of Colorado Hospital	54[b]	Sharing health information, 22 (41)	Announcing or reporting research activities, 9 (17)	Recognizing employees, 4 (7)
Shepherd Center	84	Sharing patient success stories, 23 (27)	Sharing health information, 17 (20)	Announcing or reporting event-charity, 14 (17)
Cedars-Sinai Medical Center	68	Sharing health information, 37 (54)	Announcing or reporting research activities, 8 (12)	Announcing or reporting events-health information, 7 (10)
Yale-New Haven Hospital	27[b]	Sharing health information, 10 (37)	Recognizing special days, 6 (22)	Recognizing employees, 5 (19)
Northwestern Memorial Hospital	139	Sharing health information, 42 (30.2)	Recognizing special days, 21 (15.1)	Sharing patient success stories, 17 (12.2)
Thomas Jefferson University Hospital	175	Sharing health information, 68 (51.1)	Recognizing employees, 38 (21.7)	Recognizing special days, 26 (14.9)
Massachusetts General Hospital	12[b]	Sharing health information, 6 (50)	Announcing or reporting events-charity, 2 (17)	Recognizing special days, 1 (8)
NYU Langone Medical Center	68	Sharing health information, 21 (30)	Recognizing employees, 11 (16)	Announcing or reporting research activities, 11 (16)
Florida Hospital	70	Sharing health information, 24 (34)	Announcing or reporting events-health information, 18 (26)	Recognizing special days, 11 (16)
Magee-Women's Hospital of UPMC	78	Sharing health information, 34 (44)	Recognizing special days, 22 (28)	Recognizing employees, 15 (19)

[a]The hospitals are ordered in the table based on the number of their followers (ie, page likes) on Facebook.

[b]The numbers are associated with the posts made only in May 2014 because the March and April posts of those hospitals were analyzed during the first phase of the study.

Table 5. Theme co-occurrence rates. N=number of posts tagged with the theme. N/A: not applicable.

Theme[a]	Sharing Health Information, n (%)	Recognizing Special Days, n (%)	Recognizing Employees, n (%)	Event Charity, n (%)	Event Social, n (%)
Sharing Health Information (N=424)	N/A	45 (10.6)	8 (1.9)	0 (0.0)	0 (0.0)
Recognizing Special Days (N=177)	45 (25.4)	N/A	21 (11.9)	1 (0.6)	4 (2.3)
Recognizing Employees (N=140)	8 (5.7)	21 (15.0)	N/A	1 (0.7)	1 (0.7)
Organizational News New Facilities and Services (N=90)	11 (12)	3 (3)	6 (7)	0 (0)	0 (0)
Research Activities (N=85)	32 (38)	2 (2)	5 (6)	1 (1)	0 (0)
Event Charity (N=86)	0 (0)	1 (1)	1 (1)	N/A	10 (12)
Event Social (N=53)	0 (0)	4 (8)	1 (2)	10 (19)	N/A
Organizational News Achievements (N=43)	2 (5)	2 (5)	8 (19)	0 (0)	0 (0)
Donation Organ (N=35)	3 (9)	4 (11)	1 (3)	4 (11)	0 (0)
Event Health Screening (N=17)	0 (0)	3 (18)	0 (0)	0 (0)	0 (0)

[a]To improve clarity and readability of the table, rows and columns with no percentages > 10% have been removed from the table. The full table including all the themes on the columns and rows is provided in Multimedia Appendix 1.

The other most commonly used themes were *recognizing-special days*, *recognizing-employees*, and *sharing-patient success and feel good stories*. Those themes existed in the top 3 themes of 11, 8, and 5 hospitals, respectively. Moreover, in 8 cases, event-related themes were among the commonly used themes by the hospitals. The most frequently used event-related theme was *announcing and reporting events-health information*, which appeared in the top themes of 4 hospitals. *Announcing and reporting events-charity* was the other event-related theme used commonly by 3 of the sampled hospitals.

Announcing and reporting research activities was also one of the top 3 themes for 4 hospitals including Brigham and Women's Hospital, National Jewish Health, Denver-University of Colorado Hospital, Cedars-Sinai Medical Center, and NYU Langone Medical Center. This is consistent with what we expected from hospitals such as Brigham and Women's Hospital, which is an academic medical center affiliated with Harvard Medical School. Finally, University of Pittsburgh Medical Center included *announcing and reporting-organ donation* information in 9 of its 111 posts, which was unique among the sampled hospitals. Overall, the results showed differences in the ways hospitals used their Facebook pages to communicate with their audience, although some themes were commonly used by several hospitals.

Finally, we examined co-occurrence patterns between themes by calculating the rate at which pairs of themes co-occurred in the sample. Table 5 and Multimedia Appendix 1 present the short and full versions of the co-occurrence analysis results, respectively. Each cell in Table 5 and Multimedia Appendix 1 shows *the percentage of the posts tagged with the theme in the corresponding row that was also tagged with the theme in the corresponding column*. For example, 85 posts in the sample were classified as *announcing and reporting-research activities*. Of those posts, approximately 38% (32/85) were also classified as *sharing-health information*. This means hospitals and clinics frequently use social media to make followers aware of the institutions' research activities and to inform the audience of

the health-related outcomes and issues related to that research. Similarly, 25.4% (45/177) of the posts classified as *recognizing-special days* were at the same time used for *sharing-health information*. It is worth mentioning that, to improve clarity and readability of Table 5, we only included the rows and columns with at least one co-occurrence percentage greater than or equal to 10%. The full table containing the co-occurrence percentages associated with all the ordered pairs of themes is available in Multimedia Appendix 1.

Discussion

Principal Findings

The results of this study showed most posts in the sample could be classified as serving 3 purpose groups: *announcing and reporting*, *recognizing*, and *sharing* activities. Within the *announcing and reporting* purpose group, we found the sampled institutions used Facebook to broadcast information relating to *donations* opportunities, upcoming *events*, *research activities*, and *organizational news*. We further found that *recognizing* posts were used to *acknowledge employees* and *special days*, whereas *sharing* posts were used to disseminate *health information* as well as *patient success and feel good stories*.

Over one-quarter of the content posted by the sampled institutions focused on *sharing health-related information*. Accordingly, we believe health care organizations perceive social media as a tool for disseminating health-related information. This is most likely because of information being disseminated very quickly at no cost and remaining persistent on the institution's Facebook page. This finding is in line with the results of the extant literature (eg, [4,13]) that indicates educating health consumers such as patients and caregivers is among the main reasons health care institutions adopt social media. Furthermore, our findings substantiate CSC's Global Institute for Emerging Healthcare Practices' conclusion [5] that health professionals perceive the primary reason for institutional social media use is to promote community wellness and healthy

XSL•FO

RenderX

behaviors. Accordingly, health care institutions strive to raise awareness about diseases, medications, nutrition, physical activities, and other health and wellness-related topics to benefit their patients and improve the health of the surrounding community.

The second and third most frequently occurring themes in our sample were *recognizing special days* and *recognizing employees*. These themes were quite interesting, as most previous research has focused on how health care institutions use social media to interact with external stakeholders such as patients, family members, and the surrounding community (eg, [4,13]). The recognizing posts we examined were primarily geared toward stakeholders who were internal to the organization. For example, Nurses Recognition Week fell within our sample timeframe and all of the included institutions made several posts recognizing the nurses in their organizations. In addition, many of the sampled institutions used posts to recognize the contributions and accomplishments of specific individuals and departments within the organization. We believe these acknowledgments increase employees' satisfaction with the organization and may indirectly promote the quality of health care services provided to the public.

In addition to Nurses Recognition Week, the sampled posts recognized a range of other special days. By acknowledging these days, the organizations may aim to promote health in the society (eg, through promoting National Walking Day), raise awareness about diseases and medical conditions (eg, World No Tobacco Day), or acknowledge the role of people who play major roles in providing health care services to the community (eg, National Doctor's Day). We believe this finding is an important contribution to the literature as *recognizing employees* and *recognizing special days* were not adequately emphasized in previous studies as main purposes of using social media by health care organizations.

Another difference between our findings and the results of the related studies is although *recruiting employee* and *recruiting volunteers* emerged as 2 themes in our study, only 3 of the 1184 posts the researchers achieved consensus on (<0.3%) were related to recruiting employees or volunteers. This finding is in contrast to the survey of 36 US hospitals and health systems conducted by CSC's Global Institute for Emerging Healthcare Practices in 2012 [5], which revealed 47% of the respondents indicated they used social media for workforce recruitment. This deviation may imply health care institutions have realized that Facebook is not an effective tool for communicating with health-related job seekers. Instead, those organizations may use professional social media platforms such as LinkedIn for recruitment purposes.

Finally, we found certain themes co-occurred significantly more frequently than others did. For example, 38% of the posts related to *announcing and reporting-research activities* were used to also *share health information* with community members and 12% of the posts *announcing and reporting health screening events* linked the posts back to *organizational research activities*. This may indicate that health care organizations purposely link the content they share to specific initiatives within the organization.

From a practical point of view, our results demonstrated leading hospitals use social media platforms primarily for educating patients on health and wellness topics, announcing and reporting on different types of events, and recognizing employees and special days. These results can be used as a benchmark for the health care institutions that want to establish a social media presence to communicate with the public audience and for the smaller clinics and hospitals that want to further expand and improve their activities on social media websites.

From a theoretical standpoint, our results add to the literature on the uses of social media in health care. Previous studies have reported general intentions for health care institutions' social media use, and our findings revealed both similarities and differences between the content hospitals share and those intentions. Our results suggest health care organizations' use of Facebook for recognizing employees, sharing research activities, and announcing organizational news and achievements may be for the purpose of brand recognition. This is in line with previous research in the health care context [5] and other domains such as travel and fashion [6,7] that have found brand recognition to be one of the major reasons organizations use Facebook and other social media platforms. Moreover, to the best of our knowledge, this is among the first studies that empirically and inductively developed a hierarchical thematic web of content of general care health care institutions' postings on social media websites. This novel research method can further be adopted in different research areas within the medical informatics and health care information systems domains.

Limitations and Future Research

This study has limitations. For example, we collected data from the Facebook posts made by a sample of 17 hospitals during a 3-month period. Those hospitals and their posts may not represent the activities of all the health care institutions and all social media websites. Future studies can fill this gap by expanding the scope of the sampled hospitals and social media platforms. Collecting post data associated with a longer period will also enable future researchers to perform time series analysis and examine trends of the content generated by clinics and hospitals on their Facebook pages. Another limitation of our study is that the data were collected in 2014. Considering health care institutions may change their social media strategies and activities over time, future studies can collect more recent data to compare and contrast the results with our findings. This will enhance understanding of the evolution of health social media activities by medical institutions.

Building on our results, researchers of future studies can examine the efficacy of different themes by investigating questions such as the following:

1. Does announcing events on Facebook increase attendance?
2. Does acknowledging employees increase their organizational commitment, morale, and satisfaction?
3. Is sharing health information via Facebook an effective approach in raising awareness about diseases and medical issues in the community?

Furthermore, future research can examine user engagement in different types of posts to understand the extent to which each

content type can draw users' attentions and trigger their reactions in terms of liking a post, leaving a comment on it, or sharing it with others on Facebook. Understanding user engagement is important because it can transform one-way, provider-to-consumer information dissemination activities into two-way or many-to-many communication processes. In this way, the "social" aspect of such platforms as Facebook can be realized more meaningfully, adding value to the health care organizations' activities in virtual environments.

Another limitation of this study is that our sample only included US-based organizations. To enhance the generalizability of results, future studies should include data from a larger sample of international-based health care institutions on different websites such as Twitter, YouTube, LinkedIn, and Yelp during a wider range of time. In addition, health care organization types other than hospitals and clinics could be included in future samples.

Furthermore, we used the thematic network developed in the first phase as a foundation for coding the posts in the second phase of the study. At the same time, we were open to adding new themes that might emerge in the second phase and to removing the themes represented in an infrequent number of posts analyzed. Nevertheless, the coders' judgments in the coding process during the second phase might still be subject to a level of bias toward the themes identified during the first phase. This limitation could be mitigated in future studies by using automatic, computer-based coding algorithms and tools to objectively validate the hierarchical thematic model developed in this study. Another avenue for future research is to examine the content generation process in terms of who is responsible for posting contents on health care organizations' Facebook pages, and whether their expertise is in health care (eg, physicians, dentists, and nurse practitioners) or in social media

marketing. Such research should investigate how the content provider impacts the type of content shared, the manner in which it is shared, and the resulting level of audience engagement. This can ultimately influence the overall effectiveness of health care organizations' activities on Facebook and other social media platforms.

Conclusions

Previous studies in the health care social media domain have predominantly focused on staff members' perceptions regarding the purpose of social media use for the institution. In this study, we took an alternative approach and focused on thematically analyzing the actual content shared by 17 major health care institutions to understand the different types of content shared by those organizations. Our results provide a robustly validated set of standard content themes as well as information concerning the content themes most frequently shared by the various institutions. Frequency analysis of those themes revealed 2 interesting and unexpected findings. First, a large number of the sampled posts contained information targeted toward stakeholders internal to the organization (nurses, staff, and physicians). This was unexpected as previous studies have focused on how health care institutions use social media to connect with patients, caregivers, and the larger community. Second, very few of the analyzed posts (<0.3%) were geared toward employee recruitment, which previous studies have noted as a key reason staff members believe the institution uses social media. Finally, we identified several co-occurrence patterns within the identified themes, indicating health care institutions often leverage the information in a post to serve more than one objective. We believe these findings can provide a framework that other health care institutions can use to assess their own social media activities to benchmark their activities against a set of nationally recognized institutions.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Theme co-occurrence rates (full version).

[PNG File, 91KB - jmir_v20i5e190_app1.PNG]

References

1. Kordzadeh N. Social media in health care. In: Contemporary Consumer Health Informatics. USA: Springer; 2016:101-123.
2. Thomas L, Woodside J. Social media maturity model. Int J Healthc Manag 2016;9(1):67-73. [doi: 10.1080/20479700.2015.1101940]
3. Bermúdez-Tamayo C, Alba-Ruiz R, Jiménez-Pernett J, García GJ, Traver-Salcedo V, Yubraham-Sánchez D. Use of social media by Spanish hospitals: perceptions, difficulties, and success factors. Telemed J E Health 2013 Feb;19(2):137-145 [FREE Full text] [doi: 10.1089/tmj.2012.0066] [Medline: 23368890]
4. Thaker SI, Nowacki AS, Mehta NB, Edwards AR. How U.S. hospitals use social media. Ann Intern Med 2011 May 17;154(10):707-708. [doi: 10.7326/0003-4819-154-10-201105170-00021] [Medline: 21576547]
5. Scottsdale Institute. 2012. Ready to Interact: Social Media Use by U.S. Hospitals and Health Systems URL: https://www.scottsdaleinstitute.org/docs/sponsors/csc/2012-09-28.Social-Media-survey-findings.pdf [WebCite Cache ID 6yzbe19FA]
6. Laroche M, Habibi M, Richard M. To be or not to be in social media: how brand loyalty is affected by social media? Int J Inf Manage 2013 Feb 01;33(1):76-82. [doi: 10.1016/j.ijinfomgt.2012.07.003]

7. Kang JL, Tang L, Fiore AM. Enhancing consumer–brand relationships on restaurant Facebook fan pages: maximizing consumer benefits and increasing active participation. Int J Hosp Manag 2014;36:145-155. [doi: 10.1016/j.ijhm.2013.08.015]

8. Bennett E. Mayo Clinic Social Media Network. Health Care Social Media List URL: https://socialmedia.mayoclinic.org/hcsml-grid [WebCite Cache ID 6vEB4rxC1]

9. Facebook. Johns Hopkins Facebook Page URL: https://www.facebook.com/unsupportedbrowser [accessed 2018-04-27] [WebCite Cache ID 6yzbzA0qf]

10. YouTube. Johns Hopkins YouTube Page URL: https://www.youtube.com/user/JohnsHopkinsMedicine [accessed 2018-04-27] [WebCite Cache ID 6yzc3aDrx]

11. Twitter. Johns Hopkins Twitter Page URL: https://twitter.com/HopkinsMedicine [accessed 2018-04-27] [WebCite Cache ID 6yzc7sE80]

12. Van de Belt TH, Engelen L, Berben S, Teerenstra S, Samsom M, Schoonhoven L. Internet and social media for health-related information and communication in health care: preferences of the Dutch general population. J Med Internet Res 2013 Oct 02;15(10):e220 [FREE Full text] [doi: 10.2196/jmir.2607] [Medline: 24088272]

13. Richter JP, Muhlestein DB, Wilks CE. Social media: how hospitals use it, and opportunities for future use. J Healthc Manag 2014;59(6):447-460. [Medline: 25647968]

14. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

15. Guest G, MacQueen K, Namey E. Validity and reliability (credibility and dependability) in qualitative research and data analysis. In: Applied thematic analysis. London: Sage Publications; 2012:79-106.

16. Hart M, Stetten N, Islam S, Pizarro K. Twitter and public health (Part 2): qualitative analysis of how individual health professionals outside organizations use microblogging to promote and disseminate health-related information. JMIR Public Health Surveill 2017 Oct 04;3(4):e54 [FREE Full text] [doi: 10.2196/publichealth.6796] [Medline: 28978500]

17. Rice P, Ezzy D. Qualitative research methods: A health focus. Melbourne, Australia: OUP Australia and New Zealand; 1999.

18. 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;5(1):80-92. [doi: 10.1177/160940690600500107]

19. Kneafsey R, Brown S, Sein K, Chamley C, Parsons J. A qualitative study of key stakeholders' perspectives on compassion in healthcare and the development of a framework for compassionate interpersonal relations. J Clin Nurs 2016 Jan;25(1-2):70-79. [doi: 10.1111/jocn.12964] [Medline: 26403949]

20. Holm A, Severinsson E. Surviving depressive ill-health: a qualitative systematic review of older persons' narratives. Nurs Health Sci 2014 Mar;16(1):131-140. [doi: 10.1111/nhs.12071] [Medline: 23692267]

21. Hickey N, Sumsion J, Harrison L. Why nursing? Applying a socio-ecological framework to study career choices of double degree nursing students and graduates. J Adv Nurs 2013 Aug;69(8):1714-1724. [doi: 10.1111/jan.12029] [Medline: 23057783]

22. Amann J, Rubinelli S. Views of community managers on knowledge co-creation in online communities for people with disabilities: qualitative study. J Med Internet Res 2017 Oct 10;19(10):e320 [FREE Full text] [doi: 10.2196/jmir.7406] [Medline: 29017993]

23. Rishi M, Gaur S. Emerging sales and marketing challenges in the global hospitality industry: a thematic analysis of customer reviews from the world's top two tourist destinations. WHATT 2012;4(2):131-149. [doi: 10.1108/17554211211217316]

24. Peng G, Gala C. Cloud ERP: a new dilemma to modern organisations? J Comput Inform Syst 2014;54(4):22-30. [doi: 10.1080/08874417.2014.11645719]

25. Boyatzis R. Transforming qualitative information: Thematic analysis and code development. California, USA: Sage; 1998.

26. Lincoln YS, Guba EG. Naturalistic Inquiry. California, USA: Sage; 1985.

27. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968 Oct;70(4):213-220. [Medline: 19673146]

28. Altman D. Practical Statistics for Medical Research. United States: Chapman & Hall/CRC; 1990.

[XSL•FO]
RenderX

XSL•FO

RenderX

Original Paper

Developing a Mental Health eClinic to Improve Access to and Quality of Mental Health Care for Young People: Using Participatory Design as Research Methodologies

Laura Ospina-Pinillos[1], MD; Tracey A Davenport[1], BA (Hons), eMBA; Cristina S Ricci[1], BSc (Hons), LLB, BEd, MEd; Alyssa C Milton[1], BSc (Psych), MAppSc (Health Psychology), PhD; Elizabeth M Scott[2], MD, FRANZCP; Ian B Hickie[1], AM, MD, FRANZCP, FASSA

[1]Brain and Mind Centre, The University of Sydney, Sydney, Australia
[2]School of Medicine, University of Notre Dame, Sydney, Australia

Corresponding Author:
Laura Ospina-Pinillos, MD
Brain and Mind Centre
The University of Sydney
94 Mallett Street, Camperdown
Sydney, 2050
Australia
Phone: 61 0286276946
Email: laura.ospinapinillos@sydney.edu.au

Abstract

Background: Each year, many young Australians aged between 16 and 25 years experience a mental health disorder, yet only a small proportion access services and even fewer receive timely and evidence-based treatments. Today, with ever-increasing access to the Internet and use of technology, the potential to provide all young people with access (24 hours a day, 7 days a week) to the support they require to improve their mental health and well-being is promising.

Objective: The aim of this study was to use participatory design (PD) as research methodologies with end users (young people aged between 16 and 25 years and youth health professionals) and our research team to develop the Mental Health eClinic (a Web-based mental health clinic) to improve timely access to, and better quality, mental health care for young people across Australia.

Methods: A research and development (R&D) cycle for the codesign and build of the Mental Health eClinic included several iterative PD phases: PD workshops; translation of knowledge and ideas generated during workshops to produce mockups of webpages either as hand-drawn sketches or as wireframes (simple layout of a webpage before visual design and content is added); rapid prototyping; and one-on-one consultations with end users to assess the usability of the alpha build of the Mental Health eClinic.

Results: Four PD workshops were held with 28 end users (young people n=18, youth health professionals n=10) and our research team (n=8). Each PD workshop was followed by a knowledge translation session. At the conclusion of this cycle, the alpha prototype was built, and one round of one-on-one end user consultation sessions was conducted (n=6; all new participants, young people n=4, youth health professionals n=2). The R&D cycle revealed the importance of five key components for the Mental Health eClinic: a home page with a visible triage system for those requiring urgent help; a comprehensive online physical and mental health assessment; a detailed dashboard of results; a booking and videoconferencing system to enable video visits; and the generation of a personalized well-being plan that includes links to evidence-based, and health professional–recommended, apps and etools.

Conclusions: The Mental Health eClinic provides health promotion, triage protocols, screening, assessment, a video visit system, the development of personalized well-being plans, and self-directed mental health support for young people. It presents a technologically advanced and clinically efficient system that can be adapted to suit a variety of settings in which there is an opportunity to connect with young people. This will enable all young people, and especially those currently not able or willing to connect with face-to-face services, to receive best practice clinical services by breaking down traditional barriers to care and making health care more personalized, accessible, affordable, and available.

XSL•FO
RenderX

Introduction

Background

The Internet and emerging technologies have long been identified as having the potential to significantly expand the reach of quality mental health care by addressing geographical, economical, and human resource barriers [1,2]. Over the past decade, evidence suggests that the Internet is considered not only a major source of information about health (including mental health) and well-being [3-5], but that it is also useful for mental health promotion, screening, prevention, early intervention, and referral processes [6-9]. Web-based platforms are also being used to deliver real-time stepped care services [10], providing information and low-intensity treatment for those in the early stages of help-seeking, with the capacity to increase scope and intensity of treatments as illnesses progress [11].

In Australia, new and emerging mental health technologies are urgently needed as 1 in 5 young Australians aged between 16 and 24 years experience a mental health disorder each year, yet only 1 in 4 receives professional help [12]. Of those who do receive help, only a small proportion receives timely and evidence-based treatments [13]. Access to quality care is especially difficult for disadvantaged and vulnerable populations, including children; Aboriginal and Torres Strait Islander young peoples; young people from culturally and linguistically diverse backgrounds; and young people living in regional, rural, and remote areas [14,15].

Today, with ever-increasing availability and use of technology, the potential to reach young people is especially promising [7]. For example, in 2014-15, 85% of people aged 15 years and over in Australia were Internet users, and young people in the age group of 15 to 17 years had the highest proportion of use (99%) [16]. Importantly, there is also a growing body of literature supporting the use of Internet-based treatment approaches for a range of mental health problems in adults and adolescents [6,17-19]. Although there is far less research focusing on systems that integrate real-time Web-based stepped care support in mental health services [20], some research examining comprehensive Web-based support systems is starting to emerge [21,22].

Web-Based Health Information

Young people are increasingly relying on the Internet to find answers about their health concerns. For example, in Australia, a national survey in 2012-13 revealed that 53% of young people aged between 16 and 25 years with moderate and very high levels of psychological distress sought Web-based information related to a mental health and/or alcohol or other substance use problem; the majority found this information to be helpful [23]. In other surveys, young people reported feeling comfortable accessing Web-based mental health tools because they felt they were anonymous, welcoming, less stigmatizing, and, for the most part, trustworthy [24]. Similarly, in the United States, a survey in 2013 found that 59% of patients had gone online to look for health information in the previous year and 35% had gone online to self-diagnose or diagnose someone else's condition. Of the "online diagnosers," 46% concluded that they needed to see a health professional, whereas 38% preferred a self-management option [25].

Similarly, a small pilot study evaluating an online triage platform also found that, for every user requiring a general practitioner response through an e-consultation, 5 users required online self-help only (WebGP, 2014) [26,27]. The Internet, therefore, offers people seeking information with a useful and easy gateway to answers and solutions that respond to their needs.

Internet-Based Screening

Individuals often show a preference for computerized screening over face-to-face interviews when the subject matter is sensitive in nature [28-32]. Importantly, Internet-based screening for common mental health problems has been shown to be reliable and effective [9,33,34]. In relation to screening for suicidal ideation, it is argued that the implementation of standardized, self-reported, computer-based assessments (with stringent suicide response protocols) may be a strategy that is both accurate and viable if followed up with health professional support [20,35]. In general, there is, however, the potential that symptom, Web-based assessment tools may increase demand on services as the tools are often risk averse (due to medico-legal concerns), recommending professional care when self-management is an appropriate alternative [36,37].

Videoconferencing

The provision of mental health services though videoconference systems have been widely used since the 1960s [38]. Videoconferencing is viewed as more advantageous than telephone support as the health professional can gauge important visual cues to inform their assessment such as appearance, facial expression, motor activity, movements, and mannerisms [39]. Supportive Web-based conversations with a health professional and referral to appropriate resources have also been found to negate risky behaviors, such as suicide and violence, in highly distressed people [40]. Importaly, videoconferencing has been found to be as reliable as face-to-face assessments, and more cost-effective [41].

In Australia, telepsychiatry (videoconferencing) has been practiced since the 1990s and its benefits have been translated into a wide range of populations including rural populations, Indigenous communities [42], the defense force [43], and children and adolescents [44]. Surprisingly, despite the reliability and benefits of videoconferencing, use of videoconference systems remains low in the different settings [45]. Embedding videoconferencing in comprehensive Web-based support

systems would enable a more systematic use of this technology and a greater number of people to access mental health services.

Comprehensive Web-Based Mental Health Care

The augmentation of traditional videoconferencing with online self-reported health questionnaires or screening results has now emerged in the research literature [20,46]. In a study conducted by Williams and colleagues, participants who screened positive for major depression or suicidality were given the opportunity to schedule a videoconference via Skype with a psychiatrist within a 2-week period [46]. In a recent study conducted by our team, an initial Web-based clinical assessment (with an embedded suicidality escalation protocol) was completed as a routine part of the assessment process (either before a video or face-to-face visit with a health professional) for a subset of participants. After the assessment was completed, the severity of mental illness was determined using a clinical staging model [47], and those who were considered to be high-risk (according to the suicidality escalation protocol) were escalated to the youth health service, bringing forward their initial face-to-face appointment [20].

Mental Health Interventions for Young People

In a systematic review of Web-based mental health interventions for young people, however, limited uptake, engagement, adherence, and dropout rates have been identified as significant problems [48,49]. Some researchers, including ourselves, believe that the involvement of young people and youth health professionals in the design, development, and delivery of youth services could lead to better engagement and outcomes [50-52].

Participatory Design and Technology-Based Mental Health Interventions

Participatory design (PD) methodologies, which were developed in the late 1960s and early 1970s, emphasize the importance of involving all stakeholders (including end users, developers, and researchers) during the development of products to help ensure the end product meets everyone's needs; improve usability; and increase engagement of users [53-55]. The process involves iterative design cycles in which end users and researchers contribute to knowledge production and the development of the end product [50,56]. Importantly, end users should participate in all stages of development [57], not as consultants or controllers of the process, but sharing equal responsibility with the research team for the outcomes [50]. Some researchers [58], including us, consider PD as key research methodologies that overtly put the end user at the center of research and here, the development of the MHeC [59].

The use of PD is expanding in the development of technology-based mental health and well-being interventions for young people. In 2015, a systematic review [50] described the development of these interventions in areas such as prevention, screening [60], and treatment programs [61]. In the majority of cases, however, end users assumed more of a consultative role. Despite the fact that uptake of PD in the development of technology-based mental health and well-being interventions has increased, evidence is still needed to assess the impact of these research methodologies in the outcomes of these interventions [50].

As consumers have the opportunity to share their preferences before the development of expensive, and potentially helpful, systems, the rationale behind the use of PD in the development of mental health technologies could mean that an active engagement of end users could reduce the 17-year gap in translational research [62].

This Study

In Australia, Web-based mental health services include health promotion, self-directed, and low intensity mental health support (eg, ReachOut! [63]; beyondblue [64,65]); national online counseling services (eg, eheadspace [66]); structured self-directed online therapy (eg, MoodGYM [67]); and those offering a combination of assessment and structured online therapy, including additional therapist support (eg, MindSpot Clinic [68]).

What is missing is a Web-based mental health clinic that includes access to all the components and services necessary to meet the needs of all young people (those between the ages of 16 and 25 years), regardless of where they are on the spectrum of mental health or ill-health, or where they reside in Australia.

The aim of this study was to use PD with end users (young people and youth health professionals) and our research team to codesign and build a Mental Health eClinic (MHeC) to improve timely access to, and better quality, mental health care for young people across Australia.

Methods

Participants

Participants included young people attending *headspace* Camperdown and *headspace* Campbelltown, and youth health professionals from both *headspace* Camperdown and *headspace* Campbelltown (*headspace* is the national youth mental health foundation dedicated to improving the well-being of young Australians; Camperdown and Campbelltown are two different sociodemographic areas of Sydney).

The University of Sydney's Human Research Ethics Committee approved the study (Protocol No. 2014/689). For all phases, participants (young people aged between 16 and 25 years and youth health professionals) who expressed an interest to participate in the study were provided with the participant information statement and participant consent form before providing consent and participating in the study. Parental consent was also obtained for participants aged between 16 and 17 years. At the conclusion of each PD workshop and one-on-one end user consultation, each young person was provided with a gift voucher to thank them for their participation and sharing their expertise; young people were also reimbursed for any travel-related expenses to attend the workshop or session. Youth health professionals were not provided with gift vouchers as they participated in stages 1 and 3 during work time. All workshops were catered.

Recruitment Strategy

Recruitment strategies for young people included the posting of flyers at each *headspace* center inviting young people to be involved in the study, and informing youth health professionals

and reception staff at each *headspace* center about the study so they too could assist with recruitment. All young people belonging to *headspace* Camperdown and *headspace* Campbelltown Youth Reference Group were also invited to participate in the study. Inclusion criteria for the study were as follows: young person attending either *headspace* Camperdown or *headspace* Campbelltown; aged between 16 and 25 years; and with access to the Internet through a mobile phone, tablet, desktop, or laptop.

In relation to the recruitment of youth health professionals, senior management at *headspace* Camperdown and *headspace* Campbelltown informed all staff about the study and called for expressions of interest to participate.

Research and Development Cycle

The research and development (R&D) cycle for the codesign and build of the MHeC included several iterative PD phases (Figure 1): PD workshops (phase 1); translation of knowledge and ideas generated during workshops ("knowledge translation"; KT) to produce mockups of webpages (either hand-drawn sketches or wireframes; phase 2); and rapid prototyping and one-on-one consultations with end users, including assessing the usability of the online alpha build of the MHeC (phase 3). The remaining phases of PD for the MHeC (rapid prototyping and user [acceptance] testing of the beta build [phase 4], and real-world study of the delta build [phase 5]) are currently underway and will be reported separately.

Participatory Design Workshops (Phase 1)

The PD employed in this study was informed by the research methodologies developed by the Young and Well Cooperative Research Centre for the development of evidence-based online youth mental health promotion, intervention, and treatment [58]. PD workshops were developed to accommodate a maximum of 12 participants per workshop. The same group of researchers facilitated all PD workshops. Importantly, a mental health professional was present on-site for the duration of all workshops involving young people in case anyone experienced psychological distress as a result of the subject matter. A scribe was present to take hand written notes at each workshop.

PD workshops were held in two stages with young people and youth health professionals attending separate workshops (stage 1) or a combined workshop (stage 2; Textbox 1). Following each workshop, the knowledge and ideas generated during the workshop were translated to produce mockups of webpages, either as hand-drawn sketches or wireframes produced using Balsamiq (Balsamiq Solutions, LLC, Sacramento, California, United States: a rapid wireframing tool that reproduces the experience of sketching on a whiteboard but using a computer) [69]. The mockups were then presented at the next workshop, enabling content and broad design ideas to be discussed, critically analyzed, and further developed.

Figure 1. PD research methodologies used during the design and build of the MHeC where end users participate in all stages of development, are at the center of the R&D cycle, and share equal responsibility with the research team for the outcomes. PD: participatory design; MHeC: Mental Health eClinic; R&D: research and development.

Textbox 1. Phases 1 and 2 workflow.

- Phase 1, stage 1 (young people only, Camperdown): One 1-day participatory design (PD) workshop with young people attending *headspace* Camperdown; workshop held in *headspace* Camperdown

Knowledge translation for young people only (phase 2)

- Phase 1, stage 1 (young people only, Campbelltown): One 1-day PD workshop with young people attending *headspace* Campbelltown; workshop held in *headspace* Campbelltown.

Knowledge translation for youth health professionals only (phase 2)

- Phase 1, stage 1 (youth health professionals only): One 1-day PD workshop with youth health professionals from *headspace* Camperdown and *headspace* Campbelltown; workshop held in *headspace* Campbelltown.

Knowledge translation combined workshop (phase 2)

- Phase 1, stage 2 (young people and youth health professionals combined): One half-day PD workshop with both young people and youth health professionals from *headspace* Camperdown and *headspace* Campbelltown (all participants must have previously participated in a stage 1 workshop); workshop held in *headspace* Camperdown.

Topics for stages 1 and 2 workshops included the MHeC home page; important website functions; look and feel of the website; online physical and mental health assessment; provision of assessment results online (including consideration of a dashboard of results); a video visit system (utilizing live interactive videoconferencing); and the development of personalized well-being plans based on assessment results. Importantly, technology was not used in any PD workshop; instead, several artifacts and design activities (Figure 2) were used in each workshop to facilitate discussions and the design process. The activities included:

1. The use of propositions to explore and communicate the concept of the MHeC
2. End user sketching [70] (hand-drawn sketches by young people and youth health professionals, either individually or in groups) to enhance the feedback process and generation of new ideas
3. Analysis of mockups of webpages (hand-drawn sketches and wireframes) to test designs and provide feedback about the look, feel, content, and behavior of the MHeC.

Knowledge Translation (Phase 2)

At the conclusion of each workshop, the transcript and all visual artifacts were independently analyzed by a KT Team (three 2nd-and 3rd-year psychology students [AI, ML, and ED], all females aged between 20 and 23 years) who were interns at the time at The University of Sydney's Brain and Mind Centre). Observations were tallied, and those observations with three or more tallies were considered for inclusion in the next generation of wireframes for discussion and analysis at the following workshop. Information was compiled until saturation point was reached (defined as the point where no new information was attained) [71].

One-on-One Consultations With End Users, Including Assessing the Usability of the Alpha Build of the Mental Health eClinic (Phase 3)

Phase 3 involved in-depth one-on-one consultations with new end users (young people and youth health professionals who had not participated in Phase 1). The inclusion of new participants aimed to reduce biased responses due to habituation or familiarity with the topic as a result of prior participation in the study. In each 90-min one-on-one end user consultation, a researcher was paired with a participant (end user) and an observer took notes. Sessions involved the use of laptops where participants had access to the alpha build of the MHeC website. Employing a think-aloud protocol [72], participants were observed as they navigated the MHeC and responded to questions posed by the researcher about the main components of the MHeC; responses were recorded by the observer. The initial effectiveness of the system was then assessed by asking participants to complete 3 usability tasks: (1) create an account and login; (2) find the "need help now" button; and (3) book an appointment. Task completion time was recorded to assess the efficiency of the system. No instructions or clues were provided, and comments in relation to navigation were recorded.

Data Analysis

Qualitative data were interpreted using thematic analysis techniques [73] according to the following themes: general elements of the MHeC; general look and feel; privacy and data sharing; and interaction of the MHeC with social networks. Records of all tallies obtained in Phases 1-3 were then grouped and interpreted by a team of researchers (LOP, TD, and 2nd- and 3rd-year psychology students [AH and FY] who were interns at the time at The University of Sydney's Brain and Mind Centre). Each theme and associated content was discussed by the group, and differences of opinion were discussed until consensus was reached.

Figure 2. Samples of visual artefacts. (A) Hand-drawn sketch by end users during a PD workshop; (B) Hand-drawn sketch of a webpage generated following a KT session; and (C) Wireframe generated using Balsamiq following a KT session. PD: participatory design; KT: knowledge translation.

Results

A total of 4 PD workshops, 4 knowledge translation sessions, and 1 round of one-on-one end user consultation sessions were conducted between October 2014 and June 2015 (Figure 3).

Participant Characteristics

PD workshops were held with young people attending *headspace* Camperdown (n=7) and *headspace* Campbelltown (n=11), and youth health professionals working at those services (*headspace* Camperdown n=5; *headspace* Campbelltown n=5). A total of 18 young participants participated in stage 1 (young people only) PD workshops: there was equal gender participation, and 78% (14/18) were aged between 18 and 25 years. Ten youth health professionals participated in stage 1 (youth health professionals only) PD workshop: the majority of these participants were female (70%, 7/10) and aged between 20 and 30 years (70%, 7/10). The group comprised 4 psychologists, 2 occupational therapists, 1 medical student, 1 general practitioner, 1 social worker, and 1 Aboriginal youth worker.

Nine participants participated in stage 2 (young people and youth health professionals combined): young people (n=5) and youth health professionals (n=4). The majority of the participants were male (56%, 5/9); the youth health professional's group included 3 psychologists and 1 occupational therapist.

Six people participated in Phase 3 (consultation with end users, including usability assessment of the online alpha build). The majority of these participants were female (67%, 4/6). The group contained 1 clinical psychologist, 1 psychology student, and 4 young people.

Main Components of the Alpha Build of the Mental Health eClinic

The iterative R&D cycle revealed the importance of five main components of the MHeC. These informed the alpha build of the MHeC (Figure 4-8): a welcoming home page with a visible triage system; a comprehensive physical and mental health assessment; a detailed dashboard of results; a booking and a video visit system; and the generation of a personalized well-being plan that includes links to evidence-based, and health professional-recommended, apps and etools. The five components will have different functionalities depending on whether the user is a young person or a health professional.

Element 1: Home Page and Triage

Young people suggested the home page should be a "welcoming space," where young people can feel "comfortable," without compromising the authenticity and professionalism of the site. As such they indicated a banner with the institutional logos should be visible at all times at the footer of the home page. Young people also wanted the following features to be included on the home page: testimonials from young people about their experiences with the MHeC ("a young person explaining why they are there and how to use it"); reliable mental health information (resources); information about how to help a friend; frequently asked questions; and a brief explanation about how the MHeC works. Young people also suggested that as this information would be of interest to a wide range of people, the home page needs to be available to everyone, even if they have not signed up for the MHeC. Youth health professionals and young people suggested that, before any login process, a triage system needs to be in place to ensure that young people in distress can access immediate care and a "need help now" button should be clearly visible for those in crisis.

Figure 3. Study Gantt chart. KT: knowledge translation; MHeC: Mental Health eClinic.
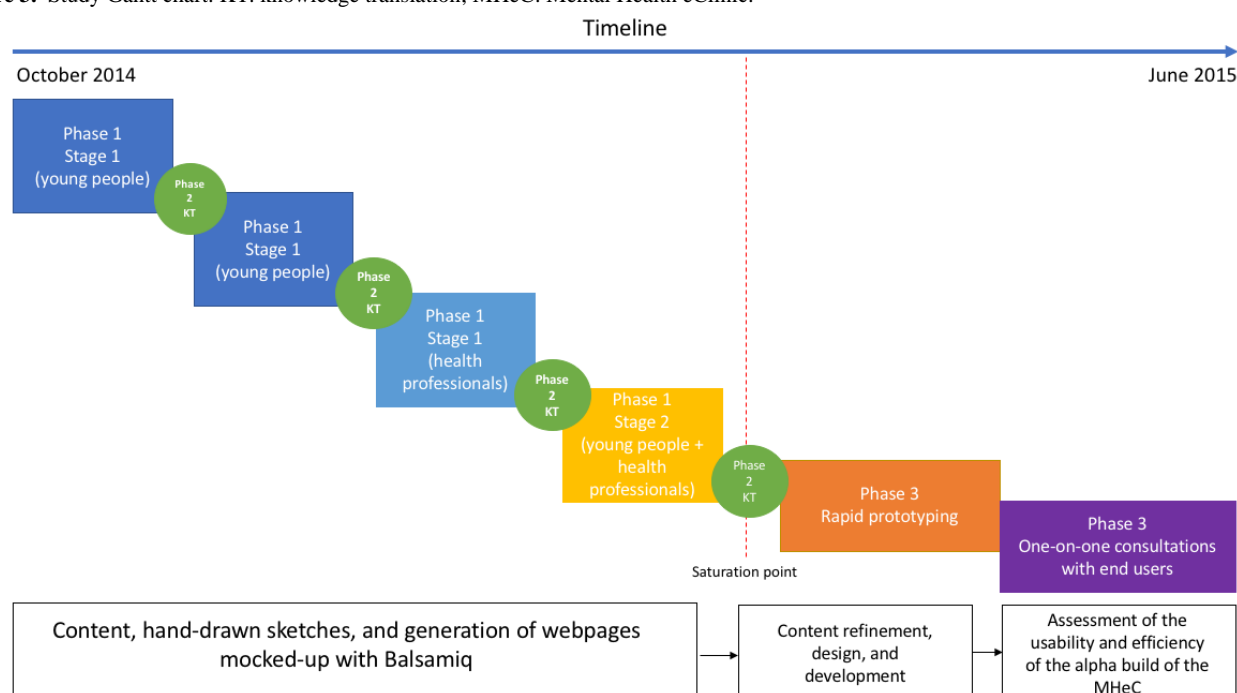
Figure 4. Home page with a clearly visible triage system for those requiring urgent help.

XSL•FO
RenderX
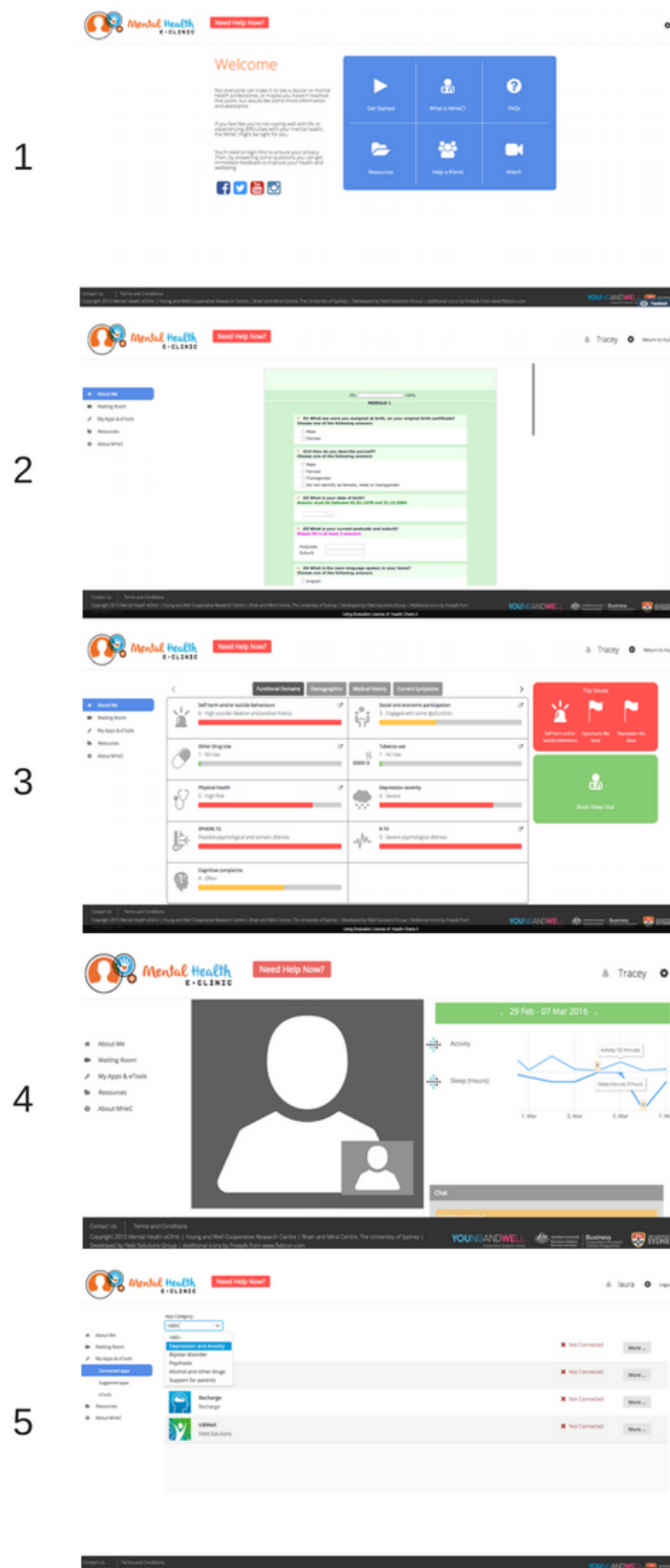
Figure 5. Comprehensive online physical and mental health assessment.

Figure 6. Dashboard of results and progress report.

XSL•FO

RenderX
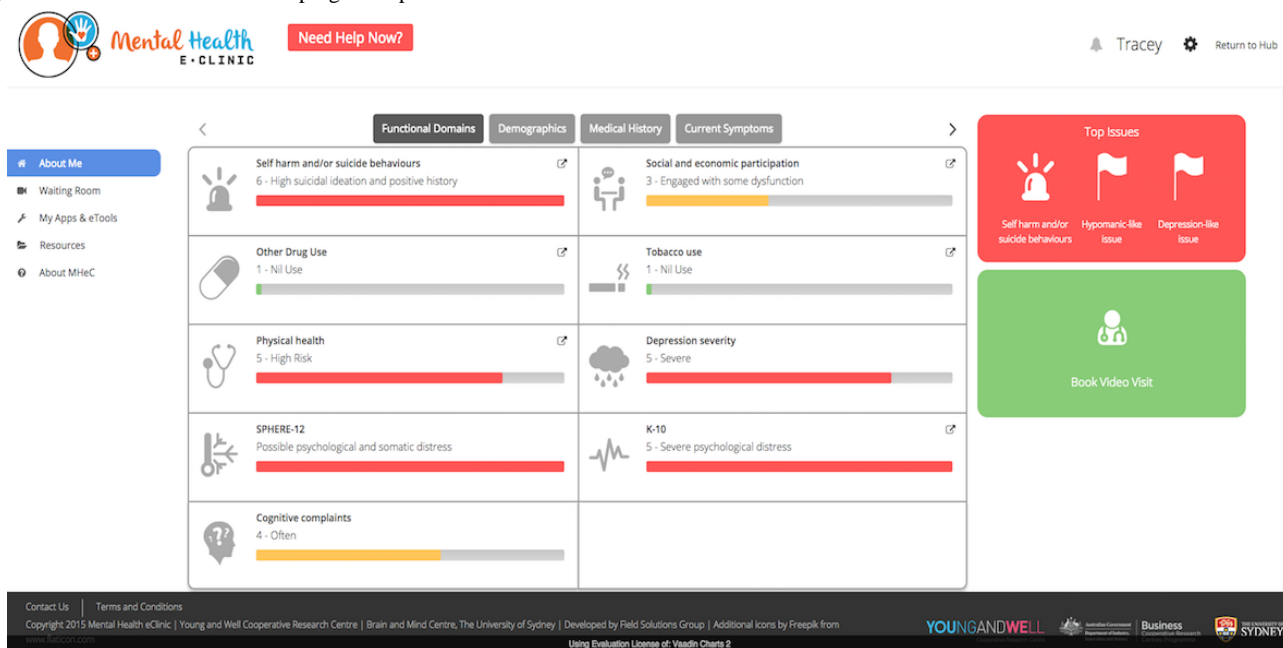
Figure 7. Booking and video visit system.
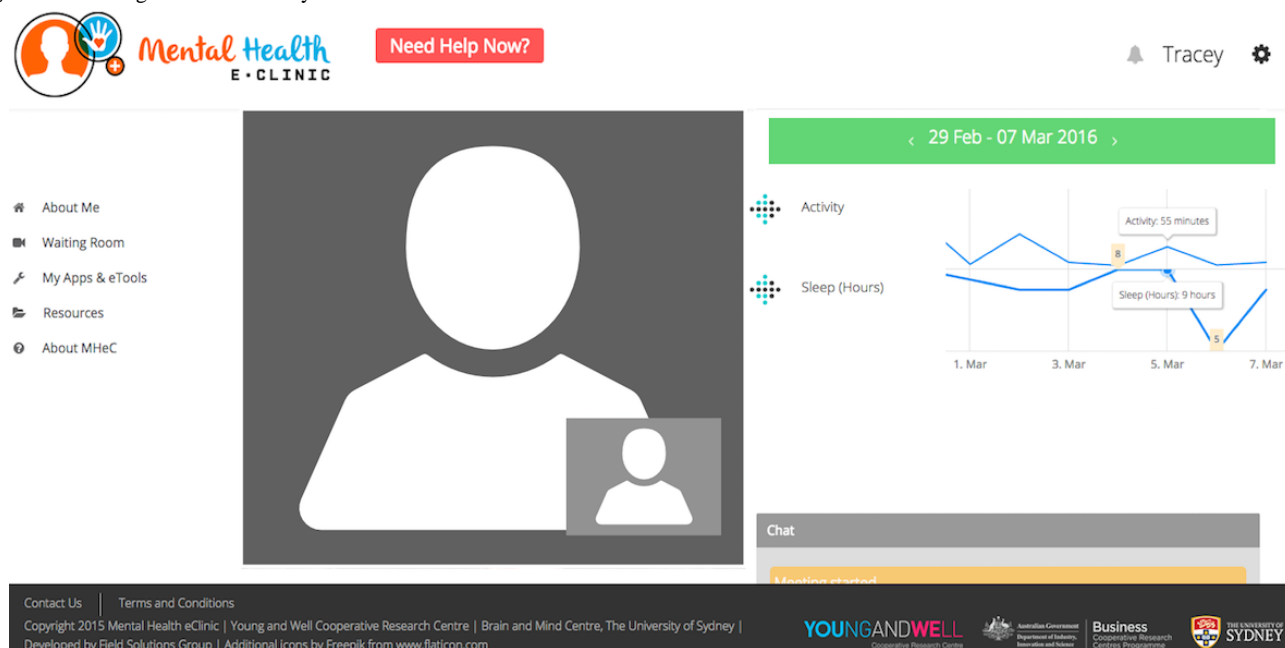
Figure 8. Personalized well-being plan, including recommended apps and etools.
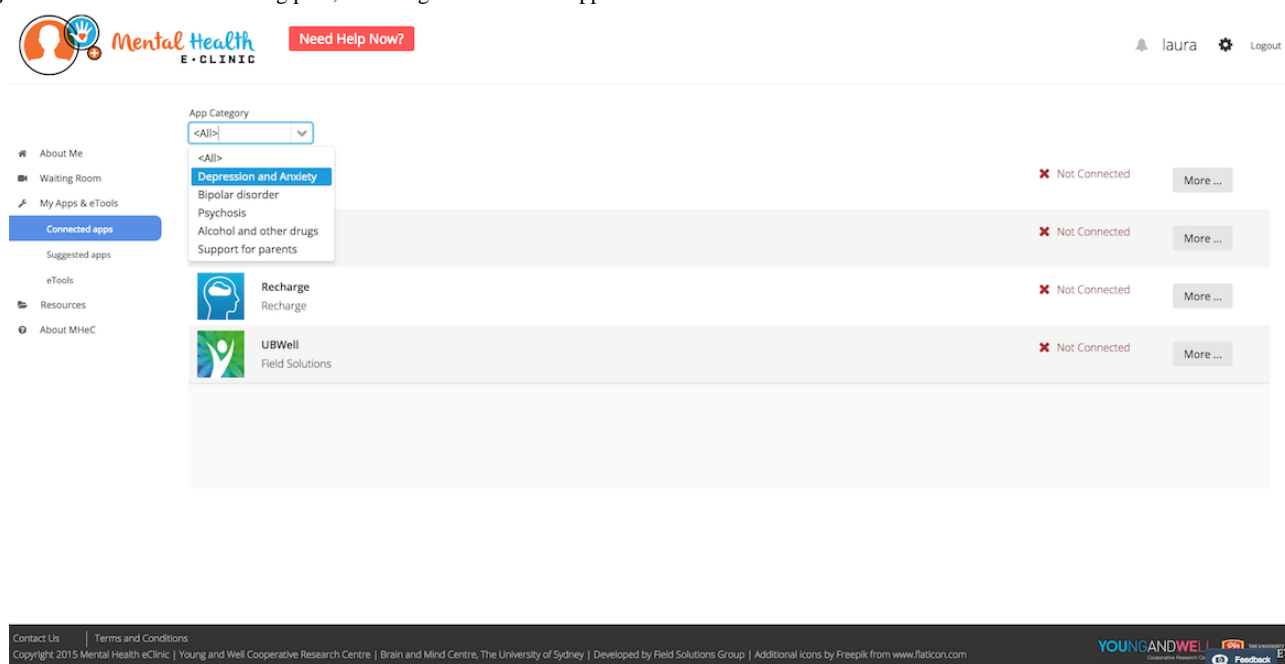
Element 2: Online Physical and Mental Health Assessment

Young people said they felt comfortable about completing an online physical and mental health assessment and receiving immediate feedback of their results. The majority also initially reported that they preferred short questionnaires (approximately 15 min in duration). After an explanation was provided, however, about the number and types of questionnaires that would be included to fully assess their physical and mental health (including the range of standardized assessments that are currently being used in *headspace* centers and completed using paper and pencil or iPad) and that a comprehensive assessment would enable a detailed dashboard of results and well-being plan to be generated (including recommended apps and etools),

young people understood the need to complete a more comprehensive assessment (up to 1-hour duration). Young people suggested that a pause and "resume later" feature would be helpful with a longer questionnaire to ensure completion. In addition, young people reported that they preferred Likert-scale questions (typically a 5-, 7-, or 9-point agreement scale), two-way close-ended question (no/yes) or multiple-choice questions rather than free-response ("enter text") questions. The MHeC's assessment includes medical history, physical health, mental health symptomatology, and health behaviors. For youth health professionals, the questionnaire functionality enables them to write notes or complete/add additional information for relevant questions in relation to their young person's progress or specific clinical observations (eg, provide a score in the Social

and Occupational Functioning Assessment Scale [74] or allocate participants to a clinical stage [47]).

Element 3: Dashboard of Results and Progress

Young people reported that after completing the assessment, immediate feedback about their results was essential. They also reported that knowing this would occur would improve motivation to respond as best as possible and answer all questions. Young people said they wanted accurate feedback about their results and for this to be represented in a detailed dashboard ("visual concept"), with the option of printing or saving the file as a PDF for their records, or to share with a health professional in the future. Simple bar graphs, colored icons, and traffic light representations were seen as the most acceptable and understandable options for the presentation of results. Line graphs were the preferred option to represent changes over time and to track progress. Young people also suggested the availability of a "customize option" would be useful to enable individuals to select multiple variables of interest to explore if, and how, they influence another. Youth health professionals would have access to a young person's results and progress as they reported that a dashboard of results would help guide care before and during video and face-to-face visits.

Element 4: Waiting Room, Booking System, and Video Visit System

Young people requested that a booking system be available to make timely appointments with youth health professionals attached to the clinic. Young people also suggested a "waiting room" be included in the MHeC where individuals wait for their video visit to begin. While on this page, young people suggested various activities could be available, for example, breathing exercises or watching selected videos until the video visit icon changes color, signaling their video visit is about to commence. In relation to the youth health professionals' functionality of this component, they suggested that it would be important to have access to a booking system ("to add or cancel a booking as well as block timeslots").

Acknowledging cultural preferences, and that some people do not feel comfortable seeing their image ("seeing myself distracts me, I would be looking at myself"), it was suggested by some young people that at the moment of booking an appointment the option of hiding their image should be available. It was also suggested that in the event of communication cut-offs or a young person wanting to say something sensitive, a chat box would allow a fluid conversation. All participants agreed that the video visit system needs to be embedded in the MHeC as "it would be secure." As it happens in regular practice, the youth health professionals are in charge of inviting a young person to come to their office; therefore, in the MHeC, the youth health professional has the possibility to start and end the video visit. Importantly, this video visit system includes a "share" functionality, where relevant information can be shared and made visible on a young person's screen (eg, "explain their dashboard of results during the session").

Element 5: Personalized Well-Being Plan That Includes Links to Evidence-Based, and Health Professional–Recommended, Apps and Etools

All participants reported that they would like to be provided with a personalized well-being plan (generated from their assessment results) that included tailored automatic recommendations on how to improve their health and well-being. Young people believe that apps and etools targeting their main concerns/issues would be beneficial. However, young people believe that "health professional recommended" and free apps/etools are more likely to be downloaded over apps/etools that have to be purchased or that are not recommended by health professionals. The areas of main interest to young people were apps/etools that help them with their sleep; improve their memory; assist with mood tracking; and help with tracking their progress over time. Ideally, apps and wearable devices should be integrated with the MHeC. Youth health professionals felt confident about the MHeC suggesting apps and etools to young people "only if they are from a reputable source" and had greater confidence in those that had been independently rated using the Mobile App Rating Scale [75]. In terms of functionality, all end users suggested that the MHeC should have a list of apps sorted by categories such as sleep, mood, anxiety, and physical health, among others. Participants suggested that each app should have an Apple Store and Google Play link to facilitate the download process.

User Interface

In relation to the look and feel of the MHeC, young people and youth health professionals agree that the system should be "clean and tidy," displaying short and concise information and making good use of space. Most of the participants preferred to have icons instead of text and preferred to self-explore the system rather than having lengthy instructions. The use of stereotyped photos with "happy" or "sad people" was highly discouraged by young people. The participants wanted to see videos about how to use the system, but indicated that these should be short (no more than 30 s) and not auto-play due to data download size concerns and privacy issues (eg, a video commencing while sitting in a public space). Soothing colors were preferred for the background and brighter colors for the functionalities, suggesting combinations such as blue and orange for consideration. Young people also suggested the possibility of a "customizing" option to enable users to choose between two to three text types and background color options.

Privacy and Data Sharing

Young males were more concerned about privacy issues than young females. For example, young male participants said "sharing data is OK, but it must say it's confidential" or "I wouldn't share my location with the system." On the other hand, young female participants said, for example, "I have no privacy concerns specifically if going to a clinician." All young people, however, said they would share their data as they believe the MHeC would be a professional and trustworthy site and this would enable them "to get most help." Young people, however, emphasized that they want to be informed about how the data will be handled before giving permission for data sharing and to have the opportunity to withdraw this permission at any time.

Integration With Social Networks

Young people believe that the MHeC should be part of social networks as they would like to be able to share the MHeC in their profile or with a friend through a private message. Moreover, they wanted to be able to like the MHeC or write a comment on it. One of the suggested preferred features of the MHeC was "tips" and for these to be displayed throughout all components of the MHeC. Young people also thought it would be valuable if they were able to share these tips via their social networks.

Usability

Information about the efficiency and effectiveness of the alpha build of the MHeC was obtained during the one-on-one consultation sessions. All participants (n=6) completed usability tasks. All participants found that "creating a MHeC account" and "login" were relatively simple processes. Half of the participants thought that giving the option of login with their social media details (single sign on) was a good idea as it would speed up this process ("super easy"). The other half were against social media login as they were concerned about the MHeC sharing information with their friends ("I don't want my friends seeing this information"). Finding the "need help now" button was a simple task; all participants found it in less than 5 s. Booking an appointment was a straightforward task as well, and participants were able to navigate the MHeC and complete the process almost immediately.

Discussion

Principal Findings

Our study utilized an innovative approach to the development of a Web-based mental health clinic for young people (the MHeC). The PD employed ensured that end users (young people and youth health professionals) had an active and equal involvement in all phases of the design and development process. By engaging these stakeholders, we attempted to respond to end users' expectations about the MHeC and what was achievable in terms of technology and what has been proven to be effective in the mental health field. The use of several design activities (propositions, hand-drawn sketches, and mockups) and the combination of different PD methodologies (workshops and one-on-one consultations with end users, including assessing usability of the online alpha build of the MHeC) enhanced feedback processes and the generation of new ideas. This combination of research methodologies also accelerated the refinement of the MHeC to achieve the build of the alpha prototype.

In the past decades, the development and use of eHealth solutions in mental health care have expanded; however, these solutions have been developed to address specific problems or to replace different components of the traditional health care system. As an example, the majority of self-triage tools rely on people actively searching for these tools on the Internet; however, some health services provide self-triage tools on their websites, particularly when booking appointments online [26]. Telepsychiatry (videoconferencing), as another example, has also been a particularly effective way of providing support as

it allows real-time interaction while negating barriers such as cost, geographical location, and stigma concerns associated with face-to-face support [46]. To the best of our knowledge, our MHeC is one of a kind as it integrates triage, online assessments, online provision of results with easy-to-interpret graphic representations (a dashboard), enables video visits with youth health professionals, provides a personalized well-being plan with immediate interventions, and tracks progress. We believe this is more sophisticated and technologically advanced than traditional telemedicine.

One of the strengths of our system is that it integrates new and emerging technologies with the traditional face-to-face process. Several studies emphasize that screening alone is insufficient for connecting end users with the necessary resources for effective treatment [9,76-78]. Instead, it is argued that online screening should supplement, and be integrated when necessary, with additional support and assessment, which can include face-to-face or online assessments with health professionals within mental health services [9,78]. In line with these recommendations, Internet-based clinical assessments have been implemented and trialed in various Web-based clinics as a means of rapid assessment and referral to appropriate online interventions [79,80] and integrated with face-to-face and online support [20].

Important considerations need to be made in relation to an end user's health literacy as it is argued that Web-based health care communication demands a higher level of health literacy from end users, including the knowledge and skills that enable them to navigate the health system [81]. National datasets report that 60% of Australians have health literacy scores at the lowest two levels of proficiency (scoring 1 or 2), whereas only 6% attain the highest two levels (scoring 4 or 5) [82]. Furthermore, if an end user is experiencing active psychosis or is in crisis, for example, the ability to accurately read or reflect upon information may prove challenging [83,84].

Video visits (as provided in the MHeC with "share" functionality), and face-to-face support provided by health professionals, may provide end users with the support they need to help them navigate the online physical and mental health assessment, dashboard of results, and the health system. These noted challenges highlight the need for health professionals to remain involved in some capacity in the online assessment process, particularly to prevent the unnecessary funneling of end users into the health system.

In 2017, it was estimated 95% of young people in Australia had access to a mobile phone [85,86], making Internet access more widely available to individuals. Mobile phones also provide individuals with higher levels of privacy (compared with shared computers) when participating in a video visit or completing assessments. Consequently, these devices could serve as great facilitators of health care provision, playing a crucial role in health reform.

Implications

In 2014, the Australian Government asked the National Mental Health Commission (NMHC) to review existing mental health services and programs and "…assess the efficiency and

effectiveness of programs and services in supporting individuals experiencing mental ill health and their families and other support people to lead a contributing life and to engage productively in the community" [87]. Similar to previous reviews, the NMHC, called for an overhaul of the Australian mental health system, including an integration of e-mental health with face-to-face services. Innovations in the use of technology during the assessment process hold potential as they can go some way to addressing documented youth service capacity issues [84,88].

In alignment with the NMHC's three key components (person-centered design principles; new system architecture; and shifting funding to more efficient and effective "upstream" of services and supports), we envisage that the MHeC will result in a usable system, with high engagement rates, and could finally improve access for all young people requiring assistance across Australia. We see the MHeC as a real-time primary care clinic integrated with current face-to-face services, offering end users at a minimum immediate online clinical assessment; immediate generation of a dashboard of results and personalized well-being plan with tailored interventions; and inbuilt triage and escalation protocols to accurately respond to severe and risky cases. The MHeC offers timely support and interventions through the development of an individual and health professional shared treatment plan. It also provides an opportunity for young people to have a video visit with a health professional despite their current geographical location. Although developed for Australian settings, the MHeC has the potential to be customized for use in developed and developing countries (and especially in those countries where Internet connectivity is high).

Information obtained from this alpha build (more specifically it's screening and triage protocols; online physical and mental health assessment, video visit system; and self-directed mental health support features) are now informing the development and build of the Synergy Online System (Synergy). Synergy is a Web-based modular platform that links integrated and interoperable resources (eg, apps, etools, data sharing, and access to online and in-clinic health services). Synergy operates through existing health providers to promote access to high-quality and cost-effective mental health services. This system does not deliver services or compete with existing service providers; rather, it aims to complement them by linking with other services. In addition, it enables real-time health and social outcomes tracking, thus providing high-quality and personalized service recommendations to the person seeking help. Synergy has been configured to permit the transfer of individual-level data (allowing for other ethical, consent, governance, and privacy considerations) between it and other existing record systems. The actual capacity to do this in any specific service setting depends on the ways in which existing service systems

are configured and governed. Synergy aims to enable Australian mental health service transformation for better outcomes for individual users, their families and supportive others, health professionals, and service providers.

Currently, a series of collaborative research trials are planned to evaluate the use of Synergy across the lifespan. This research is funded by a 3-year agreement (2017-20) between the Australian Government Department of Health and InnoWell Pty Ltd (a joint venture between The University of Sydney and PricewaterhouseCoopers) to the value of Aus $30 M. Results from this research will be reported separately.

Limitations

Our sample size in Phase 3 (one-on-one consultation with end users) was in the lower range of the average numbers for this type of study (between 6 and 12 participants) [89]. However, our sample size still enabled us to collect sufficient information for analysis in the framework and reach a saturation point. Further research is needed to understand the acceptability and usability of the system, as well as to validate all the components in real-world settings.

Conclusions

The MHeC has been designed to be accessible, affordable, available, and applicable to the many and varied settings in which there is the opportunity to connect with young people experiencing mental health concerns at anytime, anywhere across Australia, regardless of geographic location. The MHeC especially represents a solution for regional and rural/remote locations and the developing world where the need for mental health care far outstrips the number of health professionals available.

To the best of our knowledge, it is the first time that a Web-based clinic has been conceptualized and codesigned through an iterative cycle with end users. The close collaboration between all different stakeholders has the potential to effectively increase accessibility to mental health services for young people, the quality of care provided to young people, and also increase engagement with, and usability of, the final product. The main components of this system ensure that young people seeking help can find what they need within one system. The alpha prototype of the MHeC not only provides end users with accurate health information but also provides each person with the opportunity to have a comprehensive clinical (physical and mental health) assessment as well as real-time tailored interventions, including a video visit with an appropriate heath professional. Further research is needed to assess the acceptability of the MHeC as well as its applicability in large-scale studies.

Conflicts of Interest

LOP, TAD, CSR, and ACM have no conflicts of interest to declare. EMS is the Medical Director of the Young Adult Mental Health Unit, St Vincent's Hospital, Darlinghurst; Discipline Leader of Adult Mental Health, School of Medicine, University of Notre Dame; Research Affiliate, The University of Sydney; and Consultant Psychiatrist. She has received honoraria for educational seminars related to the clinical management of depressive disorders supported by Servier and Eli-Lilly pharmaceuticals. She has participated in a national advisory board for the antidepressant compound Pristiq, manufactured by Pfizer. She was the National Coordinator of an antidepressant trial sponsored by Servier. IBH has been a Mental Health Commissioner at the National Mental Health Commission since 2012. He is the Co-Director, Health and Policy at the Brain and Mind Centre (BMC), University of Sydney. The BMC operates an early-intervention youth service at Camperdown under contract to *headspace*. IBH has previously led community-based and pharmaceutical industry–supported (Wyeth, Eli Lily, Servier, Pfizer, AstraZeneca) projects focused on the identification and better management of anxiety and depression. He is a member of the Medical Advisory Panel for Medibank Private, a Board Member of Psychosis Australia Trust, and a member of the Department of Veterans' Affairs Veterans Mental Health Clinical Reference Group. He is the Chief Scientific Advisor to, and an equity shareholder in, Innowell Pty Ltd. Innowell has been formed by The University of Sydney and PricewaterhouseCoopers (PwC) to deliver the Aus $30 M Australian Government-funded "Project Synergy." Project Synergy is a 3-year program for the transformation of mental health services through the use of innovative technologies.

References

1. McHugh RK, Barlow DH. The dissemination and implementation of evidence-based psychological treatments. A review of current efforts. Am Psychol 2010;65(2):73-84. [doi: 10.1037/a0018121] [Medline: 20141263]

2. Judd FK, Jackson H, Davis J, Cockram A, Komiti A, Allen N, et al. Improving access for rural Australians to treatment for anxiety and depression: The University of Melbourne Depression and Anxiety Research and Treatment Group-Bendigo Health Care Group initiative. Aust J Rural Health 2001 Apr;9(2):91-96. [Medline: 11259963]

3. 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. [Medline: 12020305]

4. Kauer SD, Mangan C, Sanci L. Do online mental health services improve help-seeking for young people? A systematic review. J Med Internet Res 2014;16(3):e66 [FREE Full text] [doi: 10.2196/jmir.3103] [Medline: 24594922]

5. Wood FB, Benson D, LaCroix EM, Siegel ER, Fariss S. Use of Internet audience measurement data to gauge market share for online health information services. J Med Internet Res 2005 Jul 01;7(3):e31 [FREE Full text] [doi: 10.2196/jmir.7.3.e31] [Medline: 15998622]

6. Griffiths KM, Farrer L, Christensen H. The efficacy of internet interventions for depression and anxiety disorders: a review of randomised controlled trials. Med J Aust 2010 Jun 7;192(11 Suppl):S4-11. [Medline: 20528707]

7. Burns JM, Davenport TA, Christensen H, Luscombe GM, Mendoza JA, Bresnan A, et al. Game On: Exploring the Impact of Technologies on Young Men's Mental Health and Wellbeing: Findings from the first Young and Well National Survey. Melbourne: Young and Well Cooperative Research Centre; 2013.

8. Oh E, Jorm AF, Wright A. Perceived helpfulness of websites for mental health information: a national survey of young Australians. Soc Psychiatry Psychiatr Epidemiol 2009 Apr;44(4):293-299. [doi: 10.1007/s00127-008-0443-9] [Medline: 18830552]

9. Kim EH, Coumar A, Lober WB, Kim Y. Addressing mental health epidemic among university students via web-based, self-screening, and referral system: a preliminary study. IEEE Trans Inf Technol Biomed 2011 Mar;15(2):301-307. [doi: 10.1109/TITB.2011.2107561] [Medline: 21257386]

10. Cross SP, Hickie I. Transdiagnostic stepped care in mental health. Public Health Res Pract 2017 Apr 27;27(2):- [FREE Full text] [doi: 10.17061/phrp2721712] [Medline: 28474049]

11. Kauer SD, Mangan C, Sanci L. Do online mental health services improve help-seeking for young people? A systematic review. J Med Internet Res 2014;16(3):e66 [FREE Full text] [doi: 10.2196/jmir.3103] [Medline: 24594922]

12. Slade T, Johnston A, Oakley Browne MA, Andrews G, Whiteford H. 2007 National Survey of Mental Health and Wellbeing: methods and key findings. Aust N Z J Psychiatry 2009 Jul;43(7):594-605. [doi: 10.1080/00048670902970882] [Medline: 19530016]

13. Libby AM, Brent DA, Morrato EH, Orton HD, Allen R, Valuck RJ. Decline in treatment of pediatric depression after FDA advisory on risk of suicidality with SSRIs. Am J Psychiatry 2007 Jun;164(6):884-891. [doi: 10.1176/ajp.2007.164.6.884] [Medline: 17541047]

14. Clarke G, Yarborough BJ. Evaluating the promise of health IT to enhance/expand the reach of mental health services. Gen Hosp Psychiatry 2013;35(4):339-344 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.013] [Medline: 23701698]

XSL•FO

RenderX

15. Dudgeon P, Walker R, Scrine C, Shepherd C, Calma T, Ring I. Effective Strategies to Strengthen the Mental Health and Wellbeing of Aboriginal and Torres Strait Islander People. Canberra: Australian Institute of Health and Welfare (AIHW); 2014:1-72.

16. Australian Bureau of Statistics. 2014. 8146.0 Household Use of Information Technology, Australia, 2016-17 URL: http://www.abs.gov.au/ausstats/abs@.nsf/mf/8146.0 [WebCite Cache ID 6vY5o0tpY]

17. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]

18. Naslund JA, Marsch LA, McHugo GJ, Bartels SJ. Emerging mHealth and eHealth interventions for serious mental illness: a review of the literature. J Ment Health 2015;24(5):321-332. [doi: 10.3109/09638237.2015.1019054] [Medline: 26017625]

19. Tate DF, Finkelstein EA, Khavjou O, Gustafson A. Cost effectiveness of internet interventions: review and recommendations. Ann Behav Med 2009 Aug;38(1):40-45 [FREE Full text] [doi: 10.1007/s12160-009-9131-6] [Medline: 19834778]

20. Iorfino F, Davenport TA, Ospina-Pinillos L, Hermens DF, Cross S, Burns J, et al. Using new and emerging technologies to identify and respond to suicidality among help-seeking young people: a cross-sectional study. J Med Internet Res 2017 Jul 12;19(7):e247 [FREE Full text] [doi: 10.2196/jmir.7897] [Medline: 28701290]

21. Frost J, Okun S, Vaughan T, Heywood J, Wicks P. Patient-reported outcomes as a source of evidence in off-label prescribing: analysis of data from PatientsLikeMe. J Med Internet Res 2011;13(1):e6 [FREE Full text] [doi: 10.2196/jmir.1643] [Medline: 21252034]

22. Faurholt-Jepsen M, Vinberg M, Frost M, Christensen EM, Bardram J, Kessing LV. Daily electronic monitoring of subjective and objective measures of illness activity in bipolar disorder using smartphones--the MONARCA II trial protocol: a randomized controlled single-blind parallel-group trial. BMC Psychiatry 2014;14:309 [FREE Full text] [doi: 10.1186/s12888-014-0309-5] [Medline: 25420431]

23. Burns JM, Birrell E, Bismark M, Pirkis J, Davenport TA, Hickie IB, et al. The role of technology in Australian youth mental health reform. Aust Health Rev 2016 Dec;40(5):584-590. [doi: 10.1071/AH15115] [Medline: 26934382]

24. Powell J, Clarke A. Investigating internet use by mental health service users: interview study. Stud Health Technol Inform 2007;129(Pt 2):1112-1116. [Medline: 17911888]

25. Fox S, Duggan M. Pew Research Center. 2013. Health Online 2013 URL: http://www.pewinternet.org/2013/01/15/health-online-2013/ [WebCite Cache ID 6vY5MVtrg]

26. Castle-Clarke S, Imison C. The digital patient: transforming primary care?. London: Nuffield Trust; 2016 Nov. URL: http://www.informatics.manchester.ac.uk/media/1691/nuffield-report.pdf [WebCite Cache ID 6v8FaoQ9s]

27. Madan A. GP Access. 2014. webGP: the virtual general practice URL: http://gpaccess.uk/wordpress/wp-content/uploads/2015/08/e-consult-pilot-report-2014.pdf [WebCite Cache ID 6v8G5Lqo5]

28. Greist JH, Gustafson DH, Stauss FF, Rowse GL, Laughren TP, Chiles JA. A computer interview for suicide-risk prediction. Am J Psychiatry 1973 Dec;130(12):1327-1332. [doi: 10.1176/ajp.130.12.1327] [Medline: 4585280]

29. Hines DA, Douglas EM, Mahmood S. The effects of survey administration on disclosure rates to sensitive items among men: a comparison of an Internet panel sample with a RDD telephone sample. Comput Human Behav 2010 Nov 01;26(6):1327-1335 [FREE Full text] [doi: 10.1016/j.chb.2010.04.006] [Medline: 20948972]

30. Heeren T, Edwards EM, Dennis JM, Rodkin S, Hingson RW, Rosenbloom DL. A comparison of results from an alcohol survey of a prerecruited Internet panel and the National Epidemiologic Survey on Alcohol and Related Conditions. Alcohol Clin Exp Res 2008 Feb;32(2):222-229. [doi: 10.1111/j.1530-0277.2007.00571.x] [Medline: 18162076]

31. Tourangeau R, Yan T. Sensitive questions in surveys. Psychol Bull 2007 Sep;133(5):859-883. [doi: 10.1037/0033-2909.133.5.859] [Medline: 17723033]

32. Bradford S, Rickwood D. Acceptability and utility of an electronic psychosocial assessment (myAssessment) to increase self-disclosure in youth mental healthcare: a quasi-experimental study. BMC Psychiatry 2015 Dec 01;15:305 [FREE Full text] [doi: 10.1186/s12888-015-0694-4] [Medline: 26627041]

33. Cuijpers P, Boluijt P, van Straten A. Screening of depression in adolescents through the Internet : sensitivity and specificity of two screening questionnaires. Eur Child Adolesc Psychiatry 2008 Feb;17(1):32-38. [doi: 10.1007/s00787-007-0631-2] [Medline: 17876508]

34. Farvolden P, McBride C, Bagby RM, Ravitz P. A Web-based screening instrument for depression and anxiety disorders in primary care. J Med Internet Res 2003;5(3):e23 [FREE Full text] [doi: 10.2196/jmir.5.3.e23] [Medline: 14517114]

35. Lawrence ST, Willig JH, Crane HM, Ye J, Aban I, Lober W, et al. Routine, self-administered, touch-screen, computer-based suicidal ideation assessment linked to automated response team notification in an HIV primary care setting. Clin Infect Dis 2010 Apr 15;50(8):1165-1173 [FREE Full text] [doi: 10.1086/651420] [Medline: 20210646]

36. Lupton D, Jutel A. 'It's like having a physician in your pocket!' A critical analysis of self-diagnosis smartphone apps. Soc Sci Med 2015 May;133:128-135. [doi: 10.1016/j.socscimed.2015.04.004] [Medline: 25864149]

37. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. Br Med J 2015;351:h3480 [FREE Full text] [Medline: 26157077]

38. Shore J. The evolution and history of telepsychiatry and its impact on psychiatric care: current implications for psychiatrists and psychiatric organizations. Int Rev Psychiatry 2015;27(6):469-475. [doi: 10.3109/09540261.2015.1072086] [Medline: 26397182]

39. Cukor P, Baer L, Willis BS, Leahy L, O'Laughlen J, Murphy M, et al. Use of videophones and low-cost standard telephone lines to provide a social presence in telepsychiatry. Telemed J 1998;4(4):313-321. [doi: 10.1089/tmj.1.1998.4.313] [Medline: 10220471]

40. Barak A. Emotional support and suicide prevention through the Internet: a field project report. Comput Hum Behav 2007 Mar;23(2):971-984. [doi: 10.1016/j.chb.2005.08.001]

41. Hubley S, Lynch SB, Schneck C, Thomas M, Shore J. Review of key telepsychiatry outcomes. World J Psychiatry 2016 Jun 22;6(2):269-282 [FREE Full text] [doi: 10.5498/wjp.v6.i2.269] [Medline: 27354970]

42. Alexander J, Lattanzio A. Utility of telepsychiatry for Aboriginal Australians. Aust N Z J Psychiatry 2009 Dec;43(12):1185. [doi: 10.3109/00048670903279911] [Medline: 20001419]

43. Wallace D, Rayner S. Telepsychiatry services in the Australian Defence Force. Australas Psychiatry 2013 Jun;21(3):278-279. [doi: 10.1177/1039856213476123] [Medline: 23720468]

44. Wood J, Stathis S, Smith A, Krause J. E-CYMHS: an expansion of a child and youth telepsychiatry model in Queensland. Australas Psychiatry 2012 Aug;20(4):333-337. [doi: 10.1177/1039856212450756] [Medline: 22773412]

45. Douglas MD, Xu J, Heggs A, Wrenn G, Mack DH, Rust G. Assessing telemedicine utilization by using Medicaid claims data. Psychiatr Serv 2017 Feb 01;68(2):173-178 [FREE Full text] [doi: 10.1176/appi.ps.201500518] [Medline: 27691381]

46. Williams A, Larocca R, Chang T, Trinh N, Fava M, Kvedar J, et al. Web-based depression screening and psychiatric consultation for college students: a feasibility and acceptability study. Int J Telemed Appl 2014;2014:580786 [FREE Full text] [doi: 10.1155/2014/580786] [Medline: 24799895]

47. Cross SP, Hermens DF, Scott EM, Ottavio A, McGorry PD, Hickie IB. A clinical staging model for early intervention youth mental health services. Psychiatr Serv 2014 Jul;65(7):939-943. [doi: 10.1176/appi.ps.201300221] [Medline: 24828746]

48. Clarke AM, Kuosmanen T, Barry MM. A systematic review of online youth mental health promotion and prevention interventions. J Youth Adolesc 2015 Jan;44(1):90-113. [doi: 10.1007/s10964-014-0165-0] [Medline: 25115460]

49. Doherty G, Coyle D, Sharry J. Engagement with Online Mental Health Interventions: An Exploratory Clinical Study of a Treatment for Depression. 2012 Presented at: ACM Annual Conference on Human Factors in Computing Systems - CHI '12; May 5-10, 2012; Austin URL: https://www.scss.tcd.ie/Gavin.Doherty/papers/DohertyCS12.pdf

50. Orlowski SK, Lawn S, Venning A, Winsall M, Jones GM, Wyld K, et al. Participatory research as one piece of the puzzle: a systematic review of consumer involvement in design of technology-based youth mental health and well-Being interventions. JMIR Hum Factors 2015;2(2):e12 [FREE Full text] [doi: 10.2196/humanfactors.4361] [Medline: 27025279]

51. Lyles CR, Altschuler A, Chawla N, Kowalski C, McQuillan D, Bayliss E, et al. User-centered design of a tablet waiting room tool for complex patients to prioritize discussion topics for primary care visits. JMIR Mhealth Uhealth 2016 Sep 14;4(3):e108 [FREE Full text] [doi: 10.2196/mhealth.6187] [Medline: 27627965]

52. Bilsbury CD, Richman A. A staging approach to measuring patient-centred subjective outcomes. Acta Psychiatr Scand Suppl 2002(414):5-40. [Medline: 12366461]

53. Bjerkan J, Hedlund M, Hellesø R. Patients' contribution to the development of a web-based plan for integrated care - a participatory design study. Inform Health Soc Care 2015 Mar;40(2):167-184. [doi: 10.3109/17538157.2014.907803] [Medline: 24786524]

54. Rothgangel A, Braun S, Smeets R, Beurskens A. Design and development of a telerehabilitation platform for patients with phantom limb pain: a user-centered approach. JMIR Rehabil Assist Technol 2017 Feb 15;4(1):e2 [FREE Full text] [doi: 10.2196/rehab.6761] [Medline: 28582249]

55. Peters D, Davis S, Calvo RA, Sawyer SM, Smith L, Foster JM. Young people's preferences for an asthma self-management app highlight psychological needs: a participatory study. J Med Internet Res 2017 Apr 11;19(4):e113 [FREE Full text] [doi: 10.2196/jmir.6994] [Medline: 28400353]

56. Muller MJ, Kuhn S. Participatory design. Commun ACM 1993;36(6):24-28. [doi: 10.1145/153571.255960]

57. Sjöberg C, Timpka T. Participatory design of information systems in health care. J Am Med Inform Assoc 1998;5(2):177-183 [FREE Full text] [Medline: 9524350]

58. Hagen P, Collin P, Metcalf A, Nicholas M, Rahilly K, Swainston N. Participatory design of evidence-based online youth mental health promotion, intervention and treatment: A Young and Well Cooperative Research Centre Innovative Methodologies Guide. Melbourne: Young and Well Cooperative Research Centre; 2012. URL: https://www. westernsydney.edu.au/__data/assets/pdf_file/0005/476330/Young_and_Well_CRC_IM_PD_Guide.pdf [WebCite Cache ID 6z7mdPfmB]

59. Sullivan P. Beyond a narrow conception of usability testing. IEEE Trans Profess Commun 1989;32(4):256-264. [doi: 10.1109/47.44537]

60. Gordon M, Henderson R, Holmes JH, Wolters MK, Bennett IM. Participatory design of ehealth solutions for women from vulnerable populations with perinatal depression. J Am Med Inform Assoc 2016 Jan;23(1):105-109. [doi: 10.1093/jamia/ocv109] [Medline: 26342219]

61. Monshat K, Vella-Brodrick D, Burns J, Herrman H. Mental health promotion in the Internet age: a consultation with Australian young people to inform the design of an online mindfulness training programme. Health Promot Int 2012 Jun;27(2):177-186 [FREE Full text] [doi: 10.1093/heapro/dar017] [Medline: 21398335]

62. Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. J R Soc Med 2011 Dec;104(12):510-520 [FREE Full text] [doi: 10.1258/jrsm.2011.110180] [Medline: 22179294]

63. Burns J, Morey C, Lagelée A, Mackenzie A, Nicholas J. Reach Out! Innovation in service delivery. Med J Aust 2007 Oct 1;187(7 Suppl):S31-S34. [Medline: 17908022]

64. Harman G, Heath J. Australian country perspective: the work of beyondblue and SANE Australia. In: Gaebel W, Rössler W, Sartorius N, editors. The Stigma of Mental Illness - End of the Story?. Cham, Switzerland: Springer International Publishing; 2017:289-315.

65. Hickie I. Can we reduce the burden of depression? The Australian experience with beyondblue: the national depression initiative. Australas Psychiatry 2004;12 Suppl:S38-S46. [doi: 10.1111/j.1039-8562.2004.02097.x] [Medline: 15715829]

66. Rickwood D, Webb M, Kennedy V, Telford N. Who are the young people choosing web-based mental health support? Findings from the implementation of Australia's national web-based youth mental health service, eheadspace. JMIR Ment Health 2016 Aug 25;3(3):e40 [FREE Full text] [doi: 10.2196/mental.5988] [Medline: 27562729]

67. Christensen H, Griffiths KM, Korten A. Web-based cognitive behavior therapy: analysis of site usage and changes in depression and anxiety scores. J Med Internet Res 2002;4(1):e3 [FREE Full text] [doi: 10.2196/jmir.4.1.e3] [Medline: 11956035]

68. Titov N, Dear BF, Staples LG, Bennett-Levy J, Klein B, Rapee RM, et al. MindSpot Clinic: an accessible, efficient, and effective online treatment service for anxiety and depression. Psychiatr Serv 2015 Oct;66(10):1043-1050. [doi: 10.1176/appi.ps.201400477] [Medline: 26130001]

69. Balsamiq. 2017. Balsamiq Studios URL: https://balsamiq.com/ [accessed 2017-12-04] [WebCite Cache ID 6vRtbhPus]

70. Buxton B. Sketching User Experiences: Getting the Design Right and the Right Design. Burlington: Morgan Kaufmann; 2007:1-448.

71. 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]

72. Boren T, Ramey J. Thinking aloud: reconciling theory and practice. IEEE Trans Profess Commun 2000;43(3):261-278. [doi: 10.1109/47.867942]

73. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

74. Goldman HH, Skodol AE, Lave TR. Revising axis V for DSM-IV: a review of measures of social functioning. Am J Psychiatry 1992 Sep;149(9):1148-1156. [Medline: 1386964]

75. 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]

76. Eisenberg D, Gollust SE, Golberstein E, Hefner JL. Prevalence and correlates of depression, anxiety, and suicidality among university students. Am J Orthopsychiatry 2007 Oct;77(4):534-542. [doi: 10.1037/0002-9432.77.4.534] [Medline: 18194033]

77. Garlow SJ, Rosenberg J, Moore JD, Haas AP, Koestner B, Hendin H, et al. Depression, desperation, and suicidal ideation in college students: results from the American Foundation for Suicide Prevention College Screening Project at Emory University. Depress Anxiety 2008;25(6):482-488. [doi: 10.1002/da.20321] [Medline: 17559087]

78. Gilbody S, Sheldon T, House A. Screening and case-finding instruments for depression: a meta-analysis. CMAJ 2008 Apr 08;178(8):997-1003 [FREE Full text] [doi: 10.1503/cmaj.070281] [Medline: 18390942]

79. Donker T, Griffiths KM, Cuijpers P, Christensen H. Psychoeducation for depression, anxiety and psychological distress: a meta-analysis. BMC Med 2009;7:79 [FREE Full text] [doi: 10.1186/1741-7015-7-79] [Medline: 20015347]

80. Klein MC, Ciotoli C, Chung H. Primary care screening of depression and treatment engagement in a university health center: a retrospective analysis. J Am Coll Health 2011;59(4):289-295. [doi: 10.1080/07448481.2010.503724] [Medline: 21308589]

81. Baum F, Lawless A, Delany T, Macdougall C, Williams C, Broderick D, et al. Evaluation of health in all policies: concept, theory and application. Health Promot Int 2014 Jun;29 Suppl 1:i130-i142. [doi: 10.1093/heapro/dau032] [Medline: 25217350]

82. Australian Bureau of Statistics. 2008. Adult Literacy and Life Skills Survey, Summary Results, Australia, 2006 (Reissue) URL: http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4228.0Main+Features12006%20(Reissue) [WebCite Cache ID 6v8Ca3cQ3]

83. Suler J. Assessing a person's suitability for online therapy: the ISMHO clinical case study group. International Society for Mental Health Online. Cyberpsychol Behav 2001 Dec;4(6):675-679. [doi: 10.1089/109493101753376614] [Medline: 11800174]

84. Rickwood DJ, Telford NR, Parker AG, Tanti CJ, McGorry PD. headspace - Australia's innovation in youth mental health: who are the clients and why are they presenting? Med J Aust 2014 Feb 3;200(2):108-111. [Medline: 24484115]

85. Roymorgan. 9 in 10 Aussie teens now have a mobile (and most are already on to their second or subsequent handset) URL: http://www.roymorgan.com/findings/6929-australian-teenagers-and-their-mobile-phones-june-2016-201608220922 [accessed 2018-02-19] [WebCite Cache ID 6xL5Ulht8]

86. Deloitte. Smart everything, everywhere Mobile Consumer Survey 2017 The Australian cut URL: https://www2.deloitte.com/au/mobile-consumer-survey [accessed 2018-02-19] [WebCite Cache ID 6xL6Bkpeq]

XSL•FO
RenderX

87.   National Mental Health Commission. Report of the National Review of Mental Health Programmes and Services: Summary. Sydney: National Mental Health Commission; 2014.

88.   Rickwood D, Van Dyke N, Telford N. Innovation in youth mental health services in Australia: common characteristics across the first headspace centres. Early Interv Psychiatry 2015 Feb;9(1):29-37. [doi: 10.1111/eip.12071] [Medline: 23827044]

89.   Dumas JS, Redish J. A Practical Guide to Usability Testing. Exeter: Intellect Books; 1999.

Abbreviations

KT: knowledge translation
MHeC: Mental Health eClinic
NMHC: National Mental Health Commission
PD: participatory design
R&D: research and development

XSL•FO
RenderX

What Do Germans Want to Know About Skin Cancer? A Nationwide Google Search Analysis From 2013 to 2017

Stefanie Seidl[1], BSc; Barbara Schuster[1], BA; Melvin Rüth[1]; Tilo Biedermann[1], MD; Alexander Zink[1], MD, MPH

Klinikum rechts der Isar, Department of Dermatology and Allergy, Technical University of Munich, Munich, Germany

Corresponding Author:
Alexander Zink, MD, MPH
Klinikum rechts der Isar
Department of Dermatology and Allergy
Technical University of Munich
Biedersteiner Str. 29
Munich,
Germany
Phone: 49 8941403572
Email: alexander.zink@tum.de

Abstract

Background: Experts worldwide agree that skin cancer is a global health issue, but only a few studies have reported on world populations' interest in skin cancer. Internet search data can reflect the interest of a population in different topics and thereby identify what the population wants to know.

Objective: Our aim was to assess the interest of the German population in nonmelanoma skin cancer and melanoma.

Methods: Google AdWords Keyword Planner was used to identify search terms related to nonmelanoma skin cancer and melanoma in Germany from November 2013 to October 2017. The identified search terms were assessed descriptively using SPSS version 24.0. In addition, the search terms were qualitatively categorized.

Results: A total of 646 skin cancer-related search terms were identified with 19,849,230 Google searches in the period under review. The search terms with the highest search volume were "skin cancer" (n=2,388,500, 12.03%), "white skin cancer" (n=2,056,900, 10.36%), "basalioma" (n=907,000, 4.57%), and "melanoma" (n=717,800, 3.62%). The most searched localizations of nonmelanoma skin cancer were "nose" (n=93,370, 38.99%) and "face" (n=53,270, 22.24%), and the most searched of melanoma were "nails" (n=46,270, 70.61%) and "eye" (n=10,480, 15.99%). The skin cancer related category with the highest search volume was "forms of skin cancer" (n=10,162,540, 23.28%) followed by "skin alterations" (n=4,962,020, 11.36%).

Conclusions: Our study provides insight into terms and fields of interest related to skin cancer relevant to the German population. Furthermore, temporal trends and courses are shown. This information could aid in the development and implementation of effective and sustainable awareness campaigns by developing information sources targeted to the population's broad interest or by implementing new Internet campaigns.

KEYWORDS

skin cancer; melanoma; nonmelanoma skin cancer (NMSC); Google; search analysis; population

Introduction

The incidence of skin cancer, including nonmelanoma skin cancer (NMSC) and melanoma, is a major public health issue [1]. In fact, NMSC is the most common cancer among Caucasians worldwide [2]. In Germany alone, approximately 180,000 new cases are reported each year [3]. The main risk factor for NMSC is exposure to solar ultraviolet radiation (UVR) [4]. Thus, sun-exposed skin areas, such as the nose, neck, head, and face, are particularly at risk [5,6]. Less commonly, NMSC can occur on skin areas that are not exposed to sunlight owing to other etiopathogenetic factors, such as alterations in lymphatic circulation [7,8]. Currently, NMSC has an enormous socioeconomic impact with a continuously increasing incidence within the last few years [9]. In Germany, the incidence is estimated to double by 2030 [3]. Not as frequent, but with a substantially higher mortality than NMSC, is melanoma, which is diagnosed in approximately 21,000 new patients each year in Germany [10]. Recently, new diagnostic and treatment

XSL•FO
RenderX

options, especially for melanoma, have had a significant impact in the global medical community [11-14]. At the same time, NMSC has been recognized as an occupational disease in some countries, and new techniques and products for UVR protection, including apps and gadgets, have been developed [15,16]. In summary, experts agree about the global problem of skin cancer, but few studies have focused on the interest of the population [17], which might help lower the global burden. One option for analyzing the interest of the population is an Internet search analysis. This is a promising approach to reflecting the population's interest in a certain topic [18]. It is a novel tool to estimate the impact of disease in a population where traditional methods are inadequate or in the absence of data sources [19]. Analysis of the Internet search volume for different search terms provides insight into the population's general interest. Therefore, the term "search volume" concentrates on the number of searches of a particular topic or search term. This procedure has been used by communication media for several years [20]. The Internet's emerging role as a main source of health information for the population [21,22] has created a corresponding value as a novel and informative method in the medical field. Recently, Huang et al reported a small association between online cancer-related information searching behaviors and skin cancer incidences [23], and Wehner et al successfully demonstrated that Internet search volume positively correlated with the incidence and mortality rates of common cancers, including melanoma, in the United States [18]. This indicates that the unconventional method of Internet search analysis can provide useful data on the characteristics of a disease including incidence and mortality. Hence, the Internet search data reflecting the population's interest can provide information about real-life skin cancer incidence as well as medical needs. For this purpose, the Google search engine is promising for Germany because of its 95% market share [24]. Moreover, the population prefers the use of search engines such as Google over specialized websites when searching for health information online [25-27].

The aim of this study was to analyze the interest of the German population in skin cancer by analyzing Google searches of terms related to NMSC and melanoma.

Methods

Study Design

In this retrospective longitudinal study, we used Google AdWords Keyword Planner to identify the search volume of terms related to NMSC and melanoma. Google AdWords Keyword Planner is usually used by advertisers to optimize Google marketing campaigns. The most important part of Google AdWords is keywords. With their help, an advertiser can specify in advance that an ad should be displayed only in the result for a search for the mentioned terms or thematically appropriate pages. Therefore, the tool can search for keyword ideas; new keywords can be included or lists of keywords can be uploaded [28]. However, because this technology indicates the monthly search volume estimated by Google, it can also be used for scientific questions. In our study, the search terms related to NMSC and melanoma were identified using a keyword cluster including the following 13 common terms for skin cancer in German: "skin cancer," "black skin cancer," "white skin cancer," "light skin cancer," "nonmelanocytic skin cancer," "nonmelanoma skin cancer," "NMSC," "melanoma," "malignant melanoma," "basalioma," "basal cell carcinoma," "spinalioma," and "squamous cell carcinoma." On the basis of this cluster, Google AdWords Keyword Planner identified search terms to be analyzed regarding their search volume. The search volume data included only those from Google users with a German Internet protocol address and were analyzed from November 2013 to October 2017.

Figure 1. Content categorization of search terms identified by Google AdWords Keyword Planner. The search terms were manually screened and categorized into 6 skin cancer–related categories and 2 other categories. "No category" was used for search terms not fitting in any category. The figure includes absolute numbers and percentages of search volume and an example search term for each category.
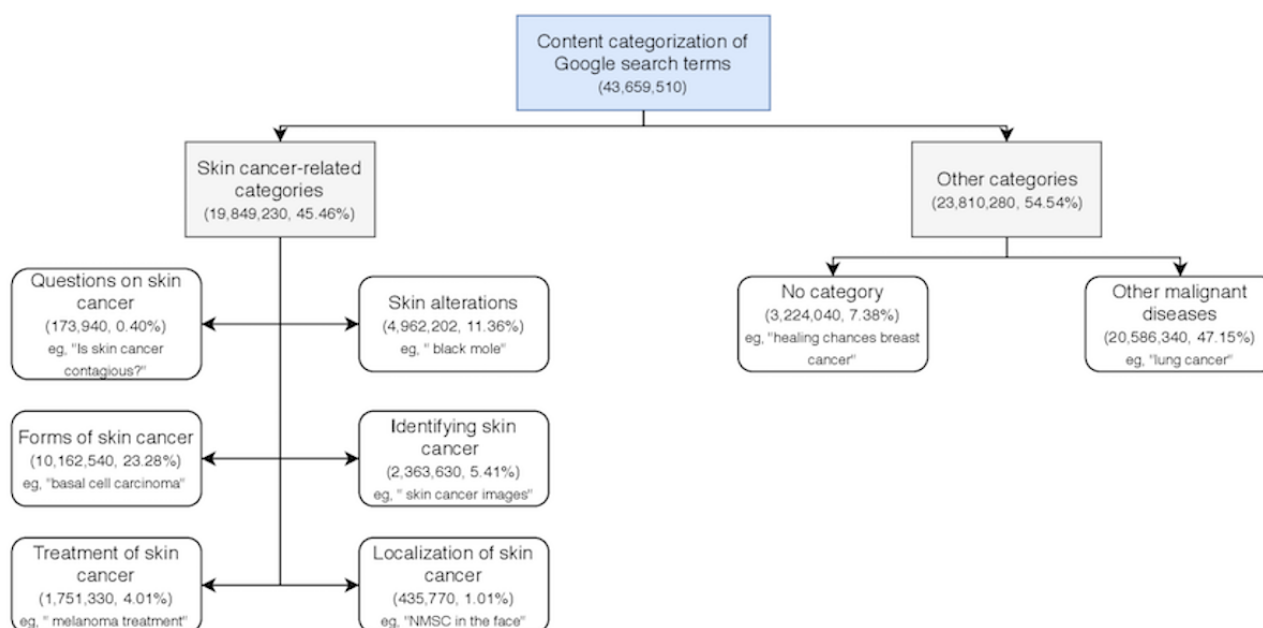
Statistical Analysis

The search volume data from the identified search terms were assessed descriptively using SPSS version 24.0. Furthermore, the search terms were qualitatively categorized based on their content after having been read carefully. In the first step, the content of all search terms was analyzed, and 6 skin cancer-related categories were identified (Figure 1). A further category for "other malignant diseases" (eg, "lung cancer") was also identified. Search terms that did not fit in any of these categories were placed in the "no category" (eg, "healing chances breast cancer"; Figure 1). Each search term was only assigned to one category. In the second step, we categorized and analyzed the search terms within the category "localization of skin cancer" according to the exact localization of NMSC and melanoma separately.

Results

In total, Google AdWords Keyword Planner identified 714 search terms related to NMSC and melanoma with a search volume of 43,659,510 in Germany from November 2013 to October 2017. Sixty-eight search terms referred to other malignant diseases or were not assignable terms, while 646 search terms directly referred to skin cancer (Figure 1). Of the skin cancer–related search terms, the most common terms were "skin cancer" (n=2,388,500, 12.03%) followed by "white skin cancer" (n=2,056,900, 10.36%), "basalioma" (n=907,000, 4.57%), "melanoma" (n=717,800, 3.62%), and "black skin cancer" (n=649,400, 3.27%; Table 1 and Multimedia Appendix 1).

Time Analyses

Figure 2 shows the Google search volume of terms related to NMSC and melanoma from November 2013 to October 2017 with annual increases. The lowest search volume was in December 2013 (n=208,400, 1.05%), while the highest volume was in June 2017 (n=398,590, 2.01%). Every year, the search volume increased in April and May, remained high over June and July, and decreased again in August. The month with the highest search volume over all consecutive years was July, except in 2017, when it was June. The largest increase was measured from April 2014 to May 2014, when the search volume increased 32.9% from 266,970 to 354,780 (Figure 2).

Table 1. Most frequently searched skin cancer–related terms in Germany from November 2013 to October 2017 (N=19,849,230).

Ranking	Skin cancer–related term	Search volume, n (%)
1	Skin cancer	2,388,500 (12.03)
2	White skin cancer	2,056,900 (10.36)
3	Basalioma	907,000 (4.57)
4	Melanoma	717,800 (3.62)
5	Black skin cancer	649,400 (3.27)
6	Actinic keratosis	482,500 (2.43)
7	Squamous cell carcinoma	433,300 (2.19)
8	Skin cancer pictures	366,700 (1.85)
9	Malignant melanoma	336,900 (1.70)
10	Basal cell carcinoma	292,500 (1.47)

Figure 2. Nationwide Google searches of terms related to nonmelanoma skin cancer and melanoma in Germany from November 2013 to October 2017. Only skin cancer–related search terms are included; search terms from the categories "no category" and "other malignant diseases" are not.
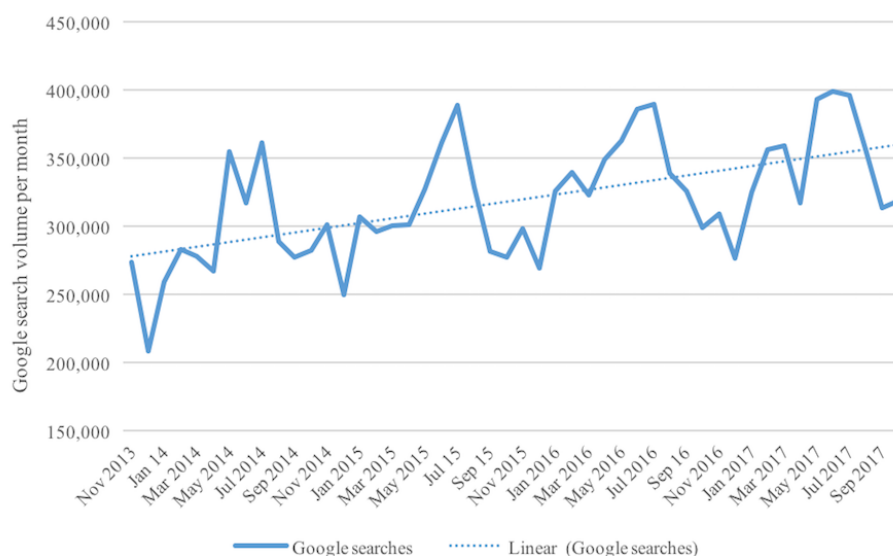
Nonmelanoma Skin Cancer and Melanoma

In a total of 8,953,870 searches of terms related to NMSC and melanoma, NMSC-related terms (eg, "white skin cancer," "nonmelanoma skin cancer") had a search volume of 4,421,480 (49.39%), which was categorized as referring to basal cell carcinoma or squamous cell carcinoma or not classifiable to either. Melanoma-related terms (eg, "black skin cancer," "malignant melanoma") had a search volume of 2,014,130 (22.49%; Figure 3). A total of 2,518,260 (28.12%) searches referred to skin cancer overall, including search terms such as "cancer skin."

A total of 927,090 searches referred to precursor lesions of skin cancer: 895,710 (96.62%) on NMSC precursor lesions and 7080 (0.76%) on melanoma precursor lesions. A total of 24,300 (2.62%) searches referred to precursor lesions of skin cancer overall (eg, "skin cancer precursor lesion").

Few searches referred to skin cancer stages (n=204,510) or metastasis of skin cancer (n=58,210); however, melanoma (n=79,830, 39.03% and n=28,170, 48.39%, respectively) had a larger search volume than NMSC (n=63,000, 30.81% and n=15,880, 27.18%, respectively).

Search Term Categories

In total, 92.62% of the search volume fit into a specific category, whereas 7.38% was summarized into "no category" because of nonspecificity (eg, "cancer") or the topic (eg, "healing chances breast cancer").

Nearly half of the search volume (N=43,659,510) was in the category "other malignant diseases" (n=20,586,340, 47.15%) followed by the categories "forms of skin cancer" (n=10,162,540, 23.28%) and "skin alterations" (n=4,962,020, 11.36%; Table 2). Few searches referred to "treatment of skin cancer" (n=1,751,330, 4.01%), "localization of skin cancer" (n=435,770, 1.01%), or "questions on skin cancer" (n=173,940, 0.40%).

Localization

In a total of 435,770 searches of terms related to skin cancer localization, 239,530 (54.97%) referred to NMSC and 65,530 (15.04%) referred to melanoma. A total of 130,710 searches (29.99%) did not fit into either NMSC or melanoma because the search terms contained only "skin cancer"; these were included in "skin cancer overall."

The most commonly searched localizations of NMSC were "nose" (n=93,370, 38.99%) followed by "face" (n=53,270, 22.24%) and "eye" (n=33,320; 13.9%). There were considerably less NMSC searches for "legs" (n=2730, 1.14 %) or "hands" (n=1860, 0.78%). For melanoma, interestingly, the most commonly searched localization was "nails" (n=46,270, 70.61%). With a search volume of 10,480 (15.99%), "melanoma eye," which probably refers to uveal melanoma, was the second most searched term, followed by melanoma of the "face" (n=3720, 5.68%) and melanoma of the "head" (n=3420, 5.22%). Few searches were done for melanoma of the "back" (n=1640, 2.50%). For "skin cancer overall," "face" (n=33,830, 25.88%), "head" (n=23,740, 18.16%), and "nose" (n=20,960, 16.04%) were the most frequently searched localizations (Figure 4).

Figure 3. Google search volume of terms related to nonmelanoma skin cancer (NMSC) and melanoma in Germany from November 2013 to October 2017. "NMSC overall" includes search terms such as "white skin cancer", "BCC", and "SCC". "Melanoma" includes search terms such as "malignant melanoma". (BCC=basal cell carcinoma; SCC=squamous cell carcinoma).

XSL•FO

RenderX
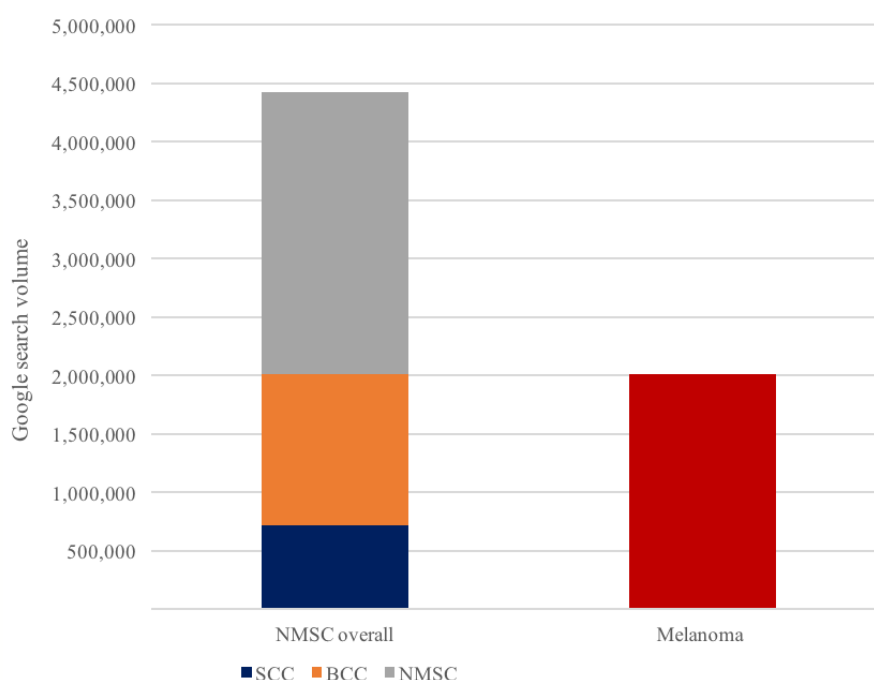
Table 2. Content categorization of the identified search terms related to nonmelanoma skin cancer and melanoma in Germany from November 2013 to October 2017 (N=43,659,510).

Search term category	n (%)
Other malignant diseases	20,586,340 (47.15)
Forms of skin cancer	10,162,540 (23.28)
Skin alterations	4,962,020 (11.36)
No category	3,224,040 (7.38)
Identifying skin cancer	2,363,630 (5.41)
Treatment of skin cancer	1,751,330 (4.01)
Localization of skin cancer	435,770 (1.01)
Questions on skin cancer	173,940 (0.40)

Figure 4. Google search volume of localization of nonmelanoma skin cancer, melanoma, and skin cancer overall in Germany from November 2013 to October 2017. Search terms without an exact skin cancer term (eg, "skin cancer nose") are included in "skin cancer overall.".
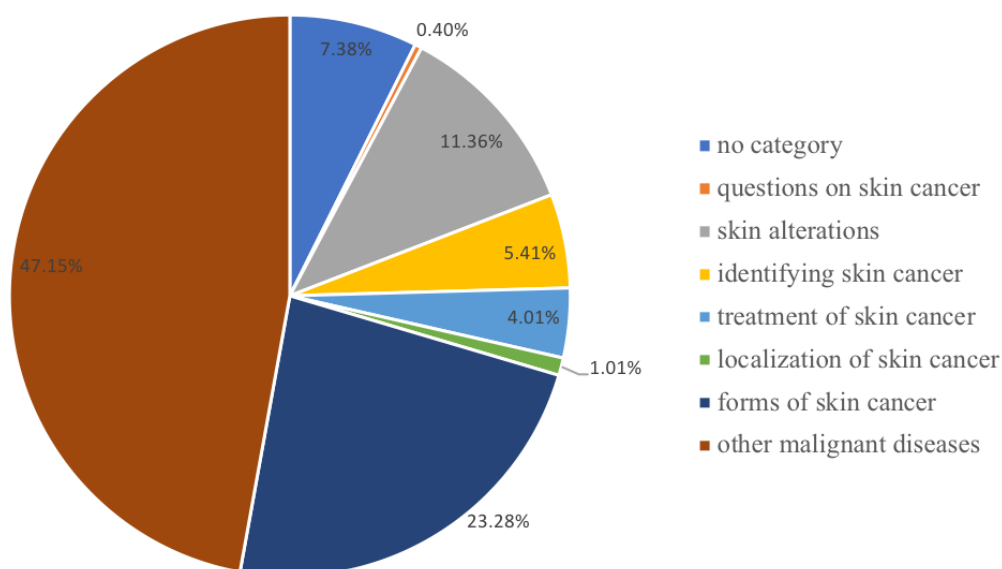
Discussion

Principal Considerations

In total, approximately 19.85 million skin cancer-related Google searches were conducted in Germany between November 2013 and October 2017. The search terms with the highest search volume were "skin cancer," "white skin cancer," "basalioma," "melanoma," and "black skin cancer." The course of time shows an overall increase in Google search volume from 2013 through 2017, with a higher search volume during each summer.

The overall increase in Google search volume of "NMSC" and "melanoma" in the course of time might be explained, in part, by the continuously increasing incidence of skin cancer [10,29] and public awareness campaigns [30]. This might have also led to an overall increase in public interest in skin cancer. Particularly, the recognition of NMSC as an occupational disease in Germany in 2015 could have led to an increased interest [15,29,31-33], which might be a reason for the higher search volume over summer 2015 compared with summer 2014.

Due to the fact that UVR is the main risk factor for NMSC and a significant risk factor for melanoma [10], it is not surprising that the summer months had the highest search volume each year. In addition, the media report on sun safety and many skin safety campaigns run during this season [34,35]. Moreover, Figure 2 demonstrates that the public interest increases in late spring, when people more frequently wear short-sleeved clothing due to rising temperatures and decreases in late summer when temperatures begin to decrease in Germany [36]. The seasonal variation may also be affected by diagnoses of NMSC and melanoma, which are reported to be significantly higher in late spring and early summer [37,38]. These seasonal differences in search volume were also reported by Bloom et al with regard to skin cancer and melanoma in the United States [17].

The incidence of NMSC is several times higher than that of melanoma [39]. The population's interest, as measured by the Google search volume, does not reflect this; the search volume of NMSC was not even three times higher than that of melanoma (Figure 3). The small difference might be traced back to the fact that melanoma is associated with a higher mortality, leading to a relatively higher interest. A striking finding is the search volume of precursor lesions of NMSC compared with

melanoma. This is probably related to the clearer definition of precursor lesions of NMSC. Moreover, the search volume of metastasis of NMSC is appreciably lower than that of melanoma, which reflects the lower incidence of NMSC metastasis relative to melanoma metastasis [40,41].

The large share of the category "other malignant diseases" can be explained by the fact that it covered several cancers (eg, lung cancer, cervical cancer), while the category "forms of skin cancer" referred only to skin cancer. These other malignant diseases were identified because Google AdWords Keyword Planner also shows search terms that are not primarily relevant to the topic of interest. The search volumes of localizations of NMSC and melanoma reflect the typical localizations where skin cancer is diagnosed [42-44]. The second most searched localization of melanoma, "eye," gives rise to the question of whether the search was intended for uveal melanoma. Because uveal melanoma is rare [45], our study shows a conspicuously high overall interest, which might reflect people's special fear or dismay regarding melanoma of the eye.

Limitations

This study has some limitations. Although Google has a market share of 95% in Germany [24], it is not used by all population groups to the same extent [46]. Younger-age groups use the Internet more frequently than older ones [46], who are more often affected by NMSC [10]. This could have led to underestimation of specific terms that would be searched by affected people. In addition, Google AdWords Keyword Planner gives only estimations regarding the monthly search volume of search terms based on a Google algorithm without any further information. Therefore, we do not know how precise the estimates are. However, it has been shown that Google data do correlate with skin cancer prevalence [18]. Moreover, the automatic completion of search terms, which is suggested by Google, may influence people's search behavior. It is possible that often-searched terms are even more frequently searched, while less frequently searched terms are discarded. However, this should not have a great impact on our assessment of the overall interest in a particular topic, as interest was mostly assessed in categories rather than by individual search terms.

Conclusion

This analysis of Google search data reflects the German population's interest in NMSC and melanoma. Despite the limitations, our study provides insight into terms and fields of interest related to skin cancer relevant to the German population. This information could aid in the development and implementation of effective and sustainable awareness campaigns. Therefore, future sources of information should be developed on the basis of general interest, such as the interest in melanoma of the eye. Educators should also use this technique to identify keywords to use when buying Google ads for Internet campaigns. In addition, prevention and screening programs should take advantage of the increasing role of the Internet as a source for skin cancer related information. To facilitate better understanding of Internet search data, the authors' future studies will focus on the correlation between search volume and real-life characteristics of skin cancer.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The 50 most frequent skin cancer-related search terms in German.

[PDF File (Adobe PDF File), 20KB - jmir_v20i5e10327_app1.pdf]

References

1. Madan V, Lear JT, Szeimies R. Non-melanoma skin cancer. Lancet 2010 Feb 20;375(9715):673-685. [doi: 10.1016/S0140-6736(09)61196-X] [Medline: 20171403]
2. Lomas A, Leonardi-Bee J, Bath-Hextall F. A systematic review of worldwide incidence of nonmelanoma skin cancer. Br J Dermatol 2012 May;166(5):1069-1080 [FREE Full text] [doi: 10.1111/j.1365-2133.2012.10830.x] [Medline: 22251204]
3. Leiter U, Keim U, Eigentler T, Katalinic A, Holleczek B, Martus P, et al. Incidence, Mortality, and Trends of Nonmelanoma Skin Cancer in Germany. J Invest Dermatol 2017 Sep;137(9):1860-1867. [doi: 10.1016/j.jid.2017.04.020] [Medline: 28487088]
4. Stockfleth E, Terhorst D, Hauschild A, Zouboulis CC, Schlaeger M, Oster-Schmidt C, et al. [Actinic keratoses]. J Dtsch Dermatol Ges 2012 Dec;10 Suppl 7:S1-23. [doi: 10.1111/j.1610-0379.2012.08019.x] [Medline: 23110516]
5. Diepgen TL, Mahler V. The epidemiology of skin cancer. Br J Dermatol 2002 Apr;146 Suppl 61:1-6. [Medline: 11966724]
6. Zink A, Hänsel I, Rotter M, Spinner CD, Böhner A, Biedermann T. Impact of Gliding on the Prevalence of Keratinocyte Carcinoma and its Precursors: A Cross-sectional Study Among Male Pilots in Bavaria. Acta Derm Venereol 2017 Dec 10;97(3):393-394 [FREE Full text] [doi: 10.2340/00015555-2547] [Medline: 27868143]
7. Fabbrocini G, Ruocco E, De Vita V, Monfrecola G. Squamous cell carcinoma arising in long-standing hidradenitis suppurativa: An overlooked facet of the immunocompromised district. Clin Dermatol 2017;35(2):225-227. [doi: 10.1016/j.clindermatol.2016.10.019] [Medline: 28274364]
8. De Vita V, Ruocco E. Lymph stasis promotes tumor growth. J Dermatol Sci 2018 Feb 15:-. [doi: 10.1016/j.jdermsci.2018.02.010] [Medline: 29519705]

9.   Zink A. [Non-melanoma skin cancer : Pathogenesis, prevalence and prevention]. Hautarzt 2017 Nov;68(11):919-928. [doi: 10.1007/s00105-017-4058-5] [Medline: 29018888]

10.   Robert Koch-Institut (Hrsg), Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (Hrsg). Krebs in Deutschland für 2013/2014. Berlin; 2017. URL: https://www.krebsdaten.de/Krebs/DE/Content/Publikationen/ Krebs_in_Deutschland/krebs_in_deutschland_inhalt.html [WebCite Cache ID 6z47xphlF]

11.   Weber J, Mandala M, Del Vecchio M, Gogas HJ, Arance AM, Cowey CL, CheckMate 238 Collaborators. Adjuvant Nivolumab versus Ipilimumab in Resected Stage III or IV Melanoma. N Engl J Med 2017 Dec 09;377(19):1824-1835. [doi: 10.1056/NEJMoa1709030] [Medline: 28891423]

12.   Stiegel E, Xiong D, Ya J, Funchain P, Isakov R, Gastman B, et al. Prognostic value of sentinel lymph node biopsy according to Breslow thickness for cutaneous melanoma. J Am Acad Dermatol 2018 May;78(5):942-948. [doi: 10.1016/j.jaad.2018.01.030] [Medline: 29408526]

13.   Thomas L, Puig S. Dermoscopy, Digital Dermoscopy and Other Diagnostic Tools in the Early Detection of Melanoma and Follow-up of High-risk Skin Cancer Patients. Acta Derm Venereol 2017 Jul 05;- [FREE Full text] [doi: 10.2340/00015555-2719] [Medline: 28676882]

14.   Fabbrocini G, De Vita V, Pastore F, D'Arco V, Mazzella C, Annunziata MC, et al. Teledermatology: from prevention to diagnosis of nonmelanoma and melanoma skin cancer. Int J Telemed Appl 2011:125762 [FREE Full text] [doi: 10.1155/2011/125762] [Medline: 21776252]

15.   Diepgen TL. New developments in occupational dermatology. J Dtsch Dermatol Ges 2016 Sep;14(9):875-889. [doi: 10.1111/ddg.13128] [Medline: 27607027]

16.   Blanco X. The little UV Sense fits on your nail and tracks your sunlight exposure. L'Oreal's tiny wearable at CES 2018 helps you protect your skin URL: https://www.cnet.com/news/loreal-uv-sense-ces-2018/ [accessed 2018-03-06] [WebCite Cache ID 6xiFChRfJ]

17.   Bloom R, Amber KT, Hu S, Kirsner R. Google Search Trends and Skin Cancer: Evaluating the US Population's Interest in Skin Cancer and Its Association With Melanoma Outcomes. JAMA Dermatol 2015 Aug;151(8):903-905. [doi: 10.1001/jamadermatol.2015.1216] [Medline: 26061357]

18.   Wehner MR, Nead KT, Linos E. Correlation Among Cancer Incidence and Mortality Rates and Internet Searches in the United States. JAMA Dermatol 2017 Sep 01;153(9):911-914. [doi: 10.1001/jamadermatol.2017.1870] [Medline: 28658470]

19.   Paules CI, Fauci AS. Yellow Fever - Once Again on the Radar Screen in the Americas. N Engl J Med 2017 Dec 13;376(15):1397-1399. [doi: 10.1056/NEJMp1702172] [Medline: 28273000]

20.   Mele C. The New York Times 2017. Is Wisconsin Really That Hard to Spell? URL: https://www.nytimes.com/2017/05/ 31/us/misspelled-words-states.html [accessed 2018-02-09] [WebCite Cache ID 6xiEyhaaA]

21.   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]

22.   Beck F, Richard J, Nguyen-Thanh V, Montagni I, Parizot I, Renahy E. Use of the internet as a health information resource among French young adults: results from a nationally representative survey. J Med Internet Res 2014 May 13;16(5):e128 [FREE Full text] [doi: 10.2196/jmir.2934] [Medline: 24824164]

23.   Huang X, Baade P, Youlden DR, Youl PH, Hu W, Kimlin MG. Google as a cancer control tool in Queensland. BMC Cancer 2017 Dec 04;17(1):816 [FREE Full text] [doi: 10.1186/s12885-017-3828-x] [Medline: 29202718]

24.   Statista. Statistiken zu Suchmaschinen URL: https://de.statista.com/themen/111/suchmaschinen/ [accessed 2018-01-19] [WebCite Cache ID 6xiF5yoMF]

25.   Mazloomdoost D, Kanter G, Chan RC, Deveaneau N, Wyman AM, Von Bargen EC, et al. Social networking and Internet use among pelvic floor patients: a multicenter survey. Am J Obstet Gynecol 2016 Nov;215(5):654.e1-654.e10. [doi: 10.1016/j.ajog.2016.06.011] [Medline: 27319368]

26.   Lee K, Hoti K, Hughes JD, 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 Dec 02;16(12):e262 [FREE Full text] [doi: 10.2196/jmir.3706] [Medline: 25470306]

27.   Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002 Mar 09;324(7337):573-577 [FREE Full text] [Medline: 11884321]

28.   Google AdWords. Reach the right consumers with the right keywords URL: https://adwords.google.com/intl/en/home/tools/ keyword-planner/ [accessed 2018-04-05] [WebCite Cache ID 6ychA7KF0]

29.   John SM, Trakatelli M, Gehring R, Finlay K, Fionda C, Wittlich M, et al. CONSENSUS REPORT: Recognizing non-melanoma skin cancer, including actinic keratosis, as an occupational disease - A Call to Action. J Eur Acad Dermatol Venereol 2016 Apr;30 Suppl 3:38-45. [doi: 10.1111/jdv.13608] [Medline: 26995022]

30.   van der Leest RJT, de Vries E, Bulliard J, Paoli J, Peris K, Stratigos AJ, et al. The Euromelanoma skin cancer prevention campaign in Europe: characteristics and results of 2009 and 2010. J Eur Acad Dermatol Venereol 2011 Dec;25(12):1455-1465. [doi: 10.1111/j.1468-3083.2011.04228.x] [Medline: 21951235]

XSL•FO
RenderX

31. Zink A, Wurstbauer D, Rotter M, Wildner M, Biedermann T. Do outdoor workers know their risk of NMSC? Perceptions, beliefs and preventive behaviour among farmers, roofers and gardeners. J Eur Acad Dermatol Venereol 2017 Oct;31(10):1649-1654. [doi: 10.1111/jdv.14281] [Medline: 28414878]

32. Tizek L, Krause J, Biedermann T, Zink A. Satisfaction of mountain guides with high sun protection as a tool to prevent non-melanoma skin cancer. J Eur Acad Dermatol Venereol 2017 Nov;31(11):1825-1827. [doi: 10.1111/jdv.14405] [Medline: 28609550]

33. Zink A, Koch E, Seifert F, Rotter M, Spinner CD, Biedermann T. Nonmelanoma skin cancer in mountain guides: high prevalence and lack of awareness warrant development of evidence-based prevention tools. Swiss Med Wkly 2016;146:w14380 [FREE Full text] [doi: 10.4414/smw.2016.14380] [Medline: 27922162]

34. Wakefield MA, Loken B, Hornik RC. Use of mass media campaigns to change health behaviour. Lancet 2010 Oct 09;376(9748):1261-1271 [FREE Full text] [doi: 10.1016/S0140-6736(10)60809-4] [Medline: 20933263]

35. Reinau D, Meier CR, Blumenthal R, Surber C. Skin Cancer Prevention, Tanning and Vitamin D: A Content Analysis of Print Media in Germany and Switzerland. Dermatology 2016;232(1):2-10 [FREE Full text] [doi: 10.1159/000435913] [Medline: 26278913]

36. Statista. Monatliche Durchschnittstemperatur in Deutschland von Dezember 2016 bis Dezember 2017 URL: https://de. statista.com/statistik/daten/studie/5564/umfrage/monatliche-durchschnittstemperatur-in-deutschland/ [accessed 2108-01-19] [WebCite Cache ID 6xiFHTHz9]

37. Quatresooz P, Piérard-Franchimont C, Piérard GE. Space-time clustering and seasonality in diagnosing skin cancers in Wallonia (south-east Belgium). Dermatology 2008;217(1):48-51. [doi: 10.1159/000121853] [Medline: 18367841]

38. Bianconi F, Masanotti GM, Liso A, La Rosa F, Duca E, Stracci F. Seasonal Variation in Skin Cancer Diagnosis. Front Public Health 2016;4:78 [FREE Full text] [doi: 10.3389/fpubh.2016.00078] [Medline: 27200331]

39. Eisemann N, Waldmann A, Geller AC, Weinstock MA, Volkmer B, Greinert R, et al. Non-melanoma skin cancer incidence and impact of skin cancer screening on incidence. J Invest Dermatol 2014 Jan;134(1):43-50 [FREE Full text] [doi: 10.1038/jid.2013.304] [Medline: 23877569]

40. Schouten LJ, Rutten J, Huveneers HAM, Twijnstra A. Incidence of brain metastases in a cohort of patients with carcinoma of the breast, colon, kidney, and lung and melanoma. Cancer 2002 May 15;94(10):2698-2705 [FREE Full text] [Medline: 12173339]

41. Ali Z, Yousaf N, Larkin J. Melanoma epidemiology, biology and prognosis. EJC Suppl 2013 Sep;11(2):81-91 [FREE Full text] [doi: 10.1016/j.ejcsup.2013.07.012] [Medline: 26217116]

42. Narayanan DL, Saladi RN, Fox JL. Ultraviolet radiation and skin cancer. Int J Dermatol 2010 Sep;49(9):978-986. [doi: 10.1111/j.1365-4632.2010.04474.x] [Medline: 20883261]

43. Chevalier V, Barbe C, Le Clainche A, Arnoult G, Bernard P, Hibon E, et al. Comparison of anatomical locations of cutaneous melanoma in men and women: a population-based study in France. Br J Dermatol 2014 Sep;171(3):595-601. [doi: 10.1111/bjd.13052] [Medline: 24725117]

44. Katalinic A, Kunze U, Schäfer T. Epidemiology of cutaneous melanoma and non-melanoma skin cancer in Schleswig-Holstein, Germany: incidence, clinical subtypes, tumour stages and localization (epidemiology of skin cancer). Br J Dermatol 2003 Dec;149(6):1200-1206. [Medline: 14674897]

45. Schicher N, Edelhauser G, Harmankaya K, Schindler K, Gleiss A, Pehamberger H, et al. Pretherapeutic laboratory findings, extent of metastasis and choice of treatment as prognostic markers in ocular melanoma- a single centre experience. J Eur Acad Dermatol Venereol 2013 Mar;27(3):e394-e399. [doi: 10.1111/jdv.12006] [Medline: 23057648]

46. Koch W, Frees B. Ergebnisse aus der Studienreihe 'Medien und ihr Publikum' (MiP) ARD/ZDF-Onlinestudie 2017: Neun von zehn Deutschen online. Media Perspektiven 2017:434-446.

Abbreviations

NMSC: nonmelanoma skin cancer
UVR: ultraviolet radiation

XSL•FO

RenderX

<u>Original Paper</u>

Public Opinions Toward Diseases: Infodemiological Study on News Media Data

Ming Huang[1], PhD; Omar ElTayeby[2], MSc; Maryam Zolnoori[1], PhD; Lixia Yao[1], PhD

[1]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, United States
[2]Department of Computer Science, University of North Carolina at Charlotte, Charlotte, NC, United States

Corresponding Author:
Lixia Yao, PhD
Department of Health Sciences Research
Mayo Clinic
200 First Street SW
Rochester, MN,
United States
Phone: 1 507 293 7953
Fax: 1 507 284 1516
Email: Yao.Lixia@mayo.edu

Abstract

Background: Society always has limited resources to expend on health care, or anything else. What are the unmet medical needs? How do we allocate limited resources to maximize the health and welfare of the people? These challenging questions might be re-examined systematically within an infodemiological frame on a much larger scale, leveraging the latest advancement in information technology and data science.

Objective: We expanded our previous work by investigating news media data to reveal the coverage of different diseases and medical conditions, together with their sentiments and topics in news articles over three decades. We were motivated to do so since news media plays a significant role in politics and affects the public policy making.

Methods: We analyzed over 3.5 million archive news articles from Reuters media during the periods of 1996/1997, 2008 and 2016, using summary statistics, sentiment analysis, and topic modeling. Summary statistics illustrated the coverage of various diseases and medical conditions during the last 3 decades. Sentiment analysis and topic modeling helped us automatically detect the sentiments of news articles (ie, positive versus negative) and topics (ie, a series of keywords) associated with each disease over time.

Results: The percentages of news articles mentioning diseases and medical conditions were 0.44%, 0.57% and 0.81% in the three time periods, suggesting that news media or the public has gradually increased its interests in medicine since 1996. Certain diseases such as other malignant neoplasm (34%), other infectious diseases (20%), and influenza (11%) represented the most covered diseases. Two hundred and twenty-six diseases and medical conditions (97.8%) were found to have neutral or negative sentiments in the news articles. Using topic modeling, we identified meaningful topics on these diseases and medical conditions. For instance, the smoking theme appeared in the news articles on other malignant neoplasm only during 1996/1997. The topic phrases HIV and Zika virus were linked to other infectious diseases during 1996/1997 and 2016, respectively.

Conclusions: The multi-dimensional analysis of news media data allows the discovery of focus, sentiments and topics of news media in terms of diseases and medical conditions. These infodemiological discoveries could shed light on unmet medical needs and research priorities for future and provide guidance for the decision making in public policy.

KEYWORDS

XSL•FO
RenderX

Introduction

Advances in biomedical research have greatly improved public health and human longevity; however, many biomedical problems remain unsolved, and many diseases still need feasible treatments or cures. Limited societal resources, such as funding and scientific research, are available for health care or anything else. Allocating resources and prioritizing scientific research to maximize health and welfare are important and challenging problems. Systematic investigations of allocation problems would help address unmet medical needs and promote public health. Appropriate understanding of resource allocation mechanisms in health care would also greatly benefit governmental and private funding agencies, biotechnology and pharmaceutical companies, and scientists and clinicians.

Scientists have gradually paid more attention to the notable problems of resource allocation and research prioritization in health care and have explored different methods to better understand resource allocation mechanisms. Measurement of resource allocation typically requires determination of disease burden and its association with available resources. Measures of disease burden are often quality-adjusted life years (QALYs), disability-adjusted life years (DALYs), or other health-related quality-of-life metrics. In 1999 and 2011, Gross et al [1] and Gillum et al [2] investigated the relationship between National Institutes of Health (NIH) funding and disease burden by studying the number of deaths, disease prevalence and incidence, and length of hospital stay for 29 common diseases and medical conditions. Both studies indicated a correlation between NIH funding and disease burden as calculated with DALYs and suggested that some diseases are overfunded (such as HIV, diabetes, and breast cancer) and some are underfunded (for example, depression and injuries). Sampat et al [3] investigated 107 diseases and medical conditions and determined a strong correlation between clinical trial funding from NIH and deaths or length of hospital stay. These studies examined the resource allocation problem by relating 1 or 2 resource factors to disease burden measured by DALYs and others. The practical calculation of DALYs or QALYs is usually expensive, since it requires surveying the preferences of the target population [4,5]. These works were hindered by data scarcity and examined up to 107 common diseases at arbitrary single time points, which limits the rigorousness and transferability of these studies.

In addition, scientists have proposed using conventional survey methods to include the opinions of citizens, patients, and experts when defining unmet medical needs and research priorities. For example, Dowsett et al [6] interviewed 420 oncologists about the most important research questions in breast oncology. Annals of Oncology launched a new section called Research Needs, which reports on 10 to 15 high-priority research topics. The list of high-priority topics is selected and discussed among biomedical experts using the Delphi method. Cardoso et al [7] published the first article in Annals of Oncology that used this strategy to report research needs in breast cancer. However, the conventional survey method is qualitative and does not efficiently or systematically detect unmet health needs.

In 2015, we introduced the research opportunity index (ROI) and public health index (PHI) for gauging resource allocation based on the assumption that the resources should be proportionally aligned with the burden of each disease [8]. ROI quantitatively measures resource allocation for a particular disease, and PHI assesses resource allocation for all medical conditions as a whole by considering public health needs (as measured by disease prevalence and financial cost), biomedical research (as calculated by the number of published articles), and clinical developments (as determined by the number of clinical trials). These quantitative indices are very flexible and can integrate many other quantifiable factors for research prioritization and resource allocation in biomedicine. Our analysis showed that resource allocation is influenced by previous research and that the current resource allocations are far greater than health needs, thereby resulting in a massive imbalance between research and development investments and US health needs.

More recently, we applied an infodemiological approach [9] to investigate if disease burden for thousands of diseases and medical conditions can be approximated by Internet usage data, including search volume on Google and page view counts on Wikipedia [10]. We found strong correlations between search volume on Google and disease burden measured by prevalence and treatment cost for 39 diseases.

In this study, we extended our previous work by investigating millions of news articles to determine the coverage of various diseases and medical conditions as well as the sentiments and topics in news articles over the last 3 decades. The derived information could reflect the health problems and concerns of news media and the public in a similar way that social media have shown us [11-15]. In addition, news media influences our daily lives [16], has a vital role in politics, and affects public policy making [17]. Thus, our study could provide useful information on unmet medical needs and research priorities for the future and aid in decision making in public policy.

Methods

Workflow

Our workflow to mine Reuters news data is illustrated in Figure 1 and consists of 2 main processing components: data collection and preprocessing, and natural language processing (NLP) and analysis. In the component of data collection and preprocessing, the 3 Reuters news corpora (Reuters Corpus Volume 1 [RCV1], Thomson Reuters Text Research Collection [TRC2], and Reuters Corpus 2016 [RC16]) were collected and cleaned due to heterogeneities and noise in the news articles. The cleaned data were then imported into Apache Solr (The Apache Software Foundation) for information indexing and searching. Solr is a popular, open-source, enterprise search platform built on Apache Lucene. The Reuters corpora were filtered in Solr to retrieve news articles that mentioned a list of disease synonyms from the Unified Medical Language System (UMLS) [18,19].

XSL•FO

RenderX

Figure 1. Workflow for mining news media data. Concat: concatenate; dx: disease; PheWAS: phenome-wide association study.
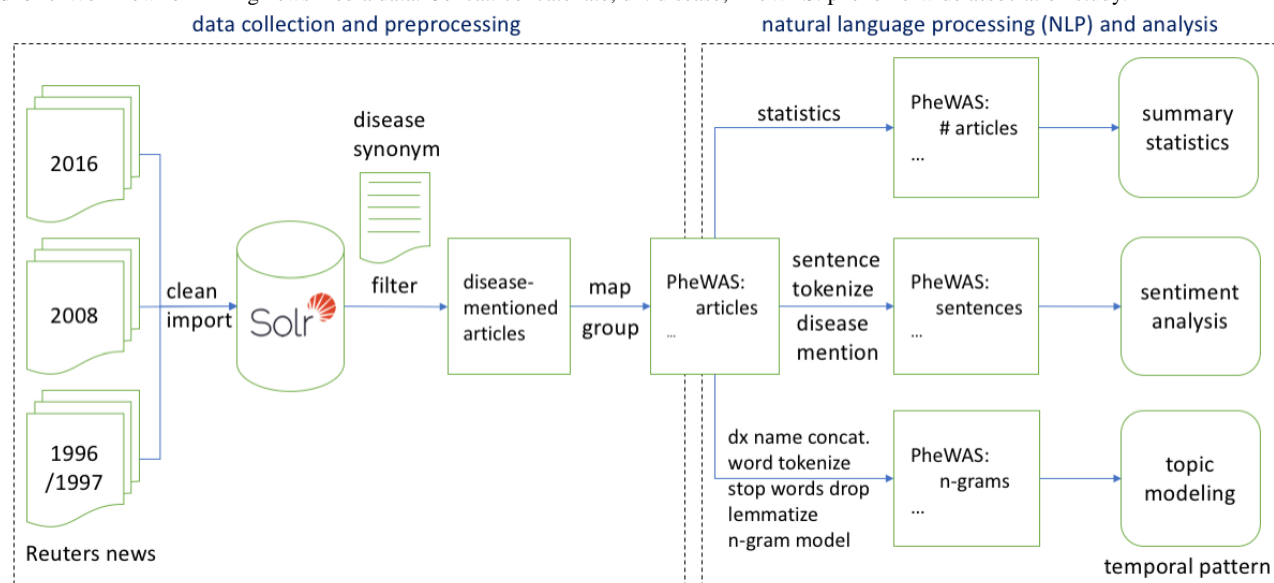
The UMLS assigns each disease concept a Concept Unique Identifier (CUI) and provides a disease synonym lookup table. It also provides the mapping between different controlled vocabularies, such as from the CUIs to the *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM). Therefore we were able to group various names of the same disease concept to UMLS CUIs and then convert the CUIs to ICD-9-CM [20,21]. Then we aggregated different ICD-9-CM codes to phenome-wide association study (PheWAS) codes using their mapping table [22].

In NLP and analysis, we performed summary statistics, sentiment analysis, and topic modeling for the news articles grouped by PheWAS codes. To summarize the data, we calculated the percentage of news articles mentioning each PheWAS disease concept by dividing the number of articles mentioning each PheWAS concept by the total number of articles that mentioned all diseases and medical conditions in a year, that is 3516 in 1996/1997, 8835 in 2008, and 9633 in 2016. For sentiment analysis, articles that mentioned diseases and medical conditions were tokenized at the sentence level, and only sentences that mentioned diseases and medical conditions were selected to compute the sentiment scores. For topic modeling, all compound disease names in the news articles were replaced by their concatenated, single-word forms, in order to preserve the compound concepts as potential candidates of meaningful topics. For example, "lung cancer" was converted to "lung_cancer" before topic modeling. The news articles were then tokenized at the word level with stop words removed. A stop word is a commonly used word such as "the," "and," or "off" that is usually discarded in the text analysis. All words were lemmatized to reduce inflectional forms to a single base form. Finally, a n-gram (n=1-4) model was used to track meaningful short phrases in topic modeling. Sentiment analysis and topic modeling were performed using Valence Aware Dictionary and sEntiment Reasoner (VADER) in Natural Language Toolkit (NLTK) [23] and the Gensim library in Python [24].

Additional details about data collection, data cleaning, summary statistics, sentiment analysis, and topic modeling are discussed in the following sections.

Data Collection

In this work, 3 news corpora in English were collected from the online newswire achieves of Reuters media (Reuters Group before 2008 and Thomson Reuters after 2008). Reuters media is a leading global information and news agency and the world's largest international text and television news provider. The 3 Reuters news corpora used in this study, each in a span of a 1-year period, represent 3 decades (1990s, 2000s, and 2010s) and allow us to study the temporal trends of major topics in news media. Table 1 summarizes the dates and statistics of the 3 data sets including the numbers of all news articles, the numbers and percentages of articles that mentioned diseases and medical conditions, and the numbers of the mapped PheWAS codes [25].

RCV1 is the first official Reuters corpus [26,27] and is a large collection of over 800,000 news articles in English that were published from August 20, 1996, to August 19, 1997. In 2000, Reuters made this corpus available for free use by researchers. RCV1 has been extensively used for information retrieval and the development of NLP and machine learning tools [26].

TRC2 is another large newswire corpus of Reuters media. TRC2 includes more than 1,800,000 news articles produced by Reuters journalists from January 1, 2008, to February 28, 2009. This data set was initially provided to the participants of the 2009 blog track at the Text Retrieval Conference. In this study, we used the news articles from January 1, 2008, to December 31, 2008, which includes more than 1,500,000 news articles.

RCV1 and TRC2 are distributed free of charge by the National Institute of Standards and Technology and can be downloaded on request [28]. To analyze the latest news articles, we also created a third news article corpus from the Reuters archive websites [29-31] for 3 available English-speaking countries: United States, United Kingdom, and India.

Table 1. Summary statistics of Reuters historical news data.

Data set	Date	Article total, n	Articles mentioning diseases, n (%)	Mapped PheWAS[a] codes
RCV1[b]	8/20/1996 to 8/19/1997	806,791	3516 (0.44)	342
TRC2[c]	1/1/2008 to 12/31/2008	1,546,350	8835 (0.57)	311
RC16[d]	1/1/2016 to 12/31/2016	1,182,761	9633 (0.81)	375

[a]PheWAS: phenome-wide association study.

[b]RCV1: Reuters Corpus Volume 1.

[c]TRC2: Thomson Reuters Text Research Collection.

[d]RC16: Reuters Corpus 2016.

The new Reuters news corpus is called RC16 for the purposes of this study. RC16 includes more than 1,100,000 news articles from January 1, 2016, to December 31, 2016.

Data Cleaning

News articles we collected are mostly heterogeneous and noisy and need to be cleaned before text analysis. In practice, automatic data cleaning is relatively difficult and data cleaning needs to be performed on a case-by-case basis because heterogeneities and noises vary among raw text corpora [11]. For example, Reuters tags and editorial information exist in Reuters corpora, and some sentences are delimited by ">" or "*" in the news articles. These are different from tweet data used in our previous study on disease burden. We designed and used the following steps to clean the data retrieved from the 3 Reuters corpora.

- Removed Reuters tags such as "(Reuters)" and "(Reuters Life)"
- Removed editorial information. All sentences starting with "Editing by," "Reporting by," "Written by," or "Page editor" were deleted
- Removed comments from news articles. Comments and remarks were discarded (eg, "click on the codes in brackets to see stories," "the following statement was released by the rating agency," and "note to subscribers: this top news page will change its name to healthcare from August 4 to better describe the sector it covers")
- Replaced hyperlinks such as "http://topnews.session.rservices.com" with the word "link"
- Removed repeating special characters, for example, "---------------"
- Replaced the sentence delimiters "*" and ">" using a period. Replaced all uppercase letters with lowercase letters

The above steps cleaned up irrelevant metadata and special characters and are necessary for meaningful topic modeling and sentiment analysis. For example, the tag of Reuters appears at the end of each Reuters news article. If we do not remove it, the topic modeling results would be biased. In another case, the repeating special characters and sentence delimiters such as "*" and ">" can confuse the sentence segmentation process and impact the sentiment analysis results.

Summary Statistics

We counted the total numbers of the news articles that mentioned all diseases and medical conditions in terms of a PheWAS disease concept. Due to different sizes of the 3 Reuters corpora, we adopted the coverage in percentage instead of the raw count of articles mentioning a PheWAS disease concept from 1 year to another. Therefore, we employed the normalized coverage (in percentage) to assess the coverage of different diseases and their temporal trends in the last 3 decades. In other word, we normalized the counts of disease-specific news articles by the total number of all news articles covering all diseases and medical conditions.

Sentiment Analysis

Sentiment analysis refers to the computational treatment of subjectivity in texts to evaluate opinions, sentiments, attitudes, and emotions. It quantifies the sentiment contents in the given texts in a continuum scale (eg, [−1, 1]) [14,15].

We used the VADER module in NLTK for sentiment analysis. VADER is a lexicon- and rule-based sentiment analysis tool that is specifically optimized for sentiments expressed in social media. VADER reports a normalized and weighted compound score as a unidimensional measure of sentiment for a given sentence. The compound score of a sentence is computed by summing the predefined score of each word that exists in the sentiment lexicon. The compound score is then adjusted according to the embedded rules and normalized to be between −1.0 (the most negative) and 1.0 (the most positive), with 0.0 indicating neutral. The sentiment of each PheWAS disease concept was computed by averaging all sentence sentiments related to the PheWAS disease concept.

Topic Modeling

Topic modeling automatically detects abstract topics in a collection of documents. Here, a topic refers to a repeating pattern of co-occurring terms in a text corpus. Intuitively, a set of words that appear most frequently represent the topic of a collection of documents. Topic modeling helps us to gain insights from large document collections by identifying hidden topics [11-13].

In this study, we used latent Dirichlet allocation (LDA), a state-of-art unsupervised topic-modeling technique [32]. LDA is a probabilistic model with a hierarchical structure for its components, which include documents, topics, and words. LDA assumes that a given document is generated from a mixture of topics and these topics produce the words in the documents according to their probabilistic distributions. LDA backtracks and derives the hidden topics that create those documents on the basis of the statistics of the included words.

XSL•FO

RenderX

More specifically, given a corpus of documents (D) with a vocabulary (V) and a preselected number of topics (K), the LDA model tries to infer 2 types of probability distributions, the probability of a topic t in a document d, P(t|d), and the probability of a word w associated with a topic t, P(w|t), where d, t, and w denote a document, a topic, and a word, respectively. The prior distributions of P(t|d) and P(w|t) are estimated from the Dirichlet distribution with 2 given hyperparameters, α and β. Because of the high number (K×D+V×K) of unknown probability densities, a Gibbs sample is applied to make iterative inference until these probability densities converge. In each inference iteration, P(t|d) is updated with the proportion of words currently assigned to a topic t in a document d, and P(w|t) is set to the proportion of a word linked to a topic t over all documents. Topic assignment of a word w in a document d is then updated with the topic t if the product of P(w|t)×P(t|d) achieves a maximal probability.

We used the LDA module in Gensim to identify the topics from the group of news articles that mentioned the same PheWAS disease concept. Gensim is a robust, open-source, topic modeling toolkit written in Python. Since we are mainly concerned with the specific contents of the disease-related topics and their temporal trends rather than the distribution of an arbitrary number of topics in each document, the number of topics in the LDA model is set to 1. The top 50 topic words were set to output, whereas all other parameters kept default values. Each topic model was iterated for at least 500 steps.

Results

Phenome-Wide Association Study Disease Concepts

After data preprocessing, 342 PheWAS disease concepts were found in RCV1, 311 PheWAS disease concepts in TRC2, and 375 PheWAS disease concepts in RC16, with 231 PheWAS disease concepts found in all 3 corpora. Summary statistics, sentiment analysis, and topic analysis were used to evaluate the news articles related to these 231 common PheWAS disease concepts. These results illustrate the temporal patterns that occurred in the last 3 decades.

Summary Statistics

Table 1 summarizes the numbers and percentages of articles that mentioned diseases and medical conditions and the numbers of the mapped PheWAS disease concepts in the 3 Reuters corpora. The percentages of articles that mentioned diseases and medical conditions were 0.44% (3516/806,791), 0.57% (8835/1,546,350), and 0.81% (9633/1,182,761) in the 3 study periods, suggesting an increasing reporting coverage on diseases and medical conditions by Reuters media.

We computed the coverage percentages of the 231 PheWAS disease concepts, and the results of coverage percentages can be found in Multimedia Appendix 1. Of the 231 PheWAS disease concepts, 53 have mean coverage percentages greater than 1% in the past 3 decades (1996/1997, 2008, and 2016), as shown in descending order in Figure 2. In total, 53 PheWAS disease concepts account for 90.27% (3174/3516) of the disease-mentioning news articles in RCV1, 94.76% (8372/8835)

of the disease-mentioning news articles in TRC2, and 87.72% (8450/9633) of the disease-mentioning news articles in RC16, suggesting that these diseases were the major disease subjects reported by Reuters media. Some diseases, such as other malignant neoplasm, other infectious diseases, and influenza have dominating mean coverage percentages of 34%, 20%, and 11%, respectively, in the past 3 decades.

In addition, we analyzed the temporal trends of coverages of the 231 PheWAS disease concepts in the past 3 decades, and 4 major temporal trends were observed:

1. Some diseases maintained steady coverage percentages with small standard deviations in the last 3 decades. Examples include other malignant neoplasm (34.0% [SD 7.1%]), pain (10.3% [SD 1.8%]), neoplasm of uncertain behavior (7.1% [SD 0.6%]), asthma (3.2% [SD 0.5%]), convulsion (2.6% [SD 0.3%]), and multiple sclerosis (1.8% [SD 0.1%]).
2. Some diseases showed large fluctuations (ie, upward or downward swings) in their coverage percentages during the 3 decades. For instance, other infectious diseases, unknown original fever, other headache syndromes, other paralytic syndromes, and nausea and vomiting showed much lower coverage percentages in 2008 than in 1996/1997 and 2016. Some diseases such as arthropathy not otherwise specified, other anemias, leukemia, osteoporosis, epilepsy, and schizophrenia had very high coverage percentages in 2008 compared with 1996/1997 and 2016.
3. Steady increase refers to the overall growing of coverage percentages in the 3 study periods. For example, the coverage percentages of diabetes mellitus, obesity, heart failure, attention-deficit/hyperactivity disorder, and rheumatoid arthritis continually increased in the last 30 years. Influenza showed a much larger coverage percentage since 2008. The coverage percentages of depression, hypertension, and rash and skin eruption considerably increased in 2016.
4. Some diseases showed steady decreases: the coverage percentage of meningitis showed a continual decrease, for instance. The coverage percentage of other brain disease largely decreased since 2008. Cystic fibrosis showed a much lower coverage percentage in 2016 compared with 1996/1997 and 2008.

Sentiment Analysis

Sentiment analysis was performed for all 231 PheWAS disease concepts, and the results were summarized in Multimedia Appendix 1. The sentiment scores of the 231 PheWAS disease concepts ranged from –0.94 to 0.73, with 226 of them having neutral or negative sentiments in news articles.

Figure 2 illustrates the sentiments of the 53 PheWAS disease concepts. While 43 diseases had neutral sentiments, the other 10 diseases, other malignant neoplasm, depression, heart attack, breast cancer, prostate cancer, heart failure, other skin cancer, ovarian cancer, renal failure not otherwise specified, and joint pain, had moderate negative sentiments with mean scores of approximately –0.59.

Figure 2. Coverage percentages and sentiments of the top 53 phenome-wide association study disease concepts. Blue, white, and red in a diverging color map denote the most negative (–1.0), right neutral (0.0), and the most positive (1.0) sentiments, respectively. The 53 phenome-wide association study disease concepts are put into 3 buckets based on coverage percentages for better resolution and comparison. From top to bottom, there are 10, 17, and 26 phenome-wide association study disease concepts in each budget with mean coverage range of 34% to 5%, 5% to 2%, and 2% to 1%, respectively. ADHD: attention-deficit/hyperactivity disorder; ASCVD: atherosclerotic cardiovascular disease; COPD: chronic obstructive pulmonary disease; dx: disease; mal neo: malignant neoplasm; NOS: not otherwise specified; other nerv sys d/o: other and unspecified disorders of the nervous system; unc behave: uncertain behaviour; unkn orig: unknown origin; unsp: unspecified; syn: syndrome.
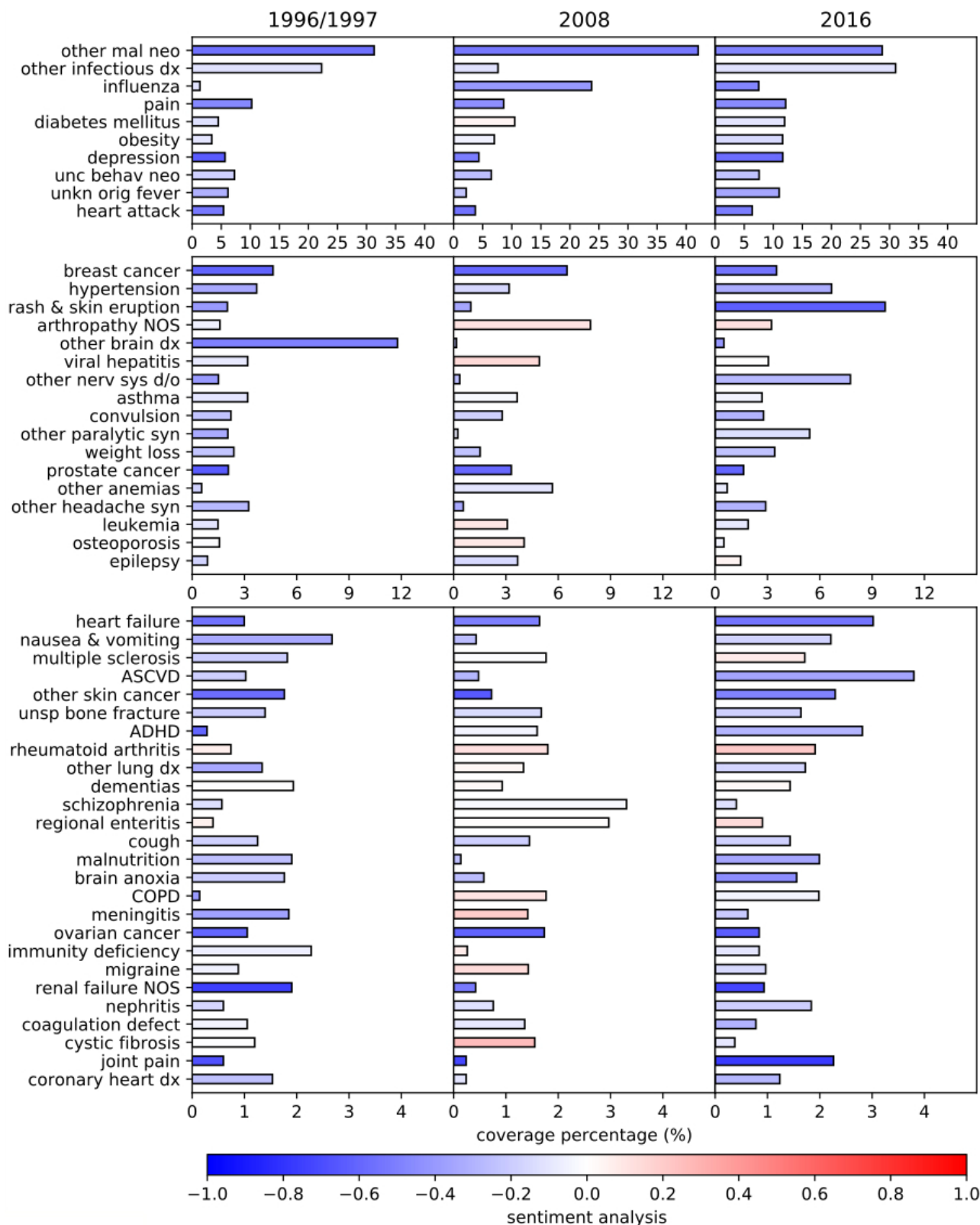
Figure 3. Top 30 topic keywords associated with the top 6 disease concepts in 3 decades. The rings represent the study periods of 1996/1997 (inner circle), 2008 (middle circle), and 2016 (outer circle). The size of the circles located on each ring denotes the permille of each term or phrase in a topic. dx: disease; mal neo: malignant neoplasm.
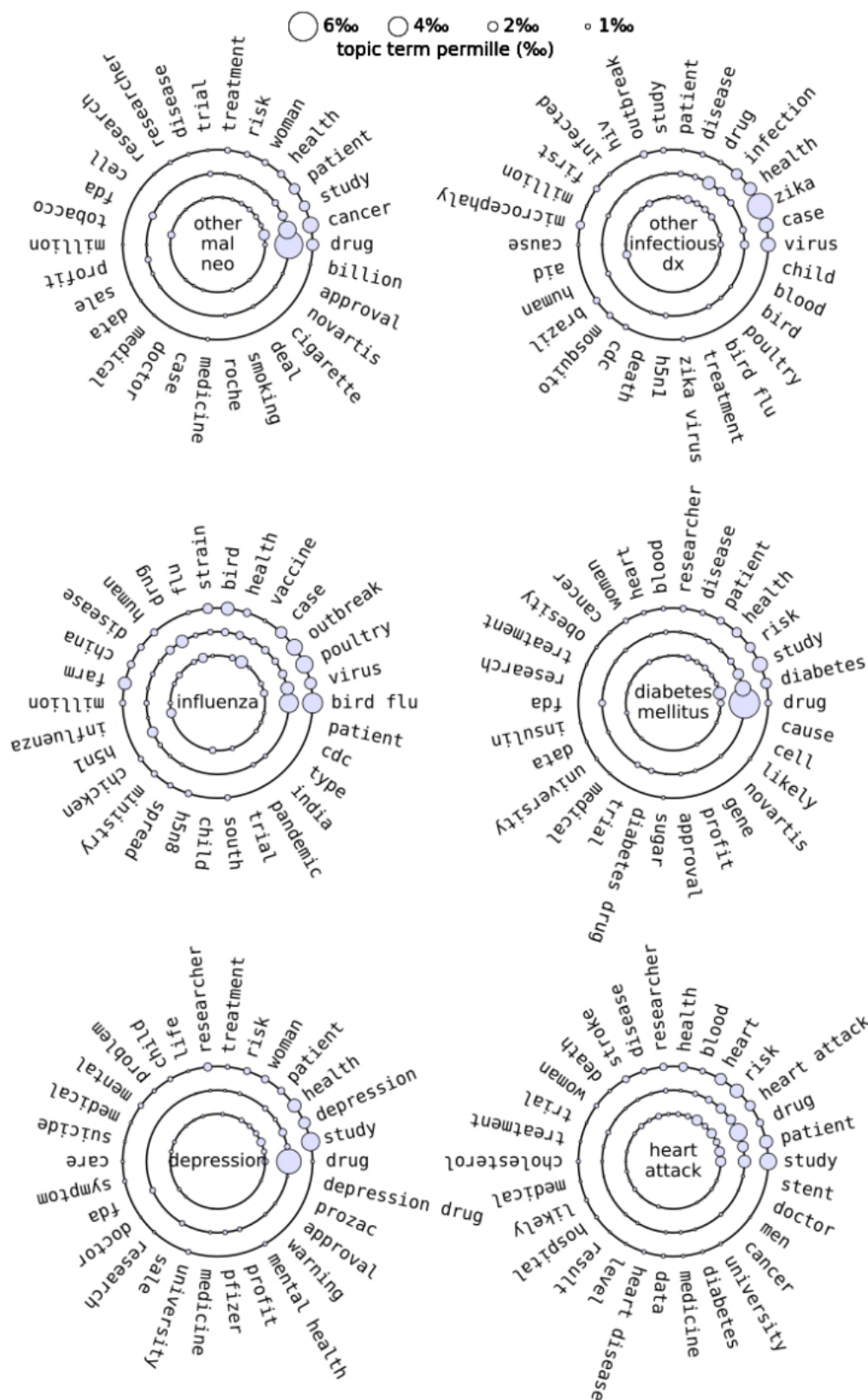
Interestingly, some diseases maintained steady sentiments from 1996 to 2016 (eg, other infectious diseases, pain, heart attack, prostate cancer, and ovarian cancer). A few diseases such as diabetes mellitus, depression, and hypertension showed fluctuations in sentiment. The sentiments of other malignant neoplasm and breast cancer gradually increased. Influenza had

more negative sentiments in 2008 and 2016 compared with 1996/1997.

Topic Modeling

Topic modeling results of the 231 PheWAS disease concepts are summarized in Multimedia Appendix 1. On the whole, the top 10 topic words occurring in the news articles about the 231 PheWAS disease concepts were drug, patient, health, study, risk, disease, treatment, researcher, trial, and FDA (US Food and Drug Administration). Figure 3 illustrates the topic contents of the 6 representative PheWAS disease concepts (other malignant neoplasm, other infectious diseases, influenza, diabetes mellitus, depression, and heart attack) in the 3 different study periods. The disease name for each PheWAS disease concept is shown in the center of the circle. The 30 most common topic words or phrases for each PheWAS disease concept are placed along small circles whose sizes represent the permille of each topic word or phrase in the topic content.

Two typical temporal patterns were found among the topics of the 231 PheWAS disease concepts in the past 3 decades.

1. There was no major change in the topic contents of some diseases over time or the permilles of key topic terms remained steady. Examples are diabetes mellitus, obesity, heart attack, breast cancer, and hypertension.
2. Some diseases showed considerable variation. Here considerable variation denotes that different or special terms arose in the topic contents of a certain disease in 1 or 2 study periods or the permilles of key topic words or phrases exhibited large variation. For example, the word smoking occurred in the topic contents of other malignant neoplasm only during 1996/1997. In the topic content of other infectious diseases, the permille of the word HIV has decreased by more than 50% since 2008.

Discussion

Combined Analysis

The top PheWAS disease concept was other malignant neoplasm. Its coverage remained steady at 34.0% in the last 3 decades (Figure 2). There are 4 other cancer-related diseases, breast cancer, prostate cancer, other skin cancer, and ovarian cancer, among the top 53 PheWAS disease concepts. Reuters media had a negative attitude toward all of these cancers. The large coverage percentages and negative sentiments indicate that Reuters media was strongly concerned with cancer, possibly because cancer is a primary global public health problem and is the second leading cause of death worldwide [33,34]. But Figure 2 shows that the overall sentiment scores have slightly increased since 1996. A 2017 report on cancer statistics noted that the overall cancer death rate declined by 25% from 1991

to 2014 in the United States as a result of a series of regulations and actions taken by the US government, such as tobacco control initiatives since the 1960s and cancer screenings [33]. Tobacco use is the most important risk factor for cancer and linked to about 80% to 90% of lung cancers [35]. Other malignant neoplasm was associated with the topic words tobacco, smoking, and cigarette in 1996/1997 but not in 2008 and 2016, which possibly reflects the progress in tobacco control in the United States (Table 2). Overall smoking prevalence decreased from 42.4% in 1965 to 16.8% in 2016, and the mortality rate due to lung cancer declined by 43% in men from 1990 to 2014 and by 17% in women from 2002 to 2014 [33]. In addition, breast cancer and prostate cancer are the top 2 and 3 cancers shown in Figure 2, which is consistent with a recent study that reported prostate cancer as the leading cancer in men and breast cancer as the leading cancer in women [33].

Other infectious diseases were the second most common PheWAS disease concept reported by Reuters media. Its coverage was 22.2% in 1996/1997, 7.6% in 2008, and 31.5% in 2016, which showed a large fluctuation in the last 30 years. Reuters media has neutral sentiment toward it, probably because infectious disease is no longer a major cause of death [36]. The topics associated with other infectious diseases varied with time. Two interesting terms that stand out in its topic content are HIV in 1996/1997 and Zika in 2016, suggesting that HIV and Zika virus probably attributed to the large fluctuation in news media coverage in the past 3 decades. The decrease in attention to HIV corresponded to a decline in the HIV/AIDS mortality rate after the introduction of antiretroviral therapy in 1995 [37]. Topics in 2016 included Zika-related words such as outbreak, microcephaly, Brazil, mosquito, and CDC (Centers for Disease Control and Prevention). The appearance of these topic words coincided with 2 events: the Zika virus outbreak in Brazil in 2015 (transmitted primarily by mosquitoes) and an association found at the same time between Zika virus infection and microcephaly [38,39].

The third most reported PheWAS disease concept was influenza. The coverage percentage of influenza dramatically increased from 1996 (1.3%) to 2008 (23.7%) and moderately decreased in 2016 (7.5%). The sentiment scores also largely declined from –0.11 (1996/1997) to –0.40 (2008) and slightly decreased to –0.48 (2016). In 2008 and 2016, the topic of influenza consisted of many terms related to avian influenza (eg, poultry, chicken, strain, H5N1, H5N8, outbreak, farm, China, India, and CDC). The outbreaks of avian influenza in 2008 and 2016 contributed to the high coverage percentages of influenza and received a large amount of attention from Reuters media [40-43]. The high mortality rate most likely resulted in the negative sentiment toward this disease [41].

Table 2. The permille of tobacco-related topic terms associated with other malignant neoplasm.

Topic term	1996/1997	2008	2016
Tobacco	1.5	0.0	0.0
Smoking	0.8	0.0	0.0
Cigarette	0.7	0.0	0.0

XSL•FO

RenderX

In addition to these top 3 covered PheWAS disease concepts, depression demonstrated some interesting findings. Its coverage percentage showed a 2-fold increase in 2016 compared with 1996/1997 and 2008, which is in accordance with a 2017 report published by the CDC [44]. The sentiment toward depression was negative in all 3 study periods, possibly because depression is the leading cause of disability worldwide and a major contributor to the overall global burden of disease [45]. Topics in news articles related to depression contained the words women, children, and suicide in 2016. The recent depression incidences have grown among adolescents and young adults, in particular among girls and young women [46], and suicide is often led by depression at its worst [47].

Here we discussed interesting observations and associations for 4 PheWAS disease concepts. However, further studies are needed to carefully and rigorously investigate the underlying causal events.

Limitations

The study has some limitations. First, the data source from Reuters media in this study is limited to 3 sparse time periods: 1996/1997 (RCV1), 2008 (TRC2), and 2016 (RC16). These datasets do not include all the news articles from 1996 to 2016 and might create biases in the temporal trend analysis for the last 3 decades. However, fundamental biomedical changes do not occur in a day, a month, or a year for most diseases and medical conditions due to the slow development of the health care industry. The exception is infectious diseases, which come and go quickly.

Second, although Reuters media is a leading information and news agency and the largest international text and television news provider in the world, the news data used in this study might not be representative of all English-language news articles in the world. Reuters could have biases in their news report that were not taken into account in our analysis.

Third, we identified disease-mentioning news articles based on a disease synonym lookup table generated from UMLS. Although the table contains more than 20,000 disease synonyms for about 1800 PheWAS disease concepts, the disease synonym list may not guarantee all the mentions of disease and medical conditions in the Reuters news articles. Disease names may have a lot of morphologic variations and abbreviations. In addition, we converted disease synonyms to the PheWAS codes via a chain of mapping. In this process, the percentage of definitive mappings (one-to-one and multiple-to-one) from disease synonyms to PheWAS codes is 94%, which indicates that the upper bound of errors caused by ambiguous mappings (one-to-multiple or multiple-to-multiple mappings) might be 6%.

Fourth, the sentiment analysis and topic modeling step may also introduce bias. We used VADER for sentiment analysis in this study. It is a lexicon- and rule-based model. Although VADER has been used to analyze *New York Times* articles with a performance close to trained human raters, it has not been evaluated or tuned for Reuters news articles. On the other hand, LDA for topic modeling requires manual tuning parameters, including the number of topics, which can be more art than science. The perplexity plot shows the relationship between the perplexity and the topic numbers and thus provides limited hints on determining the optimal number of topics [48]. However, it still requires further topic visualization and input from domain experts to determine the optimal number of topics, which is not practical to do for 231 PheWAS disease concepts. We kept the number of topics to 1 in this study for all 231 PheWAS disease concepts, and other parameters were set to default values for practicability.

Finally, it would be valuable to perform correlation analysis between funding data and focuses in news articles to understand resource allocation. However, there is no comprehensive funding data for the large number of diseases and medical conditions.

Conclusions

News media not only affects our daily lives but also plays a significant role in politics by influencing public policy making. In our study, we applied statistical analysis, sentiment analysis, and topic modeling techniques to over 3.5 million news articles from Reuters media in the past 3 decades in order to discover statistical and temporal patterns of coverage, sentiments, and topics for a large number of diseases and medical conditions. Our results show that Reuters media has gradually increased its coverage on diseases and medical conditions since 1996. Diseases and medical conditions like other malignant neoplasm, other infectious diseases, and influenza were the main focus of Reuters media with large coverage percentages, and 97.8% of diseases and medical conditions showed neutral or negative sentiments. Meaningful topics were identified in the news articles mentioning diseases and medical conditions. Combined analysis indicates statistical and temporal patterns of diseases and medical conditions that correspond to relevant policies and published research.

The infodemiological study expands our previous work on health resource allocation and disease burden estimation using online data. The multidimensional analysis of news media data enables the discoveries of disease-related focus, sentiments, and topics of news media. These discoveries could provide valuable information on unmet medical needs and research priorities and offer guidance for future decision making in public policy.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Coverage percentages, topics, and sentiments of 231 phenome-wide association study disease concepts.

[XLSX File (Microsoft Excel File), 338KB - jmir_v20i5e10047_app1.xlsx]

References

1. Gross CP, Anderson GF, Powe NR. The relation between funding by the National Institutes of Health and the burden of disease. N Engl J Med 1999 Jun 17;340(24):1881-1887. [doi: 10.1056/NEJM199906173402406] [Medline: 10369852]
2. Gillum LA, Gouveia C, Dorsey ER, Pletcher M, Mathers CD, McCulloch CE, et al. NIH disease funding levels and burden of disease. PLoS One 2011 Feb 24;6(2):e16837 [FREE Full text] [doi: 10.1371/journal.pone.0016837] [Medline: 21383981]
3. Sampat BN, Buterbaugh K, Perl M. New evidence on the allocation of NIH funds across diseases. Milbank Q 2013 Mar;91(1):163-185 [FREE Full text] [doi: 10.1111/milq.12005] [Medline: 23488714]
4. Normand C. Measuring outcomes in palliative care: limitations of QALYs and the road to PalYs. J Pain Symptom Manage 2009 Jul;38(1):27-31. [doi: 10.1016/j.jpainsymman.2009.04.005] [Medline: 19615624]
5. Duru G, Auray JP, Béresniak A, Lamure M, Paine A, Nicoloyannis N. Limitations of the methods used for calculating quality-adjusted life-year values. Pharmacoeconomics 2002;20(7):463-473. [Medline: 12093302]
6. Dowsett M, Goldhirsch A, Hayes DF, Senn HJ, Wood W, Viale G. International Web-based consultation on priorities for translational breast cancer research. Breast Cancer Res 2007;9(6):R81 [FREE Full text] [doi: 10.1186/bcr1798] [Medline: 18034879]
7. Cardoso F, Harbeck N, Barrios CH, Bergh J, Cortés J, El Saghir N, et al. Research needs in breast cancer. Ann Oncol 2017 Feb 01;28(2):208-217. [doi: 10.1093/annonc/mdw571] [Medline: 27831505]
8. Yao L, Li Y, Ghosh S, Evans JA, Rzhetsky A. Health ROI as a measure of misalignment of biomedical needs and resources. Nat Biotechnol 2015 Aug;33(8):807-811 [FREE Full text] [doi: 10.1038/nbt.3276] [Medline: 26252133]
9. 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]
10. Qiu R, Hadzikadic R, Yao L. Estimating disease burden using Google Trends and Wikipedia data. In: Advances in Artificial Intelligence: From Theory to Practice. New York: Springer; 2017:374-385.
11. Tapi Nzali MD, Bringay S, Lavergne C, Mollevi C, Opitz T. What patients can tell us: topic analysis for social media on breast cancer. JMIR Med Inform 2017 Jul 31;5(3):e23 [FREE Full text] [doi: 10.2196/medinform.7779] [Medline: 28760725]
12. Kagashe I, Yan Z, Suheryani I. Enhancing seasonal influenza surveillance: topic analysis of widely used medicinal drugs using Twitter data. J Med Internet Res 2017 Sep 12;19(9):e315 [FREE Full text] [doi: 10.2196/jmir.7393] [Medline: 28899847]
13. Surian D, Nguyen DQ, Kennedy G, Johnson M, Coiera E, Dunn AG. Characterizing Twitter discussions about HPV vaccines using topic modeling and community detection. J Med Internet Res 2016 Aug 29;18(8):e232 [FREE Full text] [doi: 10.2196/jmir.6045] [Medline: 27573910]
14. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. J Med Internet Res 2013 Nov 1;15(11):e239 [FREE Full text] [doi: 10.2196/jmir.2721] [Medline: 24184993]
15. Metwally O, Blumberg S, Ladabaum U, Sinha SR. Using social media to characterize public sentiment toward medical interventions commonly used for cancer screening: an observational study. J Med Internet Res 2017 Jun 07;19(6):e200 [FREE Full text] [doi: 10.2196/jmir.7485] [Medline: 28592395]
16. 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. [doi: 10.1126/science.aao1100] [Medline: 29123065]
17. Arceneaux K, Johnson M, Lindstädt R, Vander Wielen RJ. The influence of news media on political elites: investigating strategic responsiveness in Congress. Am J Pol Sci 2015 Feb 02;60(1):5-29. [doi: 10.1111/ajps.12171]
18. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004 Jan 01;32(Database issue):D267-D270 [FREE Full text] [doi: 10.1093/nar/gkh061] [Medline: 14681409]
19. Griffon N, Chebil W, Rollin L, Kerdelhue G, Thirion B, Gehanno JF, et al. Performance evaluation of Unified Medical Language System's synonyms expansion to query PubMed. BMC Med Inform Decis Mak 2012 Feb 29;12:12 [FREE Full text] [doi: 10.1186/1472-6947-12-12] [Medline: 22376010]
20. Schuyler PL, Hole WT, Tuttle MS, Sherertz DD. The UMLS Metathesaurus: representing different views of biomedical concepts. Bull Med Libr Assoc 1993 Apr;81(2):217-222 [FREE Full text] [Medline: 8472007]
21. UMLS Reference Manual. Bethesda: National Library of Medicine; 2009. Chapter 2: Metathesaurus URL: https://www.ncbi.nlm.nih.gov/books/n/nlmumls/ch02/ [accessed 2018-04-29]

XSL•FO

RenderX

22. Phecode Map 1.2 with ICD-9 Codes. URL: https://phewascatalog.org/phecodes [accessed 2018-04-11]

23. Hutto CJ, Gilbert E. Vader: a parsimonious rule-based model for sentiment analysis of social media text. 2014 Presented at: Eighth international AAAI conference on weblogs and social media; 2014; Ann Arbor URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8109

24. Řehůřek R, Sojka P. Software framework for topic modelling with large corpora. 2010 Presented at: Proceedings of LREC 2010 workshop New Challenges for NLP Frameworks; 2010; Valletta p. 46-50 URL: https://is.muni.cz/publication/884893/en

25. Hebbring SJ. The challenges, advantages and future of phenome-wide association studies. Immunology 2014 Feb;141(2):157-165 [FREE Full text] [doi: 10.1111/imm.12195] [Medline: 24147732]

26. Lewis DD, Yang Y, Rose TG, Li F. RCV1: a new benchmark collection for text categorization research. J Mach Learn Res 2004 Dec 1;5:361-397 [FREE Full text]

27. Rose T, Stevenson M, Whitehead M. Reuters Corpus Volume 1: from yesterday's news to tomorrow's language resources. Luxembourg: European Language Resources Association; 2002 May Presented at: Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002); 2002; Las Palmas URL: http://www.lrec-conf.org/proceedings/lrec2002/pdf/80.pdf

28. Reuters Corpora (RCV1, RCV2, TRC2). National Institute of Standards and Technology. 2018. URL: http://trec.nist.gov/data/reuters/reuters.html [accessed 2018-02-05] [WebCite Cache ID 6x1EdJEyC]

29. Reuters. Reuters site archive, United States. 2018. URL: https://www.reuters.com/resources/archive/us/ [accessed 2018-02-05] [WebCite Cache ID 6x1HXJodv]

30. Reuters. Reuters site archive, United Kingdom. 2018. URL: https://www.reuters.com/resources/archive/uk/ [accessed 2018-02-05] [WebCite Cache ID 6x1HuLl9v]

31. Reuters. Reuters site archive, India. 2018. URL: https://www.reuters.com/resources/archive/in/ [accessed 2018-02-05] [WebCite Cache ID 6x1I9kfxY]

32. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003 Jan;3:993-1022 [FREE Full text]

33. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. CA Cancer J Clin 2017 Jan;67(1):7-30 [FREE Full text] [doi: 10.3322/caac.21387] [Medline: 28055103]

34. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al, editors. GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012 v1.0. France: International Agency for Research on Cancer; 2012.

35. GBD 2013 Risk Factors Collaborators, Forouzanfar M, Alexander L, Anderson H, Bachman V, Biryukov S, et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015 Dec 5;386(10010):2287-2323 [FREE Full text] [doi: 10.1016/S0140-6736(15)00128-2] [Medline: 26364544]

36. Dye C. After 2015: infectious diseases in a new era of health and development. Philos Trans R Soc Lond B Biol Sci 2014 May 12;369(1645) [FREE Full text] [doi: 10.1098/rstb.2013.0426] [Medline: 24821913]

37. Hansen V, Oren E, Dennis LK, Brown HE. Infectious disease mortality trends in the United States, 1980-2014. JAMA 2016 Nov 22;316(20):2149-2151. [doi: 10.1001/jama.2016.12423] [Medline: 27893119]

38. Bogoch II, Brady OJ, Kraemer MUG, German M, Creatore MI, Kulkarni MA, et al. Anticipating the international spread of Zika virus from Brazil. Lancet 2016 Jan 23;387(10016):335-336 [FREE Full text] [doi: 10.1016/S0140-6736(16)00080-5] [Medline: 26777915]

39. Heukelbach J, Alencar CH, Kelvin AA, de Oliveira WK, Pamplona de Góes Cavalcanti L. Zika virus outbreak in Brazil. J Infect Dev Ctries 2016 Feb 28;10(2):116-120 [FREE Full text] [Medline: 26927450]

40. Nagarajan S, Murugkar HV, Tosh C, Behera P, Jain R, Tripathi S, et al. Avian influenza virus (H5N1) in chickens in India. Vet Rec 2009 Jan 24;164(4):128. [Medline: 19168888]

41. Lai S, Qin Y, Cowling B, Ren X, Wardrop NA, Gilbert M, et al. Global epidemiology of avian influenza A H5N1 virus infection in humans, 1997-2015: a systematic review of individual case data. Lancet Infect Dis 2016 Jul;16(7):e108-e118 [FREE Full text] [doi: 10.1016/S1473-3099(16)00153-5] [Medline: 27211899]

42. Lee DH, Bertran K, Kwon JH, Swayne DE. Evolution, global spread, and pathogenicity of highly pathogenic avian influenza H5Nx clade 2.3.4.4. J Vet Sci 2017 Aug 31;18(S1):269-280 [FREE Full text] [doi: 10.4142/jvs.2017.18.S1.269] [Medline: 28859267]

43. Li XL, Liu K, Yao HW, Sun Y, Chen WJ, Sun RX, et al. Highly pathogenic avian influenza H5N1 in mainland China. Int J Environ Res Public Health 2015 May 08;12(5):5026-5045 [FREE Full text] [doi: 10.3390/ijerph120505026] [Medline: 26006118]

44. Pratt LA, Brody DJ, Gu Q. Antidepressant use among persons aged 12 and over: United States, 2011-2014. NCHS Data Brief 2017 Aug(283):1-8 [FREE Full text] [Medline: 29155679]

45. World Health Organization. 2004. The global burden of disease: update, Table A2: Burden of disease in DALYs by cause, sex, and income group in WHO regions URL: http://www.who.int/healthinfo/global_burden_disease/GBD_report_2004update_full.pdf [WebCite Cache ID 6z349tc72]

46. Mojtabai R, Olfson M, Han B. National trends in the prevalence and treatment of depression in adolescents and young adults. Pediatrics 2016 Dec;138(6):e20161878 [FREE Full text] [doi: 10.1542/peds.2016-1878] [Medline: 27940701]

XSL•FO

RenderX

47.   Kaur H. Clinical depression (facts and statistics): a short paper. Res Rev: J Comp Biol 2017;6(2):26-32 [FREE Full text]
48.   Sun X, Lin H. Topical community detection from mining user tagging behavior and interest. J Am Soc Inf Sci Tec 2012
       Dec 21;64(2):321-333. [doi: 10.1002/asi.22740]

Abbreviations

CDC: Centers for Disease Control and Prevention
CUI: Concept Unique Identifier
DALY: disability-adjusted life years
FDA: US Food and Drug Administration
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
LDA: latent Dirichlet allocation
NIH: National Institutes of Health
NLP: natural language processing
NLTK: Natural Language Toolkit
PheWAS: phenome-wide association study
PHI: Public Health Index
QALY: quality-adjusted life years
RC16: Reuters Corpus 2016
RCV1: Reuters Corpus Volume 1
ROI: Research Opportunity Index
TRC2: Thomson Reuters Text Research Collection
UMLS: Unified Medical Language System
VADER: Valence Aware Dictionary and sEntiment Reasoner

XSL•FO
RenderX

Predicting Depression From Language-Based Emotion Dynamics: Longitudinal Analysis of Facebook and Twitter Status Updates

Elizabeth M Seabrook[1], BPsych (Hons); Margaret L Kern[2], PhD; Ben D Fulcher[1], PhD; Nikki S Rickard[1,2], PhD

[1]Monash Institute of Cognitive and Clinical Neurosciences, School of Psychological Sciences, Monash University, Melbourne, Australia

[2]Centre for Positive Psychology, The Melbourne Graduate School of Education, The University of Melbourne, Melbourne, Australia

Corresponding Author:
Elizabeth M Seabrook, BPsych (Hons)
Monash Institute of Cognitive and Clinical Neurosciences
School of Psychological Sciences
Monash University
18 Innovation Walk, Clayton Campus
Melbourne, 3800
Australia
Phone: 61 0413170668
Email: elizabeth.seabrook@monash.edu

Abstract

Background: Frequent expression of negative emotion words on social media has been linked to depression. However, metrics have relied on average values, not dynamic measures of emotional volatility.

Objective: The aim of this study was to report on the associations between depression severity and the variability (time-unstructured) and instability (time-structured) in emotion word expression on Facebook and Twitter across status updates.

Methods: Status updates and depression severity ratings of 29 Facebook users and 49 Twitter users were collected through the app *MoodPrism*. The average proportion of positive and negative emotion words used, within-person variability, and instability were computed.

Results: Negative emotion word instability was a significant predictor of greater depression severity on Facebook ($r_s(29)$=.44, P=.02, 95% CI 0.09-0.69), even after controlling for the average proportion of negative emotion words used (partial $r_s(26)$=.51, P=.006) and within-person variability (partial $r_s(26)$=.49, P=.009). A different pattern emerged on Twitter where greater negative emotion word variability indicated lower depression severity ($r_s(49)$=−.34, P=.01, 95% CI −0.58 to 0.09). Differences between Facebook and Twitter users in their emotion word patterns and psychological characteristics were also explored.

Conclusions: The findings suggest that negative emotion word instability may be a simple yet sensitive measure of time-structured variability, useful when screening for depression through social media, though its usefulness may depend on the social media platform.

Introduction

Extending How Social Media Language Predicts Depression

"With as much as we have learned about emotions, it is as if we have been taking still photos of a dance" [1].

Social media is used in different ways by different people, but for many individuals, status updates provide snapshots of their lived experience. Studies to date have primarily considered how the relative frequency of words indicating positive and negative emotion relate to other characteristics such as mental health status, or which words (or set of words) best predict different outcomes. Such studies indicate that the frequent expression of negative emotion words in status updates can accurately identify individuals experiencing symptoms of depression [2-6]. However, an individual's mental health is reflected by more than just the average frequency or the type of words used; variability in emotional expression over time might also provide significant insights. In the current research, fluctuations in

emotional expression over time is explored as another window of insight into the psychological health of social media users.

Depression in Status Updates on Social Media

Depression, including major depressive disorder (MDD) and dysphoria, are characterized by persistent low mood (including sadness or emptiness) or anhedonia (inability to experience pleasure from activities that are usually enjoyable) [7]. At a broad level, the frequent expression of negative affect within social media status updates has been associated with higher levels of depression symptoms [2,3,5,8-11]. Frequently expressing positive affect, on the other hand, tends to be associated with lower levels of depression and greater levels of well-being [9,12,13]. The link between expressed emotion in status updates and mental health is unsurprising considering that expressing current emotion and venting frustration have been reported to be a primary purpose for many users posting on Facebook [14]. Indeed, negative and positive emotional language has been observed to occur in approximately 34% and 55% of status updates on Facebook, respectively [15]. Adding to this, depressed individuals have also been shown to post content more frequently than nondepressed persons [16], and changes in depression severity may be signaled by increases in posting behavior on social media [17]. Combined, the time-structured features and emotional features of status updates may provide insights into the depression status of social media users.

Several studies have sought to code the content of social media posts for depression disclosures [6]. For instance, Moreno et al [3] demonstrated that status updates on Facebook with references to depression symptoms such as hopelessness were positively correlated with self-reported depression symptoms. Others extended this by describing the linguistic characteristics of depression in posts and developing coding schemes to identify depression-indicative tweets or status updates [2,4,8,18]. Although specific topics, keywords, and linguistic features (especially negative emotions) are able to identify depression-indicative posts with high accuracy, many of these features may also be present in posts that are nonindicative of depression (low specificity). For example, Mowery et al [18] found considerable signal discrepancies—over 70% of tweets identified in their sample containing words related to depression were not actually indicative of depression. Thus, although negative emotion words correlate with the presence of depressive symptoms, it is a noisy and imprecise metric.

This highlights the need to move beyond the frequency of emotional language alone toward other online behavioral indices that may better differentiate depressed and nondepressed individuals. Due to the time-sensitive nature of social media data, examining the dynamic movement of emotion across status updates may provide an additional avenue to tap into the nuanced cognitive-emotional processes underlying depression and may provide a more specific index of maladaptive emotional functioning.

The Emotion Dynamics of Depression

A major change in functioning associated with the onset of depression is the ability to effectively regulate emotion.

Although the capacity for emotion to vary over time is adaptive and may contribute to psychological well-being, higher levels of emotion variability, especially of negative emotions, have been linked to depression [1,19]. For example, individuals who experience intense negative affect reactivity in response to daily stressors are at greater risk of developing depression [20-22]. This experience is supported by young people's qualitative accounts of depression, where depression is reported to "[take] over during times of vulnerability such as stress or fatigue" [23]. Negative cognitive biases also contribute to emotion variability in depression. Excessive focus on personal distress (rumination) may lead to persistent experiences of severe negative affect and difficultly regulating mood away from negative states [24,25]. The combination of cognitive-emotional processes results in emergent emotion patterns that can manifest at inappropriate times and in inappropriate ways in response to internal and external events. Maladaptive patterns of emotion build over time to place the individual at an increased risk for depression onset and maintenance [19,26-28].

The emotion variability in depression described above has predominantly been operationalized in two ways. First, variability may be operationalized as within-individual variability as *iSD*, an individual's SD of emotion expression. Like the mean, variability may best be viewed as a *trait-like* measure of emotion expression, as it provides a single number that summarizes the overall variability in affect for an individual across their recording period but ignores time-structured information [29].

A second operationalization of variability describes emotional instability and uses the mean squared successive difference (MSSD) statistic [30] that quantifies differences between consecutive observations of emotion [1]. This time-structured measure of variability uses the temporal ordering of measurements to quantify the magnitude of incremental changes in emotion [30-32]. Crucially, unlike *iSD*, where the same result would be obtained if the same set of emotion expression observations are shuffled through time, the MSSD is sensitive to the time-ordering of observations. For example, for the same distribution of negative emotion values (and thus the same *iSD*), negative emotion increasing in small incremental steps from mild to severe would result in a small MSSD value, whereas negative emotion alternating (or swinging) between mild to severe would result in a large MSSD value. In this way, MSSD captures the temporal instability of positive or negative affect.

Negative affect instability, as measured by the MSSD, has been linked to more severe depression symptoms across several studies and has been identified as a concomitant and early indicator of depression [25,31,33-36]. It has been shown to be a significant risk factor for more frequent and severe suicidal ideation [35] and may be a unique underlying emotion pattern in depression. Negative affect instability has been shown to continue to predict depression when average negative affect and the frequency of negative event exposure are held constant [25,36]. In addition, reductions in negative affect have been shown to be greater for depressed individuals in response to positive events when compared with those who are not depressed, further contributing to potential moment-to-moment variability [37].

Table 1. Definitions and conceptual overlap of variability, instability, and inertia.

Name	Definition	Operationalization	Conceptual overlap
Variability	The amplitude of an individual's emotion. This is time-unstructured, referring to the "general dispersion" of scores.	Within-person SD (*iSD*)	Variance
Instability	The amplitude of moment-to-moment changes in emotion. This is time-structured, where higher scores indicate greater variance and less positively correlated between observations.	Mean squared successive difference	Variance, time-dependency
Inertia	How well a previous emotional state predicts the next emotional state. This is time-structured, where greater correlation coefficient indicates increased temporal dependency between observations.	Autocorrelation coefficient	Time-dependency

Bowen et al [33] recently aggregated 4 studies to examine the differences in mood instability between depressed and nondepressed individuals. Participants completed daily mood diary ratings of negative and positive mood upon awakening and before bedtime for 1 week. Depressed individuals experienced greater negative mood instability over the course of the week compared with nondepressed individuals. Depressed individuals also reported greater severity in negative mood than the nondepressed group, suggesting depression is characterized by both persistent low mood and more extreme daily variation in its severity.

Although depression has also been associated with a blunted emotional response to stimuli and smoother emotional experiences from day to day (inertia) [32,38,39], variability and instability span major categories of emotion dynamics as they relate to depression and are the focus of this study. Table 1 outlines the definitions of variability, instability, and inertia and describes their conceptual overlap. To best examine the unique associations that emotion dynamic patterns have with mental health, it has been recommended that the conceptual overlap between these measures be taken into account and controlled for in analyses [32], as is done in this study.

Social Media and Emotion Dynamics

Emotion dynamics may provide important insights into the "building blocks" of depression [28], but it is also challenging and time-intensive to collect adequate longitudinal emotion data. Current approaches rely on experience sampling methods (ESMs), where an individual inputs emotion information throughout the day [1,28,40]. Although the potential burden and invasiveness of real-time data collection has been significantly reduced by incorporating new and familiar technologies into ESM design (eg, smartphones) [41,42], the need to respond to automated prompts creates a context that is different than normal daily activities. Furthermore, these methodologies may not be practical for large-scale monitoring of public mental health.

Social media may be a powerful complementary tool. Considering the frequent use of emotion language in status updates that relate to current experiences [14,15], for a large proportion of the population, social media can provide unobtrusive access to time-sensitive and ecologically valid samples of expressed emotion [2,43-45]. Diurnal and seasonal variation in depression severity have been observed at a population [2] and individual level [4] on social media. In these studies, an increase in the linguistic features predictive of

depression risk was observed from day to night and from summer to winter months. Using the social media platform Reddit, De Choudhury et al [46] considered transitions from mental health subreddits only to also using a suicide support subreddit. Findings suggested that a shift from commonly expressed sentiment (ie, the average) may represent a change toward better or poorer mental health, particularly where the magnitude of the change is more pronounced. Although observations of emotion variability and instability are yet to be applied to social media as a means of automatically screening for individuals at risk of depression, it is likely that in addition to the ability to track macro-level changes in depression on social media, microlevel changes in emotion (emotion variability) relevant to mental health may also be observable.

This Study

Evidence is mounting to suggest that emotion patterns, including variability and instability, are early indicators for depression risk [19], and there is a need to utilize scalable and unobtrusive means of collecting emotion data to effectively apply these insights to monitoring public mental health. Targeting emotion variability and instability as indicators of maladaptive emotional functioning in depression is a clear area in need of further research on social media. To date, most studies examining emotion language on social media and depression have provided a static view of emotion by compressing the variation of social media language over time into an overall average, stripping the data of what could be meaningful patterns in temporal variation of emotion expression. Although the average emotion that individuals express on platforms such as Facebook and Twitter can provide accurate and sensitive insights into the presence of depression, the variability in emotion across posts has yet to be examined as a legitimate individual difference (rather than measurement error) that may be indicative of depression severity.

Taking advantage of the time-sensitive and naturally occurring data available from status updates on Facebook and Twitter, the major aim of this study was to demonstrate the feasibility of using status update emotion variability and instability as an indicator of depression severity (measured by the Patient Health Questionnaire-9, PHQ-9) [47]. It also aimed to examine if emotion instability was related to depression when controlling for its conceptual overlap with variability.

It was hypothesized that (1) Self-reported depression severity would be positively related to negative emotion word variability and instability across status updates, (2) Self-reported depression

severity would be positively related to the average proportion of negative emotion words used and negatively related to the average proportion of positive emotion words used in status updates on Facebook and Twitter, (3) Negative emotion word instability would remain positively associated with depression severity when controlling for negative emotion word variability, (4) The emotion word patterns and their association with depression would be consistent across Facebook and Twitter, and (5) Depression severity would be positively associated with the average number of status updates per day and negatively associated with the time interval between consecutive status updates (ie, shorter periods of time between posts).

Methods

Participants

This study used a subset of users from the *MoodPrism* project. *MoodPrism* is a mood-tracking app that collects data and provides engaging feedback to its users on their mood, mental health, and well-being [42]. *MoodPrism* is available for download on the iPhone operating system (iOS, Apple Inc) and Android stores for smartphone. All procedures were approved by the Monash University Human Research Ethics Committee.

Participants were recruited by convenience sampling, community engagement, and targeted online advertising (smartphone owner, interested in mental health, lives in Australia). To be included in this study, participants had to download the *MoodPrism* smartphone app, complete the depression severity index available in the app, and opt in to contribute their Facebook or Twitter data, which were automatically collected by the *MoodPrism* app. A minimum of 10 status updates over a minimum period of 7 days was required for the inclusion of a participant, to allow robust calculation of emotion word variability and instability over time. Additionally, status updates were only included if they occurred within the 12 months before the administration of the PHQ-9. For the Twitter data, only original tweets (not retweets) were used. Although retweets may reflect values or interests of a user and include topics similar to self-authored tweets [48], they also introduce ambiguity about the author's sentiments [49,50]. Furthermore, the Facebook data did not have a similar repost function, such that self-authored tweets provide a more direct behavioral comparison.

Of the 1518 users who downloaded the *MoodPrism* app from April 2016 to May 2017, 223 (14.70%) provided permission to access their social media data. After applying the inclusion criteria outlined above, 3 participants were found to have contributed both Facebook and Twitter data. These participants had a greater number of language samples on Facebook than on Twitter, and thus, were allocated to the Facebook group. A final sample of 29 Facebook users (11 males, 17 females, 1 missing) with a mean age of 32.77 years (SD 8.40, range=19-45, n=22) and 49 Twitter users (16 males, 32 females, 1 missing) with a mean age of 35.03 years (SD 12.33, range=16-57, n=39) was obtained. Participants were well educated, with 35% (10/29; Facebook) and 41% (20/49; Twitter) of participants having completed tertiary education. Chi-square tests revealed no significant differences gender or education between the included

samples and those who had opted in to contribute social media data but did not meet the inclusion criteria. Independent samples *t* tests revealed no significant differences between groups in age. There were also no significant differences between the included Facebook (n= 29) and Twitter (n= 49) samples in age, gender, or education.

Procedure

After downloading and opening *MoodPrism,* participants read an explanatory statement and provided their consent to participate. They then provided an additional opt-in consent to share their Facebook or Twitter data. If consent was provided, the *MoodPrism* app then automatically extracted the participant's previous status updates on Facebook or Twitter and repeated this extraction for all new status updates posted while *MoodPrism* was installed on the participant's smartphone. Status updates were processed locally on the participant's smartphone through the app, pulling out the time, total word count, and number of positive and negative emotion words, and then these summaries were uploaded to a secure server every 24 hours, at which point the status update content was permanently deleted from *MoodPrism* 's memory. Thus, the app provided summaries of how often emotion words were expressed, but the actual status updates were unavailable for analysis.

Participants additionally completed several blocks of questionnaires on *MoodPrism*. These blocks included demographic items collecting gender and age information and measures assessing mental health, personality, and other psychological characteristics (see [40] for the full list of measures). Blocks could be completed in any order at a time of the participants' convenience and collectively took an average of 37 min 14 s (SD 11 min 33 s) to complete.

Measures

All data for this study was collected via the *MoodPrism* app. Depression symptom severity was measured by the PHQ-9 [47], a 9-item self-report measure for depression that indicates the severity of symptoms experienced over the previous 2 weeks. Each item on the PHQ-9 (eg, "Feeling down, depressed, or hopeless") is rated from 0="Not at all" to 3="Nearly every day." These ratings are summed to create a total score ranging from 0 to 27, where higher scores indicate greater severity of depression symptoms. The PHQ-9 has been validated for use in the general population (Cronbach alpha=.87) [51] and in primary care settings (Cronbach alpha=.89) [47]. The internal reliability of the PHQ-9 was good for both the Facebook (Cronbach alpha=.87) and Twitter (Cronbach alpha=.90) samples.

Language samples from Facebook and Twitter were obtained by *MoodPrism* via the Facebook and Twitter application programming interfaces, as detailed in Rickard et al [40]. The period of posts sampled per participant between their first status update and the administration of the PHQ-9 ranged from 9 to 365 days (Facebook mean 170.69, SD 116.05; Twitter mean 145.61, SD 124.97).

MoodPrism 's automated scripts identified the total number of *words* and positive and negative emotion words in the status

updates. *Words* on social media include both normal words and variants (eg, misspellings, emoticons, and abbreviations) that are common on social media [43]. The scripts incorporated the positive and negative emotion dictionaries of the *Linguistic Inquiry and Word Count 2007* (*LIWC 2007*) [52], a widely used corpus of dictionaries commonly used for language analysis. The *LIWC 2007* dictionaries were supplemented by common emoji's and internet slang that indicated positive or negative emotion (see Multimedia Appendix 1). Although not definitive, these inclusions were made to better reflect the language used on social media (for further discussion, see [43]).

MoodPrism also collected data on the psychological characteristics of participants, which included personality, self-esteem, and social desirability. Multimedia Appendix 2 presents additional analyses, complementary to the findings presented here, exploring Facebook and Twitter user differences across these characteristics that may inform the patterns of emotion expressed over time.

Data Analysis

Within person variability, instability, and the average proportion of positive and negative emotion words in status updates on Facebook and Twitter were calculated for each participant. All equations are presented in Figure 1 and are defined below. Correlations with PHQ-9 scores were calculated. The distributions of all Facebook and Twitter variables were non-normal; consequently Spearman rho was selected for computing correlations. Exploratory post hoc comparison between the Twitter and Facebook samples on their psychological characteristics were also performed using Mann-Whitney *U* tests because of non-normal distributions. All analyses were performed in SPSS statistics, version 24 (IBM Corp) [53].

Average Proportion

Equation 1 in Figure 1 shows the relative proportion of positive and negative emotion words, which was calculated for each status update collected to adjust for the total number of words expressed, as described in Kern et al [43]. An average of these proportions was taken for each participant, resulting in the

average proportion of positive emotion words and *average proportion of negative emotion words* across all status updates (range: 0-1), where *count(word)* refers to the total number of positive emotion words (or negative emotion words; the *LIWC 2007* category) contained in a status update, and *N_words* is the total number of words in that status update.

Variability

The within person variability (*iSD*) was computed for each participant across their status updates as shown in equation 2 in Figure 1, where the sum is taken over posts, i, s_i indicates deviations from the mean in an individual's proportion of positive (or negative) words used in status updates, and n refers to the number of status updates for that individual. This resulted in the *positive emotion word variability* and *negative emotion word variability* across status updates for each participant.

Instability

The MSSD is defined for an individual in equation 3 in Figure 1, where x_i indicates the observation at index i, x_{i+1} refers to the next consecutive observation, and n refers to the total number of observations for that individual.

A major challenge in applying measures of time-structured variability to social media data is managing the irregularly spaced time intervals between posts. As observations on Facebook and Twitter occur in a natural setting, they often occur at irregular intervals spanning, for example, between hours and months. Thus, in addition to considering time order, the time elapsed between successive observations also needs to be considered. Emotion instability, when operationalized as MSSD, assumes even sampling of observations to be computed meaningfully [30,31]. Where this is not possible, adjustments can be applied to the data to provide a weighted estimate of time-structured variability [31]. As in the study by Jahng et al [31], a time-adjusted MSSD, which accounts for an uneven sampling of observations through time, was applied. This is shown in equation 4 in Figure 1, where *median(Δt)* is the median of incremental time differences across the whole recording period.

Figure 1. Equations for average proportion, within-person variability, and instability.

$$p(category) = \sum_{word \in category} p(word) = \frac{\sum_{word \in category} count(word)}{N\_words} \quad [1]$$

$$iSD = \sqrt{\frac{\sum_i s_i^2}{n-1}} \quad [2]$$

$$MSSD = \frac{\sum_{i=1}^{n-1}(x_{i+1}-x_i)^2}{n-1} \quad [3]$$

$$\text{time-adjusted } MSSD = \frac{median(\Delta t)}{(n-1)} \sum_{i=1}^{n-1} \frac{(x_{i+1}-x_i)^2}{(t_{i+1}-t_i)} \quad [4]$$

XSL•FO

RenderX

Figure 2. A simulated time series showing the proportion of negative emotion words used in status updates over 14 days. This irregularity of status updates (ie, missing observations on days 4-8 above) can be accounted for by reweighting pairs of observations by the time elapsed between them, resulting in a lower weight for the pair of points (points C and D). The observations within the box show similar levels of negative emotion word expression but occur 6 days apart and may appear to be temporally correlated if their relative temporal distance is not accounted for. The red points show the hypothetical unobserved fluctuations in negative affect that may have occurred during the intermediate 6 days.
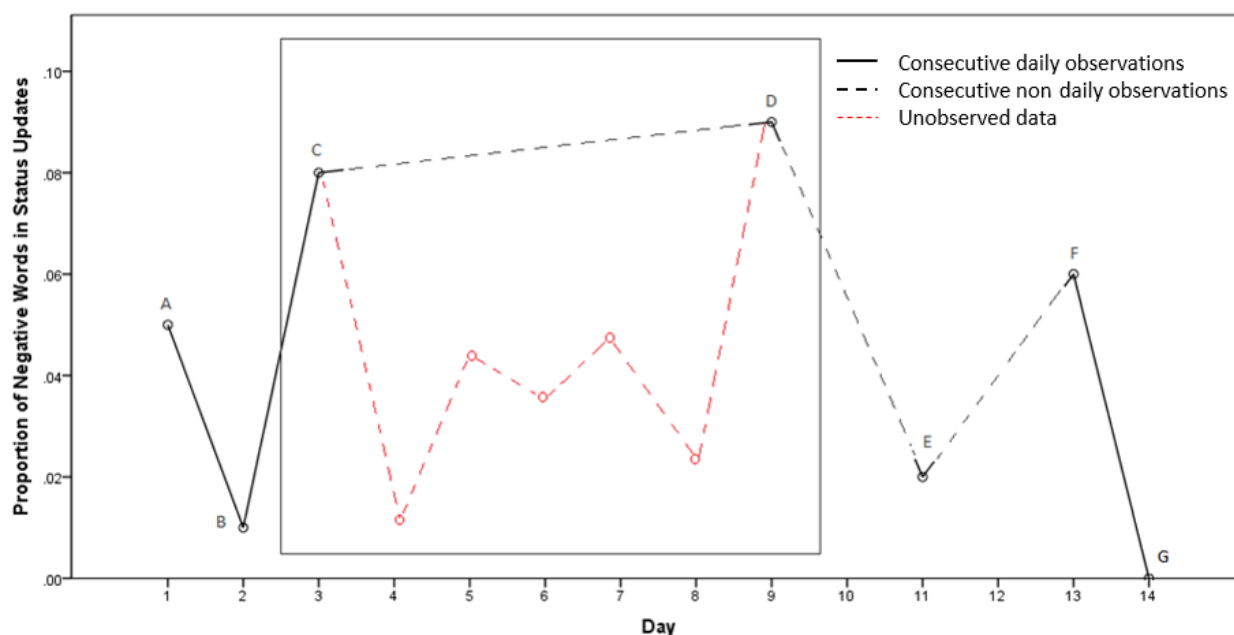
This effectively makes observations closer together in time *more important* and those further apart *less important* to the reweighted MSSD statistic, relative to a participant's median time increment between posts. Importantly, the time-adjusted MSSD, equation 4, reduces to the standard MSSD, equation 3, in the case that samples are spaced equally through time. The benefit of this adjustment in relation to social media data is the ability to utilize every observation without imposing strict inclusion criteria on the data (eg, a status update each day). As illustrated in Figure 2, this means that the overall variability contributed by all points can be included, and the potential contribution from points that may appear temporally correlated, if assumed to have occurred near in time (points C and D), is reduced when observations are in fact distant.

Applied to the Facebook and Twitter data for each participant, incremental time differences $(t_{i+1} - t_i)$ between each status update were computed. The median of these time differences was then taken for each participant and applied to each incremental time difference and successive difference $(x_{i+1} - x_i)$ in the proportion of positive or negative emotion words in a status update, as shown in equation 4. The average of the squared reweighted successive differences was then computed, resulting in the time-adjusted MSSD or the *positive emotion word instability* or *negative emotion word instability* across status updates. Here, greater values of time-adjusted MSSD indicate a greater magnitude of change in the proportion of emotion words expressed between all consecutive pairs of status updates relative to their median temporal separation.

Results

Sample Description

In total, 1856 status updates were collected (Facebook=538; Twitter=1,318) with 29,809 words expressed (Facebook=10,373; Twitter=19,436). In the Facebook sample, participants posted an average 18.55 (SD 10.01) status updates across the recording period; 55.8% (300/538) of the collected status updates contained positive emotion words, and 29.2% (157/538) contained negative emotion words. In the Twitter sample participants posted an average 26.90 (SD 11.71) status updates across the recording period; 63.6% (838/1318) of the collected tweets contained positive emotion words, and 56.2% (741/1318) contained negative emotion words.

Table 2 report the means, SDs, median, and interquartile range of the PHQ-9 scores and all Facebook and Twitter variables. It also presents descriptive statistics for the temporal aspects of posting status updates in the sample. Mann-Whitney *U* tests revealed no significant differences between Facebook and Twitter groups in the length of recording period sampled (*U*=590.50, *P*=.22), though there were differences in the median time difference (*U*=344.00, *P*<.001) and average number of status updates per day (*U*=509, *P*=.04), where Twitter users posted status updates more frequently and, based on their individual median, had smaller intervals in minutes between status updates.

Tables 3 and 4, respectively, report the two-tailed Spearman correlations (alpha level=.05) between the PHQ-9 and the positive emotion variables and the negative emotion variables from Facebook (above the diagonal) and Twitter (below the diagonal).

Facebook Emotion Variability and Depression

Facebook users reported an average depression rating of 11.48 (SD 6.38) on the PHQ-9 and expressed 9.5% positive emotion words and 3.5% negative emotion words on average across their status updates. Depression severity was not significantly related to the average proportion of positive or negative emotion words expressed, positive or negative emotion word variability (*iSD*), or positive emotion word instability (time-adjusted MSSD). Negative emotion word instability did, however, show a significant positive association with depression severity ratings, sharing 19% of the variability. This indicates that successive status updates differed more in their proportion of negative emotion words used for individuals with higher self-reported depression symptoms.

Table 2. Descriptive statistics of the Patient Health Questionnaire-9 (PHQ-9), status update frequency, and the emotion features expressed in status updates on Facebook (n=29) and Twitter (n=49).

Variable	Facebook			Twitter		
	Range	Mean (SD)	Median (IQR[a])	Range	Mean (SD)	Median (IQR)
Depression severity (PHQ-9[b])	1-22	11.48 (6.38)	10 (5.5-17)	0-26	9.80 (6.81)	9 (4-14)
Status update frequency						
Recording period (days)[c]	22-356	170.69 (116.05)	134 (54-290)	9-365	145.61 (124.97)	74.00 (33.50-272.00)
Status updates per day	0.03-1.72	0.03 (0.36)	0.16 (0.07-0.51)	0.03-4.56	0.79 (1.09)	0.40 (0.09-0.90)
Interval difference (min) between status updates[d]	661-34827	8446.65 (8724.25)	3818.00 (1877.75-13522.75)	4.0-28428.5	3939.79 (6616.84)	1037 (206.25-4571.25)
Positive emotion words						
Average proportion	0.02-0.57	0.10 (0.10)	0.08 (0.05-0.11)	0.01-0.14	0.07 (0.03)	0.08 (0.05-0.09)
Variability (*iSD*)[e]	0.04-0.47	0.13 (0.09)	0.10 (0.07-0.16)	0.03-0.17	0.07 (0.03)	0.08 (0.05-0.09)
Instability (time-adjusted MSSD)[f]	0.003-11.54	1.14 (2.94)	0.11 (0.02-0.47)	0.0002-26.80	1.49 (4.40)	0.12 (0.02-0.83)
Negative emotion words						
Average proportion	0.00-0.17	0.04 (0.04)	0.02 (0.01-0.05)	0.01-0.26	0.09 (0.06)	0.09 (0.04-0.12)
Variability (*iSD*[e])	0.00-0.31	0.07 (0.08)	0.03 (0.02-0.09)	0.02-0.14	0.08 (0.03)	0.08 (0.06-0.11)
Instability (time-adjusted MSSD[f])	0.00-1.23	0.11 (0.24)	0.01 (0.002-0.14)	0.0006-37.99	1.31 (5.43)	0.15 (0.03-0.49)

[a]IQR: interquartile range.

[b]PHQ-9: Patient Health Questionnaire-9.

[c]Recording period refers to the range of days between the first status update collected and the administration of the PHQ-9.

[d]The median interval differences between status updates.

[e]*iSD* refers to within-person variability.

[f]MSSD: mean squared successive difference.

Table 3. Spearman rho correlation analyses between depression severity (as rated by the Patient Health Questionnaire-9, PHQ-9) and the positive emotion features expressed in status updates on Facebook (n=29) and Twitter (n=49). Twitter correlations are shown *below* the diagonal; Facebook correlations are shown *above* the diagonal. CIs are reported at 95% and shown in brackets.

Variable	1	2	3	4
1. PHQ-9[a]	–	.04 (−0.33 to 0.40)	.17 (−0.21 to 0.51)	−.04 −0.40 to 0.33
2. Average proportion	.02 (−0.26 to 0.30)	–	.79[b] (0.60 to 0.90)	.48[c] (0.14 to 0.72)
3. Variability (*iSD*[d])	−.09 (−0.36 to 0.20)	.49[b] (0.24 to 0.68)	–	.61[b] (0.31 to 0.80)
4. Instability (time-adjusted MSSD[e])	−.20 (−0.46 to 0.09)	.31[c] (0.03 to 0.54)	.48[b] (0.23 to 0.67)	–

[a]PHQ-9: Patient Health Questionnaire-9.

[b]*P*<.001.

[c]*P*<.05.

[d]*iSD* refers to within-person variability.

[e]MSSD: mean squared successive difference.

XSL•FO
RenderX

Table 4. Spearman rho correlation analyses between depression severity (as rated by the Patient Health Questionnaire-9, PHQ-9) and the negative emotion features expressed in status updates on Facebook (n=29) and Twitter (n=49). Twitter correlations are shown below the diagonal; Facebook correlations are shown above the diagonal. CIs are reported at 95% and shown in brackets.

Variable	1	2	3	4
1. PHQ-9[a]	–	.12 (–0.26 to 0.46)	.20 (–0.18 to 0.53)	.44[b] (0.09 to 0.69)
2. Average proportion	–.14 (–0.41 to 0.15)	–	.95[c] (0.90 to 0.98)	.72[b] (0.48 to 0.86)
3. Variability (iSD[d])	–.36[b] (–0.58 to 0.09)	.57[c] (0.34 to 0.73)	–	.82[c] (0.65 to 0.91)
4. Instability (time-adjusted MSSD[e])	–.20 (–0.46 to 0.09)	.28 (–0.001 to 0.52)	.49[c] (0.24 to 0.68)	–

[a]PHQ-9: Patient Health Questionnaire-9.

[b]$P<.05$.

[c]$P<.001$.

[d]iSD refers to within-person variability.

[e]MSSD: mean squared successive difference.

When controlling for the average proportion of negative emotion words expressed in status updates, negative emotion word instability remained strongly associated with depression severity (partial Spearman correlation: $r_s(26)=.51$, $P=.006$). Similarly, when controlling for negative emotion word variability, negative emotion word instability remained strongly associated with depression severity ($r_s(26)=.49$, $P=.009$). To illustrate this effect, Figure 3 shows samples of the pattern of negative emotion word instability from 2 participants; one with low and one with high self-reported depression symptoms. As can be seen in Figure 3, the individual with low depression severity shows small magnitude changes in their use of negative emotion words across status updates. In contrast, the individual with high depression severity exhibits greater magnitude spikes in negative emotion word expression. Here, instability is independent of variability (see Multimedia Appendix 3 for consideration of instability under fixed variability conditions).

Twitter Emotion Variability and Depression

Twitter users reported an average depression rating of 9.80 (SD 6.81) on the PHQ-9 and expressed 7.4% positive emotion words and 9.2% negative emotion words on average across their status updates. Depression severity was not significantly related to the average proportion of positive or negative emotion words expressed, positive emotion word variability (iSD), or positive or negative emotion word instability (time-adjusted MSSD). Negative emotion word variability, however, was significantly negatively associated with depression severity ratings, sharing 13% of the variability. That is, a greater general dispersion of negative emotion across status updates on Twitter was associated with lower depression severity. When controlling for the average proportion of negative emotion words expressed in status updates, negative emotion word variability retained its association with depression severity in a partial Spearman correlation $r_s(46)=-.35$, $P=.01$.

To illustrate this effect, Figure 4 shows samples of the pattern of negative emotion word variability from 2 participants; one with low and one with high self-reported depression symptoms. As can be seen in Figure 4, the individual with low depression severity shows larger overall variability in their use of negative emotion words across status updates. In contrast, the individual with high depression severity exhibits more restricted variability in negative emotion word expression. It is important to note that in Figure 4, tweets often occurred on the same day, which accounts for the clustering in the figure.

Facebook and Twitter Status Update Frequency and Depression

Descriptive statistics for the average number of status updates per day and the median time interval between status updates are presented in Table 2. Spearman correlations revealed a significant positive association between the average number of status updates per day and depression severity for Facebook users, $r_s(29)=.48$, $P=.008$. There was also a significant negative association between the median time interval between status updates and depression severity for Facebook users, $r_s(29)= -.61$, $P<.001$. Depression severity was not significantly related to the average number of status updates per day or the median interval between status updates for Twitter users.

Differences in Emotion Language Patterns

As indicated in Tables 3 and 4, the pattern of relationships between depression and emotion language use varied between Facebook and Twitter users. To explore this further, comparisons of the social media emotion language variables between the samples were conducted. As all variables were nonnormally distributed, Mann-Whitney U tests were used to compare the mean rank differences between Facebook and Twitter users in their emotion language patterns. The Twitter sample expressed more negative language ($U=256.00$, $P<.001$) that was more variable ($U=400.00$, $P=.001$) and unstable ($U=379.00$, $P=.001$) than did the Facebook group across the recording period. Twitter users also expressed greater variability in their positive emotion compared with Facebook users ($U=413.00$, $P=.002$).

Figure 3. Graphs showing the proportion of negative emotion words used in individual status updates on Facebook across 35 days. (a) Shows an individual with low self-reported depression severity (Patient Health Questionnaire-9, PHQ-9 score=9) who demonstrated little post-to-post variation in the proportion of negative emotion words used, with the maximum difference of .03. The horizontal trend line shows the median proportion of negative emotion words used (.022) and interpolation lines link consecutive status updates. (b) Shows an individual with high self-reported depression severity (PHQ-9 score=22), who demonstrates large post-to-post changes in the proportion of negative emotion words used in status updates with the largest difference being .21. The horizontal trend line shows the median proportion of negative emotion words used (.01) and interpolation lines link consecutive status updates.
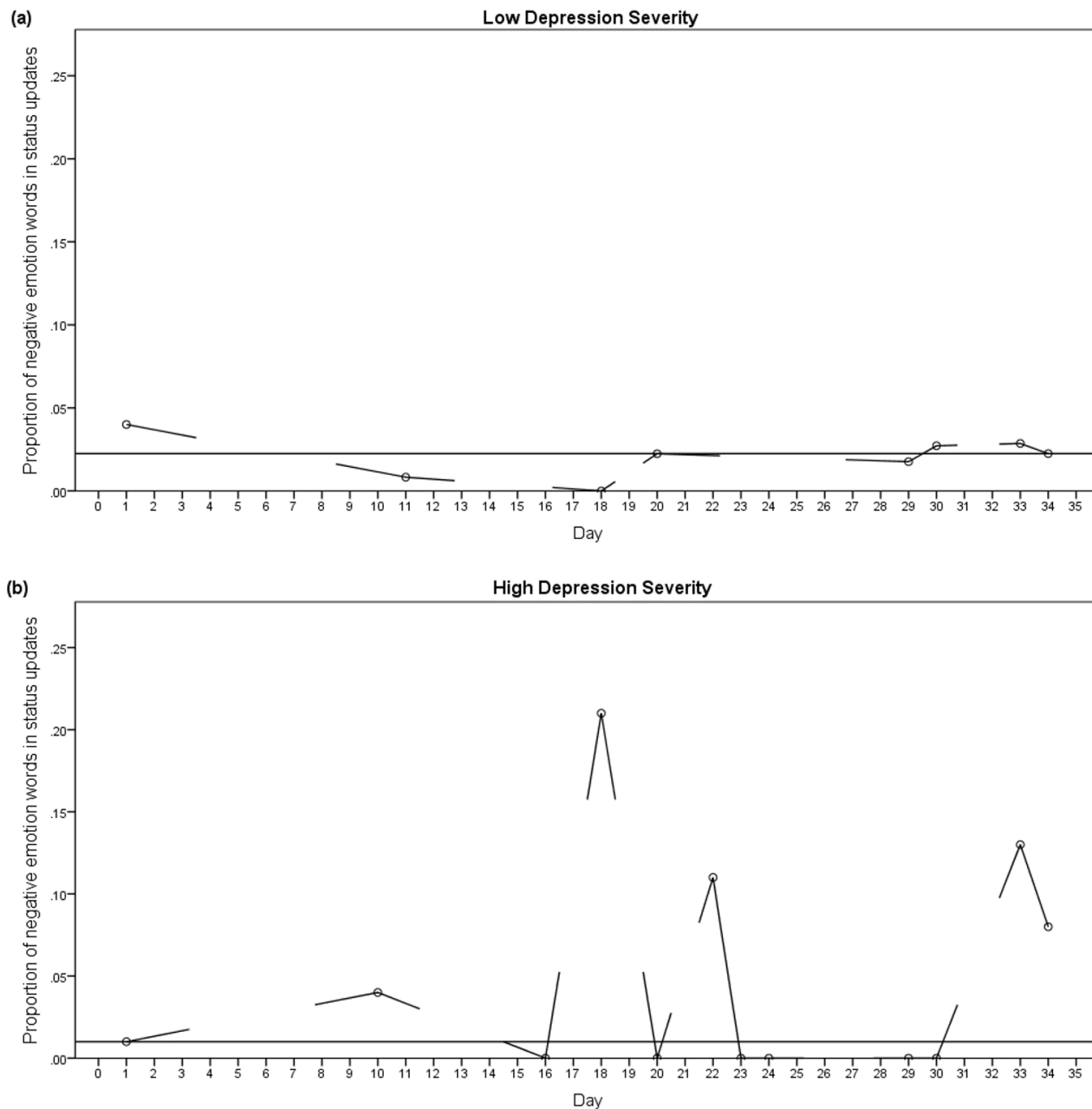
Figure 4. Graphs showing the proportion of negative emotion words used in individual status updates across (a) 160 and (b) 182 days. (a) Shows an individual with low self-reported depression severity (Patient Health Questionnaire-9, PHQ-9 score=8) and high variability in the proportion of negative emotion words used across their recording period. The horizontal trend line shows the median proportion of negative emotion words (.17) and interpolation line links status updates. (b) Shows an individual with high self-reported depression severity (PHQ-9 score=16) and low variability in the proportion of negative emotion words used across their recording period. The median proportion of negative words used was .00 and is therefore not shown. The interpolation line links status updates.
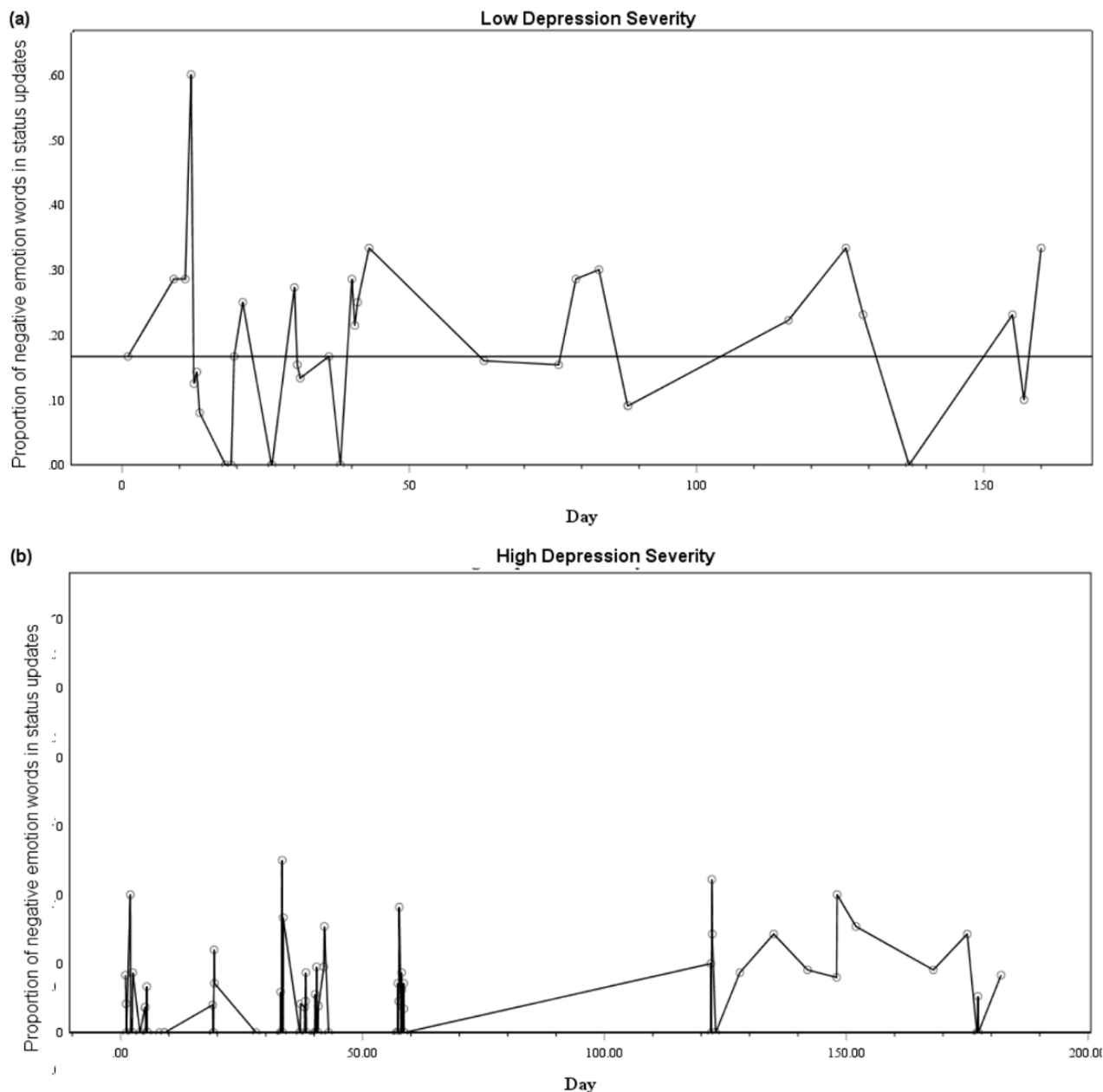
Discussion

Principal Findings

This study aimed to determine whether emotion variability and instability across status updates on Facebook and Twitter are useful indicators of depression. Differences between the social media platforms were also explored. The findings suggest that instability in the negative emotion content across Facebook status updates may indeed be a useful indicator for depression and that the time-adjusted MSSD is an effective index of instability that accounts for the uneven temporal sampling of social media posts. As hypothesized, negative emotion word instability retained its association with depression severity when the average proportion of negative emotion word use and negative emotion word variability were controlled. This index may provide additional sensitivity over basic frequency indices that are typically used in social media and depression studies. However, negative emotion word instability did not emerge as a predictor of depression on Twitter. Rather, in contrast to expectations, negative emotion word variability was negatively associated with depression severity. Furthermore, the average proportions of negative and positive emotion word use were not significantly associated with depression severity on either Facebook or Twitter. Other temporal features, the average

number of status updates per day, and the median time interval between status updates were also associated with depression severity, but only for Facebook users.

Negative Affect Instability on Facebook

Greater negative emotion word instability on Facebook was associated with individuals experiencing greater depression severity. The time-adjusted MSSD scores were driven by the pattern of frequent, high magnitude changes in negative emotion word use between status updates, not variability alone. This finding is consistent with previous studies measuring negative affect through self-report over time that have demonstrated negative affect instability to be predictive of depression [25,31,33-36].

Many users post on Facebook to broadcast emotion [14], and emotion words are often present in posts [15]. Individuals with depression are more likely to produce more content on social media when experiencing more severe symptoms [16,17], and this often relates to the disclosure of symptoms, negative experiences, or posting to seek social support [3,8,10,11]. Indeed, this was reflected in the current sample where Facebook users with greater depression severity ratings posted more status updates per day, more frequently (ie, there was a smaller median time interval between consecutive status updates). Large changes in the proportion of negative emotion words used between consecutive status updates could reflect patterns of Facebook use that mirror the inherent variation in the severity of depression symptoms over time. In this way, the negative emotion word instability in the status updates on Facebook may reveal the ebb and flow of depression symptoms and emotion dysregulation in daily life [36].

Negative emotion word instability on Facebook may also be tied to specific events, capturing momentary responses to internal and external stressors. Individuals exposed to an emotional event generally post status updates in a mood congruent way (eg, happy or sad) [54]. The proportion of negative emotion words used in a status update may reflect the extremity with which an event is perceived as negative or positive. In this light, status updates could provide insight into emotional reactivity to events. A depression-specific pattern of instability in status update expression on Facebook may exist that reflects the amplification of negative emotion in response to ambiguous or negative events [55] and a mood brightening effect in response to positive stimuli, where there is a large reduction in expressed negative affect [36,56].

The unique fluctuating pattern of negative emotion expression in individuals with more severe depression symptoms was further supported by negative emotion word instability, which remained associated with greater depression symptoms when controlling for the average proportion of negative emotion words used in status updates. This suggests that the time-structured patterns of emotion expressed on Facebook may provide better differentiation between individuals with and without depression where they express similar levels of negative emotion words. This study suggests that the poor hit rate in some keyword approaches to classifying depression in status updates, as described by Mowery et al [18], may be enhanced by including

measures of moment-to-moment variability in emotion word use.

Within-Person Variability in Emotional Expression on Twitter

Contrary to expectations, Twitter users who had lower variability in their use of negative emotion words across the recording period were more likely to have greater self-reported depression severity. This sits in contrast with a recent meta-analysis that showed negative emotion variability shares a positive association with depression [1].

It could be that the greater variability in emotion expressed by individuals lower in depression on Twitter reflects adaptive emotional functioning. In addition to personal disclosures and using Twitter to talk about daily events [57], people turn to Twitter to post content about politics, world events, and to share information [58]. Expressing a wide range of negative emotion in response to these diverse personal and community-related events may be appropriate to the context or be a part of effective emotion regulation strategies. Indeed, expressive emotional writing has been linked to better psychological and physical outcomes in offline and online settings [59-62].

On the other hand, Twitter users with higher levels of depression expressed a more clustered spread of negative emotion. Emotion appraisals of internal and external events and their subsequent expression in status updates may be more restricted or blunted for Twitter users with higher levels of depression. This is consistent with studies indicating that MDD is associated with reduced emotion reactivity [63].

Variability Differences Between Facebook and Twitter

Two divergent emotion patterns relating to depression emerged from the Facebook and Twitter samples. This highlights the importance of collecting data from multiple social media platforms, as differences in the communication mechanisms and population demographics across social media sites greatly impacts on the generalizability of findings [49]. In terms of emotion expression, Twitter users expressed more negative emotion that was more variable across the recording period than Facebook users. This may be because of the 140-character restriction placed on tweets (recently increased to 280 characters) [64] compared with the 63,206-character limit on Facebook [65], which may impact on the total proportion of emotion words expressed and the magnitude of change observed between posts. On Twitter, when an emotion word is used, it is likely to occur in the context of fewer total words and will result in a greater proportion emotion expressed per Tweet. In contrast, when a Facebook user expresses emotion, it may occur in the context of more total words, potentially reducing the overall proportion of emotion words expressed.

Other confounding variables may also create differences between negative emotion expression on Facebook and Twitter. For example, Twitter allows users to generate anonymous accounts, whereas Facebook accounts are likely to be linked to a real name. The anonymity may release the user from social norms and increase expression of negative emotion [66,67]. Twitter also is less symmetrical, with weaker relational ties, and less dense network structures, which impacts on the emotion

people express to their networks [68,69]. These different social contexts and related norms are a fruitful area for future research.

Averages of Negative and Positive Emotion Word Use Are Not Associated With Depression

Inconsistent with many previous findings [2,3,5,9], the average proportion of positive and negative emotion words used across status updates on both Facebook and Twitter were not significantly associated with depression. Other approaches using the *LIWC 2007* positive and negative dictionaries have found that as negative emotion expression increases, so does the ratings of self-reported depression severity (eg, [5]). This could be due in part to the small sample used here; language is noisy [43], and with only 29 and 49 participants in the Facebook and Twitter samples, respectively, the signal may not be enough to counteract that noise (see Kern et al [43] for further consideration of language and sample size considerations). Among this noise, it is notable with the small number of participants that a robust association between negative affect instability and depression on Facebook was found, suggesting a strong relationship between these two quantities. Although this result needs to be replicated in other samples, this suggests that when a smaller number of participants are available, instability may be a more sensitive measure than frequency in detecting depression severity.

The null findings between depression and the average proportion of words in status updates may also reflect the lack of precision that frequency measures provide. As shown by Mowery et al [18], using a keyword approach to identifying depression in social media posts results in a large proportion of false-positives, reducing the specificity with which depression can be identified through the average emotion expressed over time. Context matters [43], such that the use of a word may not directly link to an experienced emotion (eg, "I went to visit Happy Valley" does not indicate positive emotion). It is important to acknowledge also that negative emotion expression is not the exclusive domain of individuals with higher levels of depressive symptoms. It is also possible that in this study, the amended negative emotion word dictionary of the *LIWC 2007* alone was not sufficient in identifying the words most indicative of depression. Indeed, the dictionaries were recently updated [70], and future studies should examine whether the updated *LIWC 2015* dictionaries offer a better indication of depression. Personality, gender, and age have all previously been shown to impact on the number of negative emotion words people use online (cf [71]), and this complexity might also be considered in future research.

Limitations and Future Directions

There are several limitations to this study. First, although emotion scores were calculated, the actual posts were not available (because of privacy considerations), such that the context of their content could not be considered. It is therefore possible that some posts may have obtained a negative emotion word count where a positive message was conveyed. Future research should seek to apply more sophisticated open vocabulary approaches or postprocessing of status updates [18,43] to provide greater detail and accuracy of the language use context.

Second, the sample analyzed was small, and this may have impacted on the power to detect significant associations between variables. This may have obscured potential associations between the expression of negative emotion words and depression severity. It is also likely that, because of sample size, the findings obtained here may not be generalizable to the Facebook and Twitter populations. Replication is required in larger samples.

Third, only original posts were used, with retweets or shared posts excluded. Although retweets may be an indirect indication of a person's emotions, beliefs, values, and behaviors, the intentions underlying reposting are unclear. Furthermore, at the time of data collection, reposting updates was less common in Facebook, so excluding retweets provided a clearer comparison. Future studies might explore the extent to which reposts (retweets and the sharing of posts) reflect a user's values and emotions and indicate depression status.

It is also important to note that it was unknown if the emotion words expressed on Facebook or Twitter accurately reflected same-day subjective changes in mood. Further research should seek to link consecutively measured mood ratings with social media data to strengthen the assumption that interpretation of social media content reflects real-world emotion experience.

Finally, studies should seek to explicitly consider inertia in the emotion expressed in status updates as a predictor of depression and consider how the sensitivity and accuracy of frequency and instability metrics changes across different sample sizes. Such analyses will, however, require adjustments be made to calculations to account the sparseness and irregularity of social media data.

Conclusions

This study suggests that instability in the negative emotion expressed on Facebook provides insight into the presence of depression symptoms for social media users, and greater variability of negative emotion expression on Twitter may be protective for mental health.. If replicated in other samples, emotion dynamics might be applied to big data approaches for depression screening at a population level, providing insight into the emotion processes underlying depression and improving the specificity of depression identification above using language averages alone. The time-adjusted MSSD appropriately accounts for the uneven temporal sampling of real-world social media data, providing a sensitive measure of emotion instability that may be used as an early indicator of (or identified as a risk factor for) depression. Variability is often seen as a nuisance factor that creates noise and obscures other associations. Treating emotion variability as a legitimate individual difference may be an important step in better describing the microprocesses that lead to psychopathology. The findings also point to possible differences across the online culture created by a particular social media platform, such that different platforms may provide different insights into mental health.

The widespread and frequent use of social media has generated considerable concern around its impact on mental health. Yet, social media is also revealing itself to be a valuable avenue for the ongoing monitoring of depression. This study contributes

to understanding the best approaches for using the technology   to help users suffering from depression.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Emoji and internet slang supplements to the LIWC 2007 dictionaries.

[PDF File (Adobe PDF File), 38KB - jmir_v20i5e168_app1.pdf]

Multimedia Appendix 2

Supplementary analyses: personality, self-esteem, and social desirability.

[PDF File (Adobe PDF File), 48KB - jmir_v20i5e168_app2.pdf]

Multimedia Appendix 3

Illustrative participant data showing instability at fixed levels of variability.

[PDF File (Adobe PDF File), 125KB - jmir_v20i5e168_app3.pdf]

References

1. Houben M, Van Den Noortgate W, Kuppens P. The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. Psychol Bull 2015 Jul;141(4):901-930. [doi: 10.1037/a0038822] [Medline: 25822133]
2. De Choudhury M, Counts S, Horvitz E. Social media as a measurement tool of depression in populations. In: Proceedings of the 5th Annual ACM Web Science Conference. New York, NY, USA: ACM; 2013 Presented at: 5th Annual ACM Web Science Conference WebSci'13; May 02-04, 2013; Paris, France. [doi: 10.1145/2464464.2464480]
3. Moreno MA, Christakis DA, Egan KG, Jelenchick LA, Cox E, Young H, et al. A pilot evaluation of associations between displayed depression references on Facebook and self-reported depression using a clinical scale. J Behav Health Serv Res 2012 Jul;39(3):295-304 [FREE Full text] [doi: 10.1007/s11414-011-9258-7] [Medline: 21863354]
4. Schwartz H, Eichstaedt J, Kern M. Towards assessing changes in degree of depression through Facebook. In: Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality.: Association for Computational Linguistics; 2014 Presented at: Conference of the Association for Computational Linguistics; June 27, 2014; Baltimore, Maryland, USA p. 118-125. [doi: 10.3115/v1/w14-3214]
5. Settanni M, Marengo D. Sharing feelings online: studying emotional well-being via automated text analysis of Facebook posts. Front Psychol 2015;6:1045 [FREE Full text] [doi: 10.3389/fpsyg.2015.01045] [Medline: 26257692]
6. Guntuku SC, Yaden DB, Kern ML, Ungar LH, Eichstaedt JC. Detecting depression and mental illness on social media: an integrative review. Curr Opin Behav Sci 2017 Dec;18:43-49. [doi: 10.1016/j.cobeha.2017.07.005]
7. American Psychiatric Association. Depressive disorders. In: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Washington, DC, USA: American Psychiatric Association; 2013.
8. Cavazos-Rehg PA, Krauss MJ, Sowles S, Connolly S, Rosas C, Bharadwaj M, et al. A content analysis of depression-related Tweets. Comput Human Behav 2016 Jan 1;54:351-357. [doi: 10.1016/j.chb.2015.08.023] [Medline: 26392678]
9. Locatelli SM, Kluwe K, Bryant FB. Facebook use and the tendency to ruminate among college students: Testing mediational hypotheses. J Educ Comput Res 2012 Sep 7;46(4):377-394. [doi: 10.2190/EC.46.4.d]
10. Moreno MA, Jelenchick LA, Egan KG, Cox E, Young H, Gannon KE, et al. Feeling bad on Facebook: depression disclosures by college students on a social networking site. Depress Anxiety 2011 Jun;28(6):447-455 [FREE Full text] [doi: 10.1002/da.20805] [Medline: 21400639]
11. Park J, Lee DS, Shablack H, Verduyn P, Deldin P, Ybarra O, et al. When perceptions defy reality: the relationships between depression and actual and perceived Facebook social support. J Affect Disord 2016 Aug;200:37-44. [doi: 10.1016/j.jad.2016.01.048] [Medline: 27126138]
12. Bollen J, Gonçalves B, Ruan G, Mao H. Happiness is assortative in online social networks. Artif Life 2011;17(3):237-251. [doi: 10.1162/artl_a_00034] [Medline: 21554117]

XSL•FO

RenderX

13. Schwartz HA, Eichstaedt JC, Kern M. Characterizing geographic variation in well-being using Tweets. In: Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media (ICWSM).: Association for the Advancement of Artificial Intelligence; 2013 Presented at: Seventh International AAAI Conference on Weblogs and Social Media (ICWSM); July 8-11, 2013; Boston, MA, USA URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6138/6398

14. Manago AM, Taylor T, Greenfield PM. Me and my 400 friends: the anatomy of college students' Facebook networks, their communication patterns, and well-being. Dev Psychol 2012 Mar;48(2):369-380. [doi: 10.1037/a0026338] [Medline: 22288367]

15. Kramer ADI. The spread of emotion via Facebook. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM; 2012 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 05-10, 2012; Austin, Texas, USA p. 767-770. [doi: 10.1145/2207676.2207787]

16. Shaw AM, Timpano KR, Tran TB, Joormann J. Correlates of Facebook usage patterns: the relationship between passive Facebook use, social anxiety symptoms, and brooding. Comput Hum Behav 2015 Jul;48:575-580. [doi: 10.1016/j.chb.2015.02.003]

17. Park S, Kim I, Lee SW, Yoo J, Jeong B, Cha M. Manifestation of depression and loneliness on social networks: a case study of young adults on Facebook. In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing. New York, NY, USA: ACM; 2015 Presented at: 18th ACM Conference on Computer Supported Cooperative Work & Social Computing; March 14-18, 2015; Vancouver, BC, Canada p. 557-570. [doi: 10.1145/2675133.2675139]

18. Mowery D, Smith H, Cheney T, Stoddard G, Coppersmith G, Bryan C, et al. Understanding depressive symptoms and psychosocial stressors on Twitter: a corpus-based study. J Med Internet Res 2017 Feb 28;19(2):e48 [FREE Full text] [doi: 10.2196/jmir.6895] [Medline: 28246066]

19. Kuppens P, Verduyn P. Emotion dynamics. Curr Opin Psychol 2017 Dec;17:22-26. [doi: 10.1016/j.copsyc.2017.06.004] [Medline: 28950968]

20. Shapero BG, Black SK, Liu RT, Klugman J, Bender RE, Abramson LY, et al. Stressful life events and depression symptoms: the effect of childhood emotional abuse on stress reactivity. J Clin Psychol 2014 Mar;70(3):209-223 [FREE Full text] [doi: 10.1002/jclp.22011] [Medline: 23800893]

21. Wichers M, Geschwind N, Jacobs N, Kenis G, Peeters F, Derom C, et al. Transition from stress sensitivity to a depressive state: longitudinal twin study. Br J Psychiatry 2009 Dec;195(6):498-503 [FREE Full text] [doi: 10.1192/bjp.bp.108.056853] [Medline: 19949197]

22. Wichers M, Jacobs N, Derom C, Thiery E, van Os J. Depression: too much negative affect or too little positive affect? Twin Res Hum Genet 2012 Feb 21;10(S1):19-20. [doi: 10.1375/twin.10.supp.19]

23. Dundon EE. Adolescent depression: a metasynthesis. J Pediatr Health Care 2006 Nov;20(6):384-392. [doi: 10.1016/j.pedhc.2006.02.010] [Medline: 17071369]

24. Nolen-Hoeksema S, Wisco BE, Lyubomirsky S. Rethinking rumination. Perspect Psychol Sci 2008 Sep;3(5):400-424. [doi: 10.1111/j.1745-6924.2008.00088.x] [Medline: 26158958]

25. Thompson RJ, Berenbaum H, Bredemeier K. Cross-sectional and longitudinal relations between affective instability and depression. J Affect Disord 2011 Apr;130(1-2):53-59 [FREE Full text] [doi: 10.1016/j.jad.2010.09.021] [Medline: 20951438]

26. Bradley B, DeFife JA, Guarnaccia C, Phifer J, Fani N, Ressler KJ, et al. Emotion dysregulation and negative affect: Association with psychiatric symptoms. J Clin Psychiatry 2011 May;72(5):685-691 [FREE Full text] [doi: 10.4088/JCP.10m06409blu] [Medline: 21658350]

27. Gross JJ, Muñoz RF. Emotion regulation and mental health. Clin Psychol Sci Pract 1995;2(2):151-164. [doi: 10.1111/j.1468-2850.1995.tb00036.x]

28. Wichers M. The dynamic nature of depression: a new micro-level perspective of mental disorder that meets current challenges. Psychol Med 2014 May;44(7):1349-1360. [doi: 10.1017/S0033291713001979] [Medline: 23942140]

29. Ram N, Gerstorf D. Time-structured and net intraindividual variability: tools for examining the development of dynamic characteristics and processes. Psychol Aging 2009 Dec;24(4):778-791 [FREE Full text] [doi: 10.1037/a0017915] [Medline: 20025395]

30. von Neumann J, Kent RH, Bellinson HR, Hart BI. The mean square successive difference. Ann Math Stat 1941 Jun;12(2):153-162. [doi: 10.1214/aoms/1177731746]

31. Jahng S, Wood PK, Trull TJ. Analysis of affective instability in ecological momentary assessment: indices using successive difference and group comparison via multilevel modeling. Psychol Methods 2008 Dec;13(4):354-375. [doi: 10.1037/a0014173] [Medline: 19071999]

32. Koval P, Pe ML, Meers K, Kuppens P. Affect dynamics in relation to depressive symptoms: variable, unstable or inert? Emotion 2013 Dec;13(6):1132-1141. [doi: 10.1037/a0033579] [Medline: 23914765]

33. Bowen R, Peters E, Marwaha S, Baetz M, Balbuena L. Moods in clinical depression are more unstable than severe normal sadness. Front Psychiatry 2017;8:56 [FREE Full text] [doi: 10.3389/fpsyt.2017.00056] [Medline: 28446884]

XSL•FO

RenderX

34. Gershon A, Eidelman P. Inter-episode affective intensity and instability: Predictors of depression and functional impairment in bipolar disorder. J Behav Ther Exp Psychiatry 2015 Mar;46:14-18 [FREE Full text] [doi: 10.1016/j.jbtep.2014.07.005] [Medline: 25164093]

35. Palmier-Claus JE, Taylor PJ, Gooding P, Dunn G, Lewis SW. Affective variability predicts suicidal ideation in individuals at ultra-high risk of developing psychosis: an experience sampling study. Br J Clin Psychol 2012 Mar;51(1):72-83. [doi: 10.1111/j.2044-8260.2011.02013.x] [Medline: 22268542]

36. Thompson RJ, Mata J, Jaeggi SM, Buschkuehl M, Jonides J, Gotlib IH. The everyday emotional experience of adults with major depressive disorder: examining emotional instability, inertia, and reactivity. J Abnorm Psychol 2012 Nov;121(4):819-829 [FREE Full text] [doi: 10.1037/a0027978] [Medline: 22708886]

37. Peeters F, Berkhof J, Rottenberg J, Nicolson NA. Ambulatory emotional reactivity to negative daily life events predicts remission from major depressive disorder. Behav Res Ther 2010 Aug;48(8):754-760. [doi: 10.1016/j.brat.2010.04.008] [Medline: 20537317]

38. Koval P, Sütterlin S, Kuppens P. Emotional inertia is associated with lower well-being when controlling for differences in emotional context. Front Psychol 2015;6:1997 [FREE Full text] [doi: 10.3389/fpsyg.2015.01997] [Medline: 26779099]

39. Kuppens P, Sheeber LB, Yap MB, Whittle S, Simmons JG, Allen NB. Emotional inertia prospectively predicts the onset of depressive disorder in adolescence. Emotion 2012 Apr;12(2):283-289. [doi: 10.1037/a0025046] [Medline: 21988744]

40. Larson R, Csikszentmihalyi M. The experience sampling method. In: Csikszentmihalyi M, editor. Flow and the Foundations of Positive Psychology. Dordrecht, Netherlands: Springer; 2014:21-34.

41. Miller G. The smartphone psychology manifesto. Perspect Psychol Sci 2012 May;7(3):221-237 [FREE Full text] [doi: 10.1177/1745691612441215] [Medline: 26168460]

42. Rickard N, Arjmand HA, Bakker D, Seabrook E. Development of a mobile phone app to support self-monitoring of emotional well-being: a mental health digital innovation. JMIR Ment Health 2016 Nov 23;3(4):e49 [FREE Full text] [doi: 10.2196/mental.6202] [Medline: 27881358]

43. Kern ML, Park G, Eichstaedt JC, Schwartz HA, Sap M, Smith LK, et al. Gaining insights from social media language: methodologies and challenges. Psychol Methods 2016 Dec;21(4):507-525. [doi: 10.1037/met0000091] [Medline: 27505683]

44. Schwartz HA, Ungar LH. Data-driven content analysis of social media: a systematic overview of automated methods. Ann Am Acad Pol Soc Sci 2015 Apr 09;659(1):78-94. [doi: 10.1177/0002716215569197]

45. Larsen ME, Boonstra TW, Batterham PJ, ODea B, Paris C, Christensen H. We feel: mapping emotion on Twitter. IEEE J Biomed Health Inform 2015 Jul;19(4):1246-1252. [doi: 10.1109/JBHI.2015.2403839] [Medline: 25700477]

46. De Choudhury M, Kiciman E, Dredze M, Coppersmith G, Kumar M. Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM Press; 2016 May Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose, California, USA p. 2098-2110 URL: http://europepmc.org/abstract/MED/29082385 [doi: 10.1145/2858036.2858207]

47. 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] [Medline: 11556941]

48. Geva H, Oestreicher-Singer G, Saar-Tschansky M. Using retweets to shape our online persona: a topic modeling approach. Rochester, NY, USA: Social Science Research Network; 2016 Presented at: Thirty Seventh International Conference on Information Systems (ICIS); December 11-14, 2016; Dublin, Ireland. [doi: 10.2139/ssrn.2759811]

49. Tufekci Z. Big questions for social media big data: representativeness, validity and other methodological pitfalls. : The AAAI Press; 2014 Presented at: 8th International Conference on Weblogs and Social Media, ICWSM; June 1-4, 2014; Ann Arbor, USA p. 505-514.

50. Boyd D, Golder S, Lotan G. Tweet, Tweet, Retweet: Conversational aspects of Retweeting on Twitter. In: Proceedings of the 2010 43rd Hawaii International Conference on System Sciences. Washington, DC, USA: IEEE Computer Society; 2010 Presented at: 2010 43rd Hawaii International Conference on System Sciences; January 5-8, 2010; Hawaii, USA p. 1-10. [doi: 10.1109/HICSS.2010.412]

51. Kocalevent RD, Hinz A, Brähler E. Standardization of the depression screener patient health questionnaire (PHQ-9) in the general population. Gen Hosp Psychiatry 2013;35(5):551-555. [doi: 10.1016/j.genhosppsych.2013.04.006] [Medline: 23664569]

52. Pennebaker JW, Chung CK, Ireland M, Gonzales A, Booth RJ. 2007. The LIWC 2007 Application URL: http://liwc.wpengine.com/ [accessed 2018-03-15] [WebCite Cache ID 6xwsId0OV]

53. IBM Corp. IBM SPSS Statistics for Windows. Armonk, NY: IBM Corp; 2016.

54. Kreitler CM, Stenmark CK, Serrate J, Winn N. The role of individual differences and emotion in Facebook activity. J Psychol Behav Sci 2016;4(1):1-10. [doi: 10.15640/jpbs.v4n1a1]

55. Gotlib IH, Joormann J. Cognition and depression: current status and future directions. Annu Rev Clin Psychol 2010;6:285-312 [FREE Full text] [doi: 10.1146/annurev.clinpsy.121208.131305] [Medline: 20192795]

56. Bylsma LM, Taylor-Clift A, Rottenberg J. Emotional reactivity to daily events in major and minor depression. J Abnorm Psychol 2011 Feb;120(1):155-167. [doi: 10.1037/a0021662] [Medline: 21319928]

XSL•FO

RenderX

57. Java A, Song X, Finin T, Tseng B. Why we Twitter: an analysis of a microblogging community. In: Advances in Web Mining and Web Usage Analysis. Lecture Notes in Computer Science. Heidelberg, Germany: Springer; 2009:118-138.

58. Park CS. Does Twitter motivate involvement in politics? Tweeting, opinion leadership, and political engagement. Comput Hum Behav 2013 Jul;29(4):1641-1648. [doi: 10.1016/j.chb.2013.01.044]

59. Baikie KA, Wilhelm K. Emotional and physical health benefits of expressive writing. Adv Psychiatr Treat 2005 Sep 01;11(5):338-346. [doi: 10.1192/apt.11.5.338]

60. Smyth JM. Written emotional expression: effect sizes, outcome types, and moderating variables. J Consult Clin Psychol 1998 Feb;66(1):174-184. [Medline: 9489272]

61. Pennebaker JW. Theories, therapies, and taxpayers: on the complexities of the expressive writing paradigm. Clin Psychol Sci Pract 2004;11(2):138-142. [doi: 10.1093/clipsy.bph063]

62. Baker JR, Moore SM. Blogging as a social tool: a psychosocial examination of the effects of blogging. Cyberpsychol Behav 2008 Dec;11(6):747-749. [doi: 10.1089/cpb.2008.0053] [Medline: 19072151]

63. Bylsma LM, Morris BH, Rottenberg J. A meta-analysis of emotional reactivity in major depressive disorder. Clin Psychol Rev 2008 Apr;28(4):676-691. [doi: 10.1016/j.cpr.2007.10.001] [Medline: 18006196]

64. Rosen A, Ihara I. 2017 Sep 26. Giving you more characters to express yourself URL: https://blog.twitter.com/official/en_us/topics/product/2017/Giving-you-more-characters-to-express-yourself.html [accessed 2017-10-26] [WebCite Cache ID 6tm7Re3Bg]

65. Taylor C. Mashable. 2011 Dec 01. Your new Facebook status: 63,206 characters or less URL: http://mashable.com/2011/11/30/facebook-status-63206-characters/#oOlxKAXmjSqu [accessed 2017-10-26] [WebCite Cache ID 6xwszZkqm]

66. Peddinti ST, Ross KW, Cappos J. On the Internet, nobody knows you're a dog: A Twitter case study of anonymity in social networks. In: Proceedings of the Second ACM Conference on Online Social Networks. New York, NY, USA: ACM; 2014 Presented at: Second ACM Conference on Online Social Networks; October 1-2, 2014; Dublin, Ireland p. 83-94. [doi: 10.1145/2660460.2660467]

67. Correa D, Silva LA, Mondal M, Benevenuto F, Gummadi KP. The many shades of anonymity: Characterizing anonymous social media content. In: Proceeding of the Ninth International Conference on Web and Social Media, ICWSM.: AAAI Press; 2015 Presented at: Ninth International Conference on Web and Social Media, ICWSM; May 26-29, 2015; Oxford, England p. 71-80 URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10596

68. Lerman K, Arora M, Gallegos L, Kumaraguru P, Garcia D. Emotions, demographics and sociability in Twitter interactions. In: Proceedings of the Tenth International AAAI Conference on Web and Social Media.: AAAI Press; 2016 Presented at: Tenth International AAAI Conference on Web and Social Media; May 17-20, 2016; Cologne, Germany p. 201-210 URL: http://arxiv.org/abs/1510.07090

69. Myers S, Sharma A, Gupta P, Lin J. Information network or social network?: The structure of the Twitter follow graph. In: Companion Proceedings of the 23rd International Conference on World Wide Web. New York, NY, USA: ACM Press; 2014 Presented at: 23rd International Conference on World Wide Web; April 7-11, 2014; Seoul, Korea p. 493-498. [doi: 10.1145/2567948.2576939]

70. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The Development and Psychometric Properties of LIWC2015. Austin, TX, USA: University of Texas at Austin; 2015.

71. Schwartz HA, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones SM, Agrawal M, et al. Personality, gender, and age in the language of social media: the open-vocabulary approach. PLoS One 2013;8(9):e73791 [FREE Full text] [doi: 10.1371/journal.pone.0073791] [Medline: 24086296]

Abbreviations

ESM: experience sampling method
LIWC 2007: Linguistic Inquiry and Word Count 2007
MDD: major depressive disorder
MSSD: mean squared successive difference
PHQ-9: Patient Health Questionnaire-9

XSL•FO

RenderX

<u>Original Paper</u>

Increasing Interest of Mass Communication Media and the General Public in the Distribution of Tweets About Mental Disorders: Observational Study

Miguel Angel Alvarez-Mon[1], MD; Angel Asunsolo del Barco[2,3], MD, PhD; Guillermo Lahera[3,4,5], MD, PhD; Javier Quintero[6], MD, PhD; Francisco Ferre[7], MD, PhD; Victor Pereira-Sanchez[1], MD; Felipe Ortuño[1], MD, PhD; Melchor Alvarez-Mon[3,4], MD, PhD

[1]Department of Psychiatry, Clinica Universidad de Navarra, University of Navarra, Pamplona, Spain

[2]Department of Surgery, Medical and Social Sciences, University of Alcala, Madrid, Spain

[3]Instituto Ramón y Cajal de Investigaciones Sanitarias, Madrid, Spain

[4]Department of Medicine and Medical Specialities, Hospital Universitario Príncipe de Asturias, University of Alcala, Madrid, Spain

[5]Center for Biomedical Research in the Mental Health Network, Madrid, Spain

[6]Department of Psychiatry, Hospital Universitario Infanta Leonor, Complutense University, Madrid, Spain

[7]Department of Psychiatry, Hospital Universitario Gregorio Marañón, Complutense University, Madrid, Spain

Corresponding Author:
Miguel Angel Alvarez-Mon, MD
Department of Psychiatry
Clinica Universidad de Navarra
University of Navarra
Avda Pio XII, 36
Pamplona, 31008
Spain
Phone: 34 948255400
Fax: 34 948296500
Email: malvarezdem@unav.es

Abstract

Background: The contents of traditional communication media and new internet social media reflect the interests of society. However, certain barriers and a lack of attention towards mental disorders have been previously observed.

Objective: The objective of this study is to measure the relevance of influential American mainstream media outlets for the distribution of psychiatric information and the interest generated in these topics among their Twitter followers.

Methods: We investigated tweets generated about mental health conditions and diseases among 15 mainstream general communication media outlets in the United States of America between January 2007 and December 2016. Our study strategy focused on identifying several psychiatric terms of primary interest. The number of retweets generated from the selected tweets was also investigated. As a control, we examined tweets generated about the main causes of death in the United States of America, the main chronic neurological degenerative diseases, and HIV.

Results: In total, 13,119 tweets about mental health disorders sent by the American mainstream media outlets were analyzed. The results showed a heterogeneous distribution but preferential accumulation for a select number of conditions. Suicide and gender dysphoria accounted for half of the number of tweets sent. Variability in the number of tweets related to each control disease was also found (5998). The number of tweets sent regarding each different psychiatric or organic disease analyzed was significantly correlated with the number of retweets generated by followers (1,030,974 and 424,813 responses to mental health disorders and organic diseases, respectively). However, the probability of a tweet being retweeted differed significantly among the conditions and diseases analyzed. Furthermore, the retweeted to tweet ratio was significantly higher for psychiatric diseases than for the control diseases (odds ratio 1.11, CI 1.07-1.14; *P*<.001).

Conclusions: American mainstream media outlets and the general public demonstrate a preferential interest for psychiatric diseases on Twitter. The heterogeneous weights given by the media outlets analyzed to the different mental health disorders and conditions are reflected in the responses of Twitter followers.

Introduction

Mental health disorders occur frequently in the general population. In 2015, approximately 44 million Americans suffered from some type of mental illness, with depression and anxiety representing the most prevalent forms [1]. Mental health disorders lead to a poor quality of life and patient disability [2,3]. Furthermore, mortality is significantly higher among people with mental health disorders than it is among comparable populations, with a 10-year median of potential life lost [4,5]. Additionally, mental health diseases commonly provoke self-stigma, societal stigma, or both, which negatively affect patients' disclosure of these psychiatric disorders [6,7]. Social regard for mental health disorders appears to be a key factor for the adequate consideration of these diseases, for the understanding and support received by psychiatric patients, and for the funding provided for medical investment and research of these disorders [8]. Thus, measurement of the social relevance of mental disorders is a fundamental objective for progressing the field of psychiatry [9].

Access to, and the diffusion of news information, has dramatically changed in recent years. In addition to traditional media, the internet and social media have become pivotal instruments for sharing knowledge [10-12]. Accordingly, the internet has radically modified how most people find, communicate, and share information regarding health and medical conditions, and its use and popularity have increased considerably [13]. Its relevance is further exemplified by the growing reliance on the internet as a source of information and health advice [14]. Social media is a relatively new health communication channel that enables interactions among large groups of people suffering from the same afflictions and promotes the ability to find and share information about health and medical conditions and receive health messages [15].

For example, Twitter is a social networking site that is one of the most popular and widely used forms of social media [16,17] in which users ("tweeters") post status updates (ie, "tweets") that are distributed to "followers." These tweets are also made available to the public. This form of largely public conversation in which "short bursts" of inconsequential information are relayed in 140-character "tweets" seems an unlikely source for lifelong learning [18,19].

Mainstream media outlets, such as television, radio, newspapers, and online journals are considered to be sensors and drivers of society [20-22]. These media outlets use Twitter as a tool for news distribution and thus subsequently influence large groups of people in real time [23]. The analysis of distributed tweets could represent an effective indicator of "real-world performance" [24-26]. Furthermore, because Twitter has become more popular, different players in health and medicine have begun to realize its potential for acquiring and distributing medical information [27]. Moreover, the qualitative and quantitative relevance of tweets has been shown in various investigations, including analyses of the interests and feelings of the general population with respect to health and disease, the interactions between patients and doctors or health care providers, and the promotion of the scientific impact of medical research in the news media. However, the validity of Twitter as a reflection of public opinion has been challenged [28-32]. Furthermore, patient attitudes toward various medical topics, including vaccines, illnesses, pain, drug use, and oncological and cardiovascular disease have been analyzed [33-41]. Consequently, the analysis of distributed tweets about mental health disease by primary media channels and the frequency of retweets they generate may be an effective tool for assessing social and individual interest toward psychiatric diseases.

In this paper, we investigated the distribution of tweets about mental health diseases from highly recognized and relevant American communication media sources. More specifically, the study cites periodicals and various television and radio channels, which are used as sensors of societal attitudes towards psychiatric disease throughout the first decade of Twitter's networking activity. Furthermore, we analyzed the interest generated among followers through the quantification of subsequent retweets. As a control, we simultaneously studied the number of tweets distributed by our selected social media platforms about diseases considered to be the main causes of death within the United States of America (USA) as well as tweets about HIV because of its recognized social relevance.

Methods

Communication Media Analyzed

In this study, we focused on tweets sent among a representative sample of primary American communications media outlets. We selected 15 general media outlets among those with the highest number of followers on Twitter, as estimated by their individual accounts, and ranked among those with highest social influence during the study duration [42-46]. Furthermore, we selected representative samples from different categories of media outlets to avoid potential bias. We included 6 newspapers (*New York Times*, *Washington Post*, *Los Angeles Times*, *USA Today*, *Chicago Tribune*, and *New York Post*), 5 TV or radio channels (*NBC*, *CBS*, *Fox*, *CNN*, and *ABC*), 1 general magazine (*Time*), 1 news agency (*AP*), and 2 online news outlets (*BuzzFeed* and *Huffington Post*).

Search Strategy

Our research strategy focused on searching for tweets that referred to common psychiatric terms of interest. We investigated all tweets sent from Twitter accounts, filtering them according to specific criteria using the following list of keywords: anxiety, phobias, posttraumatic stress disorder (PTSD), panic disorder, generalized anxiety disorder (GAD), obsessive compulsive disorder (OCD), depressive disorder, suicide, bipolar disorder, insomnia, schizophrenia, attention

XSL•FO

RenderX

deficit hyperactivity disorder or hyperactivity (ADHD), alcoholism, drug addiction, gambling disorder, anorexia nervosa, bulimia, dysthymia, addictions, addictive, Asperger syndrome, autism, personality disorder, and gender dysphoria. Additionally, as controls, we used tweets focused on the main causes of death in the USA (prostate, lung, colorectal, and breast cancer, stroke, diabetes mellitus, and chronic obstructive pulmonary disease [COPD]), the main causes of chronic neurologic degenerative disease (Alzheimer and Parkinson diseases) [47,48], and HIV infection.

Search Tool Utilized

In this study, we used the Twitter Firehose data stream, which is managed by Gnip and allows access to 100% of all public tweets that match a set of "search" criteria (query) [49]. In our study, the search criteria were the previously indicated keywords, and the following is an example of a query: "depression -economic -great -tropical from:nytimes OR from:washingtonpost OR from:nypost OR from:latimes OR from:USATODAY OR from:chicagotribune OR from:CNN OR from:ABC OR from:NBCNews OR from:CBSNews OR from:FoxNews OR from:AP OR from:TIME OR from:HuffingtonPost OR from:BuzzFeed until:2017-01-01. Tweet Binder, the search engine we employed, uses automatic machine learning text analysis algorithms, and it also uses node.js and the PHP language, which enables an analysis of tweets in the json format (used by Gnip).

Next, all the collected tweets were individually inspected by 3 members of the research team to identify tweets deemed irrelevant for the purpose of this study. Tweets that included keywords not related to psychiatric content were excluded, such as those referring to suicide attacks, economic depression, etc. The content of the tweets was then specifically analyzed by 3 separate blinded members of the research team, and those with at least 2 coincidences were excluded. This process led to the creation of a more concise database that we could easily reference. Moreover, the number of tweets generated was stratified by month and year beginning in January 2007 and terminating in December 2016. We also analyzed the number of retweets that each tweet generated, which yielded a total database of 19,117 tweets and 1,455,787 retweets.

Statistical Analysis

A descriptive analysis of the number of tweets and retweets was performed for both the mental health and control conditions. The correlations among the observation time units (months) were evaluated using the Spearman rank test. To analyze the retweets generated by the disease-related tweets, odds ratios (ORs) were calculated for each of the studied diseases. The odds of the sum of all conditions (retweet to tweet ratio) was used as the baseline and confidence intervals were calculated using a Bonferroni-adjusted significance level (alpha) of .001. To evaluate the annual changes within and differences between the two groups, a multivariable generalized linear model (negative binomial regression) was performed for both tweets and retweets. Finally, seasonality was studied through the Seasonal Decomposition procedure of a multiplicative time series model.

All statistical analyses were performed using SPSS v22 and STATA v14.

Results

Media Outlets Showed a Marked Interest in Mental Health Diseases and Tweet Patterns Generated Responses From Followers

We first analyzed the number of tweets generated by 15 mainstream American media outlets related to mental health disorders beginning in 2007 (soon after the launch of Twitter) through December 2016. As a control, we also included a parallel analysis of tweets related to the primary causes of death in the USA (prostate, lung, colorectal, and breast cancer; stroke, diabetes mellitus, and COPD), the two most relevant chronic neurologic degenerative diseases (Alzheimer and Parkinson disease), and HIV infection.

As shown in Table 1, 13,119 tweets were generated by the media about mental health disorders. The number of tweets about each of the analyzed diseases follows a heterogeneous pattern of distribution, with a preferential accumulation for a select number of conditions. Suicide and gender dysphoria accounted for half of the total number of tweets. The tweets related to highly prevalent anxiety and its different clinical forms only accounted for 11.39% (1494/13119) of the total number, and it was followed by depression, which accounted for 10.66% (1399/13119) of tweets. Mental health diseases characterized by child and adolescent incidence, such as autism, Asperger syndrome, ADHD, anorexia, and bulimia, accounted for 13.87% (1819/13119) of the total tweets generated. Additionally, 9.39% (1232/13119) of all tweets were related to addictive disorders, specifically alcoholism, drug abuse, and gambling disorders. Less than 8% of the analyzed tweets referred to the eight other diseases included in the study. Of note, bipolar disorder and schizophrenia, both of which are highly prevalent and disabling, only accounted for 0.63% (82/13119) and 1.33% (174/13119) of all generated tweets, respectively.

In the parallel control study, we measured the tweets distributed by American media on the diseases that are considered to be the main causes of death in the USA and paradigmatic examples of diseases with a demonstrated level of social interest (Table 2). In total, only 5998 tweets were generated by social media on this group of prevalent and severe diseases. The number of tweets focused on each individual disease analyzed also followed a heterogeneous pattern of distribution. A predominance of tweets was observed for a select number of conditions. In total, 31.06% (1863/5998) of the tweets referred to the four most lethal forms of cancer, although they mainly focused on breast cancer (1321/5998, 22.02%). HIV infection and Alzheimer disease received 22.79% (1367/5998) and 17.56% (1053/5998) of the tweets about organic disease generated by social media, respectively. Additionally, 28.59% (1715/5998) of the tweets were related to diabetes mellitus, stroke, Parkinson disease, and COPD. However, despite its prevalence, COPD only accounted for 0.08% (5/5998) of the tweets.

XSL•FO

RenderX

Table 1. Number of tweets about mental health diseases distributed by American media and the retweets they generated. Percentages (%) were calculated with respect to the total number of tweets distributed about the mental diseases group and the retweets generated. Spearman rank correlation coefficients (rho) between the tweets and retweets are shown for each condition or disease along with the level of statistical significance.

Mental health condition or disease	Tweet, n (%)	Retweet, n (%)	Spearman rho	*P* value
Suicide	4124 (31.44)	268,395 (26.03)	0.876	<.001
Gender dysphoria	2555 (19.48)	238,298 (23.11)	0.941	<.001
Total for anxiety disorders	1494 (11.39)	134,726 (13.07)	0.907	<.001
Anxiety	984 (7.50)	92,042 (8.93)	0.872	<.001
PTSD[a]	453 (3.45)	39,243 (3.81)	0.991	<.001
Phobias	34 (0.26)	1018 (0.10)	0.886	<.001
GAD[b]	22 (0.17)	2386 (0.23)	0.172	.064
Panic disorder	1 (0.01)	37 (<0.01)	–0.008	.927
Depression	1399 (10.66)	11,067 (11.26)	0.785	<.001
Autism spectrum disorders	1337 (10.19)	129,066 (12.52)	0.870	<.001
Autism	1253 (9.55)	117,955 (11.44)	0.860	<.001
Asperger syndrome	84 (0.64)	11,111 (1.08)	0.875	<.001
Addictive disorders	1232 (9.39)	83,809 (8.13)	0.822	<.001
Addictions	933 (7.11)	67,114 (6.51)	0.798	<.001
Alcoholism	146 (1.11)	7392 (0.72)	0.865	<.001
Drug addiction	143 (1.09)	8997 (0.87)	0.865	<.001
Gambling disorder	10 (0.08)	306 (0.03)	0.933	<.001
Anorexia and bulimia	274 (2.09)	11,792 (1.14)	0.852	<.001
ADHD[c]	208 (1.59)	12,103 (1.17)	0.853	<.001
Schizophrenia	174 (1.33)	15,232 (1.48)	0.839	<.001
Insomnia	128 (0.98)	10,014 (0.97)	0.825	<.001
Bipolar disorder	82 (0.63)	6946 (0.67)	0.867	<.001
OCD[d]	81 (0.62)	3564 (0.35)	0.907	<.001
Personality disorder	31 (0.24)	962 (0.09)	0.038	.684
Dysthymia	0 (0)	0 (0)	N/A[e]	N/A
Total for mental health disorders	13,119 (100)	1,030,974 (100)	0.915	<.001

[a]PTSD: posttraumatic stress disorder.

[b]GAD: generalized anxiety disorder.

[c]ADHD: attention deficit hyperactivity disorder.

[d]OCD: obsessive-compulsive disorder.

[e]N/A: not applicable.

Next, we investigated the impact of tweets about mental health and disease control among social media followers by analyzing the responses based on the number of retweets. In total, 1,030,974 retweets were related to the studied mental health diseases and 424,813 were related to the control organic diseases (Tables 1 and 2). We observed a significant correlation between the number of tweets referring to each individual mental health disorder and the number of subsequent retweets generated. The statistical significance of the correlations was similar for the control organic diseases. The percentages of tweets and retweets generated for each of the control diseases, mental health conditions, and psychiatric diseases are shown in a figure in Multimedia Appendix 1. A scatterplot of the tweets about mental health conditions, psychiatric diseases and control diseases as well as the number of retweets that they subsequently generated is also shown in the Multimedia Appendix 2.

We also investigated the retweets of disease-related tweets by analyzing the retweet to tweet ratio and absolute numbers for the mental health disorders and control diseases. We found that the retweet to tweet ratio for the psychiatric diseases was significantly higher than that found for the control diseases (OR 1.11, CI 1.07-1.14, *P*<.001). The analysis of the probabilities of retweeting a tweet related to a specific disease showed a marked heterogeneity between mental health and organic

disorders (Figure 1). Among the mental health conditions and diseases, the tweets about suicide, addictive disorders, anorexia and bulimia, and ADHD had a statistically significantly lower probability of being retweeted. In contrast, the probability of being retweeted was significantly higher for tweets related to gender dysphoria, anxiety, and autism spectrum disorders. For the control diseases, we also found a heterogeneous pattern of retweet responses, with the highest statistically significant probability of being retweeted found for Parkinson disease. In contrast, the tweets about cancer, diabetes, and stroke had significantly lower probabilities of being retweeted.

Table 2. Number of tweets about control diseases distributed by American media and the retweets they generated. Percentages (%) were calculated with respect to the total number of tweets or retweets distributed in the control group of diseases. Spearman rank correlation coefficients between the tweets and retweets are shown for each condition or disease along with the level of statistical significance.

Control disease	Tweet, n (%)	Retweet, n (%)	Spearman rho	P value
Total for cancers	1863 (31.06)	109,697 (25.82)	0.715	<.001
Breast cancer	1321 (22.02)	79,152 (18.63)	0.763	<.001
Prostate cancer	326 (5.44)	13,675 (3.22)	0.648	<.001
Lung cancer	196 (3.27)	16,425 (3.87)	0.733	<.001
Colorectal cancer	20 (0.33)	445 (0.10)	0.845	<.001
HIV	1367 (22.79)	110,919 (26.11)	0.812	<.001
Alzheimer disease	1053 (17.56)	82,334 (19.38)	0.828	<.001
Diabetes	760 (12.67)	47,354 (11.15)	0.734	<.001
Stroke	701 (11.69)	44,328 (10.43)	0.796	<.001
Parkinson disease	249 (4.15)	30,160 (7.10)	0.873	<.001
COPD[a]	5 (0.08)	21 (<0.01)	0.624	<.001
Total for control diseases	5998 (100)	424,813 (100)	0.869	<.001

[a]COPD: chronic obstructive pulmonary disease.

Figure 1. Different probabilities of retweets generated for tweets posted on mental health conditions and diseases (blue dots) and organic diseases (red dots). The odds ratios (ORs) are shown for the retweet to tweet ratios found for each individual disease with more than 100 tweets. Circles represent the calculated OR, and the vertical lines represent the CI. ADHD: attention deficit hyperactivity disorder.

	Suicide	Gender dysphoria	Total anxiety	Depression	Autism spectrum disorders	Addictive disorders	Anorexia and bulimia	ADHD	Schizophrenia	Insomnia	Total cancer	HIV	Alzheimer's	Diabetes	Stroke	Parkinson's
Max	0.9	1.31	1.29	1.19	1.39	0.98	0.69	0.96	1.48	1.38	0.84	1.17	1.14	0.93	0.94	1.97
Min	0.81	1.14	1.08	0.99	1.15	0.81	0.46	0.6	0.89	0.76	0.71	0.97	0.92	0.72	0.73	1.29
OR	0.85	1.22	1.18	1.09	1.27	0.89	0.56	0.76	1.15	1.03	0.77	1.06	1.03	0.82	0.83	1.59

XSL•FO
RenderX

Figure 2. Kinetic study of the number of tweets (left panel, continuous line) distributed by American mass media outlets on mental health conditions and diseases (blue) and control diseases (red) and retweets (right panel, dotted line) generated by their followers; Y-axis: total number of tweets or retweets.
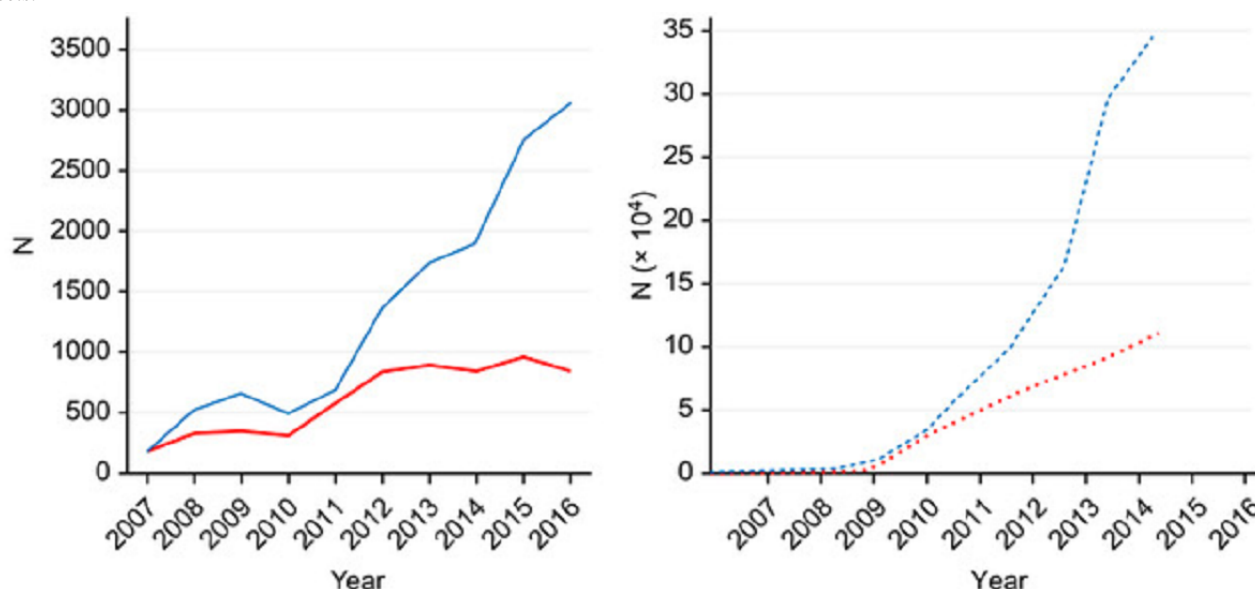
Figure 3. Left panel: monthly distribution of the tweets sent by American mass media outlets about mental health conditions and diseases (blue solid line) and control diseases (red solid line); Y-axis: percentages were calculated with respect to the total number of tweets. Top right panel: specific analyses of breast cancer and HIV infection; Y-axis: number of tweets. Bottom right panel: specific analyses of depression and gender dysphoria; Y-axis: number of tweets or retweets.
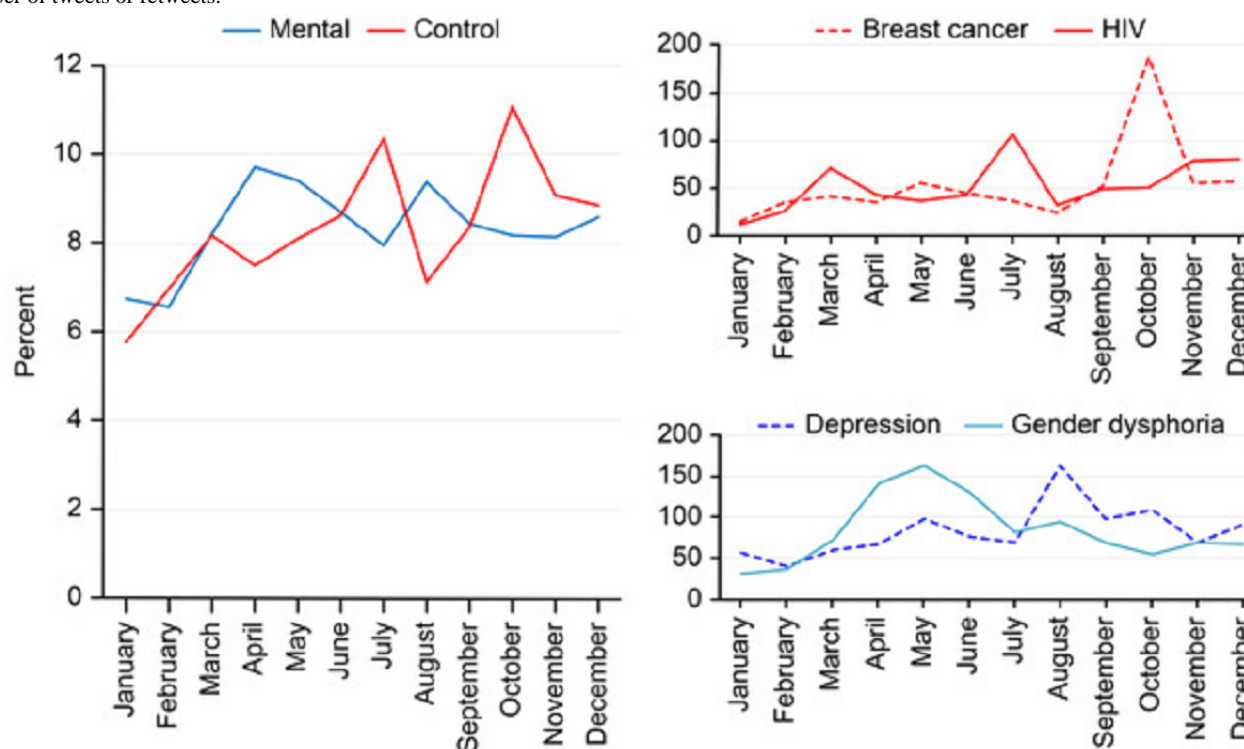
Number of Mass Media Tweets and Follower Retweets Is Continuously Increasing

We analyzed the evolution of the number of tweets about mental health conditions and control diseases that were distributed by the mainstream American media outlets along the analyzed decade. We also studied the kinetics of the retweets that these tweets generated (Figure 2); and observed a steady and progressive increase in the number of tweets generated for mental health conditions and diseases by communication media

across the analyzed years. Furthermore, there was an associated increase in the number of retweets sent by followers. Interestingly, a large increase in retweet responses was observed beginning in 2014. For the control diseases, an increase in the number of communication media generated tweets was observed between 2007 and 2012, and a steady level was reached by 2016. However, the number of generated retweets among nonpsychiatric control diseases also showed a continuous increase. To determine the effects of the year and type of

disease, we ran generalized linear models for both tweets and retweets. In both models, these variables were statistically significant (*P*<.001). The output of the negative binomial regression parameters is included in the Multimedia Appendix 3.

We also investigated the number of tweets generated over continuous months about the mental health and control diseases. Temporal variability was observed in the frequency of tweets about psychiatric disease, with a significant increase in April and August and a decrease in February (Figure 3). Monthly variability in the tweets generated about organic control diseases was also observed, with a statistically significant increase in months July and October and a decrease in January. This monthly variability was also observed in the analysis of individual mental health conditions and diseases. The results obtained for gender dysphoria, depression, breast cancer, and HIV are shown as representative cases for both the mental health conditions and control diseases.

Discussion

Principal Findings

In this paper, we showed that American outlets show preferential interest in psychiatric disorders compared with prevalent and severe organic diseases. The elevated number of tweets sent by the analyzed media outlets about mental health conditions and diseases was heterogeneously distributed between the different clinical entities studied. The relative attention of media outlets for the different mental health disorders conditioned the retweet response of followers.

The important role of communication media outlets in generating popular opinion and emotions via information distribution has been clearly established in our society [50]. In addition to traditional forms of communication media, both the internet and social media have become particularly pivotal instruments for sharing knowledge and news. Along with this change in the pattern of access to and sharing of information, communicative mass media includes the use of social media for connecting to the public. Currently, the use of social media websites, such as Facebook and Twitter, is commonplace, with approximately 65% of American adults and 66% of British adults reporting ownership of at least one active social media account [51]. Likewise, Twitter is currently considered an equally effective channel for communication [52].

Communication Media and Psychiatry

Our work demonstrates that American classic communication media outlets show a relevant interest in psychiatric diseases, as measured by the number of tweets about mental health conditions and disorders with respect to those about a group of severe and prevalent nonpsychiatric diseases, including the main causes of death in the USA. In recent decades, the stigma associated with mental health-related disorders has been widespread as evidenced by our social behaviors [53,54]. This social attitude has had major adverse effects on the lives of people with mental health problems, conditions, and diseases [55]. Therefore, the interest of traditional communication media outlets in psychiatric diseases should decrease over time.

However, our findings contradict this hypothesis. The number of tweets sent about the analyzed mental conditions and diseases was higher than that of the control group throughout the decade examined, and a continuously increasing trend was observed in recent years. Interestingly, the control diseases included the main causes of mortality in the USA, such as the most predominant malignant tumor causes of death (cancer), stroke, diabetes mellitus, chronic degenerative neurological diseases, and COPD [47,48]. The control group of diseases also included HIV infection, a disease that has maintained a high level of general interest in our society in recent decades [56,57]. In addition to the demonstrated interest in mental health conditions and diseases by mass media, we found that this interest is more focused on certain clinical entities.

Interestingly, the relative weight given to each disease as defined by the percentage of tweets received was not related to the actual prevalence of the disease (the prevalence of mental health conditions, psychiatric diseases, and control diseases are included in the Multimedia Appendix 4). Despite the low incidence of suicide and gender dysphoria, these topics accounted for half of the tweets generated by communication media. In contrast, anxiety and depression are highly prevalent in society but only accounted for a quarter of the total number of tweets. Furthermore, psychiatric diseases with a marked prevalence and associated morbidity, such as schizophrenia and bipolar disorder, only accounted for a marginal percentage of the tweets. This lack of correlation between the prevalence and the morbidity or mortality (or both) of a disease and its relative presence in the number tweets generated by communication media outlets was also observed in the control group. These results are aligned with previous results demonstrating that certain chronic diseases, such as hypertension, are "undertweeted" relative to their prevalence, whereas other chronic diseases, such as diabetes and heart failure, are "over-tweeted" relative to their prevalence [58].

Interest in Psychiatry on Twitter

The interest provoked by mental health disease-associated tweets sent by mass media organizations to the general public, as measured by the number of retweets generated by followers, is clearly relevant. The retweet frequency is a parameter that indicates the user interest in the topic of each tweet [59,60]. Our data demonstrate that the retweet to tweet ratio generated by mental health disease-related tweets was significantly higher than that of the control diseases. Thus, in addition to a correlation between the number of tweets sent about a specific disease and the retweet response provoked, the characteristics of the health disorder also modulate the interest and quantitative retweet response of the followers. This finding is clearly supported by the significantly higher possibility of retweeting a tweet on gender dysphoria, anxiety, and autism spectrum disorders and the decreased possibility of retweeting a tweet related to suicide, addictive disorders, anorexia and bulimia, and ADHD. Several reasons that are not mutually exclusive may explain this public behavior. First, the potential anonymity of Twitter might favor its use by people who present feelings of potential self-stigma. For example, tweeting about mental health conveys the notion of a "Twitter community" that allows communication to flourish, awareness to be raised, stigmas to

be fought, and support to be both offered and received [51]. Twitter use allows for anonymity; thus, it is preferred by people with real or perceived personal and/or or social restrictions [61]. The reported use of Twitter by transgender individuals and allies to discuss health and social needs supports this statement [62]. Second, Twitter is becoming more popular in our society, and the average user profile is distributed across different age groups. However, Twitter is predominantly used by younger and middle-aged demographics [63,64]. Thus, the social media pattern of Twitter might indicate a modification in attitudes toward mental health diseases among these two generations. Furthermore, the age of the person affects their general interest in health-related matters [65]. Third, high rates of social media use are observed among individuals who experience mental health problems [66,67]. Fourth, health care professionals and provider communities may also show a greater interest in mental diseases and contribute to the dissemination of this information. However, the attitudes of professionals, such as general practitioners, towards these diseases cannot be considered optimal at the present time [68]. Additionally, certain mental health conditions, such as gender dysphoria and suicide, are topics that often appear in breaking social news and may easily go viral on Twitter. The information transmitted by mass media may be selected using different criteria, including content generally considered to be of public interest [69]. According to cultural selection theory, any selection of messages from communication media outlets will have a profound effect on society at large and can contribute to the modulation of individual and societal attitudes and knowledge [70]. Based on the frequency of tweets generated about mental health disorders found in this work, we conclude that mass media outlets do not support a quantitative stigmatic exclusion of psychiatric patients.

However, the results observed for suicide should be further discussed. Suicide was one of the most frequently mentioned topics on Twitter by communication media outlets. Interestingly, the Werther effect of suicide reports in social media networks, such as Twitter, has been established [71]. Thus, the criteria applied for generating this increased frequency of suicide-related tweets by communication media outlets may require revision. Fortunately, a suicide-related tweet has a significantly reduced possibility of being retweeted by followers.

Limitations

This study has some limitations. The relevance of Twitter as a marker of social interest is a matter of controversy [24-26,28-32]. Furthermore, news media outlets do not necessarily reflect the interests of society [72]. Large media outlets can also have a different set of priorities than news media in general. The newsworthiness of health science articles has previously been reported [73-75].

Conclusions

In conclusion, our findings show a marked correlation between the number of tweets generated about a psychiatric or control disease and the number of retweets that are subsequently generated. These results could represent a coincidence between the interest of communication media outlets and the general population and/or merely the quantitative reactive response of followers to the tweets they receive. Interestingly, the frequency of retweeting a tweet related to suicide was less than expected, whereas that of gender dysphoria was greater. Moreover, there are contradictory results with respect to the association between mental health problems and social media, which indicates either the potential for harm or a significant improvement in social media engagement as previously described [71,76-78].

Conflicts of Interest

None declared.

Multimedia Appendix 1

The percentages of tweets and retweets generated for each of the control diseases, mental health conditions, and psychiatric diseases.

[PNG File, 138KB - jmir_v20i5e205_app1.png]

Multimedia Appendix 2

A scatterplot of the tweets about mental health conditions, psychiatric diseases and control diseases as well as the number of retweets that they subsequently generated.

[PNG File, 147KB - jmir_v20i5e205_app2.png]

Multimedia Appendix 3

Output table for the negative binomial regression parameters.Estimated coefficients from IRR reports transformed into incidence-rate ratios.The standard errors (SEs) reported in the table were calculated using the robust or sandwich estimator of variance.

XSL•FO

RenderX

[PNG File, 132KB - jmir_v20i5e205_app3.png]

Multimedia Appendix 4

Prevalence of mental health conditions, psychiatric diseases, and control diseases.

[PNG File, 1MB - jmir_v20i5e205_app4.png]

References

1. National Institute of Mental Health. Mental Illness URL: http://www.nimh.nih.gov/health/statistics/prevalence/any-mental-illness-ami-among-adults.shtml [WebCite Cache ID 6vSpzoaEm]

2. Evans DL, Charney DS, Lewis L, Golden RN, Gorman JM, Krishnan KRR, et al. Mood disorders in the medically ill: scientific review and recommendations. Biol Psychiatry 2005 Aug 01;58(3):175-189. [doi: 10.1016/j.biopsych.2005.05.001] [Medline: 16084838]

3. Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2197-2223. [doi: 10.1016/S0140-6736(12)61689-4] [Medline: 23245608]

4. Chang C, Hayes RD, Perera G, Broadbent MTM, Fernandes AC, Lee WE, et al. Life expectancy at birth for people with serious mental illness and other major disorders from a secondary mental health care case register in London. PLoS One 2011;6(5):e19590 [FREE Full text] [doi: 10.1371/journal.pone.0019590] [Medline: 21611123]

5. Walker ER, McGee RE, Druss BG. Mortality in mental disorders and global disease burden implications: a systematic review and meta-analysis. JAMA Psychiatry 2015 Apr;72(4):334-341 [FREE Full text] [doi: 10.1001/jamapsychiatry.2014.2502] [Medline: 25671328]

6. Dickerson FB, Sommerville J, Origoni AE, Ringel NB, Parente F. Experiences of stigma among outpatients with schizophrenia. Schizophr Bull 2002;28(1):143-155. [Medline: 12047014]

7. Drapalski AL, Lucksted A, Perrin PB, Aakre JM, Brown CH, DeForge BR, et al. A model of internalized stigma and its effects on people with mental illness. Psychiatr Serv 2013 Mar 01;64(3):264-269. [doi: 10.1176/appi.ps.001322012] [Medline: 23573532]

8. Thornicroft G, Mehta N, Clement S, Evans-Lacko S, Doherty M, Rose D, et al. Evidence for effective interventions to reduce mental-health-related stigma and discrimination. Lancet 2016 Mar 12;387(10023):1123-1132. [doi: 10.1016/S0140-6736(15)00298-6] [Medline: 26410341]

9. Prince M, Patel V, Saxena S, Maj M, Maselko J, Phillips MR, et al. No health without mental health. Lancet 2007 Sep 08;370(9590):859-877. [doi: 10.1016/S0140-6736(07)61238-0] [Medline: 17804063]

10. 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]

11. Antheunis ML, Tates K, Nieboer TE. Patients' and health professionals' use of social media in health care: motives, barriers and expectations. Patient Educ Couns 2013 Sep;92(3):426-431. [doi: 10.1016/j.pec.2013.06.020] [Medline: 23899831]

12. Melvin L, Chan T. Using Twitter in Clinical Education and Practice. J Grad Med Educ 2014 Sep;6(3):581-582 [FREE Full text] [doi: 10.4300/JGME-D-14-00342.1] [Medline: 26279790]

13. Fox S, Duggan M. Pew Research Center. Washington, DC: Pew Internet and American Life Project; 2013. Health Online 2013 URL: http://www.pewinternet.org/2013/01/15/health-online-2013/ [WebCite Cache ID 6vT1ISBAz]

14. Hesse BW, Moser RP, Rutten LJ. Surveys of physicians and electronic health information. N Engl J Med 2010 Mar 4;362(9):859-860. [doi: 10.1056/NEJMc0909595] [Medline: 20200398]

15. 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]

16. Topf JM, Hiremath S. Social media, medicine and the modern journal club. Int Rev Psychiatry 2015 Apr;27(2):147-154. [doi: 10.3109/09540261.2014.998991] [Medline: 25906989]

17. Twitter, Inc. 2015. About Twitter URL: https://about.twitter.com/en_us/company.html [accessed 2017-12-04] [WebCite Cache ID 6vSqUsqnI]

18. Peters ME, Uible E, Chisolm MS. A Twitter Education: Why Psychiatrists Should Tweet. Curr Psychiatry Rep 2015 Dec;17(12):94. [doi: 10.1007/s11920-015-0635-4] [Medline: 26463050]

19. Sano D. Twitter creator Jack Dorsey illuminates the site's founding document. 2009 URL: http://latimesblogs.latimes.com/technology/2009/02/twitter- [WebCite Cache ID 6vSrMRloz]

20. Lauridsen M, Sporrong S. How does media coverage effect the consumption of antidepressants? A study of the media coverage of antidepressants in Danish online newspapers 2010?2011. Res Soc Adm Pharm Internet 2017:30130-30134 [FREE Full text] [doi: 10.1016/j.sapharm.2017.07.011] [Medline: 28811152]

21. Hernandez JF, Mantel-Teeuwisse AK, van Thiel GJ, Belitser SV, Warmerdam J, de Valk V, et al. A 10-year analysis of the effects of media coverage of regulatory warnings on antidepressant use in The Netherlands and UK. PLoS One 2012;7(9):e45515 [FREE Full text] [doi: 10.1371/journal.pone.0045515] [Medline: 23029065]

22. Slothuus R, de Vreese CH. Political Parties, Motivated Reasoning, and Issue Framing Effects. The Journal of Politics 2010 Jul;72(3):630-645. [doi: 10.1017/S002238161000006X]

23. Jansen BJ, Zhang M, Sobel K, Chowdury A. Twitter power: Tweets as electronic word of mouth. J. Am. Soc. Inf. Sci 2009 Nov;60(11):2169-2188. [doi: 10.1002/asi.21149]

24. Asur S, Huberman B, editors. Predicting the future with social media. In: IEEE/WIC/ACM International Conference on Web IntelligenceIntelligent Agent Technology (WI-IAT).: IEEE; 2010 Presented at: IEEE/WIC/ACM International Conference on Web IntelligenceIntelligent Agent Technology (WI-IAT); 31 Aug.-3 Sept. 2010; Toronto, ON, Canada p. 2010. [doi: 10.1109/WI-IAT.2010.63]

25. Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market. Journal of Computational Science 2011 Mar;2(1):1-8. [doi: 10.1016/j.jocs.2010.12.007] [Medline: 25904163]

26. Carr A. Fast Company. Facebook, Twitter Election Results Prove Remarkably Accurate Internet URL: https://www.fastcompany.com/1699853/facebook-twitter-election-results-prove-remarkably-accurate [accessed 2018-03-27] [WebCite Cache ID 6zRD22cox]

27. Kamel Boulos MN, Anderson PF. Preliminary survey of leading general medicine journals' use of Facebook and Twitter. JCHLA 2012 Aug;33(02):38-47. [doi: 10.5596/c2012-010]

28. Mitchell A, Hitlin P. Pew Research Center. Twitter Reaction to Events Often at Odds with Overall Public Opinion Internet URL: http://www.pewresearch.org/2013/03/04/twitter-reaction-to-events-often-at-odds-with-overall-public-opinion/ [accessed 2018-05-15] [WebCite Cache ID 6zRHuZxlh]

29. Tufekci Z. Big Questions for social media big data: Representativeness, validity and other methodological pitfalls. In: Big Questions for social media big data: Representativeness, validity and other methodological pitfalls. In: Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014. The AAAI Press; 2014 Presented at: 8th International Conference on Weblogs and Social Media, ICWSM 2014; Jun 1-4 2014; Ann Arbor, United States p. 505.

30. Gayo-Avello D, Metaxas P, Mustafaraj E. Limits of Electoral Predictions using Twitter. 2011 Jan 17 Presented at: Fifth International AAAI Conference on Weblogs and Social Media; July 17–21, 2011; Barcelona, Spain URL: http://digibuo.uniovi.es/dspace/handle/10651/11899

31. Cohen R, Ruths D. Classifying Political Orientation on Twitter: It's Not Easy!. Classifying political orientation on Twitter: It?s not easy! In: Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM 2013. AAAI press; 2013 Jul 8 Presented at: Seventh International AAAI Conference on Weblogs and Social Media; July 8–11, 2013; Cambridge, Massachusetts p. 91.

32. Daniel Gayo-Avello. I wanted to predict elections with twitter and all I got was this Lousy paper. Computers and Society 2012:6441.

33. Robillard JM, Johnson TW, Hennessey C, Beattie BL, Illes J. Aging 2.0: health information about dementia on Twitter. PLoS One 2013;8(7):e69861 [FREE Full text] [doi: 10.1371/journal.pone.0069861] [Medline: 23922827]

34. Mowery J. Twitter Influenza Surveillance: Quantifying Seasonal Misdiagnosis Patterns and their Impact on Surveillance Estimates. Online J Public Health Inform 2016;8(3):e198 [FREE Full text] [doi: 10.5210/ojphi.v8i3.7011] [Medline: 28210419]

35. Quintana DS, Doan NT. Twitter Article Mentions and Citations: An Exploratory Analysis of Publications in the American Journal of Psychiatry. Am J Psychiatry 2016 Feb 01;173(2):194. [doi: 10.1176/appi.ajp.2015.15101341] [Medline: 26844798]

36. Meng Y, Elkaim L, Wang J, Liu J, Alotaibi NM, Ibrahim GM, et al. Social media in epilepsy: A quantitative and qualitative analysis. Epilepsy Behav 2017 Jun;71(Pt A):79-84. [doi: 10.1016/j.yebeh.2017.04.033] [Medline: 28554148]

37. 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]

38. Diug B, Kendal E, Ilic D. Evaluating the use of twitter as a tool to increase engagement in medical education. Educ Health (Abingdon) 2016;29(3):223-230 [FREE Full text] [doi: 10.4103/1357-6283.204216] [Medline: 28406107]

39. Sedrak MS, Cohen RB, Merchant RM, Schapira MM. Cancer Communication in the Social Media Age. JAMA Oncol 2016 Jun 01;2(6):822-823. [doi: 10.1001/jamaoncol.2015.5475] [Medline: 26940041]

40. Nastasi A, Bryant T, Canner JK, Dredze M, Camp MS, Nagarajan N. Breast Cancer Screening and Social Media: a Content Analysis of Evidence Use and Guideline Opinions on Twitter. J Cancer Educ 2018 Jun;33(3):695-702. [doi: 10.1007/s13187-017-1168-9] [Medline: 28097527]

41. Fox CS, Bonaca MA, Ryan JJ, Massaro JM, Barry K, Loscalzo J. A randomized trial of social media from Circulation. Circulation 2015 Jan 06;131(1):28-33 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.114.013509] [Medline: 25406308]

42. The Statistics Portal | Statistic Internet. Number of organizational Twitter feeds belonging to major news outlets in the United States in 2011 URL: https://www.statista.com/statistics/207367/twitter-feeds-of-major-american-news-outlets/ [accessed 2018-05-15] [WebCite Cache ID 6zRK1MPyQ]

XSL•FO

RenderX

43. Nur Bremmen. memeburn. 2010 Sep 03. The 100 most influential news media Twitter accounts URL: https://memeburn. com/2010/09/the-100-most-influential-news-media-twitter-accounts/

44. Mitchell A, Matsa K, Kiley J, Gottfried J. Pew Research Center.: Pew Research Center; 2014 Oct 21. Political Polarization & Media Habits URL: http://www.journalism.org/2014/10/21/political-polarization-media-habits/ [accessed 2018-05-15] [WebCite Cache ID 6zRLccbKp]

45. Mitchell A, Page D. Pew Research Center.: Pew Research Center; 2015 Apr 29. State of the News Media 2015 URL: http://www.journalism.org/files/2015/04/FINAL-STATE-OF-THE-NEWS-MEDIA.pdf [accessed 2018-05-15] [WebCite Cache ID 6zRLrDmtM]

46. Holcomb J, Gross K, Mitchell A. Pew Research Center.: Pew Research Center; 2011 Nov 14. How Mainstream Media Outlets Use Twitter URL: http://www.journalism.org/2011/11/14/how-mainstream-media-outlets-use-twitter/ [accessed 2018-05-15] [WebCite Cache ID 6zRM1hpPs]

47. Heron M. Deaths: Leading Causes for 2014. Natl Vital Stat Rep 2016 Jun;65(5):1-96 [FREE Full text] [Medline: 27376998]

48. Kowal SL, Dall TM, Chakrabarti R, Storm MV, Jain A. The current and projected economic burden of Parkinson's disease in the United States. Mov Disord 2013 Mar;28(3):311-318. [doi: 10.1002/mds.25292] [Medline: 23436720]

49. Morstatter F, Pfeffer J, Liu H, Carley K. Is the sample good enough? Comparing data from Twitter's streaming API with Twitter?s firehose. 2013 Aug 07 Presented at: 7th International AAAI Conference on Weblogs and Social Media, ICWSM 2013; 2013/08/07; Cambridge, MA, United States p. 2013.

50. Wakefield MA, Loken B, Hornik RC. Use of mass media campaigns to change health behaviour. Lancet 2010 Oct 9;376(9748):1261-1271 [FREE Full text] [doi: 10.1016/S0140-6736(10)60809-4] [Medline: 20933263]

51. Berry N, Lobban F, Belousov M, Emsley R, Nenadic G, Bucci S. #WhyWeTweetMH: Understanding Why People Use Twitter to Discuss Mental Health Problems. J Med Internet Res 2017 Apr 05;19(4):e107 [FREE Full text] [doi: 10.2196/jmir.6173] [Medline: 28381392]

52. Ferguson C, Inglis SC, Newton PJ, Cripps PJS, MacDonald PS, Davidson PM. Social media: a tool to spread information: a case study analysis of twitter conversation at the Cardiac Society of Australia & New Zealand 61st annual scientific meeting 2013. Collegian 2014;21(2):89-93. [Medline: 25109206]

53. Thornicroft G. Shunned: Discrimination Against People with Mental Illness. Oxford: Oxford University Press; 2006.

54. Evans-Lacko S, Courtin E, Fiorillo A, Knapp M, Luciano M, Park A, ROAMER Consortium, et al. The state of the art in European research on reducing social exclusion and stigma related to mental health: a systematic mapping of the literature. Eur Psychiatry 2014 Aug;29(6):381-389. [doi: 10.1016/j.eurpsy.2014.02.007] [Medline: 24726533]

55. Corrigan PW, Larson JE, Rüsch N. Self-stigma and the "why try" effect: impact on life goals and evidence-based practices. World Psychiatry 2009 Jun;8(2):75-81 [FREE Full text] [Medline: 19516923]

56. Acheson E. AIDS: A challenge for the public health. Lancet 1986;327(8482):A. [doi: 10.1016/S0140-6736(86)91736-8]

57. Ayers JW, Althouse BM, Dredze M, Leas EC, Noar SM. News and Internet Searches About Human Immunodeficiency Virus After Charlie Sheen's Disclosure. JAMA Intern Med 2016 Apr;176(4):552-554. [doi: 10.1001/jamainternmed.2016.0003] [Medline: 26902971]

58. Sinnenberg L, DiSilvestro CL, Mancheno C, Dailey K, Tufts C, Buttenheim AM, et al. Twitter as a Potential Data Source for Cardiovascular Disease Research. JAMA Cardiol 2016 Dec 01;1(9):1032-1036 [FREE Full text] [doi: 10.1001/jamacardio.2016.3029] [Medline: 27680322]

59. Suh B, Hong L, Pirolli P, Chi E, editors. Want to be retweeted? Large scale analytics on factors impacting retweet in twitter network. USA: IEEE; 2010 Presented at: IEEE Second International Conference on Social Computing; 2010/10/10; Minneapolis, MN, USA.

60. Kwak H, Lee C, Park H, Moon S. What is twitter, a social network or a news media? In: Proceedings of the 19th International Conference on World Wide Web Internet. 2010 Apr 26 Presented at: 19th International Conference on World Wide Web; 2010/04/26; Raleigh, NC, USA p. 2010 URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.212.1490

61. Gillespie-Lynch K, Kapp SK, Shane-Simpson C, Smith DS, Hutman T. Intersections between the autism spectrum and the internet: perceived benefits and preferred functions of computer-mediated communication. Intellect Dev Disabil 2014 Dec;52(6):456-469. [doi: 10.1352/1934-9556-52.6.456] [Medline: 25409132]

62. Simoncic TE, Kuhlman KR, Vargas I, Houchins S, Lopez-Duran NL. Facebook use and depressive symptomatology: Investigating the role of neuroticism and extraversion in youth. Computers in Human Behavior 2014 Nov;40:1-5. [doi: 10.1016/j.chb.2014.07.039]

63. Sloan L, Morgan J, Burnap P, Williams M. Who tweets? Deriving the demographic characteristics of age, occupation and social class from twitter user meta-data. PLoS One 2015;10(3):e0115545 [FREE Full text] [doi: 10.1371/journal.pone.0115545] [Medline: 25729900]

64. Pew Research Center. 2015. Social Media Usage: 2005-2015 URL: http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/ [WebCite Cache ID 6vSv9VPJ4]

65. Deeks A, Lombard C, Michelmore J, Teede H. The effects of gender and age on health related behaviors. BMC Public Health 2009;9:213 [FREE Full text] [doi: 10.1186/1471-2458-9-213] [Medline: 19563685]

XSL•FO

RenderX

66.   Gowen K, Deschaine M, Gruttadara D, Markey D. Young adults with mental health conditions and social networking websites: seeking tools to build community. Psychiatr Rehabil J 2012;35(3):245-250. [doi: 10.2975/35.3.2012.245.250] [Medline: 22246123]

67.   Birnbaum ML, Rizvi AF, Confino J, Correll CU, Kane JM. Role of social media and the Internet in pathways to care for adolescents and young adults with psychotic disorders and non-psychotic mood disorders. Early Interv Psychiatry 2017 Dec;11(4):290-295. [doi: 10.1111/eip.12237] [Medline: 25808317]

68.   Manzanera R, Lahera G, Álvarez-Mon MA, Alvarez-Mon M. Maintained effect of a training program on attitudes towards depression in family physicians. Fam Pract 2017 Jul 23:61-66. [doi: 10.1093/fampra/cmx071] [Medline: 28985370]

69.   Fog A. Mass Media. Cultural Selection. Dordrecht: Springer; 1999.

70.   Pew Research Center: Journalism and Media Staff. Washington, DC: Pew Internet & American Life Project; 2008. Health news coverage in the US media URL: http://www.journalism.org/2008/11/24/health-news-coverage-in-the-u-s-media/ [WebCite Cache ID 6vSqrvncL]

71.   Niederkrotenthaler T, Fu KW, Yip PSF, Fong DYT, Stack S, Cheng Q, et al. Changes in suicide rates following media reports on celebrity suicide: a meta-analysis. J Epidemiol Community Health 2012 Nov;66(11):1037-1042. [doi: 10.1136/jech-2011-200707] [Medline: 22523342]

72.   Mitchell A, Simmons K, Matsa K, Silver L. Pew Research Center.: Pew Research Center; 2018 Jan 11. Publics Globally Want Unbiased News Coverage, but Are Divided on Whether Their News Media Deliver URL: http://www.pewglobal.org/2018/01/11/publics-globally-want-unbiased-news-coverage-but-are-divided-on-whether-their-news-media-deliver/ [accessed 2018-05-15] [WebCite Cache ID 6zRNqEKNJ]

73.   Zhang Y, Willis E, Paul MJ, Elhadad N, Wallace BC. Characterizing the (Perceived) Newsworthiness of Health Science Articles: A Data-Driven Approach. JMIR Med Inform 2016 Sep 22;4(3):e27 [FREE Full text] [doi: 10.2196/medinform.5353] [Medline: 27658571]

74.   O'Connor EM, Nason GJ, O'Kelly F, Manecksha RP, Loeb S. Newsworthiness vs scientific impact: are the most highly cited urology papers the most widely disseminated in the media? BJU Int 2017 Dec;120(3):441-454. [doi: 10.1111/bju.13881] [Medline: 28418091]

75.   Wright D, Young A, Iserman E, Maeso R, Turner S, Haynes RB, et al. The clinical relevance and newsworthiness of NIHR HTA-funded research: a cohort study. BMJ Open 2014 May 07;4(5):e004556 [FREE Full text] [doi: 10.1136/bmjopen-2013-004556] [Medline: 24812191]

76.   Lin LY, Sidani JE, Shensa A, Radovic A, Miller E, Colditz JB, et al. Association Between Social Media Use and Depression Among U.S. Young Adults. Depress Anxiety 2016 Apr;33(4):323-331. [doi: 10.1002/da.22466] [Medline: 26783723]

77.   Rosen L, Whaling K, Rab S, Carrier L, Cheever N. Is Facebook creating ?iDisorders?? The link between clinical symptoms of psychiatric disorders and technology use, attitudes and anxiety. Comput Human Behav 2013 May 05;29(3):1243-1254.

78.   Mabe AG, Forney KJ, Keel PK. Do you "like" my photo? Facebook use maintains eating disorder risk. Int J Eat Disord 2014 Jul;47(5):516-523. [Medline: 25035882]

Abbreviations

ADHD: attention deficit hyperactivity disorder
COPD: chronic obstructive pulmonary disease
GAD: generalized anxiety disorder
OCD: obsessive-compulsive disorder
OR: odds ratio
PTSD: posttraumatic stress disorder
USA: United States of America

XSL•FO

RenderX

XSL•FO

RenderX

<u>Original Paper</u>

Reducing Negative Outcomes of Online Consumer Health Information: Qualitative Interpretive Study with Clinicians, Librarians, and Consumers

Reem El Sherif[1], MBBCh, MSc; Pierre Pluye[1], MD, PhD; Christine Thoër[2], PhD; Charo Rodriguez[1], MD, PhD

[1]Department of Family Medicine, McGill University, Montreal, QC, Canada

[2]Department of Social and Public Communication, University of Quebec in Montreal, Montreal, QC, Canada

Corresponding Author:
Reem El Sherif, MBBCh, MSc
Department of Family Medicine
McGill University
Suite 300
5858 Chemin Cotes-des-Neiges
Montreal, QC, H3S 1Z1
Canada
Phone: 1 5146323616
Email: reem.elsherif@mail.mcgill.ca

Abstract

Background: There has been an exponential increase in the general population's usage of the internet and of information accessibility; the current demand for online consumer health information (OCHI) is unprecedented. There are multiple studies on internet access and usage, quality of information, and information needs. However, few studies explored negative outcomes of OCHI in detail or from different perspectives, and none examined how these negative outcomes could be reduced.

Objective: The aim of this study was to describe negative outcomes associated with OCHI use in primary care and identify potential preventive strategies from consumers', health practitioners', and health librarians' perspectives.

Methods: This included a two-stage interpretive qualitative study. In the first stage, we recruited through a social media survey, a purposeful sample of 19 OCHI users who had experienced negative outcomes associated with OCHI. We conducted semistructured interviews and performed a deductive-inductive thematic analysis. The results also informed the creation of vignettes that were used in the next stage. In the second stage, we interviewed a convenient sample of 10 key informants: 7 health practitioners (3 family physicians, 2 nurses, and 2 pharmacists) and 3 health librarians. With the support of the vignettes, we asked participants to elaborate on (1) their experience with patients who have used OCHI and experienced negative outcomes and (2) what strategies they suggest to reduce these outcomes. We performed a deductive-inductive thematic analysis.

Results: We found that negative outcomes of OCHI may occur at three levels: internal (such as increased worrying), interpersonal (such as a tension in the patient-clinician relationship), and service-related (such as postponing a clinical encounter). Participants also proposed three types of strategies to reduce the occurrence of these negative outcomes, namely, providing consumers with reliable OCHI, educating consumers on how to assess OCHI websites, and helping consumers present and discuss the OCHI they find with a health professional in their social network or a librarian for instance.

Conclusions: We examined negative outcomes associated with using OCHI from five complementary perspectives (consumers, family physicians, pharmacists, nurses, and health librarians). We identified a construct of OCHI use–related tension that included and framed all negative outcomes. This construct has three dimensions (three interdependent levels): internal, interpersonal, and service-related tensions. Future research can focus on the implementation and effectiveness of the proposed strategies, which might contribute to reducing these tensions.

(J Med Internet Res 2018;20(5):e169)   doi:10.2196/jmir.9326

XSL•FO

RenderX

Introduction

Online Consumer Health Information

Online consumer health information (OCHI) is defined as information on health and diseases that is created for and directed to the general public [1]. General health information is available in written, audio, and video formats and freely accessible on government sites, professional organizations' websites, health journals, online forums, and blogs, among other sources. Moreover, consumers actively use social media to facilitate self-care, as well as being passively exposed to OCHI posts being shared by their social network through social media platforms such as Facebook [2,3].

American surveys of representative samples of the population have shown that the use of OCHI has increased dramatically over the last decade, and the internet is now the most popular source of consumer health information, whereas the use of other sources has decreased [4,5]. Reasons for this are the sheer volume of readily accessible health information available online, the increased engagement of people in their own health and self-care, and decreased access to health care services [6-8].

Increased access to OCHI is generally associated with increased consumer engagement in their own health care, increased empowerment of themselves and their families, and improved health outcomes [9-11]. Indeed, one of the most common ways consumers use OCHI is for consultation with health practitioners, for engagement in their own health care, compliance with or modification of management plan, or support of relatives or friends with health conditions [12]. There may, however, be some unintended negative consequences that are poorly understood and perhaps under reported [13,14]. With the increasing amount of OCHI available and the exponential increase in OCHI use, these negative consequences may also increase. The aim of this study was to identify and understand these negative outcomes from the viewpoint of primary care consumers and practitioners and try to find ways to reduce them.

Literature Review

Although there is evidence available on OCHI use and its outcomes in primary care, few studies have focused on the possible negative outcomes. In 2015, we conducted a systematic mixed studies review that examined outcomes associated with OCHI in primary care [15]. This review included 65 studies [15]. Although most of these studies reported positive outcomes (eg, reduced worries, increased satisfaction with health care services, increased involvement in decision making, and improvement of health), 23 studies described negative outcomes associated with using OCHI from either a physician or a patient perspective. One of the most well-known negative outcomes mentioned by both parties is increased anxiety, sometimes referred to as *cyberchondria* by researchers. Although physicians perceive this anxiety as *excessive or overestimated*, it is a reported consequence of looking for and using OCHI [16,17]. Another commonly reported outcome found in our review is deterioration in the patient-clinician relationship, especially after the patient shares the retrieved OCHI with a clinician [14]. A third outcome is the effect of using OCHI on the health care system resources, for example, leading to longer, unnecessary encounters with the family physician [18].

Most of the studies in our review were set in an oncology or public health setting and focused on specialized information (as opposed to general) or specific patient populations (eg, pregnant women) [19]. This is problematic as not all results from a specialist setting or tertiary health care population are transferrable to a primary care setting.

Figure 1. Conceptual framework.

XSL•FO

RenderX
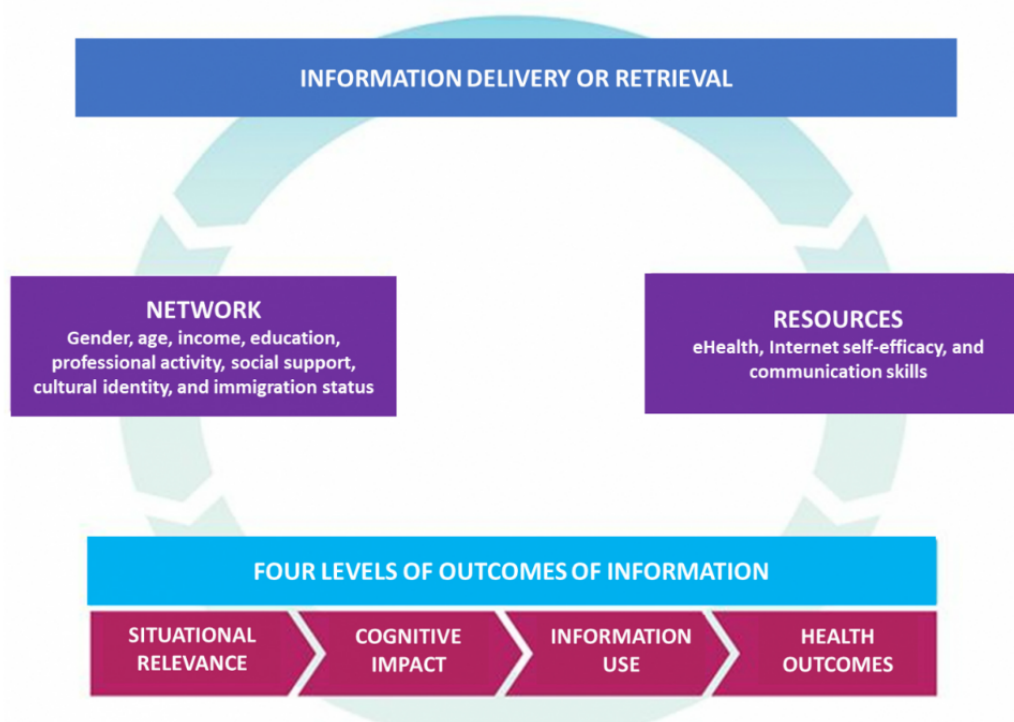
Objectives

The research question that guided this study was as follows: what are the negative outcomes of OCHI, and how can they be reduced? More specifically, we aimed to reach the following two objectives: (1) to identify and understand the meaning of the negative outcomes associated with OCHI use from consumers', health practitioners', and librarians' viewpoints in primary care and (2) to report means to reduce these negative outcomes proposed by consumers, health practitioners, and librarians.

Conceptual Framework

Our conceptual framework (Figure 1) [20] includes four levels of outcomes of information delivery and retrieval [12,21]: situational relevance, cognitive impact and use, and health or well-being outcomes of information. These levels are defined in relation to a specific information-seeking context: eg, a Web page used in a patient-clinician encounter. This framework was used to develop the interview guides and guided the deductive data analysis.

Methods

Study Design

We conducted a two-stage exploratory interpretive qualitative study that allows researchers "to obtain straight and largely unadorned answers to questions of special relevance to health care providers (HCPs) or policy makers" [22]. Methods and results are reported using the consolidated criteria for reporting qualitative studies [23]. We received ethical approval from the McGill University Faculty of Medicine's Research Ethics Office (institutional review board) before we started recruitment

Stage 1—Consumers

Sampling

We used a purposive sampling strategy to find participants who had experienced negative outcomes after using OCHI for themselves. Using a short online survey on SurveyMonkey (Multimedia Appendix 1) as a recruitment tool on social media platforms (Facebook and LinkedIn), we found a broad range of potential cases [24]. Some of the reported benefits of using Facebook to recruit participants include "reduced costs, shorter recruitment periods, better representation, and improved participant selection in young and hard to reach demographics" [25].

We received 148 completed surveys; 75 respondents indicated that they had experienced both positive and negative consequences of using OCHI and agreed to be contacted. These respondents were emailed a consent form by order of response (five per day) until data saturation was reached.

Data Collection

To accommodate the participants' time constraints and geographic dispersion, we used semistructured telephone interviews. Telephone interviews encourage a more explicit description of the participant's emotional experiences because of the absence of visual cues [26]. We developed an interview guide based on our conceptual framework, pilot-tested it for clarity, and modified it accordingly (Multimedia Appendix 2). Participants were asked to elaborate on their negative experience with OCHI use by telling their *story* and what factors they believe may have led to these outcomes. They were also asked what they think could have been done to prevent these outcomes. The 10-20 min interviews were recorded and conducted from November 2015 to February 2016. Interviews were transcribed verbatim (using pseudonyms to protect identities), and the transcripts were imported into the NVivo 11 software package (QSR International Pty Ltd, Victoria, Australia) for qualitative data analysis. Interviewing continued until data saturation was reached after 19 participants [27].

Data Analysis

A deductive-inductive analytical approach was adopted for coding [28]. We created a coding manual of the types of use and types of outcomes, both positive and negative, and the codes were discussed with the team until a consensus was reached. The codes were then progressively clustered into major themes. We also performed a secondary analysis of the interview transcripts using a story telling technique [29]. This involved viewing the interview transcript through *multiple lenses* and developing interpretive stories based on those lenses [30]. These stories, or vignettes, were created to represent each of the different types of negative outcomes identified in the literature review and the interviews. In stage 2 of this study, these stories were used to introduce the topic to the health practitioners and health librarians.

Stage 2—Health Professionals

Sampling

In stage 2, we selected seven health practitioners (three family physicians, two registered nurses, and two community pharmacists) and three health librarians because they are considered a primary source of health information for their patients [31]. These participants are considered key informants on the measures that could be taken to prevent negative consumer outcomes [27].

We used a purposeful sample of these key informants in Montreal and Ottawa. Using personal contacts in the Department of Family Medicine and the School of Information Studies at McGill, the Herzl Family Medicine Clinic, the Jewish General Hospital, and the Canadian Pharmacists Association of Canada, we invited primary care practitioners and health librarians to participate by email.

Data Collection

We conducted in-person semistructured interviews, 45 and 60 min in length, from March 2016 to April 2016. All the interviews were conducted at the participant's office (either in a clinic or hospital). We developed the interview guide based on our conceptual framework, pilot-tested it for clarity, and modified it accordingly (Multimedia Appendix 2). At the beginning of the interview, three vignettes created from the stage 1 interviews were presented to the stage 2 participants (Textbox 1).

Textbox 1. Vignettes presented at the start of the interview with health care practitioners and health librarians.

Vignette 1

A young 22-year-old woman who identifies herself as a "bit of a hypochondriac" usually goes online to look for health information when she has multiple symptoms and she is unsure if they are related or not. She uses information to decide if she needs to see a doctor or not.

On one occasion, she had pain near her ribs and pain with "breathing." After checking online, she found "scary" diagnoses of similar symptoms and decided to go to the emergency room. After waiting there for a few hours, she was told it was nothing and went home. This happened a few times.

She feels that if there was more specific information online, or lists of the "most common diagnoses" for each symptom, she would worry less about her online findings.

Vignette 2

Mark is a 32-year-old man who usually looks for health information online when he has a new unfamiliar symptom. He uses information to decide whether (or not) he needs to see a doctor and to find possible explanations for his symptoms. The information helps him reflect on his lifestyle and determine if there are any changes he needs to make.

On one occasion, after suffering from abdominal pain for months, his family physician requested an ultrasound. During the ultrasound, and after some probing from Mark, the ultrasound technician suggested it may be a polycystic kidney causing his pain. As his follow-up appointment with the specialist was weeks away, Mark decided to do as much research on the topic as possible, which led to increasing anxiety over this diagnosis. Eventually at his appointment, the specialist diagnosed him with a failed kidney and not polycystic kidneys, which while severe, was still a relief for Mark.

Although in this case Mark blames the technician for making an unfounded diagnosis, he feels doing so much research on the topic allowed him to have an educated discussion with the specialist during his appointment.

Vignette 3

Sarah is a 26-year-old woman who was diagnosed with epilepsy. Her doctor prescribed Depakene as the best treatment, and she started using it. Then she started getting side effects from this medication.

After looking online for health information on this issue, she found that there were complementary and alternative treatments for epilepsy such as reiki and yoga, as well as herbal remedies, dietary supplements, and homeopathic treatments. She read testimonies by other people who had used these alternatives other than the Depakene and decided to perhaps try following them instead.

She feels that the information she found was biased, "you find what you look for." She admits that while looking for information, she might have valued a lot less information that said she needed to take the Depakene and valued more heavily information that said epilepsy was potentially manageable alternative and homeopathic remedies.

The vignettes offered the practitioners and librarians with an opportunity to reflect on the negative outcomes resulting from OCHI use and refreshed their memories of their own experiences. They were then asked to give their opinion on the situations described in the vignettes, as well as elaborate on their own experience with consumers or patients using OCHI. Finally, practitioners were asked what strategies they used to prevent or alleviate negative outcomes in their practice, whereas health librarians were asked what strategies they believe should be used to prevent negative outcomes. The interviews were conducted, transcribed, and analyzed two at a time [27]. No new codes emerged during the analysis of the 9th and 10th interviews, so data collection stopped.

Data Analysis

Using NVivo 11, we coded the transcripts for types of negative outcomes and preventive measures. Similar to the first stage, a deductive-inductive analytical approach was adopted [27]. We created a coding manual of the types negative outcomes and preventive measures, and the codes were discussed with the team until a consensus was reached. The codes were then progressively clustered into major themes.

Results

Overview of the Results

A description of the participants of both stages is presented in Tables 1 and 2. In both stages, many positive outcomes of OCHI were found: OCHI allows consumers to be informed and involved in their health care. According to participants, consumers have different strategies for finding and assessing OCHI and that factors such as health literacy and access to an HCP influence the type of outcome that occurs. Participants of both stages described similar negative outcomes such as increased worrying and postponing a health care visit. Preventive strategies mentioned in both stages included providing reliable sources of OCHI, teaching consumers to properly assess OCHI found, and discussing OCHI with someone else.

How Consumers Find, Understand, and Use Online Consumer Health Information

Consumers described their motivations and strategies for searching for OCHI, their understanding of the information they found, and how they used it. The identified themes and subthemes are described in Table 3. In summary, participants would most commonly search for information for themselves or for someone else by entering their symptoms into a search engine ("Googling their symptoms"). They had different ways of deciding the credibility of a website, and for the most part, just wanted more information on an issue, though some participants used OCHI to decide whether or not to book an appointment with a HCP or stop a medication, as illustrated in the following quote:

> I've looked up stuff like side effects of birth control pills if I'm worried or more emotional, I'll see if that is one. I've actually gone off [pills] because of that. [Jenny]

XSL·FO

RenderX

Table 1. Participants in stage 1.

Pseudonym	Gender	Age group (years)
Alan	Male	18-24
Betty	Female	18-24
Cara	Female	18-24
Dina	Female	18-24
Ella	Female	18-24
Fred	Male	18-24
Gina	Female	18-24
Harry	Male	25-34
Isabel	Female	45-54
Jenny	Female	18-24
Karen	Female	18-24
Lara	Female	25-34
Mariah	Female	18-24
Nathan	Male	18-24
Pamela	Female	18-24
Rita	Female	18-24
Sarah	Female	18-24
Tamara	Female	18-24
Vanessa	Female	25-34

Table 2. Participants in stage 2.

Alias	Profession	Work environment
Pharmacist #1	Clinical pharmacist	A family medicine clinic attached to a teaching hospital in Montreal.
Pharmacist #2	Community pharmacist	A community pharmacy and a family medicine clinic in Ottawa.
Doctor #1	Family physician	An academic hospital and a walk-in clinic in Montreal.
Doctor #2	Family physician	An academic hospital and a walk-in clinic in Montreal.
Doctor #3	Family physician	A family medicine clinic in Ottawa.
Librarian #1	Health librarian	A hospital health sciences library in Montreal.
Librarian #2	Health librarian	A hospital health sciences library in Montreal.
Librarian #3	Health librarian	A children's hospital health sciences library in Montreal.
Nurse #1	Nurse	An academic hospital in Montreal.
Nurse #2	Nurse practitioner	A family medicine clinic affiliated with an academic hospital in Montreal.

Table 3. Finding, understanding, and using online consumer health information (OCHI).

Theme and subtheme	Example quote
1. Motivation for searching for health information online	
1.1 Searching for information for themselves	"When it is something that I cannot explain, like I have multiple symptoms and I don't know if all these symptoms are related, unrelated..." [Alan]
1.2 Hypochondria	"I'm a little hypochondriac, I mean literally last night I was feeling nauseous, so I started to Google, so anytime I'm feeling an odd symptom...anytime I feel something is abnormal and I'll look up those symptoms to see if I have anything, from nausea and headaches to weird circulatory feelings." [Rita]
1.3 Searching for information for someone else	"Last time I looked up stuff online was for my grandfather. He suffered from Parkinson's...and we were looking for alternatives." [Nathan]
2. Strategies for searching for information online	
2.1 Using a search engine (Google)	"I usually Google either my symptoms if I don't know what it is, or if I have an idea of what it might be then I'll Google that." [Betty]
2.2 Using a renowned medical website	"I just Google but the ones I usually end up in are WebMD or mayo clinic, I think if you Google something those are the first ones that show up anyway." [Rita]
2.3 Using websites or forums with patient experiences	"...there are a lot of useful forums where experienced marathoners have training advice, stuff like that. When to do icing or heat, which one is better than the other." [Ella]
2.4 Strategies for evaluating OCHI websites	"I usually avoid sites that are trying to sell you stuff or that anyone can edit." [Tamara]
3. Making sense of the information	
3.1 Understanding the information found	"I understand it, I might have to do further research for specific terms, but overall I understand what they're saying." [Alan]
3.2 Gaining general knowledge without answering a specific question	"Sometimes you don't know what is wrong or right and each case is different as well, so you have an idea globally, but you don't really have the answer I guess." [Mariah]
3.3 Not finding the answer to a specific health question	"No. I would have a symptom and it would usually end with me convincing myself that I had some sort of terminal illness." [Cara]
3.4 How health literacy influences understanding	"No, I can usually understand it. I feel like I may be more science and health literate than a lot of people since I have a Bachelor's degree in Science." [Betty]
4. Decision making after finding relevant OCHI	
4.1. Deciding whether or not to book a medical appointment	"I wouldn't say immediately but when I have a recurring kind of problem, so I'll look at it probably before calling the doctor and making an appointment." [Isabel]
4.2 Postponing a medical appointment because of limited access	"It's not even that, it's that you have to wait so long now to get an appointment that if I can home remedy it that's how I sort of look at it." [Isabel]
4.3 Making a health care decision	"Usually if it's something like I can change what I'm eating, I follow if it doesn't seem to extreme or too hard to do. If it's something that seems a bit ridiculous then..." [Tamara]
4.4 Stopping a medication	"I've looked up stuff like side effects of birth control pills if I'm worried or more emotional, I'll see if that is one. I've actually gone off [pills] because of that." [Jenny]
4.5 Discussion in a physician encounter	"Some things I'll bring up when seeing my physician and get their advice on it." [Dina]
4.6 To confirm a physician's diagnosis	"Yes, I have symptoms and look them up and if I find what I think it is I go to the doctor and I'll let the doctor suggest on their own but I'll kind of suggest that this what I think it could be, could you confirm that for me or not?" [Sarah]

XSL•FO
RenderX

Health Care Practitioners' and Health Librarians' Experience With Online Consumer Health Information

Stage 2 participants described their general opinion of OCHI, the types of OCHI they had encountered, and factors they believed influenced the outcomes of OCHI use. The identified themes and subthemes are described in Table 4.

All the participants had seen their patients, their clients, and their friends and family use OCHI and had used OCHI themselves in their own health care. They believed it was a permanent fixture and an inevitable presence in the health care system, as illustrated in the following quote:

> *For patients, "I saw this online" is the new "my friend told me," which I still see a lot with the elderly, although even the elderly are going online.* [Pharmacist #2]

Participants had experiences with different formats and topics of OCHI depending on their profession, the location of their practice, and the types of patients or consumers they saw. Physicians reported dealing with diverse topics (eg, medications and their side effects, medical conditions, and diagnostic tests) from various sources (reviewed online medical resources, patient forums, blogs, etc). Nurses, on the other hand, are traditionally more involved in patient education, and therefore tend to be more exposed to OCHI on general health information rather than specific health conditions. Pharmacists were more exposed

to OCHI covering medications and their side effects and herbal products or supplements. Health librarians are traditionally health information providers, so are not necessarily exposed to patients presenting them with OCHI. They were, however, very familiar with the different sources of OCHI available, specifically patient forums, and were aware of its pros and cons.

A recurrent theme during the interviews was the OCHI related to alternative and complementary medical treatments and therapies. Participants mentioned some specific examples where patients had read of an unconventional treatment for their condition online and wanted to find out if it was a viable alternative. Another theme that was brought up during the interviews was the antivaccine movement, and participants mentioned interactions with people about vaccinations for which there is conflicting OCHI.

Negative Outcomes of Online Consumer Health Information

We identified three negative outcomes of OCHI based on participants' use of it (Table 5). First, increased worrying as a result of finding "scary" or worse-case-scenario information that might or might not be relevant to their symptoms. Second, tension in the relationship with a family member because the latter's use of potentially harmful OCHI. Finally, postponing seeking medical help for a health problem, or to ignore their health problem altogether.

Table 4. Health care practitioners' and health librarians' experience with online consumer health information (OCHI).

Theme and subtheme	Example quote
1. General opinion of participants on OCHI	"And for patients, 'I saw this online' is the new 'my friend told me,' which I still see a lot with the elderly, although even the elderly are going online." [Pharmacist #2]
2. Types of OCHI	
2.1 General health information	"I once saw a patient who had a dry cough and nothing else and came into an appointment because her friend had posted on Facebook that she had pneumonia." [Doctor #3]
2.2 Forums and patient-sourced information	"There are a ton of forums online people talking about their personal experiences...you don't get that from your health professional, they don't know what it's like to live with a condition. So, it can be very helpful to see other people's experiences and it may give your ideas for alternative treatments." [Librarian #1]
2.3 Alternative medicine information	"I had a patient who was relatively healthy but had high blood pressure that he treated with valerian root he had read about online...and so I looked it up and there was no real evidence for its effect on blood pressure." [Nurse #1]
2.4 Antivaccination information	"It's extremely frustrating because a lot of this antivaccine stuff is focused on really small risks and you have to acknowledge there might be risks and people tend to fixate on them, like there is mercury in vaccines, yes but there's mercury in food. So, it can be extremely time consuming to combat that. I think that topic is the biggest and most harmful." [Pharmacist #2]
3. Factors influencing outcomes	
3.1 Individual characteristics	"Definitely low health literacy but there also really well-educated people who don't have a health background and can be quite susceptible to the alternative medicine stuff. In another clinic where I work we see a lot of new immigrants, a lot of them Arabic speaking, I can't work with them as much." [Pharmacist #2]
3.2 Information avoidant personalities	"I think if you know your patient and kind of know they're the type who would basically somaticize every side effect you're not going to go over them in as much detail, you will sort of down play them." [Nurse #1]
3.3 Access to health care services	"Just that in Montreal, I don't know exact numbers, but around 30-40% of people don't have a family doctor, and the more vulnerable you are the more your access to good medical care decreases so I think that yes there are flaws to internet usage to access health care, but in a system where person-to-person health care is not good or easy to access, it may be the only resource that many people have available to them." [Doctor #2]

XSL•FO

RenderX

Table 5. Negative outcomes of online consumer health information (OCHI).

Stage and subtheme	Example quote
Stage 1: consumers	
Increased worrying	"Sometimes it is anxiety inducing. If you can't find something that's a good match for what symptoms you're having or if you find something that is a good match that isn't so pleasant." [Betty]
Tension with family members	"One of my aunts takes online health info way too far, and everything online, she follows, it doesn't matter where it's from which is horrible because the internet has all sort of things...for 5 years my cousin's life had all the random health natural remedies online, never doctors, it was so bad. It was disturbing when we found that when he would have an infection she wouldn't take him to a doctor but make him drink honey...This situation caused stress between family members worried about the information she used." [Vanessa]
Postponing (not seeking help for) a health problem	"All my symptoms match a virus going around I read about it and I thought it will, blow over in a week, I don't need to miss class to go to a doctor's appointment and then I ended up going to the doctor and it ended up being an ear infection and a sinus infection and it turned into 2 months of being miserable..." [Ella]
Stage 2: practitioners and librarians	
Increased worrying	"Yeah, I think so, she was worried, she took time off work to come in to see me, and she waited in the waiting room for a while. So, I have to take her worry seriously. This applies to many patients I see, where there are no actually worrisome symptoms, if they had waited a few days whatever they had would have gone away on its own. But they had read something online either after Googling their symptoms or after accidentally stumbling on a piece of online information through social media for example, and they worry they might have that." [Doctor #3]
Spending money on nonbeneficial products	"A lot of the herbal and complementary and alternative therapy stuff, the biggest harm to a lot of people is that it costs money and might not work... BP: I think the main consequence is that they can't afford, it's common for people who are poor to have poor literacy so will believe all this stuff they read online or Dr. Oz, so they end up spending money that they shouldn't be spending." [Pharmacist #2]
Tension in the provider-patient relationship	"I want them to know that I'm aware of it, that I'm not ignorant, because a lot of time this OCHI can undermine their trust in your ability and your competence and they will say why didn't you tell me about this? And sometimes the reason we haven't told is because we think that it will just scare them which is true, and we do." [Nurse #1]
Nonadherence to management plan	"I think one of the biggest ones, the area I've had most problems with is mental health, it's a huge issue and affects a particular anxiety, a patient who is going through a lot of problems unfortunately the internet and their ability to get information is a major block to being treated. They would look up the side effects of the medications because they are more suggestible, experience every side effect of the medication and eventually stop it." [Nurse #1]
Postponing seeking medical help	"We had a gentleman come in here [health library] and he was looking for information, and he started discussing what was wrong with him and saying he felt numbness in his leg and I said immediately let me get you a wheelchair and transport you to the emergency room. He was asking me for info about something that I clearly couldn't solve, and part of my job is identifying when someone comes to me and saying you should go see a doctor or go to the emergency room." [Librarian #2]

Similar to the first stage, increased worrying was a negative outcome found in stage 2 (Table 5). This could stem from reading reliable but nonrelevant information, from reading too much information, or from finding incorrect information from unreliable websites. It was advice on these latter websites that led to the second negative outcome: the purchase of useless or potentially harmful medications online. A third negative outcome is the breakdown of trust in the patient-clinician relationship from, for example, clinicians not validating patients' information-seeking efforts. Moreover, finding information that contradicted that provided by the health practitioner also led to a breakdown in trust and lowered adherence to the management plan.

Tension: A Comprehensive and Meaningful Construct

On further examination of these outcomes, they appeared to fall under one main theme of *tensions*, with three dimensions, depending on who or what was being affected by OCHI use: internal, interpersonal, and service-related (summarized in Table 6).

Strategies for Reducing Online Consumer Health Information Negative Outcomes

Stage1 participants identified strategies that they used or believed would be helpful, such as managing expectations when searching for health information online, using reliable OCHI sources provided or reviewed by health practitioners, and discussing the information found with a health practitioner to validate its reliability and relevance to their health question. These strategies are presented in Table 7.

Table 6. Online consumer health information (OCHI) tensions.

Levels	Stage 1: consumers	Stage 2: practitioners and librarians
Internal tensions	Increased worrying	Increased worrying; Spending money on nonbeneficial products
Interpersonal tensions	Tension with family members	Tension in the provider-patient relationship
Service-related tensions	Ignoring (not seeking help) for a health problem	Nonadherence to management plan; Postponing seeking medical help

Table 7. Strategies for reducing negative outcomes of online consumer health information (OCHI).

Stage and subthemes	Example quote
Stage 1: consumers	
Be aware of limitations of OCHI	"You have to be careful, when you do a Google search you get a ton of stuff there and sometimes rewording your search you get different things, so you want to be reading the same thing and not doing something that could do more damage than good." [Isabel]
Reliable and relevant sources of OCHI	"I think there are already doctors online, but I don't know maybe something more precise because Web MD can be precise but it's not that precise, like you can have just normal back pain and it will direct you to kidney failure." [Lara]
Follow physician-provided OCHI or search parameters	"...but also, if there is a 2-week delay between getting a scan and seeing the professional about something that is serious, you should be provided with, you know, here are search parameters that you should look up that are neutral and that might give you content to reflect on so that you have an informed discussion with your doctor…" [Harry]
Discuss OCHI with physician, telehealth, or members of social network	"I keep in mind that it's on the internet, so if I was really stressed I would go talk to a real person. I am skeptical of the information so if I was worried I would go talk to a doctor." [Jenny]
Stage 2: practitioners and librarians	
Provide reliable sources of OCHI	"Look it's there, so instead of resisting it, let's provide high quality alternatives so we have a little more control." [Librarian #1]
Teach people how to evaluate OCHI	"I think that if more health care providers used the approach of showing people where they look for info and pointing out potential issues with their sources and that is very effective, but it is time consuming." [Pharmacist #2]
Discuss OCHI during a clinical encounter	"I've mostly had a more positive experience just by being open and discussing it." [Nurse #1]
Nonadherence to management plan	"I think one of the biggest ones, the area I've had most problems with is mental health, it's a huge issue and affects a particular anxiety, a patient who is going through a lot of problems unfortunately the internet and their ability to get information is a major block to being treated. They would look up the side effects of the medications because they are more suggestible, experience every side effect of the medication and eventually stop it." [Nurse #1]
Handling OCHI on alternative treatments or nonconventional therapies	"For the third vignette, we certainly don't discourage exploration into complementary and alternative treatments, we have an excellent evidence based database we can search that have knowledge synthesis of the research that shows whether a given alternative treatment is actually effective, so we could have looked at maybe different remedies to show if there is any solid evidence and if it actually works and maybe there's definitely bias." [Librarian #1]

Table 8. Preventive strategies.

Strategies	Stage 1: consumers	Stage 2: practitioners and librarians
Before OCHI[a] search: providing reliable sources	Be aware of limitations of OCHI; Reliable and relevant sources of OCHI; Follow physician-provided OCHI or search parameters	Provide reliable sources of OCHI
During the search: teaching consumers how to evaluate OCHI sources	Using OCHI sources that are of good quality	Teach people how to evaluate OCHI sources
After finding relevant OCHI: discussing the information found	Discuss OCHI with physician, telehealth, or members of social network	Discuss OCHI during a clinical encounter; Handling OCHI on alternative treatments or nonconventional therapies

[a]OCHI: online consumer health information.

Similar strategies were also proposed by stage 2 participants to reduce the occurrence of negative outcomes. First, they recommended providing reliable OCHI sources. Second, teach consumers how to assess websites, either during an appointment or by referring them to online resources. Third, encourage consumers to discuss the OCHI with an HCP during a clinical

XSL•FO
RenderX

encounter to validate its reliability and relevance. Our findings suggest this may be important for OCHI about alternative and complementary therapies, where it is important to explain to consumers the difference between regulated and unregulated therapies, as well as how to assess the reliability of the information found. Finally, our results suggest that because of the frequency of OCHI use, HCPs should be trained to deal with patients who bring OCHI to a clinical visit. These strategies are presented in Table 7.

Participants also suggested discussing OCHI with health librarians as they are well situated to provide reliable OCHI sources, teach consumers how to evaluate websites, help consumers prepare the information to discuss during a clinical encounter, and find reviewed evidence on complementary and alternative therapies.

In summary, many preventive strategies were proposed by participants to reduce the occurrence of negative outcomes, as shown in Table 8. These include providing reliable OCHI sources before consumers start the search, teaching consumers to evaluate websites, and encouraging consumers to validate the information's reliability and relevance with an HCP.

Discussion

Principal Findings

Our findings confirm that OCHI is a part of daily routines in today's health care processes. It is a common, if not the most frequent, source of health information for consumers and is an integral part of the health care decision-making process. Congruent with existing evidence on OCHI, the outcomes of using OCHI are generally positive, especially when information sources are reliable. However, negative outcomes were reported consistently in terms of tensions across this study: the literature review, consumers in stage 1, and health practitioners and librarians in stage 2.

Online Consumer Health Information Use–Related Tensions

Elaborating on the Merriam-Webster dictionary definition of tension, the term OCHI tension refers to the feeling of uneasiness people who actively search for online health information experience with themselves, with other people, as well as vis-à-vis the health care services. Therefore, we argue that tension is a comprehensive and meaningful construct that represents a variety of negative outcomes along three dimensions. [33].

Level 1: Internal Tensions

These are outcomes that affect the consumer alone because of seeking and using OCHI and are associated with an emotional state. Internal outcomes uncovered in this study include increased worrying and anxiety. It has been suggested that "challenge and confusion, and dealing with the familiar and with the contradictory, are sources or triggers of emotional behavior in information situations" [33]. Moreover, with some consumers' lack of theoretical knowledge and ability to critically evaluate the information, this will inevitably lead to misinterpretation and unnecessary fear and anxiety [34].

Consumers who have even moderate levels of health anxiety are more likely to seek higher amounts of OCHI and spend more time online for health purposes overall [35,36]. One influencing factor is the individual's proneness to worry; one study examining the relationship between *anxiety sensitivity* and OCHI use reported that there was a relationship between exposure to OCHI and the etiology and maintenance of anxiety sensitivity [37].

Level 2: Interpersonal Tensions

Interpersonal tensions include any strain in the relationship between the OCHI consumer and other individuals, such as their HCP or a family member. An example of how this strain in the patient-clinician relationship occurs is when the practitioner does not acknowledge or validate the information brought in by their patient. A lack of trust developed when patients found information online that their HCP had not mentioned during the clinical encounter. This has also frequently been reported in the literature: patients who have read health information online may give less credence to their doctor's opinion and may use the information to test their doctor's knowledge, causing damage to the patient-clinician relationship [34,38]. On the other hand, some doctors lack the communication skills or are not up to date on all the information available and thus, report difficulties in dealing with OCHI.

Level 3: Service-Related Tensions

These tensions refer to any strain in the relationship between an OCHI consumer or patient and the health care system, leading to a change in the individual's use of health care services or adherence to management plans. This is in line with results reported in other studies. In one study, over 11% of the respondents reported that finding health information online led to them refusing or discontinuing treatment recommended by a physician or dentist [39]. Other studies also reported that participants (35% and 29.9%) would use the internet as a health information source instead of getting a professional opinion [40,41]. On the other hand, it was also reported that OCHI could lead to more frequent encounters with their HCP based on the information found [41].

On the basis of this construct of tension, we conceive the *OCHI use–related tensions* as presented in Figure 2. This enriches our original conceptual framework and adds to the scientific knowledge on the outcomes of OCHI use [15]. In the literature, there is an established link between health anxiety (internal tension) and the patient-clinician relationship (interpersonal tension). Health anxious people are more prone to wrong self-diagnosis and unnecessary worries, which is likely to increase the risk of misunderstanding and frustration with their doctor [35]. They may also feel that the duration of the clinical encounter was not enough to discuss all their worries and be less satisfied with the consultation [32]. Moreover, there are studies that report a link between low trust in the physician (interpersonal tension) and nonadherence or mal-adherence to a management plan (service-related tension) [42]. Our results suggest there is a relationship between the three dimensions of tension.

Figure 2. Online consumer health information (OCHI) use–based tensions.
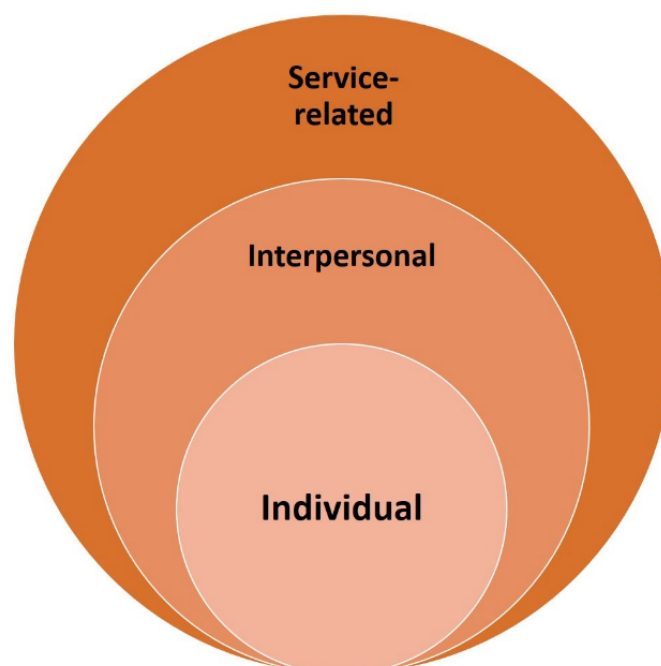
Potential Preventive Strategies

Several strategies targeting different periods of the iterative information-seeking process have been identified: before seeking the information online, while searching for information online, and after finding the information online. They can be summarized into three main preventive strategies as shown in Figure 3. During data analysis, it became clear that the health librarian participants in the second stage provided a distinct separate viewpoint and played a very different role from health practitioners.

Before Online Consumer Health Information Search: Providing Reliable Sources

Health practitioners preemptively provide patients with the names of reliable, reviewed websites during the clinical encounter rather than wait for patients to navigate on their own. This is in line with other studies that recommend that HCPs, specifically doctors, guide patients to reliable sources of OCHI [34,43-45]. Inevitability, people will try to search for health information online; however, they may not be adequately equipped to deal with the vast number of OCHI resources. In one study, even physicians expressed a need for training on how to navigate OCHI resources so that they are better able to recommend websites to their patients [46].

During the Search: Teaching Consumers How to Evaluate Online Consumer Health Information Sources

In the literature, it has been reported that evaluation interventions led to a more critical evaluation of online information [47]. This education process, however, is time-consuming and may not be a priority during the clinical encounter. Although practitioners could provide their patients with a list of criteria for reliable websites, there are also online resources available in the form of guidelines and checklists to follow while evaluating a website. However, many consumers, especially those in a lower socioeconomic strata, may not be aware of these resources or the fact that they are not correctly evaluating resources [48]. As suggested by participants in this study, an organizational effort is needed, for example, through mass media, in school curriculums, or in public libraries.

After Finding Relevant Online Consumer Health Information: Discussing the Information Found

The final strategy is discussing the information found with a health professional (eg, someone in their social network or a nurse phone line). This is supported by the literature; it was reported that patients simply need to have the information they found explained, contextualized, or validated by a health practitioner [34,40]. Studies report that discussing the information they had found with their physicians had a positive effect on the patient-clinician relationship, led to more involvement in decision making, and led to reduction of worries [32].

For health practitioners, there are ongoing initiatives to add OCHI into their continuing education (eg, workshops on dealing with their informed patients). For consumers, there are initiatives to encourage them to discuss information with their providers either through the help of a health librarian who can help organize the information and questions, or applications and websites that aid in that role (eg, the webpage Discutons Sante). There are, however, limitations to discussing OCHI; time limitations during the clinical encounter, and there may be a barrier related to practitioner attitudes. There may also be a barrier in understanding the health information (because of low health literacy or low education) or a limited social network.

Figure 3. Strategies to reduce online consumer health information (OCHI) tensions.
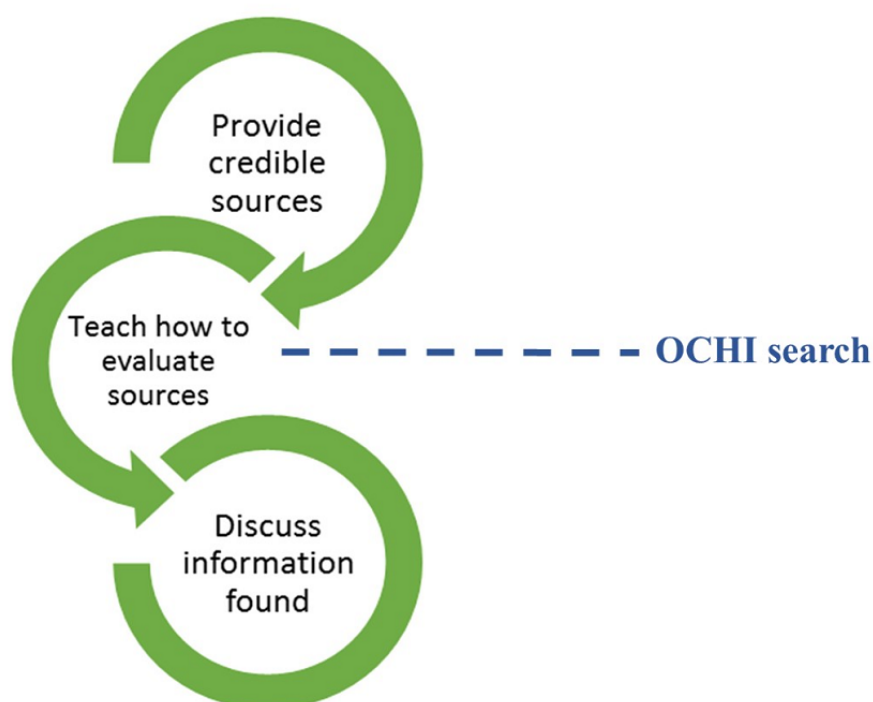
The Key Role of Librarians

Librarians are responsible for providing reliable health information and advocating the advantages of using OCHI for informed decision making. Working with consumers and health practitioners, they are well positioned to implement the preventive strategies described in this study. The integration of health librarians into the consumer's health information–seeking process may help ensure the reliability of the OCHI they find and use, as well as the appropriateness of its level of health literacy, leading to fewer internal tensions. Librarians' involvement may facilitate the discussion with health practitioners, leading to fewer interpersonal tensions. Finally, they can help consumers find situationally relevant OCHI, helping them to make more appropriate health care decisions and potentially leading to fewer service-related tensions.

Two barriers to the integration of health librarians into this information-seeking process should be noted: the lack of awareness of available health librarian services and the lack of access to health librarians by the public (because of their location inside hospitals). One potential solution would be to train community librarians working in public libraries on how to provide health information services and instruction, or at least train them to refer consumers to the local hospital-based health librarian.

Strengths and Limitations

Most respondents in stage 1 were females in the age range of 18 to 24 years. Although the lack of heterogeneity of our sample may present a limitation, studies report that the majority of individuals who search for and used OCHI are young women, which is reflected in our sample [40,49]. A recent systematic review on the use of Facebook in recruiting participants for health research purposes suggests that using social media to recruit participants may have led to this young female population sample [25]. A future study could use other recruitment tools to focus on an older population and to explore differences. One strength of our study is that our participants were all key informants, purposefully sampled for their experience and knowledge on the topic, as well as their willingness to incorporate OCHI in their practice [27,50]. No new ideas emerged in the final few interviews, and there was corroboration after triangulation of results of the review and all interviews.

Conclusions

The purpose of this investigation was to describe the negative outcomes associated with using online consumer health information, as well as to identify and reflect on any potentially preventive strategies.

This work makes two major contributions to the advancement of knowledge on OCHI. The first concerns a fine-grained identification of OCHI negative outcomes, which results from the construct OCHI use–related tension with three different and noninclusive levels of tensions (dimensions): individual, interpersonal, and service-related. This original construct enriches the original conceptual framework on outcomes of OCHI use and can serve as a foundation for future research. The second contribution, which involves clear practical implications, refers to the strategies primary care consumers, community, and health librarians and all types of primary care practitioners could adopt to prevent the risks associated with OCHI use. Exploration of these strategies and their implementation will be part of our future research.

Authors' Contributions

PP and CR were involved in all stages of the project and specifically in the data analysis. CT was a thesis committee member who provided feedback throughout the project. All authors participated in writing and revising the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey used as a recruitment tool.

[PNG File, 150KB - jmir_v20i5e169_app1.png]

Multimedia Appendix 2

Interview guides.

[PDF File (Adobe PDF File), 258KB - jmir_v20i5e169_app2.pdf]

References

1. HLWIKI Canada. 2015. Consumer health information URL: http://hlwiki.slais.ubc.ca/index.php?title=Consumer_health_information&oldid=142405 [accessed 2018-03-07] [WebCite Cache ID 6xkVoWs1D]
2. Fox S, Jones S. Pew Internet. 2011 May 12. The Social Life of Health Information, 2011 URL: http://pewinternet.org/Reports/2011/Social‐Life‐of‐Health‐Info.aspx [accessed 2018-03-07] [WebCite Cache ID 6xkQah0GD]
3. Hamm MP, Chisholm A, Shulhan J, Milne A, Scott SD, Given LM, et al. Social media use among patients and caregivers: a scoping review. Br Med J Open 2013;3(5):e002819 [FREE Full text] [doi: 10.1136/bmjopen-2013-002819] [Medline: 23667163]
4. 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 Jul;20(7):790-798. [doi: 10.1080/10810730.2015.1018590] [Medline: 26042588]
5. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: An analysis of health information seeking behavior among US adults. Cogent Soc Sci 2017 Mar 13;3(1):1302785. [doi: 10.1080/23311886.2017.1302785]
6. Amante DJ, Hogan TP, Pagoto SL, English TM, Lapane KL. Access to care and use of the Internet to search for health information: results from the US National Health Interview Survey. J Med Internet Res 2015;17(4):e106 [FREE Full text] [doi: 10.2196/jmir.4126] [Medline: 25925943]
7. Prescott J, Mackie L. "You sort of go down a rabbit hole...you're just going to keep on searching": a qualitative study of searching online for pregnancy-related information during pregnancy. J Med Internet Res 2017 Jun 05;19(6):e194 [FREE Full text] [doi: 10.2196/jmir.6302] [Medline: 28583906]
8. Case DO, Given LM. Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior: 4th Edition. UK: Emerald Group Publishing Limited; 2016.
9. Coulter A. Patient engagement--what works? J Ambul Care Manage 2012;35(2):80-89. [doi: 10.1097/JAC.0b013e318249e0fd] [Medline: 22415281]
10. Edwards M, Davies M, Edwards A. What are the external influences on information exchange and shared decision-making in healthcare consultations: a meta-synthesis of the literature. Patient Educ Couns 2009 Apr;75(1):37-52. [doi: 10.1016/j.pec.2008.09.025] [Medline: 19036550]
11. Renahy E, Chauvin P. Internet uses for health information seeking: a literature review. Rev Epidemiol Sante Publique 2006;54(3):263-275. [Medline: 16902386]
12. Pluye P, Grad R, Repchinsky C, Jovaisas B, Johnson-Lafleur J, Carrier ME, et al. Four levels of outcomes of information-seeking: A mixed methods study in primary health care. J Am Soc Inf Sci Technol 2012 Dec 10;64(1):108-125. [doi: 10.1002/asi.22793]
13. Berg K. University of Toronto Doctoral dissertations. 2011. Health Management in the Age of the Internet (Doctoral dissertation) URL: https://tspace.library.utoronto.ca/handle/1807/29663 [accessed 2018-04-03] [WebCite Cache ID 6yPPNHeQx]

14. Iverson SA, Howard KB, Penney BK. Impact of internet use on health-related behaviors and the patient-physician relationship: a survey-based study and review. J Am Osteopath Assoc 2008 Dec;108(12):699-711. [Medline: 19075034]

15. Granikov V, El Sherif R, Tang DL, Bujold M, Frati F, Bartlett G, et al. What are the outcomes associated with the use of online consumer health information in primary health care? A systematic review with framework synthesis of quantitative and qualitative evidence. 2016 May Presented at: Joint Meeting of the Medical Library Association (MLA), the Canadian Health Libraries Association/Association des bibliothèques de la santé du Canada (CHLA/ABSC), and the International Clinical Librarian Conference (ICLC), Mosaic '16. J Can Health Libr Assoc; 2016; Toronto.

16. Lewis T. Seeking health information on the internet: lifestyle choice or bad attack of cyberchondria? Media Cult Soc 2006 Jul 01;28(4):521-539. [doi: 10.1177/0163443706065027]

17. White RW, Horvitz E. Cyberchondria: studies of the escalation of medical concerns in web search. ACM Trans Inf Syst 2009 Nov 01;27(4):1-37. [doi: 10.1145/1629096.1629101]

18. Ahmad F, Hudak PL, Bercovitz K, Hollenberg E, Levinson W. Are physicians ready for patients with Internet-based health information? J Med Internet Res 2006;8(3):e22 [FREE Full text] [doi: 10.2196/jmir.8.3.e22] [Medline: 17032638]

19. Burton-Jeangros C, Hammer R. [Information seeking on the internet: what information are pregnant women seeking?]. Rev Med Suisse 2013;9(383):895-897. [Medline: 23697085]

20. Saracevic T, Kantor P. Studying the value of library and information services Part 1. Establishing a theoretical framework. J Am Soc Inf Sci Technol 1997;48(6):527-542. [doi: 10.1002/(SICI)1097-4571(199706)48:6<527::AID-ASI6>3.0.CO;2-W]

21. Pluye P, Granikov V, Bartlett G, Grad RM, Tang DL, Johnson-Lafleur J, et al. Development and content validation of the information assessment method for patients and consumers. JMIR Res Protoc 2014 Feb 18;3(1):e7 [FREE Full text] [doi: 10.2196/resprot.2908] [Medline: 24550180]

22. Schwartz-Shea P, Yanow D. Interpretive Research Design: Concepts and Processes. UK: Routledge; 2013:041587808.

23. 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]

24. Merolli M, Gray K, Martin-Sanchez F. Developing a framework to generate evidence of health outcomes from social media use in chronic disease management. Med 2 0 2013;2(2):e3 [FREE Full text] [doi: 10.2196/med20.2717] [Medline: 25075238]

25. 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]

26. Cachia M, Millward L. The telephone medium and semi‐structured interviews: a complementary fit. Qual Res Org Manag Int J 2011 Nov 15;6(3):265-277. [doi: 10.1108/17465641111188420]

27. Sandelowski M. Sample size in qualitative research. Res Nurs Health 1995 Apr;18(2):179-183. [Medline: 7899572]

28. 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 2016 Nov 29;5(1):80-92. [doi: 10.1177/160940690600500107]

29. Sandelowski M. Telling stories: narrative approaches in qualitative research. Image J Nurs Sch 1991;23(3):161-166. [Medline: 1916857]

30. McCormack C. From interview transcript to interpretive story: part 1—viewing the transcript through multiple lenses. Field Methods 2016 Jul 24;12(4):282-297. [doi: 10.1177/1525822X0001200402]

31. Medlock S, Eslami S, Askari M, Arts DL, Sent D, de Rooij SE, et al. Health information-seeking behavior of seniors who use the Internet: a survey. J Med Internet Res 2015;17(1):e10 [FREE Full text] [doi: 10.2196/jmir.3749] [Medline: 25574815]

32. Tanis M, Hartmann T, Te Poel F. Online health anxiety and consultation satisfaction: a quantitative exploratory study on their relations. Patient Educ Couns 2016 Dec;99(7):1227-1232. [doi: 10.1016/j.pec.2016.01.021] [Medline: 26873545]

33. Nahl D, Bilal D. Information and Emotion: The Emergent Affective Paradigm in Information Behavior Research and Theory. New Jersey: Information Today, Inc; 2007:1573873101.

34. Agius M, Stangeland H. How Does the Internet Influence the Doctor–Patient Relationship? In: e-Mental Health. Cham: Springer; 2016.

35. Singh K, Brown RJ. Health-related internet habits and health anxiety in university students. Anxiety Stress Coping 2014;27(5):542-554. [doi: 10.1080/10615806.2014.888061] [Medline: 24467278]

36. Eastin MS, Guinsler NM. Worried and wired: effects of health anxiety on information-seeking and health care utilization behaviors. Cyberpsychol Behav 2006 Aug;9(4):494-498. [doi: 10.1089/cpb.2006.9.494] [Medline: 16901253]

37. Norr AM, Capron DW, Schmidt NB. Medical information seeking: impact on risk for anxiety psychopathology. J Behav Ther Exp Psychiatry 2014 Sep;45(3):402-407. [doi: 10.1016/j.jbtep.2014.04.003] [Medline: 24818986]

38. Hewitt-Taylor J, Bond CS. What e-patients want from the doctor-patient relationship: content analysis of posts on discussion boards. J Med Internet Res 2012;14(6):e155 [FREE Full text] [doi: 10.2196/jmir.2068] [Medline: 23137788]

39. Weaver JB, Thompson NJ, Weaver SS, Hopkins GL. Healthcare non-adherence decisions and internet health information. Comput Human Behav 2009 Nov;25(6):1373-1380. [doi: 10.1016/j.chb.2009.05.011]

40. Fox S, Duggan M. Pew Internet. Health online 2013 URL: http://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2018-03-07] [WebCite Cache ID 6xkQQbJPP]

41.   Beck F, Richard JB, Nguyen-Thanh V, Montagni I, Parizot I, Renahy E. Use of the internet as a health information resource among French young adults: results from a nationally representative survey. J Med Internet Res 2014;16(5):e128 [FREE Full text] [doi: 10.2196/jmir.2934] [Medline: 24824164]

42.   Wroth TH, Pathman DE. Primary medication adherence in a rural population: the role of the patient-physician relationship and satisfaction with care. J Am Board Fam Med 2006;19(5):478-486 [FREE Full text] [Medline: 16951297]

43.   Silver MP. Patient perspectives on online health information and communication with doctors: a qualitative study of patients 50 years old and over. J Med Internet Res 2015;17(1):e19 [FREE Full text] [doi: 10.2196/jmir.3588] [Medline: 25586865]

44.   Iacovetto MC, Allen LA. Transitioning into a digital world: time for providers to recommend Internet health information? Am Heart J 2015 Jul;170(1):1-2. [doi: 10.1016/j.ahj.2015.04.002] [Medline: 26093858]

45.   Pluye P, Grad R, Barlow J. Look It Up!: What Patients, Doctors, Nurses, and Pharmacists Need to Know about the Internet and Primary Health Care. Montreal: McGill-Queen's University Press; 2017.

46.   van Uden-Kraan CF, Drossaert CH, Taal E, Smit WM, Seydel ER, van de Laar MA. Experiences and attitudes of Dutch rheumatologists and oncologists with regard to their patients' health-related Internet use. Clin Rheumatol 2010 Nov;29(11):1229-1236 [FREE Full text] [doi: 10.1007/s10067-010-1435-1] [Medline: 20383731]

47.   Walraven A, Brand-Gruwel S, Boshuizen HP. Fostering transfer of websearchers' evaluation skills: a field test of two transfer theories. Comput Human Behav 2010 Jul;26(4):716-728. [doi: 10.1016/j.chb.2010.01.008]

48.   Diviani N, van den Putte DB, Giani S, van Weert JC. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]

49.   Bianco A, Zucco R, Nobile CG, Pileggi C, Pavia M. Parents seeking health-related information on the Internet: cross-sectional study. J Med Internet Res 2013;15(9):e204 [FREE Full text] [doi: 10.2196/jmir.2752] [Medline: 24047937]

50.   Crouch M, McKenzie H. The logic of small samples in interview-based qualitative research. Soc Sci Inf 2006 Dec 01;45(4):483-499. [doi: 10.1177/0539018406069584]

Abbreviations

HCP: health care provider
OCHI: online consumer health information

XSL•FO

RenderX

Original Paper

Electronic Health Use in the European Union and the Effect of Multimorbidity: Cross-Sectional Survey

Francisco Lupiáñez-Villanueva[1*], PhD; Dimitra Anastasiadou[1*], PhD; Cristiano Codagnone[1,2*], PhD; Roberto Nuño-Solinís[3*], PhD; Maria Begona Garcia-Zapirain Soto[4,5*], PhD

[1]Estudís de Ciencies de la Informació i la Comunicació, Universitat Oberta de Catalunya, Barcelona, Spain
[2]Università degli Studi di Milano, Milano, Italy
[3]Deusto Business School Health, University of Deusto, Bilbao, Spain
[4]eVIDA Research Group, Facultad de Ingeniería, University of Deusto, Bilbao, Spain
[5]Facultad de Ingeniería, Universidad de Deusto, Bilbao, Spain
[*]all authors contributed equally

Corresponding Author:
Maria Begona Garcia-Zapirain Soto, PhD
Facultad de Ingeniería
Universidad de Deusto
Facultad de Ingeniería, 4ª Planta
Avda. Universidades, 24
Bilbao, 48007
Spain
Phone: 34 619 96 72 23
Email: mbgarciazapi@deusto.es

Abstract

Background: Multimorbidity is becoming increasingly common and is a leading challenge currently faced by societies with aging populations. The presence of multimorbidity requires patients to coordinate, understand, and use the information obtained from different health care professionals, while simultaneously striving to distinguish the symptoms of different diseases and self-manage their sometimes conflicting health problems. Electronic health (eHealth) tools provide a means to disseminate health information and education for both patients and health professionals and hold promise for more efficient and cost-effective care processes.

Objective: The aim of this study was to analyze the use of eHealth tools, taking into account the citizens' sociodemographic and clinical characteristics, and above all, the presence of multimorbidity.

Methods: Cross-sectional and exploratory research was conducted using online survey data from July 2011 to August 2011. Participants included a total of 14,000 citizens from 14 European countries aged 16 to 74 years, who had used an eHealth tool in the past 3 months. The variables studied were sociodemographic variables of the participants, the questionnaire items assessing the frequency of using eHealth tools, the degree of morbidity, and the eHealth adoption gradient. Chi-square tests were conducted to examine the relationship between the sociodemographic and clinical variables of participants and the group the participants were assigned to according to their frequency of eHealth use (eHealth user group). A one-way analysis of variance (ANOVA) allowed for assessing the differences in the eHealth adoption gradient average between different groups of individuals according to their morbidity level. A two-way between-groups ANOVA was performed to explore the effects of multimorbidity and age group on the eHealth adoption gradient.

Results: According to the eHealth adoption gradient, most participants (68.15%, 9541/14,000) were labeled as rare users, with the majority of them (55.1%, 508/921) being in the age range of 25 to 54 years, with upper secondary education (50.3%, 464/921), currently employed (49.3%, 454/921), and living in medium-sized cities (40.7%, 375/921). Results of the one-way ANOVA showed that the number of health problems significantly affected the use of eHealth tools ($F_{2,13996}=11.584$; $P<.001$). The two-way ANOVA demonstrated that there was a statistically significant interaction between the effects of age and number of health problems on the eHealth adoption gradient ($F_{4,11991}=7.936$; $P<.001$).

Conclusions: The eHealth adoption gradient has proven to be a reliable way to measure different aspects of eHealth use. Multimorbidity is associated with a more intense use of eHealth, with younger Internet users using new technologies for health

purposes more frequently than older groups with the same level of morbidity. These findings suggest the need to consider different strategies aimed at making eHealth tools more sensitive to the characteristics of older populations to reduce digital disadvantages.

Introduction

Multimorbidity

Multimorbidity is a growing phenomenon in ageing societies that is especially prevalent among older age groups [1-3]. Although no definition has been uniformly applied across studies, there seems to be a clear association between multimorbidity and lower quality of life (QoL), functional decline, worse health care outcomes (including adverse drug effects, rehospitalization, and mortality), increased use of health care resources, and higher health care costs [3-5]. From a clinical perspective, multimorbidity compounds the complexity of managing chronic diseases. Clinicians, in particular, are faced with a lack of a common theoretical background and guidance on the best care pathways for patients with multimorbidity [6]. In fact, clinical guidelines rarely address comorbidities, and clinical trials often exclude older patients with comorbidities [7,8]. Patients with multimorbidity struggle to make sense of and use the information obtained from different health care professionals and to distinguish between the symptoms of different diseases [9]. Both in terms of service quality and the economic impact, the increasing prevalence of multimorbidity among ageing population represents one of the key health care challenges European countries are currently facing [9-12].

Electronic Health Tools for Patients With Multimorbidity

Multimorbidity requires, among other things, that both patients and professionals effectively coordinate, synthesize, and reconcile health information from multiple providers and about different conditions [9]. In this respect, electronic health (eHealth) tools have been recommended worldwide to support this type of patient [11,12], as they can improve patient engagement and health outcomes [13,14]. For example, at the European Union level, the European Commission launched a partnership initiative in 2012 (European Innovation Partnership on Active and Healthy Ageing) that aims to join forces between European partners in the development of solutions for active and healthy ageing. International chronic care models of reference also recommend the exploitation of the potential benefits of eHealth tools for the self-management of chronic conditions, emphasizing the need to link informed and actively engaged patients with proactive and prepared health care teams; while also shifting from disease-centered models to patient-centered ones [9,15-17]. In fact, the systematic review by Jong and colleagues [18] revealed improvements in both adherence and health outcomes among chronically ill patients who were using Internet-based self-care techniques, although in other studies the benefits of implementing these applications are still to be proven [19].

To fully understand the potential of eHealth tools for patients with multimorbidity, it is useful to explore how their current usage may or may not differ between Internet users with and without chronic conditions. Results from our scoping review of studies, which aimed to estimate the effect of multimorbidity on the use of eHealth tools, seemed to point to an increasing use of them among people with comorbidities. Patients with a higher comorbidity score were more likely to use the eHealth tools available to them for ordering prescription refills, scheduling appointments, and asking medical or prescription drug related questions [20], as well as for searching, gathering, and sharing online health information with their doctor [21,22]. After controlling for education, sex, and age, Wagner and colleagues [16] found that people with depression, or three or more chronic conditions, were more likely than other chronically ill individuals to use the Internet to search for health information. In addition, qualitative research by Zulman and colleagues [9] revealed the challenges faced by people with comorbidities, including the need to manage a high volume of information from multiple providers, coordinate and synthesize this information and sometimes serve as experts, activities that were found to be associated with emotional distress. Similarly, in a study on the adoption of a personal health record (PHR) [23], patients with multiple chronic conditions were found to have a higher likelihood (25% more likely) of adopting the PHR to track their health condition, compared with those without the selected comorbidities. They were also found to make more intensive use of it, in that they were more likely to use the secure messaging functionality than other patients. Finally, the recent policy brief by Barbabella and colleagues (2016) [11] synthesized the available evidence related to the implementation, benefits, and policies regarding the adoption of eHealth tools for people with multimorbidity in Europe. Authors concluded that eHealth tools for this specific population can significantly improve their health status and QoL through increased integration, personalization, quality, and accessibility of care.

Although the use of eHealth tools may represent an avenue for improving QoL and the capacity to manage the diseases among people with multiple chronic conditions, the concern of health inequality arises in relation to the "digital divide." If socioeconomic inequalities across different demographic groups (eg, ethnicity, socioeconomic status, age, and gender) determine access to the Internet and the skills to use it effectively, then these may compound health inequality [24]. With respect to this specific dimension, the available empirical evidence is limited and inconclusive. Contradictory findings were reported in studies with samples of veterans [25,26], in which poor health status and some sociodemographic variables, but not comorbidities, were associated with greater eHealth use. The "Pew Research Center" [27,28] reported that people with chronic conditions were significantly less likely than other adults to

have Internet access. However, once online, they may more regularly take advantage of the benefits of the health information offered.

Objective

Within the context of an online survey previously developed and administered in the framework of the Joint Research Centre/Institute for Prospective Technological Studies (JRC/IPTS) "Strategic Intelligence Monitor on Personal Health Systems" (SIMPHS) project in 14 European countries [29], a secondary analysis of the information collected in the study was conducted to assess the use of eHealth tools among the Internet users, with a specific focus on individuals that self-reported multimorbidity conditions. In light of previous research, our first hypothesis was that multimorbidity would lead to greater usage of eHealth tools as a result of more complicated health care needs, a higher likelihood of suffering from adverse events related to their conditions, and greater uncertainty about the effects of treatments on people with multiple conditions. As chronic diseases are often associated with sociodemographic characteristics (ie, advanced age, lower education, and income) and with lower Internet access [30,31], the second hypothesis tested was that when controlling for participants' age, the opposite trend would be expected: the presence of multiple conditions would be associated with lower usage of eHealth tools. We also expected a gender impact on use of new technologies for health purposes, with men more likely to utilize these tools than women.

Methods

Recruitment

This study was cross-sectional and exploratory and used internationally representative survey data from the JRC/SIMPHS project [29]. The target population was composed of citizens aged 16 to 74 years, who had used the Internet in the past 3 months. The survey was carried out from July 2011 to August 2011 and consisted of a total panel of 14,000 interviews from 14 European countries. The participating countries were as follows: Austria, Belgium, Germany, Denmark, Estonia, Finland, France, Italy, Netherlands, Sweden, Slovenia, Slovakia, Spain, and the United Kingdom. It is important to note that as the surveyed population comprised individuals who were already using the Internet, they may be more advanced users than individuals in the general population.

Survey Instruments and Definition of Variables

First, information concerning the participants' sociodemographic and health characteristics was collected. Second, the study also collected information regarding eHealth tools use by focusing on the responses that interviewees provided to the question beginning with, "Regarding health, wellness and the Internet, how often have you...?" Individuals were asked about 23 activities that were carried out using eHealth tools. Items were answered on a 5-point Likert scale ranging from "never" to "every day or almost every day." A previous pilot study based on 231 interviews demonstrated the high internal consistency of the 23 items, which had a Cronbach alpha value of .96 [29].

It is important to note that the word "Internet" was used instead of "eHealth tools" to facilitate better understanding of the questionnaire by the respondents. However, each one of the 23 items of the questionnaire reflected different eHealth tools, delivered or enhanced through the Internet (eg, "Looked for information about a physical illness or condition that you or someone you know has") and other related technologies such as mobile apps or wearable devices (eg, "...Used a health or wellness app on your mobile phone" and "...Used devices to transmit clinical information, received alarms, followed-up about your health anytime, anywhere"); online platforms (eg, "...Made, canceled, or changed an appointment with your family doctor, specialist or other health professionals online"); email (eg, "...Used email to communicate with a doctor's office"); or videoconference (eg, "...Made an online consultation through videoconference with your doctor or nurse").

Next, two variables were constructed for the purpose of this study; the first one represents the individual's degree of morbidity (and multimorbidity), and the second one is an eHealth adoption gradient that has been constructed as a composite indicator of the frequency of use of the 23 activities carried out via eHealth tools.

The measure of multimorbidity was based on self-reported data and defined according to the number of health problems reported by the individual. To explore the effect of multimorbidity on eHealth use, two cut-off points were established: having two of these health problems and having three or more health problems. Although definitions of multimorbidity used in the literature vary widely, disease counting is a very common strategy. Furthermore, the cut-off points followed in this study are commonly used and accepted in the literature. That is, the cut-off of two or more chronic conditions appears to be the norm in multimorbidity prevalence studies in the general population [3], whereas a cut-off point of three or more chronic conditions may also be used in primary care settings [1,32]. Regarding the number of chronic conditions in the list of reference, 13 conditions were included, which fulfils the recommendation for a minimum of 12 conditions proposed by Fortin and colleagues [1]. The number and type of conditions considered when using a disease-counting approach were the factors found to be more critical when comparing prevalence studies on multimorbidity, and the specific list of conditions was taken from a question on reasons for long-term treatment included in a Eurobarometer survey [33]. The specific conditions were as follows: (1) chronic anxiety or depression; (2) peptic ulcer (gastric or duodenal ulcer); (3) stroke, cerebral hemorrhage; (4) osteoporosis; (5) chronic bronchitis, emphysema; (6) migraine or frequent headaches; (7) cataract; (8) cancer; (9) long-standing troubles with your muscles, bones, and joints (rheumatism, arthritis); (10) hypertension (high blood pressure); (11) asthma; (12) an allergy; and (13) diabetes.

To measure the degree of adoption of eHealth tools by the Internet users, we created a composite indicator. As defined by the Organisation for Economic Co-operation and Development/JRC handbook (2008) [34], "a composite indicator is formed when individual indicators are compiled into a single index on the basis of an underlying model. The composite indicator should ideally measure multidimensional concepts

which cannot be captured by a single indicator." Therefore, the construction of this composite indicator aimed to encompass a wide range of eHealth tools following the eHealth global definition suggested by Eysenbach (2001) [35] and supported by Pagliari and colleagues (2005) [36]. We have titled this composite indicator an "eHealth adoption gradient" designed to capture the degree of adoption (ie, usage) of eHealth tools by Internet users and conceptualized as a continuous variable ranging from 0 "no adoption" (no usage) to 1 "frequent usage."

Statistical Analysis

The statistical software package SPSS 15.0 (IBM Corp) for Windows was used [37]. To identify the level of multimorbidity, two cut-off points were established: having two of these health problems and having three or more health problems.

For the construction of the composite indicator called "eHealth adoption gradient," we followed the subsequent steps: first, an analysis was performed on the frequencies of the 23 activities carried out using eHealth tools. Second, to confirm the overlapping content of these activities, the means and their significant correlations were checked. Then, a factor analysis was carried out for data reduction purposes, so as to summarize the information related to the 23 activities reported in the questionnaire asking how often they used the Internet for a list of health-related purposes into a smaller number of unobservable factors (latent variables or constructs). The parameters of these linear functions are referred to as factor loadings. An analysis of the correlation matrix (Kaiser-Meyer-Olkin and Bartlett's test of sphericity) was carried out to check that the correlation matrixes were factorable. Data reduction was carried out by principal components analysis using the varimax option to identify possible underlying dimensions. Third, a careful and transparent definition of weights was performed, squaring and normalizing the estimated factor loadings from the factor analysis. The squared factor loadings represent the proportion of the total unit variance of a base variable that is explained by a factor. The resulting score by subdimension can be aggregated into the summary indicator of the dimension according to its relative contribution to the explanation of the overall variance of all factors.

The eHealth adoption gradient (composite indicator) was then compared between different groups of individuals according to their level of morbidity (no health problem, one health problem, or two or more health problems) using a one-way analysis of variance (ANOVA). In addition, a two-way between-groups ANOVA was performed to explore the effects of multimorbidity group and age group (16-24, 25-54, and 55-74 years) on the eHealth adoption gradient, followed by post hoc Tukey's honest significant difference (HSD) tests. Gender was also included as a covariate in an analysis of covariance (ANCOVA) of comparisons between age group and morbidity levels. Finally, a two-way between-groups ANOVA was performed to explore the effects of age group and the receival (or not) long-term medical treatment group on the eHealth adoption gradient, followed by post hoc Tukey's HSD tests. All statistical tests were conducted with a criterion for significance at $P<.05$. All bivariate and multivariate analyses were weighted.

Results

Frequency and Type of Electronic Health Use

The list of items and the frequency of each response are shown in Table 1. With regard to the activities carried out using eHealth tools, 81.60% of the surveyed individuals (11,424/14,000) used the Internet at least once a month to look for information about a physical illness or condition, and 75.30% (10,542/14,000) of the individuals used eHealth tools for wellness or lifestyle. Nevertheless, in other eHealth tools such as mobile apps or wearable devices (eg, "...Used a health or wellness app on your mobile phone" and "...Used devices to transmit clinical information, received alarms, followed-up about your health anytime, anywhere"), online platforms (eg, "...Made, canceled, or changed an appointment with your family doctor, specialist or other health professionals online"), or emails (eg, "...Used email to communicate with a doctor's office"), we observed less frequent use, with the percentage of the surveyed individuals who had never engaged in the other 21 possible uses varying between 15.82% (2215/14,000) and 87.43% (12,240/14,000). See Table 1 for frequency rates of responses.

Further Examination of the Electronic Health Adoption Construct

The factor analysis performed helped us to summarize the 23 activities in two distinctive components, which together explained 73% of the variance: (1) a factor composed of 13 items assessing the engagement in common activities related to "information and communication" (explaining 49% of the variance), which represented a basic usage of eHealth tools and (2) a factor formed by 10 items, which evaluated the use of eHealth tools related to "services and devices" (explaining 31% of the variance), which represented a more advanced type of usage. The results of this analysis are presented in Table 2.

These two factors were then used to construct a composite indicator as reported in the statistical analysis section, the "eHealth adoption gradient," which was used as a continuous variable for further comparisons performed in this study.

Then, after applying cluster analysis to the eHealth adoption index, a three-cluster solution emerged (two recommended cut-off points at 0.38 and 0.81), allowing us to divide participants into three groups (rare users, normal users, and super users) depending on how frequently they used eHealth tools [38].

According to the cut-off points recommended for the eHealth adoption gradient, 9,541 of the 14,000 participants (68.15%) were considered as rare users. Participants who sometimes used the Internet for health purposes during the week or month were classified as normal users (3538/14,000; 25.27%), and those who engaged in health-related behaviors online even more frequently were considered super users (921/14,000; 6.57%).

Table 1. Frequency rates of responses on electronic health (eHealth) uses.

Responses	Never, n (%)	Less than once a month, n (%)	At least once a month (but not every week), n (%)	At least once a week (but not every day), n (%)	Every day or almost every day, n (%)
Looked for information about a physical illness or condition that you or someone you know has	2147 (15.82)	5641 (41.57)	3505 (25.83)	1757 (12.95)	519 (3.82)
Looked for information about wellness or lifestyle	2930 (21.75)	4614 (34.26)	3419 (25.38)	1913 (14.20)	593 (4.40)
Participated in an online support group for people who are concerned about the same health or medical issue	8445 (68.50)	1758 (14.26)	1114 (9.04)	728 (5.90)	284 (2.30)
Participated in social networking sites (SNSs) talking about health and wellness	8121 (65.13)	1987 (15.94)	1197 (9.60)	817 (6.55)	347 (2.78)
Used email or went to a website to communicate with a doctor's office	8045 66.43)	2015 (16.64)	1112 (9.18)	632 (5.22)	306 (2.53)
Clicked on a health or medical website's privacy policy to read about how the site uses personal information	7291 (59.47)	2421 (19.75)	1323 (10.79)	864 (7.05)	361 (2.94)
Described a medical condition or problem online to get advice from an online doctor	8538 (69.91)	1796 (14.71)	966 (7.91)	614 (5.03)	299 (2.45)
Described a medical condition or problem online in order to get advice from other online users	8133 (64.81)	2199 (17.52)	1198 (9.55)	730 (5.82)	289 (2.30)
Kept a health website "bookmarked" or saved as a "favorite place," so you can go back to it regularly	6235 (48.72)	2763 (21.59)	1878 (14.68)	1327 (10.37)	594 (4.64)
Looked to see what company or organization is providing the advice or information that appears on a health website	6196 (48.53)	3293 (25.79)	1930 (15.12)	981 (7.68)	367 (2.87)
Looked for information about a mental health issue such as depression or anxiety	6482 (50.76)	3185 (24.94)	1691 (13.24)	980 (7.67)	433 (3.39)
Disclosed medical information on SNSs	9189 (75.64)	1175 (9.67)	899 (7.40)	596 (4.91)	290 (2.39)
Disclosed medical information on websites to share pictures, videos, movies, etc.	9325 (77.46)	1023 (8.50)	801 (6.65)	626 (5.20)	264 (2.19)

Responses	Never, n (%)	Less than once a month, n (%)	At least once a month (but not every week), n (%)	At least once a week (but not every day), n (%)	Every day or almost every day, n (%)
Made, canceled, or changed an appointment with your family doctor, specialist, or other health professionals online	9203 (70.00)	2236 (17.01)	904 (6.88)	520 (3.95)	285 (2.17)
Sent or received an email from your doctor, nurse, or health care organization	9554 (72.89)	2049 (15.63)	772 (5.89)	521 (3.97)	211 (1.61)
Made an online consultation through videoconference with your doctor or nurse	11,044 (87.43)	492 (3.89)	528 (4.18)	368 (2.91)	200 (1.58)
Received the results of your clinical or medical test online	10,476 (81.23)	1160 (8.99)	612 (4.75)	457 (3.54)	192 (1.49)
Accessed or uploaded your medical information or health record through an Internet protocol	10,801 (85.21)	661 (5.21)	589 (4.65)	438 (3.46)	187 (1.48)
Accessed or uploaded your medical information or health record through an Internet application provided by your health care organization	10,703 (84.31)	756 (5.96)	596 (4.69)	446 (3.51)	194 (1.53)
Used a game console to play games related with your health or your wellness	10,054 (77.90)	1191 (9.23)	855 (6.62)	569 (4.41)	237 (1.84)
Used a health or wellness app on your mobile phone	10,728 (82.25)	891 (6.83)	686 (5.26)	537 (4.12)	201 (1.54)
Used devices to transmit clinical information, received alarms, followed-up about your health anytime, anywhere	10,299 (79.28)	1069 (8.23)	785 (6.04)	560 (4.31)	278 (2.14)
Received any message about health promotion or health prevention	8575 (64.57)	2312 (17.41)	1292 (9.73)	765 (5.76)	337 (2.54)

Table 2. Results of factor analyses of electronic health (eHealth) uses. Rotated components matrix; sampling method: factor analysis by main components; rotation method: varimax with Kaiser-Meyer-Olkin 0.984; Bartlett's test of sphericity $P<.001$; convergence in three itineration; minimum eigenvalue 1.

Electronic health uses	Factor 1: services and devices		Factor 2: information and communication	
	Factor loadings[a]	Weights of variables in factor[b]	Factor loadings[a]	Weights of variables in factor[b]
Made an online consultation through videoconference with your doctor or nurse	0.88	0.09	0.33	0.01
Accessed or uploaded your medical information or health record through an Internet protocol	0.87	0.08	0.35	0.02
Accessed or uploaded your medical information or health record through an Internet application provided by your health care organization	0.86	0.08	0.35	0.02
Received online the results of your clinical or medical test	0.85	0.08	0.34	0.01
Used a health or wellness app on your mobile phone	0.81	0.07	0.34	0.02
Sent or received an email from your doctor, nurse, or health care organization	0.78	0.07	0.37	0.02
Used devices to transmit clinical information, received alarms, follow-up about your health anytime, anywhere	0.78	0.07	0.30	0.01
Made, canceled, or changed an appointment with your family doctor, specialist, or other health professionals online	0.77	0.07	0.35	0.025
Used a game console to play games related with your health or your wellness	0.76	0.06	0.31	0.01
Received any message about health promotion and/or health prevention	0.63	0.04	0.42	0.02
Looked for information about a physical illness or condition that you or someone you know has	0.17	0.01	0.78	0.08
Looked to see what company or organization is providing the advice or information that appears on a health website	0.32	0.01	0.77	0.08
Looked for information about wellness or lifestyle	0.16	0.01	0.75	0.07
Participated in social networking sites (SNSs) talking about health and wellness	0.46	0.02	0.74	0.07
Described a medical condition or problem online to get advice from other online users	0.50	0.03	0.72	0.07
Kept a health web site "bookmarked" or saved as a "favorite place," so you can go back to it regularly	0.28	0.01	0.71	0.07

Electronic health uses	Factor 1: services and devices		Factor 2: information and communication	
	Factor loadings[a]	Weights of variables in factor[b]	Factor loadings[a]	Weights of variables in factor[b]
Looked for information about a mental health issue such as depression or anxiety	0.35	0.01	0.71	0.07
Participated in an online support group for people who are concerned about the same health or medical issue	0.50	0.03	0.71	0.06
Clicked on a health or medical website's privacy policy to read about how the site uses personal information	0.42	0.02	0.71	0.06
Described a medical condition or problem online to get advice from an online doctor	0.54	0.03	0.69	0.06
Disclosed medical information on SNSs	0.57	0.04	0.67	0.06
Disclosed medical information on websites to share pictures, videos, movies, etc.	0.58	0.04	0.66	0.06
Used email or gone to a website to communicate with a doctor's office	0.54	0.03	0.64	0.05
Percentage of variance explained (%)	31		49	

[a]Based on rotated component matrix.

[b]Normalized squared factor loadings.

Sociodemographic and Clinical Data of Participants by User Group

Table 3 shows the descriptive statistics of the sociodemographic and clinical variables by eHealth user groups (rare users, normal users, and super users). Among the super users, 55.1% (508/921) were aged 25 to 54 years, 50.3% (464/921) had upper secondary education, 49.3% (454/921) were employed, and 40.7% (375/921) lived in medium-sized cities. Furthermore, 61.3% (565/921) also reported no health problem, a significantly higher percentage than the 47.76% registered among the normal users (1690/3538) and 55.73% (5317/9540) among the rare users (χ^2_2=156.8; P<.001). When examining the most frequently reported health problems, 3221 of the 9541 rare users (1422 of the 3538 normal users, 40.19%), and 257 of the 921 of super users (27.9%) reported allergy problems (χ^2_2=68.9; P<.001). In turn, 2728 of the 9541 rare users (28.59%), 1287 of the 3538 normal users (36.37%), and 227 of the 921 super users (24.6%) reported migraine or frequent headaches (χ^2_2=88.9; P<.001). In addition, 2110 of the 9541 rare users (22.11%), 990 of the 3538 normal users (27.98%), and 213 of the 921 super users (23.1%) reported strong, long-standing troubles with muscles and bones. Finally, 1765 of the 9541 rare users (18.49%), 908 of the 3538 normal users (25.66%), and 123 of the 921 super users (13.3%) reported chronic anxiety or depression (χ^2_2=110.0; P<.001).

Information and Communication Technology for Health by Morbidity Levels

Results of the one-way ANOVA showed that the number of health problems had a significant effect on the use of eHealth use ($F_{2,13996}$=11.584; P<.001). Post hoc comparisons using the Tukey HSD test indicated that participants with two or more health problems used new technologies for health significantly more (mean= 0.36; SD=0.23) compared with participants with no health problem (mean=0.33, SD=0.27) and participants with one health problem (mean=0.34, SD=0.25). However, there were no significant differences between participants with no health problem and those with one health problem on their use of eHealth.

A similar pattern was observed in the frequency of eHealth use with the series of one-way ANOVAs that were carried out to examine effect of health status group (no health problem, one health problem, and two or more health problems) on the mean scores of the 23 items on the questionnaire (Table 1). For example, in the question "How often have you looked for information about a physical illness or condition that you or someone you know has," participants with two or more health problems used eHealth significantly more to access health-related information (mean=1.59, SD=1.06) than participants with no health problem (mean=1.22, SD=1.04) and participants with one health problem (mean=1.36, SD=0.98; $F_{2,13996}$=164.127; P<.001).

XSL•FO
RenderX

Table 3. Sociodemographic and clinical data of participants by group of electronic health (eHealth) user.

Sociodemographic and clinical data	Rare users (N=9541), n (%)	Normal users (N=3538), n (%)	Super users (N=921), n (%)	Chi-square (degrees of freedom); P value
Age groups (years)				
16-24	1645 (17.24)	932 (26.35)	200 (21.6)	209.1 (4); <.001
25-54	6031 (63.21)	2169 (61.32)	508 (55.1)	
55-74	1861 (19.54)	436 (12.32)	214 (23.2)	
Gender				
Male	4896 (51.31)	1859 (52.54)	456 (49.4)	3213 (2); .20
Female	4645 (48.68)	1679 (47.45)	466 (50.6)	
Educational level				
Primary or lower secondary education (international standard classification of education, ISCED 3 or 4)	1415 (14.83)	529 (14.83)	184 (19.9)	45.2 (2); <.001
Upper secondary education (ISCED 3 or 4)	4306 (45.13)	1670 (47.18)	464 (50.3)	
Tertiary education (ISCED 5 or 6)	3819 (40.03)	1340 (37.86)	274 (29.8)	
Occupation				
Employed or self-employed (including family workers)	5644 (59.16)	2091 (59.10)	454 (49.3)	131.2 (6); <.001
Unemployed	892 (9.35)	315 (8.90)	128 (13.9)	
Student (not in the workforce)	1239 (12.98)	642 (18.14)	126 (13.7)	
Other not in the workforce (eg, retired or inactive)	1765 (18.50)	490 (13.84)	213 (23.7)	
Type of locality				
Densely populated area (cities and large towns)	3660 (38.36)	1469 (41.52)	338 (36.7)	48.6 (2); <.001
Intermediate area (towns)	3609 (37.82)	1423 (40.22)	375 (40.7)	
Thinly populated area (village and rural)	2272 (23.81)	646 (18.25)	208 (22.6)	
Presence of long-standing illness or health problem				
Yes	3819 (40.02)	1473 (41.63)	364 (39.5)	4566 (2); .10
No	5422 (56.82)	1918 (54.21)	510 (55.4)	
Long-term medical treatment				
Yes	3197 (33.50)	1228 (34.71)	307 (33.3)	2146 (2); .34
No	6269 (65.70)	2266 (64.05)	596 (64.7)	
Comorbidities				
No health problem	5317 (55.72)	1690 (47.76)	565 (61.3)	156.8 (4); .001
One health problem	2122 (22.24)	722 (20.4)	167 (18.2)	
Two or more health problems	2101 (22.02)	1126 (31.82)	189 (20.5)	

Table 4. Two-way between-groups analysis of variance of electronic health (eHealth) adoption gradient.

Variable	Mean (SD)			Source of variation	F value (degrees of freedom)	P value	Partial eta squared
	No health problem	One heath problem	Two or more health problems				
eHealth adoption gradient (age range, years)							
16-24	0.41 (0.30)	0.40 (0.27)	0.39 (0.22)	Age	78.248 (2)	.001	0.013
25-54	0.32 (0.26)	0.33 (0.23)	0.37 (0.23)	Comorbidities	5.6747.936 (2)	.003	0.001
55-74	0.28 (0.24)	0.33 (0.26)	0.33 (0.24)	Age x comorbidities	7.936 (2,2)	.001	0.003

In regards to the use of eHealth to participate in online support groups, in response to the question "How often have you participated in an online support group for people who are concerned about the same health or medical issue," participants with two or more health problems participated in online support groups significantly more frequently (mean=0.62, SD=1.04) than participants with no health problem (mean=0.42, SD=0.92) and participants with one health problem (mean=0.46, SD=0.93; $F_{2,13996}=$ 61.328; $P<.001$). The post hoc comparisons between individuals who did not report any health problem and those with one health problem were not significant for either of the two items described above.

The Electronic Health Adoption Gradient by Morbidity Level and Age Group

A two-way ANOVA was conducted to examine the effect of age and morbidity level on the eHealth adoption gradient (Table 4). There was a statistically significant interaction between the effects of age and number of health problems on the eHealth adoption gradient ($F_{4,11991}=7.936$; $P<.001$. Simple main effects analysis showed that people aged 16 to 24 years used new technologies for health purposes significantly more than people aged 25 to 54 years ($P<.001$) and people aged 55 to 74 years ($P<.001$) with no health problem, and also that people aged 25 to 54 years with no health problem used eHealth tools significantly more than people aged 55 to 74 years ($P=.003$). In addition, people with one health problem aged 16 to 24 years used eHealth tools significantly more than people aged 25 to 54 years ($P<.001$) and people aged 55 to 74 years ($P<.001$), but there were no statistically significant differences in the use of eHealth between people with one health problem aged 25 to 54 years and people aged 55 to 74 years ($P=.99$). Furthermore, people with one or more comorbidities (two or more health problems) aged 16 to 24 years used eHealth tools significantly more than people aged 25 to 54 years ($P=.007$) and people aged 55 to 74 years ($P<.001$), and there were also statistically significant differences in the use of eHealth between people with one or more comorbidities aged 25 to 54 years and people aged 55 to 74 years ($P<.001$).

To contrast these results, a comparison of the average eHealth adoption gradient was also performed on other health status variables included in the questionnaire of reference for this analysis. In particular, a two-way ANOVA was performed to examine the effect of age and the response to a question regarding whether the individual was receiving long-term medical treatment. There were main effects of age ($F_{2,11587}=8.350$; $P<.001$) and whether or not they were receiving long-term medical treatment ($F_{2,11587}=11.160$; $P<.001$) on the use of eHealth adoption gradient, but there was not a significant interaction between the aforementioned variables ($F_{4,11587}=1.729$; $P=.14$). Simple main effects analyses showed that the frequency of use of eHealth was greater for younger participants (16-24 > 25-54 > 55-74 years), with $P<.001$ for the three comparisons. In addition, people undergoing long-term medical treatment used eHealth tools significantly more than people without treatment ($P<.001$). The results follow a different pattern from the one described previously, with those individuals who reported long-term medical treatment using eHealth tools more frequently, independent of the age group to which they belonged.

Gender as a Covariate

When gender was included as a covariate in an ANCOVA of comparisons between age group and morbidity levels, results showed that gender had no significant effect on the eHealth adoption gradient ($F_{1,11990}=0.748$; $P=.39$).

Discussion

Principal Findings

The aim of this study was to explore the frequency of use of eHealth tools among a representative sample of European Internet users, taking into account their sociodemographic and clinical characteristics. We also examined the effect of multimorbidity on eHealth use. There was a positive and significant association between the number of health problems and the use of eHealth. An age-based digital divide was also observed, with individuals in younger age groups reporting a higher use of eHealth than older groups with the same level of morbidity. These findings add support to previous research, which has suggested that people with multiple conditions may demonstrate increased use of the available eHealth solutions. However, the full potential of eHealth tools for this population in particular has not yet been fully explored. There are only a limited number of studies and policies that have addressed people with multiple chronic conditions because of the fact that the disease-oriented approach still influences clinical research and health care organizations [11]. Therefore, it is important to move from a disease-centered to a patient-centered model [9] and increase our understanding of the experiences, specific needs, and challenges that may be faced by this population to improve the efficiency and quality of the existing eHealth services and encourage the implementation of new ones [10,11].

Main Findings

The eHealth adoption gradient proved to be a reliable way to measure different aspects of eHealth use [29]. A striking finding was that although most of the users went online to look for health information at least once a month, the majority of participants were classified as rare users, and only 921 out of the 14,000 total participants (6.57%) were super users. A closer exploration of the sociodemographic and clinical characteristics of the most frequent eHealth users reveals that the majority of them were aged 25 to 54 years and reported having zero health problems. This would suggest that those individuals who most frequently use eHealth do so for prevention, rather than as a reaction to a health concern. In addition, there were not gender differences in the use of the eHealth purposes, which is contrary to previous findings which concluded that females were more likely to look for health information online than males [27,39].

This study shows that multimorbidity, defined as the reporting of two coexisting health problems, is associated with more intense use of eHealth, both for information and communication purposes (eg, information about a physical illness or lifestyle and participation in online support groups), as well as for health-related services and devices (eg, videoconferences, Internet or mHealth apps, games, and mails to or from the general practitioner), a finding that is consistent with previous research [20-22]. It is interesting to note that the use of new technologies for health purposes was greater in questions related to seeking heath information on the Internet, compared with other uses of eHealth, such as participation in support groups, exchange of personal information online, or communication online with medical staff. These findings are similar to the ones reported by Zulman and colleagues (2015) [9]. Issues related to security and confidentiality of online personal information may explain the lower use of eHealth for the aforementioned specific purposes, as they are considered barriers for Internet use [27,28].

The existence of an age-based digital divide was observed when we included the age group of eHealth users in a two-way ANOVA. Our results showed that individuals in younger age groups reported greater use of eHealth solutions than older groups with the same level of morbidity. For example, individuals with two or more health problems aged 16 to 24 years used eHealth more frequently than those in the older age groups of 24 to 54 years and 55 to 74 years. Furthermore, healthy people aged 16 to 24 years were more likely to engage in eHealth activities than those aged 25 to 54 years and 55 to 74 years. It is also worth noting that the post hoc comparison between individuals who did not report any health problem and those with one health problem was not significant, but this may be because of the limited number of health problems considered (13), which does not preclude the presence of other conditions among those who reported no health problem. In any case, the interest of this exploratory study focuses on the effect of multimorbidity, for which the number of conditions considered should be sufficient, according to Fortin et al (2012) [40]. As expected, in line with the results of previous studies [29,31], the use of eHealth decreased with age in all morbidity groups.

Limitations

The principal limitation of this study concerns possible selection bias. That is, the surveyed population was already Internet users, and those with chronic illness may also have been less inclined to participate in an online survey or were unable to participate because of fatigue or disability. It would be interesting for future studies to test whether the relationship observed in this study between multimorbidity and eHealth use remains valid in a representative sample of multimorbid patients from the general population. Another limitation of this study is that the population studied only included citizens up to 74 years of age, whereas the prevalence of multimorbidity is particularly high in the oldest age groups, among which use of the eHealth is expected to be lower [30]. In addition, given the absence of a severity scale for the assessment of multimorbidity, it might also be worth exploring different and more complex ways of assessing it in future research. For example, some authors have criticized the simple counting of chronic diseases as oversimplistic, as it does not allow for discriminating between groups, particularly at older ages. Instead, they propose a more holistic approach that considers additional factors such as "emotional and psychological distress, and even existential or spiritual distress, all of which are socially patterned" [41]. Furthermore, self-reporting of chronic diseases may also represent a limitation for our study, as many people need further information and explanation to better understand if they have a specific chronic disease or not. In addition, the large proportion of rare users found in this study poses questions in terms of generalizability of our findings. Finally, the level of motivation and perception of barriers of the participants with multiple conditions were not evaluated in this study. Future research would benefit from identifying potential barriers (eg, low level of digital skills, lack of personalization of eHealth devices, or high perceived costs [6,11]) and assessing variables related to motivation (eg, how can eHealth be used to increase users' motivation to engage with sustainable lifestyle changes and how people can be motivated to use eHealth solutions? [6]) in this specific population.

Conclusions

In conclusion, it seems plausible that people with multimorbidity would benefit from new technologies for health-related activities as an additional means of communication. However, multimorbidity affects mostly older people who, in turn, are less likely to ask for support online. Therefore, complementary studies should test and implement new ways of making the already available eHealth solutions more accessible, attractive, and sensitive to the needs of older people to reduce the digital disadvantage which the majority of older adults may experience. For example, comprehensive training and educational campaigns addressing low digital literacy for health purposes for older people affected by multiple conditions and their informal caregivers can be offered. In addition, exploring the association between multimorbidity and eHealth use in a sample of individuals that are more representative of the general population with multimorbidity, including not only Internet users but also nonusers and persons in age groups older than 74 years, would be an interesting area for future research.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Fortin M, Hudon C, Haggerty J, Akker MV, Almirall J. Prevalence estimates of multimorbidity: a comparative study of two sources. BMC Health Serv Res 2010 May 06;10:111 [FREE Full text] [doi: 10.1186/1472-6963-10-111] [Medline: 20459621]

2. Loza E, Jover JA, Rodriguez L, Carmona L, EPISER Study Group. Multimorbidity: prevalence, effect on quality of life and daily functioning, and variation of this effect when one condition is a rheumatic disease. Semin Arthritis Rheum 2009 Feb;38(4):312-319. [doi: 10.1016/j.semarthrit.2008.01.004] [Medline: 18336872]

3. Salive ME. Multimorbidity in older adults. Epidemiol Rev 2013;35:75-83. [doi: 10.1093/epirev/mxs009] [Medline: 23372025]

4. Fried LP, Ferrucci L, Darer J, Williamson JD, Anderson G. Untangling the concepts of disability, frailty, and comorbidity: implications for improved targeting and care. J Gerontol A Biol Sci Med Sci 2004 Mar;59(3):255-263. [Medline: 15031310]

5. Gijsen R, Hoeymans N, Schellevis F, Ruwaard D, Satariano WA, van den Bos GA. Causes and consequences of comorbidity: a review. J Clin Epidemiol 2001 Jul;54(7):661-674. [Medline: 11438406]

6. Honka A, Kaipainen K, Hietala H, Saranummi N. Rethinking health: ICT-enabled services to empower people to manage their health. IEEE Rev Biomed Eng 2011;4:119-139. [doi: 10.1109/RBME.2011.2174217] [Medline: 22273795]

7. McMurdo ME, Witham MD, Gillespie ND. Including older people in clinical research. BMJ 2005 Nov 05;331(7524):1036-1037 [FREE Full text] [doi: 10.1136/bmj.331.7524.1036] [Medline: 16269472]

8. Hughes LD, McMurdo ME, Guthrie B. Guidelines for people not for diseases: the challenges of applying UK clinical guidelines to people with multimorbidity. Age Ageing 2013 Jan;42(1):62-69. [doi: 10.1093/ageing/afs100] [Medline: 22910303]

9. Zulman DM, Jenchura EC, Cohen DM, Lewis ET, Houston TK, Asch SM. How Can eHealth Technology Address Challenges Related to Multimorbidity? Perspectives from Patients with Multiple Chronic Conditions. J Gen Intern Med 2015 Aug;30(8):1063-1070. [doi: 10.1007/s11606-015-3222-9] [Medline: 25691239]

10. Orueta JF, García-Álvarez A, García-Goñi M, Paolucci F, Nuño-Solinís R. Prevalence and costs of multimorbidity by deprivation levels in the basque country: a population based study using health administrative databases. PLoS One 2014 Feb 27;9(2):e89787 [FREE Full text] [doi: 10.1371/journal.pone.0089787] [Medline: 24587035]

11. Barbabella F, Melchiorre MG, Quattrini S, Papa R, Lamura G. How can eHealth improve care for people with multimorbidity in Europe? Health Systems and Policy Analysis: Policy Brief 25 2017:1-31.

12. World Health Organization. 2016. From innovation to implementation – eHealth in the WHO European Region URL: http://www.euro.who.int/en/publications/abstracts/from-innovation-to-implementation-ehealth-in-the-who-european-region-2016 [accessed 2018-03-09] [WebCite Cache ID 6xnC0NqgP]

13. Jimison H, Gorman P, Woods S, Nygren P, Walker M, Norris S, et al. Barriers and drivers of health information technology use for the elderly, chronically ill, and underserved. Evid Rep Technol Assess (Full Rep) 2008 Nov(175):1-1422. [Medline: 19408968]

14. Finkelstein J, Knight A, Marinopoulos S, Gibbons MC, Berger Z, Aboumatar H, et al. Enabling patient-centered care through health information technology. Evid Rep Technol Assess (Full Rep) 2012 Jun(206):1-1531. [Medline: 24422882]

15. Ham C. The ten characteristics of the high-performing chronic care system. Health Econ Policy Law 2010 Jan;5(Pt 1):71-90. [doi: 10.1017/S1744133109990120] [Medline: 19732475]

16. Wagner TH, Baker LC, Bundorf MK, Singer S. Use of the Internet for health information by the chronically ill. Prev Chronic Dis 2004 Oct;1(4):A13 [FREE Full text] [Medline: 15670445]

17. World Health Organization. Geneva, Switzerland: World Health Organization; 2002. The World Health Report 2002: Reducing Risks, Promoting Healthy Life URL: http://www.who.int/whr/2002/en/whr02_en.pdf?ua=1

18. de Jong CC, Ros WJ, Schrijvers G. The effects on health behavior and health outcomes of Internet-based asynchronous communication between health providers and patients with a chronic condition: a systematic review. J Med Internet Res 2014 Jan 16;16(1):e19 [FREE Full text] [doi: 10.2196/jmir.3000] [Medline: 24434570]

19. Bower P, Cartwright M, Hirani SP, Barlow J, Hendy J, Knapp M, et al. A comprehensive evaluation of the impact of telemonitoring in patients with long-term conditions and social care needs: protocol for the whole systems demonstrator cluster randomised trial. BMC Health Serv Res 2011 Aug 05;11:184 [FREE Full text] [doi: 10.1186/1472-6963-11-184] [Medline: 21819569]

20. Hsu J, Huang J, Kinsman J, Fireman B, Miller R, Selby J, et al. Use of e-Health services between 1999 and 2002: a growing digital divide. J Am Med Inform Assoc 2005;12(2):164-171 [FREE Full text] [doi: 10.1197/jamia.M1672] [Medline: 15561786]

21. Flynn KE, Smith MA, Freese J. When do older adults turn to the internet for health information? Findings from the Wisconsin Longitudinal Study. J Gen Intern Med 2006 Dec;21(12):1295-1301 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00622.x] [Medline: 16995892]

22. Houston TK, Allison JJ. Users of Internet health information: differences by health status. J Med Internet Res 2002;4(2):E7 [FREE Full text] [doi: 10.2196/jmir.4.2.e7] [Medline: 12554554]

23. Yamin CK, Emani S, Williams DH, Lipsitz SR, Karson AS, Wald JS, et al. The digital divide in adoption and use of a personal health record. Arch Intern Med 2011 Mar 28;171(6):568-574. [doi: 10.1001/archinternmed.2011.34] [Medline: 21444847]

24. Newman L, Lupiáñez-Villanueva F. A framework to research the social determinants of ICTs for E-Health. The Journal of Community Informatics 2015;11(3):-- [FREE Full text]

25. McInnes DK, Gifford AL, Kazis LE, Wagner TH. Disparities in health-related internet use by US veterans: results from a national survey. Inform Prim Care 2010;18(1):59-68. [Medline: 20429979]

26. Nazi KM. Veterans' voices: use of the American Customer Satisfaction Index (ACSI) Survey to identify My HealtheVet personal health record users' characteristics, needs, and preferences. J Am Med Inform Assoc 2010;17(2):203-211 [FREE Full text] [doi: 10.1136/jamia.2009.000240] [Medline: 20190065]

27. Fox S. Pew Research Center [Internet]. 2011 Feb 28. Peer-to-peer Health Care URL: http://www.pewinternet.org/2011/02/28/peer-to-peer-health-care-2/ [WebCite Cache ID 6xrzaUZpH]

28. Bundorf MK, Wagner TH, Singer SJ, Baker LC. Who searches the internet for health information? Health Serv Res 2006 Jun;41(3 Pt 1):819-836 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00510.x] [Medline: 16704514]

29. Lupiañez F, Maghiros I, Abadie F. Citizens and ICT for Health in 14 European Countries: Results from an Online Panel. JRC Scientific and Policy Reports 2013:1-216. [doi: 10.2791/84062]

30. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the Internet and e-mail for health care information: results from a national survey. JAMA 2003 May 14;289(18):2400-2406. [doi: 10.1001/jama.289.18.2400] [Medline: 12746364]

31. Grant RW, Cagliero E, Chueh HC, Meigs JB. Internet use among primary care patients with type 2 diabetes: the generation and education gap. J Gen Intern Med 2005 May;20(5):470-473 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.04239.x] [Medline: 15963175]

32. Löffler C, Kaduszkiewicz H, Stolzenbach CO, Streich W, Fuchs A, van den Bussche H, et al. Coping with multimorbidity in old age--a qualitative study. BMC Fam Pract 2012 May 29;13:45 [FREE Full text] [doi: 10.1186/1471-2296-13-45] [Medline: 22639848]

33. European Commission. 2007 Dec. Health and long-term care in the European Union URL: http://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_283_en.pdf [WebCite Cache ID 6xufhELFv]

34. Handbook on Constructing Composite Indicators: Methodology and User Guide. Italy: OECD; 2008.

35. 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]

36. Pagliari C, Sloan D, Gregor P, Sullivan F, Detmer D, Kahan JP, et al. What is eHealth (4): a scoping exercise to map the field. J Med Internet Res 2005 Mar 31;7(1):e9 [FREE Full text] [doi: 10.2196/jmir.7.1.e9] [Medline: 15829481]

37. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp; 2013.

38. 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 Apr 30;16(4):e118 [FREE Full text] [doi: 10.2196/jmir.3187] [Medline: 24784798]

39. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009 Feb 20;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]

40. Fortin M, Stewart M, Poitras ME, Almirall J, Maddocks H. A systematic review of prevalence studies on multimorbidity: toward a more uniform methodology. Ann Fam Med Internet 2012;10(2):142-151. [doi: 10.1370/afm.1337]

41. Mercer SW, Smith SM, Wyke S, O'Dowd T, Watt GC. Multimorbidity in primary care: developing the research agenda. Fam Pract 2009 Apr;26(2):79-80. [doi: 10.1093/fampra/cmp020] [Medline: 19287000]

Abbreviations

ANOVA: analysis of variance
ANCOVA: analysis of covariance
eHealth: electronic health

XSL•FO

RenderX

HSD: honest significant difference
IPTS: Institute for Prospective Technological Studies
JRC: Joint Research Centre
PHR: personal health record
QoL: quality of life
SIMPHS: Strategic Intelligence Monitor on Personal Health Systems

<u>Original Paper</u>

Internet and Social Media Access Among Youth Experiencing Homelessness: Mixed-Methods Study

Lauren A Houdek VonHoltz[1*], MD, MPH; Rosemary Frasso[2*], PhD, CPH; Jesse M Golinkoff[3*], MPH; Alicia J Lozano[3*], MS; Alexandra Hanlon[3*], PhD; Nadia Dowshen[4,5*], MD, MSHP

[1]Children's Hospital of Philadelphia, Philadelphia, PA, United States

[2]College of Population Health, Thomas Jefferson University, Philadelphia, PA, United States

[3]School of Nursing, University of Pennsylvania, Philadelphia, PA, United States

[4]Craig-Dalsimer Division of Adolescent Medicine, Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA, United States

[5]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[*]all authors contributed equally

Corresponding Author:
Lauren A Houdek VonHoltz, MD, MPH
Children's Hospital of Philadelphia
3401 Civic Center Blvd, 9th Floor
Philadelphia, PA,
United States
Phone: 1 267 760 7680
Fax: 1 215 590 2768
Email: vonholtzl@email.chop.edu

Abstract

Background: Youth experiencing homelessness are at a risk for a variety of adverse outcomes. Given the widespread use of the internet and social media, these new technologies may be used to address their needs and for outreach purposes. However, little is known about how this group uses these resources.

Objective: This study investigated how homeless adolescents use these technologies for general and health-related purposes, whether the scope of their use changes with housing status, and their interest in a website dedicated to youth experiencing homelessness.

Methods: A convenience sample of youth aged 18 to 21 years was recruited from a youth-specific homeless shelter. All participants completed a 47-item survey, with 10 individuals completing a semistructured interview. Descriptive statistics, exact testing, logistic regression, and generalized estimating equation modeling was performed for quantitative data analysis. Interviews were transcribed verbatim, and NVivo 10 (QSR International) was employed to facilitate double coding and thematic analysis.

Results: A total of 87 participants completed the survey with a mean age of 19.4 (SD 1.1) years. While experiencing homelessness, 56% (49/87) accessed the internet at least once a day, with 86% (75/87) accessing once a week. Access to a smartphone was associated with a 3.03 greater odds of accessing the internet and was the most frequently used device (66% of participants, 57/87). While experiencing homelessness, subjects reported a 68% decreased odds in internet access frequency (odds ratio [OR] 0.32, $P<.001$), 75% decreased odds in spending greater amounts of time on the internet (OR 0.25, $P<.001$), and an 87% decreased odds of social media use (OR 0.13, $P=.01$). Ten participants completed the semistructured interview. Several themes were identified, including (1) changes in internet behaviors while experiencing homelessness, (2) health status as a major concern and reason for Internet use, and (3) interest in a website dedicated to youth experiencing homelessness. While experiencing homelessness, participants indicated their behaviors were more goal-oriented and less focused on leisure or entertainment activities.

Conclusions: While homeless youth experience changes in the frequency, amount of time, and specific uses of the internet and social media, study participants were able to access the internet regularly. The internet was used to search health-related topics. Given the importance of smartphones in accessing the internet, mobile-optimized websites may be an effective method for reaching this group.

XSL•FO
RenderX

KEYWORDS

adolescent; homeless youth; internet; social media; smartphone; health

Introduction

An estimated 1.6 to 2 million youth in the United States experience homelessness yearly and often suffer adverse outcomes [1,2]. Youth experiencing homelessness (YEH) have a higher prevalence of mental health conditions such as depression, conduct disorders, posttraumatic stress disorder, and suicide attempts and ideation [3-7]. They are often the victims of physical and sexual abuse, engage in risky sexual behaviors, and are frequently diagnosed with sexually transmitted diseases including HIV [8,9]. Finally, YEH have higher rates of both substance use and dependence than their housed counterparts [10-12].

Numerous interventions have been designed to improve health and well-being of YEH including housing programs, substance use, mental health issues, and basic daily needs [13-16]. Although YEH often express interest in engaging in such programs and interventions, and often initiate participation, they meet challenges when trying to identify services, and once enrolled in programs, these youth have trouble staying engaged with service providers [17,18]. It is also difficult to establish a stable mode of contact as they often move frequently between different service organizations, shelters, and cities [2]. Additionally, transient living situations complicate service provision and make it difficult to engage youth in longitudinal research and evaluation of interventions in this population [19].

Given the ubiquity of internet and social media use among adolescents in the United States, mHealth technologies represent a unique opportunity for interventions to improve health outcomes among YEH [20-23]. The utilization of these technologies as a novel approach to intervention demonstrates promise as preliminary work in the field suggests that YEH are able to access the internet with relative frequency [24,25]. A recent study indicated that approximately 80% of YEH used the internet at least once per week, with nearly 25% using the internet for at least an hour daily [25]. With regard to health-seeking behaviors, recent reports demonstrated that YEH use the internet to find information regarding HIV and other sexually transmitted infections (STIs) as well as looking for STI testing services [25].

Although these studies provide support for the use of the internet and various forms of social media as a viable method of connecting with YEH, several questions remain. More information is needed to determine if there are differences in youths' rates of access to the internet and social media before a homeless episode as compared with when they are in the midst of an episode. It is possible that those who face homelessness may have already been disenfranchised regarding access to these technologies. Additionally, understanding differential access methods between the 2 time periods (prehomelessness and during homelessness) may provide insight on how to best reach these youth during episodes of homelessness as well as help determine preferred contact and outreach methods. Furthermore, understanding what information this group is seeking and how

they seek it can inform interventions to improve health outcomes. Therefore, this study utilized a mixed-methods approach to examine rates of internet access, duration of internet use, devices used to access the internet, and characterize specific internet behaviors among a sample of youth before and during an episode of homelessness. We hypothesized that participants would have less access and less frequent use of the aforementioned technologies while experiencing homelessness. We also sought to identify health-related behaviors YEH engage in on the internet. Finally, using qualitative semistructured interviews, we explored how technology could be utilized to reach YEH to accurately and efficiently provide resources appropriate for YEH.

Methods

Setting and Sample

We engaged participants living at a nonprofit, youth homeless shelter in Philadelphia that annually serves approximately 350 youth aged 18 to 21 years. The facility provides services such as education and vocational advisement, an on-site medical clinic, and case management to the youth residing in the shelter. Study participants were 18-21 years old, housed at the shelter at the time of the study, and English speaking. Individuals were excluded if they were visibly intoxicated or unable to participate due to medical or psychosocial distress at the time of the interview. Only 1 subject was excluded for intoxication. Enrollment occurred between July 14 and September 12, 2014, Monday-Friday, 9 AM-5 PM. The team employed a convenience sampling strategy, and residents of the shelter were invited to participate in the study at on-site meal-time gatherings. All surveys and interviews were conducted in a private room in the shelter. This study was approved by the University of Pennsylvania Institutional Review Board.

Data Collection

All participants were considered homeless during this study by the investigators. Throughout the survey and interview, participants were asked to compare their internet, social media, and mobile health app usage during this period of homelessness to a time they did not consider themselves homeless and had stable housing.

This study employed a complementary qualitative and quantitative data collection approach, wherein quantitative and qualitative data were collected over the same period [26].

Quantitative Data Collection

A 47-item survey was administered to all study participants. Questions and answer choices were read aloud to minimize literacy concerns. Study data were collected and managed using a secure electronic database [27]. The survey took 10-15 min to complete and included topics such as age, gender, race, and sexual orientation. Internet and social media use and Web-based health-seeking behaviors during and before the participant's current homeless episode were assessed. Respondents were asked to indicate their interest in a website or app that focused

on the needs of YEH (see Multimedia Appendix 1 for full survey). Participants were given a US $10 gift card for participation and only completed a single survey.

Qualitative Data Collection

Following quantitative data collection, approximately every fifth participant (n=10) was invited to complete a semistructured, approximately 45-min interview in a private office at the shelter with a member of the research team (LV). We employed a constant comparison approach, that is, we reviewed and analyzed the transcripts as data were collected, to determine if new information was emerging [26]. After 10 participants were interviewed, analysis revealed that theme saturation had been reached, and no further participants were recruited. The most common reason for declining to participate was the length of the interview. The interview guide was informed by the literature and expert consultation [25,28,29]. All interviews were audio recorded and began with questions regarding the participant's current and historical access to the internet. Questions were designed to explore frequency of internet use during this episode of homelessness, specific access methods (ie, library computer, smartphone), time spent using the internet, and compare their current experience to that when they were housed. Respondents were also asked to comment on how and for what purposes they used the internet and social media, and whether they felt there were any differences between their current usage patterns and their patterns of use before this shelter stay. Other questions addressed health concerns and behaviors, the frequency of health-related thoughts, and health-related internet searches. Participants were asked whether they would use an app or website with content specifically designed to support YEH. They were asked to suggest content and resources, comment on format preference (app or website), and share any other suggestions that would make this type of resource useful to them. Participants were given a US $20 gift card for their participation in the interview.

Analyses

Survey Analysis

Descriptive statistics were used to characterize the sample. Means and standard deviations were used to describe continuous variables, and frequencies and percentages were used to describe categorical variables. Fisher exact tests were used to examine differences in demographics between participants who completed the survey and follow-up interview versus those who only completed the survey; two-sample t tests were used to compare continuous variables across the 2 groups. Generalized estimating equation (GEE) regression was used to model categorical internet use variables (internet access frequency, time spent using the internet, use of social media, access to a smartphone) at 2 time points, specifically before and while experiencing homelessness. Although surveys were conducted only once, study participants were asked to answer the same questions about access before and while experiencing homelessness. GEE methodology allows for modeling of the marginal distribution of each categorical outcome variable as a function of homelessness, while adjusting for covariates, and accounts for the likely correlations of the repeated outcome measures for each participant. Finally, 2 separate logistic

regression models were used to examine the association between having access to a smartphone on the odds of internet access at least once a day in participants before and while experiencing homelessness. All models were adjusted by age, male gender, and sexual orientation. Statistical significance was taken at $P<.05$. All statistical analyses were conducted using SAS Version 9.4 (SAS Institute Inc, Cary, NC).

Interview Analysis

Interviews were transcribed verbatim and entered in NVivo 10 software (QSR International, Melbourne, Australia) to facilitate thematic coding and systematic analysis [30]. A coding dictionary was developed inductively following careful reading of a set of representative transcripts by 3 members of the study team (LV, JG, and RF) [31]. See Multimedia Appendix 2 for coding tree and codebook. Specific codes identified in this study include internet access, cost, employment, housing, commercial resources, social media, emotions related to access, mobile phone app s, health activities, and suggestions for a YEH-specific website. Each subsequent interview was double-coded by 2 members of the study team (LV and JG). Evaluation of intercoder reliability was supported by NVivo 10 software employing the kappa coefficient. The software compares agreement and disagreement between coders in the assignment of specific codes to segments of the interview transcript. Complete agreement in coding correlates with a mean kappa of 1; near perfect agreement, a mean kappa of .81-.99; substantial agreement, a mean kappa of .61-.80; and moderate agreement, a mean kappa of .41-.60. Analysis of intercoder reliability for this study revealed substantial to almost perfect agreement (mean κ=.94; range, .76-1.00). This result was supported by percentage of agreement analysis, which yielded a mean of 99% (range, 92-100%) agreement of all codes examined in this study. After coding was complete, the study team organized the codes into thematic categories described later in the paper.

Results

Quantitative Results

Participant Demographics

Participant demographics for the sample have been summarized in Table 1. A total of 87 youth completed the survey. Respondents were 18-21 years old, with a mean age of 19.4 years (SD 1.1). The sample consisted of 60% (52/87) males, with 75% (65/87) identifying as African American or black and 14% (12/87) identifying as a sexual minority. No statistically significant differences in demographics were observed between participants who completed the survey and a follow-up interview and those who only completed the survey (Table 1).

Internet Use, Social Media Use, and Health-Seeking Behavior While in Shelter

While experiencing homelessness, 86% (75/87) participants were able to access the internet at least once a week and 56% (49/87) of youth were able to access the internet at least once a day. The majority of youth surveyed spent less than an hour using the internet each day, with smartphones (66%, 57/87) and

public computers (59%, 51/87) being the most commonly used devices for access. Participants also used the internet for job searching (86%, 75/87), checking email (83%, 72/87), searching for housing (60%, 52/87), and browsing school websites (51%, 44/87). The majority of youth utilized social media (85%, 74/87). Significant differences in internet access frequency (P<.001), amount of time spent using the internet each day (P<.001), access to a smartphone (P<.001), and use of social media (P=.01) were observed before and while experiencing homelessness (Table 2).

Comparison of Internet Use Characteristics Before and During Current Shelter Stay

Adjusted GEE model results demonstrated that current homelessness was significantly associated with a lower odds of regular internet access frequency, more time spent using the internet, use of social media, and access to a smartphone (see Multimedia Appendix 3). Specifically, participants reported a 68% decreased odds in internet access frequency of at least once a day during their current episode of homelessness as compared with before experiencing homelessness, adjusting for age, male gender, and sexual orientation (odds ratio [OR] .32, 95% CI 0.18-0.57, P<.001). Compared with before experiencing

homelessness, participants reported a 75% decreased odds in greater amounts of time spent using the internet during their current episode of homelessness, adjusting for age, male gender, and sexual orientation (OR 0.25, 95% CI 0.16-0.41, P<.001). Participants reported an 87% decreased odds in use of social media while experiencing homelessness compared with before experiencing homelessness, adjusting for age, male gender, and sexual orientation (OR 0.13, 95% CI 0.03-0.55, P=.01). Finally, participants reported a 67% decreased odds in access to a smartphone during their current episode of homelessness compared with before experiencing homelessness, adjusting for age, male gender, and sexual orientation (OR 0.33, 95% CI 0.18-0.60, P<.001).

Furthermore, adjusted logistic regression models demonstrated that although smartphone access was significantly associated with greater odds of regular internet access while experiencing homelessness (P=.02), it had no influence on a participant's ability to use the internet before experiencing homelessness (data not shown; P=.38). Specifically, the odds of internet access at least once a day among participants with access to a smartphone while experiencing homelessness was 3.03 times that of someone without a smartphone, adjusting for age, gender, and sexual orientation (OR 3.03, 95% CI 1.17-7.84).

Table 1. Participant demographics: youth living in a young adult homeless shelter in Philadelphia.

Characteristic	All participants (N=87)	Survey and follow-up interview participants (n=10)	Survey-only participants (n=77)	P value[a]
Age in years, mean (SD)	19.4 (1.1)	19.3 (1.2)	19.4 (1.1)	.71
Gender, n (%)				.12
Male	52 (60)	5 (50)	47 (61)	
Female	34 (39)	4 (40)	30 (39)	
Transgender	1 (1)	1 (10)	0 (0)	
Race, n (%)				
African American or black	65 (75)	7 (70)	58 (75)	.71
White	10 (12)	1 (10)	9 (12)	.99
American Indian or Alaska Native	5 (6)	0 (0)	5 (7)	.99
Other	19 (22)	3 (30)	16 (21)	.68
Hispanic, n (%)				.45
Yes	23 (26)	4 (40)	19 (25)	
No	64 (74)	6 (60)	58 (75)	
Sexual orientation, n (%)				.12
Heterosexual	74 (85)	8 (80)	66 (86)	
Gay	3 (3)	1 (10)	2 (3)	
Lesbian	1 (1)	1 (10)	0 (0)	
Bisexual	8 (9)	0 (0)	8 (10)	
Prefer not to answer	1 (1)	0 (0)	1 (1)	

[a]P values based on Fisher exact tests for categorical variables, and two-sample t tests for continuous variables.

Table 2. Internet use characteristics of youth living in a Philadelphia youth homeless shelter comparing before experiencing homelessness to while experiencing homelessness.

Characteristic	Before experiencing homelessness (N=87)	While experiencing homelessness (N=87)	P value[a]
Internet access frequency, n (%)			<.001
≥once/day	69 (79)	49 (56)	
<once/day	18 (21)	38 (44)	
Amount of time spent using the internet daily, n (%)			<.001
<20 min	8 (9)	23 (26)	
20 min-1 h	20 (23)	32 (37)	
1-2 h	10 (12)	11 (13)	
>2 h	49 (56)	21 (24)	
Access to smartphone, n (%)	74 (85)	57 (66)	<.001
Job searching, n (%)	81 (93)	75 (86)	.11
Use of social media (Facebook, Twitter, and Instagram), n (%)	85 (98)	74 (85)	.01
Use of email, n (%)	79 (91)	72 (83)	.07
Housing searches, n (%)	55 (63)	52 (60)	.56
School website searches, n (%)	55 (63)	44 (51)	.06

[a]P values based on unadjusted GEE regression models to account for the likely correlations of repeated measures for each participant.

Interest in a Website or App Focused on Youth Experiencing Homelessness

Survey respondents were asked if they would use a website or app if one were designed to focus specifically on issues facing YEH. A total of 91% (79/87) of the participants indicated that they would use a website or app specifically designed to support YEH.

Qualitative Results

We conducted semistructured interviews with 10 study participants that were representative of the larger sample (Table 1). Coding of interview transcripts revealed several important themes (which capture and organize constructs and experiences). The study team then met and organized the codes into the following thematic categories: (1) impact of homelessness on internet activities, (2) focus on health status, and (3) interest in a youth homelessness website. The following paragraphs summarize the findings for each of these categories followed by illustrative quotes.

Impact of Homelessness on Internet Activities

The majority of interviewed youth noted that there were striking differences in how they currently use the internet and social media compared with before experiencing homelessness. Consistent with survey findings, participants reported less frequent access of the internet and less time spent using the internet. However, the interviews revealed that the decrease in internet activity was nuanced. Specifically, participants experienced more substantial decreases in entertainment or leisure activities. internet activities became more goal oriented with participants using most of their time on the internet to look for basic needs such as housing, food, and employment. Finally, the youth indicated that time restrictions on public computers

impacted their internet activities by making them more purpose driven:

> *It's not so much for leisure or entertainment anymore. It's like I'm going to there with a purpose now. I'm really looking for, or waiting for certain information to help me get where it is I'm trying to go.* [Male, 21 years]

Several participants spent time discussing how their perception of their housing security affected how they used the internet. When they considered their living situation to be more tenuous and unstable, the internet was regarded as an essential resource. Participants indicated they would use the internet to contact people who might help them with housing/or to find food resources. When their living situation became more secure, the internet could then be used more for entertainment purposes. Conversely, one participant indicated that now that she had found stable housing through the shelter, she rarely used the internet. For this youth, the internet was exclusively used as a resource during times of crisis:

> *It's [the internet] really more important when you're struggling, because you need it. It's like you need it, because you need to keep in contact and you need to make sure you write. You need it for a lot, than when you're not struggling [with housing].* [Female, 19 years]

Focus on Health Status

Participants indicated they thought about their health frequently, with most (9/10) indicating they thought about their health daily. For many participants, health was a source of preoccupation and worry. The participants discussed being stressed about developing various acute illnesses or other diseases, worrying about current health conditions such as asthma and HIV,

focusing on their physical fitness, and finding ways to stay healthy.

Participants used the internet as a resource to find information about acute illnesses, child development, alternative medicine and remedies, cancer, exercise, nutrition and calories, provider and hospital contact information, and STIs. Although the youth were able to easily find information on the internet about these various topics, many reported having trouble interpreting the results of their searches. Often, they found the information was presented using complicated terminology that was meant for health care professionals. Some described working around this limitation by looking up synonyms for the words they did not understand. Others expressed frustration at the sheer volume of information that was available and the difficulty they experienced when trying sorting through the results of an internet search on a topic:

> *Like I'll be thinking like if I do certain things, how it's going to affect me on the inside, or like I'll be wondering if I'm sick or if I'm about to take a pill or something, I'll hurry up and Google it, to see what it's going to do to my insides. I do it all day.* [Male, 20 years]

The youth were asked about using social media as a means to connect with health care providers or clinics. The group had mixed views; some opposed the idea, whereas others thought it would be a convenient way to reach their physicians. Email and phone conversations were mentioned as preferred methods of contact:

> *I don't know. I don't feel like it's professional for me to talk to a doctor on Facebook. It's not really my way of looking at things. I feel like they should be a business or somebody I should call, not on Facebook.* [Male, 18 years]

Interest in a Youth Homelessness Website

All participants expressed interest in the development of a website or app that would aggregate resources and address topics to help YEH. They also indicated they would use this resource and recommend it to friends. The youth suggested many topics, including information about reliable housing, job opportunities, and women's and men's health issues. They also recommended that youth living at the shelter be involved in website development. A full list of potential website topics that were suggested by participants can be found in Textbox 1.

Participants also suggested using social media, YouTube, and word of mouth as ways to ensure the website or app reached the people who would be most in need of the resources:

> *I think it's [the website] cool, because, honestly, we can use all the help we can get.* [Female, 18 years]

Textbox 1. Suggested topics for youth homelessness website obtained from youth who completed an interview while living in a Philadelphia youth homeless shelter.

- Housing
    - Shelters and emergency housing
    - Legitimate housing and reliable landlords

- Food pantries and kitchens

- Classifieds and job opportunities

- Health
    - Women's and men's health sections
    - Clinic information, planned parenthood locations
    - Sexually transmitted infection information
    - Substance use information

- General Equivalency Diploma and education resources

- Important phone numbers

- Services

- Serve as a Web-based repository for important information (eg, immunization records, school transcripts)

Discussion

Principal Findings

The results of this study contribute to a small volume but growing body of literature that investigate how to best reach and engage YEH. Consistent with previous research, this study indicated that YEH are able to access the internet and social media with relative frequency, though less often than when they were housed [32]. Smartphones and public computers were heavily utilized for internet access during episodes of homelessness and shelter stays. Smartphone ownership had a strong impact on internet activities, as youth in the shelter with access to a smartphone were able to use the internet and social media more consistently than their peers who did not have smartphone access.

A significant body of research in this area has emerged from the work of Rice and colleagues in Los Angeles, CA [25,28,29]. Our study provides a perspective from a very different geographic location and captures the experience of a predominantly minority demographic. Despite the differences in study populations, several similar themes emerged. YEH are utilizing internet access for a multitude of activities including connecting with family and friends on social media, entertainment, job searching, and health-related searches. A trend among this study's population of shelter-housed youth was the use of the internet in a more goal-directed fashion. The internet was more frequently used as a tool to seek out resources related to housing and employment rather than as a source of entertainment during periods of housing insecurity. This highlights how these various technologies are vital tools for homeless youth as they work to obtain a more secure financial footing during times of housing strife. These goal-directed behaviors deserve recognition by policy makers when they are considering legislation that affects public internet access and the importance of reliable and regular internet access among this vulnerable group of young people.

The results of this study also highlight the importance of personal health among study participants. However, many homeless youth lack access to the health care system due to a lack of health insurance. Additionally, many do not utilize available services because they are addressing more pressing needs such as finding food and housing or due to perceived and actual barriers to accessing medical care [33]. As many youth are already turning to the internet for health concerns, there exists an opportunity to create a resource for homeless youth. Previous researchers created a Web-based personal health information system for YEH that allowed participating individuals a safe space to store confidential health information and direct access to public health nurses [34]. The interface was well received by study participants and created a safe environment for addressing many of the complex health needs the adolescents face. This study, like ours, identified an interest in reliable resources regarding health and a need among YEH. In our study, both the survey and qualitative data revealed the majority of participants used the internet as a primary source of information for health-related matters, and that health was very important to those interviewed. Future work in this area

can focus on dependable information about common adolescent health issues and about available local resources in a central location.

In addition to a resource that focuses on the health needs of YEH, this study identifies the need and interest in an internet-based platform that addresses the multitude of challenges faced by homeless youth. The results from the qualitative portion of the study further elucidated that limited phone data plans and computer time limits at public libraries had a strong influence on their internet behaviors. As the prevalence of smartphone use is high in this population, any internet outreach efforts must be specifically designed to be compatible with mobile devices. Websites that do not work well on mobile devices may discourage use and reduce their desired outreach effect.

This research stresses the importance of digital inclusion, especially with regard to this vulnerable population. Digital inclusion refers to the ability of groups and individuals to gain access to the internet, identify appropriate material and resources, as well as have opportunities for training to obtain the necessary skills to effectively use available material [35]. This study draws attention to significant opportunities to enhance digital inclusion among YEH. Specific opportunities include website content that is relevant to the acute needs of YEH, creating websites that reflect access patterns, and increasing access points to expand the reach of resources and services. Public policy efforts, libraries, and research have sought to understand and increase digital inclusion; however, there remains significant opportunity for improvement among YEH [36].

Limitations

There are several limitations to this exploratory study. Given that participants currently in shelter were asked to recall behavior before experiencing homelessness, responses may have been subject to recall bias. Additionally, because participants were all recruited from the same youth shelter in one city, results may have limited transferability and generalizability. Our sample size was relatively small with only 87 subjects, which led to wide confidence intervals throughout analysis. Future work can seek to examine internet use among larger samples and more diverse demographics of YEH to better establish specific types of interventions that would be the most useful. Despite these sample size limitations, among similar studies recruiting YEH, this study has a comparable sample size. Additionally, it is among one of the first and largest mixed methods studies to describe access to the internet and social media among a sample of YEH in a major Northeastern US city.

Conclusions

Although the overall time spent using the internet and utilizing social media sites was decreased compared with the time in which they were not struggling with homelessness, the young people in our sample were able to use these resources with relative frequency. Smartphones and public libraries were instrumental in providing access. Moving forward, mobile device friendly, adolescent-centered Web platforms, may serve

as an important resource to providing services to and connecting   with young people struggling with homelessness.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1

Internet use survey.

[PDF File (Adobe PDF File), 56KB - jmir_v20i5e184_app1.pdf]

Multimedia Appendix 2

Internet and youth homelessness codebook and coding tree.

[PDF File (Adobe PDF File), 118KB - jmir_v20i5e184_app2.pdf]

Multimedia Appendix 3

GEE model results.

[PDF File (Adobe PDF File), 52KB - jmir_v20i5e184_app3.pdf]

References

1. Hammer H, Finkelhor D, Sedlak AJ. Runaway/Thrownaway Children: National Estimates and Characteristics. In: US Dept. of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention. 2002 Oct Presented at: Interdisciplinary Conference on Human Trafficking; 2002; University of Nebraska p. 1-12 URL: https://scholars.unh.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1015&context=ccrc

2. Ringwalt CL, Greene JM, Robertson M, McPheeters M. The prevalence of homelessness among adolescents in the United States. Am J Public Health 1998 Sep;88(9):1325-1329. [Medline: 9736871]

3. Cauce AM, Paradise M, Ginzler JA, Embry L, Morgan CJ, Lohr Y, et al. The characteristics and mental health of homeless adolescents. J Emot Behav Disord 2016 Sep 14;8(4):230-239. [doi: 10.1177/106342660000800403]

4. Rohde P, Noell J, Ochs L, Seeley JR. Depression, suicidal ideation and STD-related risk in homeless older adolescents. J Adolesc 2001 Aug;24(4):447-460. [doi: 10.1006/jado.2001.0382] [Medline: 11549325]

5. Whitbeck LB, Hoyt DR, Bao WN. Depressive symptoms and co-occurring depressive symptoms, substance abuse, and conduct problems among runaway and homeless adolescents. Child Dev 2000;71(3):721-732. [Medline: 10953939]

6. Birmaher B, Ryan ND, Williamson DE, Brent DA, Kaufman J, Dahl RE, et al. Childhood and adolescent depression: a review of the past 10 years. Part I. J Am Acad Child Adolesc Psychiatry 1996 Nov;35(11):1427-1439. [doi: 10.1097/00004583-199611000-00011] [Medline: 8936909]

7. Thompson SJ. Factors associated with trauma symptoms among runaway/homeless adolescents. Stress Trauma Crisis 2005 Apr;8(2-3):143-156. [doi: 10.1080/15434610590956912]

8. Medlow S, Klineberg E, Steinbeck K. The health diagnoses of homeless adolescents: a systematic review of the literature. J Adolesc 2014 Jul;37(5):531-542. [doi: 10.1016/j.adolescence.2014.04.003] [Medline: 24931556]

9. Bender KA, Thompson SJ, Ferguson KM, Yoder JR, Kern L. Trauma among street-involved youth. J Emot Behav Disord 2013 Mar 10;22(1):53-64. [doi: 10.1177/1063426613476093]

10. Johnson KD, Whitbeck LB, Hoyt DR. Substance abuse disorders among homeless and runaway adolescents. J Drug Issues 2005;35(4):799-816 [FREE Full text] [Medline: 21533015]

11. Baer JS, Ginzler JA, Peterson PL. DSM-IV alcohol and substance abuse and dependence in homeless youth. J Stud Alcohol 2003 Jan;64(1):5-14. [Medline: 12608478]

12. Merscham C, Van Leeuwen JM, McGuire M. Mental health and substance abuse indicators among homeless youth in Denver, Colorado. Child Welfare 2009;88(2):93-110. [Medline: 19777794]

13. Altena AM, Brilleslijper-Kater SN, Wolf JL. Effective interventions for homeless youth: a systematic review. Am J Prev Med 2010 Jun;38(6):637-645. [doi: 10.1016/j.amepre.2010.02.017] [Medline: 20494240]

14. Tucker JS, D'Amico EJ, Ewing BA, Miles JNV, Pedersen ER. A group-based motivational interviewing brief intervention to reduce substance use and sexual risk behavior among homeless young adults. J Subst Abuse Treat 2017 May;76:20-27. [doi: 10.1016/j.jsat.2017.02.008] [Medline: 28340904]

15. Pedersen ER, Tucker JS, Kovalchik SA. Facilitators and barriers of drop-in center use among homeless youth. J Adolesc Health 2016 Aug;59(2):144-153 [FREE Full text] [doi: 10.1016/j.jadohealth.2016.03.035] [Medline: 27238839]

16. Slesnick N, Prestopnik JL, Meyers RJ, Glassman M. Treatment outcome for street-living, homeless youth. Addict Behav 2007 Jun;32(6):1237-1251 [FREE Full text] [doi: 10.1016/j.addbeh.2006.08.010] [Medline: 16989957]

17. Hobden K, Curtis Forney J, Wyszacki Durham K, Toro P. Limiting attrition in longitudinal research on homeless adolescents: what works best? J. Community Psychol 2011 Apr 04;39(4):443-451. [doi: 10.1002/jcop.20445]

18. North CS, Black M, Pollio DE. Predictors of Successful Tracking over Time in a Homeless Population. Soc Work Res 2012 Oct 09;36(2):153-159. [doi: 10.1093/swr/svs005]

19. Garvey R, Pedersen ER, D'Amico EJ, Ewing BA, Tucker JS. Recruitment and retention of homeless youth in a substance use and HIV-risk reduction program. Field Methods 2017 Sep 25;30(1):22-36. [doi: 10.1177/1525822X17728346]

20. Trent M, Thompson C, Tomaszewski K. Text messaging support for urban adolescents and young adults using injectable contraception: outcomes of the DepoText pilot trial. J Adolesc Health 2015 Jul;57(1):100-106 [FREE Full text] [doi: 10.1016/j.jadohealth.2015.03.008] [Medline: 26002432]

21. Dowshen N, Kuhns LM, Johnson A, Holoyda BJ, Garofalo R. Improving adherence to antiretroviral therapy for youth living with HIV/AIDS: a pilot study using personalized, interactive, daily text message reminders. J Med Internet Res 2012;14(2):e51 [FREE Full text] [doi: 10.2196/jmir.2015] [Medline: 22481246]

22. Devine S, Bull S, Dreisbach S, Shlay J. Enhancing a teen pregnancy prevention program with text messaging: engaging minority youth to develop TOP® Plus Text. J Adolesc Health 2014 Mar;54(3 Suppl):S78-S83 [FREE Full text] [doi: 10.1016/j.jadohealth.2013.12.005] [Medline: 24560081]

23. Dowshen N, Lee S, Matty Lehman B, Castillo M, Mollen C. IknowUshould2: feasibility of a youth-driven social media campaign to promote STI and HIV testing among adolescents in Philadelphia. AIDS Behav 2015 Jun;19 Suppl 2:106-111 [FREE Full text] [doi: 10.1007/s10461-014-0991-9] [Medline: 25563502]

24. Pollio DE, Batey DS, Bender K, Ferguson K, Thompson S. Technology use among emerging adult homeless in two U.S. cities. Soc Work 2013 Apr;58(2):173-175. [Medline: 23724580]

25. Rice E, Barman-Adhikari A. Internet and social media use as a resource among homeless youth. J Comput Mediat Commun 2014 Jan 1;19(2):232-247 [FREE Full text] [doi: 10.1111/jcc4.12038] [Medline: 25328374]

26. Creswell JW, Clark VL. Designing and Conducting Mixed Methods Research: 2nd Edition. Thousand Oaks, CA: Sage Publications; 2011.

27. 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]

28. Barman-Adhikari A, Rice E. Sexual health information seeking online among runaway and homeless youth. J Soc Social Work Res 2011 Jun;2(2):88-103 [FREE Full text] [doi: 10.5243/jsswr.2011.5] [Medline: 22247795]

29. Eyrich-Garg KM, Rice E. Cyber behavior of homeless adolescents and adults. In: Zheng Y, editor. Encyclopedia of Cyber Behavior 1st Edition. Hershey, PA, USA: IGI Global; 2012:284-291.

30. Guest G, Bunce A, Johnson L. How many interviews are enough? An experiment with data saturation and variability. Field Methods 2016 Jul 21;18(1):59-82. [doi: 10.1177/1525822X05279903]

31. MacQueen KM, McLellan E, Kay K, Milstein B. Codebook development for team-based qualitative analysis. Cult Anthropol Methods 2016 Sep 14;10(2):31-36. [doi: 10.1177/1525822X980100020301]

32. McInnes DK, Li AE, Hogan TP. Opportunities for engaging low-income, vulnerable populations in health care: a systematic review of homeless persons' access to and use of information technologies. Am J Public Health 2013 Dec;103 Suppl 2:e11-e24. [doi: 10.2105/AJPH.2013.301623] [Medline: 24148036]

33. Gelberg L, Gallagher TC, Andersen RM, Koegel P. Competing priorities as a barrier to medical care among homeless adults in Los Angeles. Am J Public Health 1997 Feb;87(2):217-220. [Medline: 9103100]

34. Dang MT, Whitney KD, Virata MC, Binger MM, Miller E. A web-based personal health information system for homeless youth and young adults. Public Health Nurs 2012;29(4):313-319. [doi: 10.1111/j.1525-1446.2011.00998.x] [Medline: 22765243]

35. Institute of Museum and Library Services, University of Washington, International City/County Management Association. Washington, DC: Institute of Museum and Library Services; 2012 Jan. Building Digital Communities: A Framework for Action URL: https://www.imls.gov/assets/1/AssetManager/BuildingDigitalCommunities_Framework.pdf [WebCite Cache ID 6x26V7jCi]

36. Jaeger PT, Bertot JC, Thompson KM, Katz SM, DeCoster EJ. The intersection of public policy and public access: digital divides, digital literacy, digital inclusion, and public lbraries. Public Library Q 2012;31(1):1-20. [doi: 10.1080/01616846.2012.654728]

Abbreviations

GEE: generalized estimating equation
OR: odds ratio
STI: sexually transmitted infection
YEH: youth experiencing homelessness

XSL•FO
RenderX

Original Paper

The Health and Retirement Study: Analysis of Associations Between Use of the Internet for Health Information and Use of Health Services at Multiple Time Points

Hyunju Shim[1*], MSW; Jennifer Ailshire[1*], PhD; Elizabeth Zelinski[1*], PhD; Eileen Crimmins[1*], PhD

USC Davis School of Gerontology, University of Southern California, Los Angeles, CA, United States

[*]all authors contributed equally

Corresponding Author:
Hyunju Shim, MSW
USC Davis School of Gerontology
University of Southern California
3715 McClintock Ave.
Los Angeles, CA, 90089-0191
United States
Phone: 1 2137405156
Email: hyunjush@usc.edu

Abstract

Background: The use of the internet for health information among older people is receiving increasing attention, but how it is associated with chronic health conditions and health service use at concurrent and subsequent time points using nationally representative data is less known.

Objective: This study aimed to determine whether the use of the internet for health information is associated with health service utilization and whether the association is affected by specific health conditions.

Methods: The study used data collected in a technology module from a nationally representative sample of community-dwelling older Americans aged 52 years and above from the 2012 Health and Retirement Study (HRS; N=991). Negative binomial regressions were used to examine the association between use of Web-based health information and the reported health service uses in 2012 and 2014. Analyses included additional covariates adjusting for predisposing, enabling, and need factors. Interactions between the use of the internet for health information and chronic health conditions were also tested.

Results: A total of 48.0% (476/991) of Americans aged 52 years and above reported using Web-based health information. The use of Web-based health information was positively associated with the concurrent reports of doctor visits, but not over 2 years. However, an interaction of using Web-based health information with diabetes showed that users had significantly fewer doctor visits compared with nonusers with diabetes at both times.

Conclusions: The use of the internet for health information was associated with higher health service use at the concurrent time, but not at the subsequent time. The interaction between the use of the internet for health information and diabetes was significant at both time points, which suggests that health-related internet use may be associated with fewer doctor visits for certain chronic health conditions. Results provide some insight into how Web-based health information may provide an alternative health care resource for managing chronic conditions.

Introduction

Background

Using the internet for health information has the potential to be a tool for health management [1]. Substantial evidence supports the idea that people improve their knowledge and ability to manage their own health by using Web-based health information [2-4]. Some studies find that more than half of patients using online tools report that they believed health information on the Web was as good as information from their primary physicians [5].

XSL•FO
RenderX

However, empirical tests of the association between the use of the Web-based health information and the use of medical care from a physician among older adults in the United States have been limited to mostly cross-sectional analyses. Previous literature suggests that the use of the internet for health information is associated with more visits to a doctor among older adults in the United States [6,7]. Research has also shown that health-related internet users are younger, more highly educated, have better cognitive functioning, better health literacy, and are sicker [7,8-11]. However, because most of these studies among older persons have been largely descriptive with a cross-sectional design, whether the effect of using the internet for health information on health service use is independent or correlated with other concurrent factors in the model remains unclear. In addition, the interactions between health-related internet use and different disease conditions on health care utilization have not been tested despite its implications for self-care management.

Due to these limitations, the evidence on the effect of using the internet for health information remains mixed. Some studies have reported that searching for health information on the Web may offer older adults an opportunity to engage in more patient-centered interactions with physicians and promote self-health management [12], such as through the use of online discussion boards or searching for information on diseases and conditions. Other researchers, however, have found that the use of the internet for health purposes may not be influential on individuals' offline interactions with their physicians and treatment decision-making [13], and have raised concerns that such activity may exacerbate health-related anxiety among some individuals [14].

Theoretical Framework

Andersen's Behavioral Model of Health Services Use is one of the most widely used models that identifies constructs related to the use of health care services [15-20]. This model has been refined and expanded over the years and remains the most highly adopted approach to modeling health service usage [21,22]. The model posits that health care use is influenced by individuals' predisposing factors, enabling factors, and need factors. Our model is derived from the Andersen model of general health service usage [16,21,22], as described in Figure 1. *Predisposing factors* describe factors that incline people to use health care services (eg, age, gender), which include personality traits in this study. Recent studies have suggested that psychological characteristics are important to include among predisposing factors [23]. *Enabling factors* facilitate health service use (eg, health insurance, regular source of care, use of Web-based health information). For instance, the availability of health-related information has recently emerged as an important enabling characteristic among members of ethnic minority groups [24]. However, applying this model to older people and including the use of health-related information on the Web specifically as an independent factor remains unexplored. Therefore, we include the use of Web-based health information as one of the enabling factors that predispose people to health service use. Sense of control is also included as a psychosocial characteristic that may be enabling, as sense of control has been shown to be a key factor influencing both optimism and discomfort in individuals' readiness to use technology [25], and those with high sense of control were found to be more likely to seek Web-based health information [26-28]. *Need factors* include both self-perceived and evaluated health needs (eg, comorbid conditions, self-reported health). This study examines reports of 5 doctor-diagnosed health conditions (ie, hypertension, diabetes, cancer, heart disease, and stroke) as an indicator of need.

Figure 1. Conceptual framework.
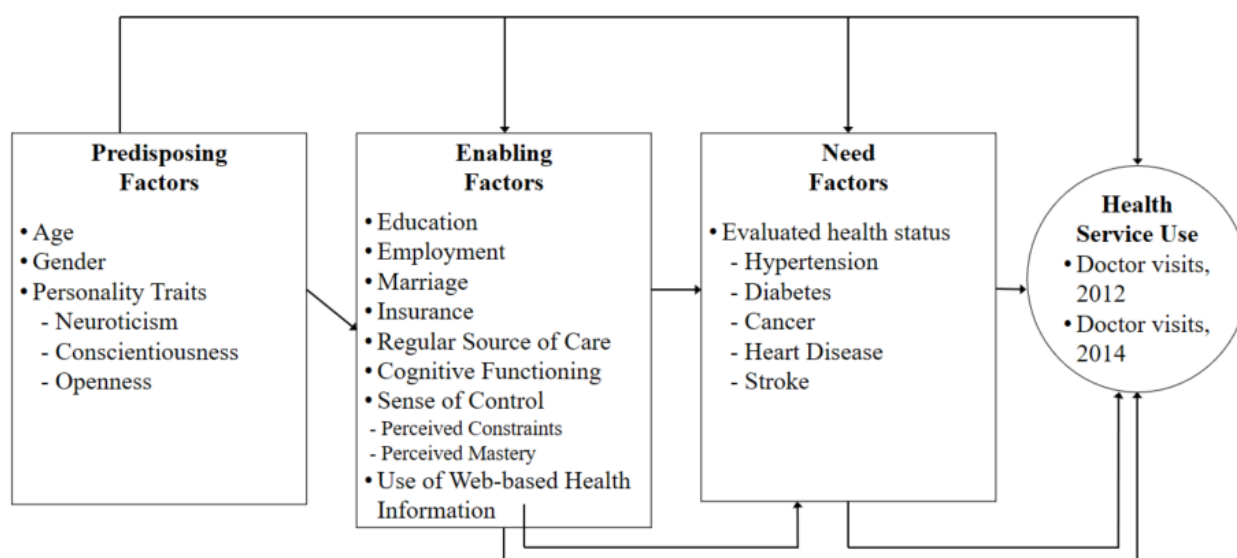
Objective

The primary purpose of this study was to examine whether the health-related internet use is linked to health service usage at 2 time points in a nationally representative survey. Examining the use at multiple time points helps to clarify whether using the internet for health information is associated with subsequent use in health care utilization or whether the association is only concurrent. Additionally, we also explore whether having a chronic health condition and the use of Web-based health information are synergistic by testing their interactions, which have not been examined in prior studies. Examination of the synergistic effect of different diseases and the use of the internet for health information on physician visits may have important implications for developing future tools for self-care management. The findings of this study provide the first evidence on the association between the use of the internet for health information and health service use at multiple time points and add to our understanding on how different diseases are related to this association. Using nationally representative data from the Health and Retirement Study (HRS), we aim to (1) test whether the use of Web-based health information and health service use are associated at concurrent and subsequent times (2) and examine whether the association is different among individuals with different chronic conditions.

Methods

Data

We used data from the HRS, a nationally representative sample of more than 20,000 older Americans aged 50 years and above. In 2012, the HRS included a technology module including a question on the use of Web-based health information that was administered to a random subsample of 1740 participants from the core participants.

Health service use is indicated by responses to a question on the number of visits to a doctor in the past 2 years. We assessed the effect of Web-based health information use on health care utilization both in the period (2012) concurrent with the reported internet use and in the period (2014) after the use, to see whether associations are present subsequent to the reported use of Web-based health information.

We also used psychosocial measures collected in a self-administered leave-behind questionnaire (SAQ) that is also administered to a random subsample of core participants on a biennial basis. The first random half-sample of households received the SAQ in 2010 and the second half in 2012. The combined 2010 and 2012 sample had an overall response rate of 71% [29].

Therefore, we limited the sample to 1737 respondents who responded to the question on the use of Web-based health information. For 2012, 73 cases were missing on visits to a doctor; 265 participants who did not have complete information in 2014 were also excluded. We also removed 251 respondents who did not meet age eligibility (under age 52 years), 34 participants who were missing on covariates, and 123 participants who revealed missing data on 5 psychosocial measures in the SAQ. The final analytic sample comprised 991

persons aged 52 years and above in 2012 who were followed up and had complete information in 2014.

Measures

Health Service Use

Health service use is measured by the number of times participants saw or talked to a medical doctor about their health in the last 2 years, reported in 2012 and 2014. Respondents were asked, "Aside from any hospital stays or outpatient surgery, how many times have you seen or talked to a medical doctor about your health, including emergency room or clinic visits in the last two years?" This question has regularly been used in the literature based on the HRS to assess health care utilization [30].

We categorized the independent variables affecting health service use according to conceptual framework in Figure 1, as predisposing, enabling, and need factors.

Predisposing Factors

Demographic Variables

Demographic variables included age and gender (male or female). Personality traits include neuroticism, conscientiousness, and openness because these 3 personality traits were found to be significant predictors of health-related internet use and health service use in prior studies [7,31]. These traits were derived from participants' rating of how well a list of words described them (1=a lot, 2=some, 3=a little, and 4=not at all). Response options were reverse coded to indicate higher agreement with each personality construct. Neuroticism was measured by moody, worrying, nervous, and calm (reversed; alpha=.71); conscientiousness with organized, responsible, hardworking, self-disciplined, careless (reversed), reckless (reversed), impulsive (reversed), cautious, thorough, and thrifty (alpha=.71); and openness to experience with creative, imaginative, intelligent, curious, broad-minded, sophisticated, and adventurous (alpha=.78). The composite scores for each trait ranged from 1 to 4, with higher scores indicating more of a given personality trait (see Multimedia Appendix 1). The scores of all items were then averaged within subdimensions of each personality trait. These personality items have been validated for comparability of each adjective item and tested for measurement invariance for assessing personality across different age groups [32,33].

Enabling Factors

Use of Web-Based Health Information

For assessing the use of Web-based health information, respondents were asked, "Do you use websites for finding medical and health information?" in the 2012 Technology Module. The response was binary (1=yes, 0=no) and coded as a dichotomous variable.

Sense of Control

We included sense of control indicated by perceived constraints and mastery (see Multimedia Appendix 2). Perceived constraints indicate one's assessment of external problems or barriers and were measured by participants' reports of agreement or

disagreement with the 5 statements on how they perceive barriers or obstacles in life. Perceived mastery indicates one's assessment of sense of competence over life circumstances and was measured by agreement or disagreement with the 5 statements. The answers for both questions used a 6-level Likert scale and the composite scores ranged from 1 (strongly disagree) to 6 (strongly agree), with higher values indicating greater agreement. The Cronbach alpha reliability coefficients were high for both constraints (alpha=.86) and mastery (alpha=.89). The scores of all items were then averaged within each of the 2 scales. All the psychosocial measures were coded following the instructions from the HRS psychosocial questionnaire guidelines [34].

Other Covariates

We included sociodemographic and other characteristics that facilitate health service uses. Educational attainment is measured with years of completed education, ranging from 0 to 17. Marital status (1=married, 0=not married) and employment (1=working, 0=not working) are dichotomized. Insurance was also a binary variable indicating whether the respondent reported having any type of public or private insurance (0=none, 1=has insurance). Having a regular source of care was assessed with the question that asked "Is there a place that you usually go to when you are sick or need advice about your health?" with a dichotomous answer (1=yes, 0=no). Cognitive functioning is measured by a summary score based on a series of tests including immediate and delayed recall tests (0-20), backward counting (0-2), and a serial 7s subtraction (0-5), with a score ranging from 0 to 27 [35]. This score assessing working memory and attention has been used in in many studies as a composite score indicating cognitive functioning [36,37].

Need Factors

Health Covariates

Health measures include self-reports of whether a doctor ever told the respondent that he or she had a set of chronic conditions: hypertension, diabetes, cancer, heart disease, or stroke. Responses were coded into binary variables indicating the presence or absence of each condition in 2012 and 2014.

Statistical Analyses

We first examined bivariate associations using χ^2 tests and F ratios to examine differences between users and nonusers of Web-based health information on predisposing, enabling, and need characteristics. We then turned to testing the predictive value of the use of Web-based health information for the number of visits to a doctor for both 2012 and 2014. We first estimated the effect of using Web-based health information on doctor visits controlling for predisposing, enabling, and need factors using negative binomial regressions. The examination of doctor visits data showed that the variances were higher than the means within each level of the use of Web-based health information, suggesting overdispersion of count data. Therefore, negative binomial regression models were used to address skewed distribution and overdispersion of the count data [38]. The likelihood ratio test also showed that the negative binomial model fitted better than a Poisson model. This method has been used in prior HRS studies that examined doctor visits as an

outcome of interest [30,39]. In addition, interaction terms for the use of Web-based health information and 5 doctor-diagnosed health conditions were included to examine whether the use differentially affects these health conditions to have synergistic effects on doctor visits.

Finally, to test the validity of the overall results and detect any potential type I error of interaction terms, we performed the False Discovery Rate (FDR)-controlling procedure tests. The FDR-controlling procedure test detects the expected proportion of false positives through dividing the number of true null hypotheses rejected by the total number of hypotheses rejected, which can also be regarded as the probability of making even one error if all null hypotheses are true [40]. We applied the Benjamini-Yekutieli method to test whether the significant results were spurious or not, using Stata's *multproc* command [41]. Stata 13 (StataCorp LP, College Station, TX) is used to estimate all models. Analytic results were weighted using Stata's survey prefix commands (SVY) to account for complex survey design in terms of differential sampling probabilities, nonresponse, population stratification, and sample weights.

Results

Sample Characteristics

Table 1 describes the sociodemographic, health, health care utilization characteristics of the internet users for health information compared with nonusers. Women were more likely to be users of Web-based health information (58.3% of users). The users were also significantly more likely to be younger, married, and in the labor force. Users were more conscientious (3.32 [SD 0.38] vs 3.22 [SD 0.43], *P*=.001) and open to experience (3.04 [SD 0.49] vs 2.87 [SD 0.51], *P*<.001). Users also had better cognitive functioning (17.25 [SD 3.48] vs 15.12 [SD 4.24], *P*<.001), but differences between users and nonusers on disease-related characteristics were mixed. The presence of hypertension and heart disease differed significantly between users and nonusers: 17.8% of users compared with 25.3% of nonusers on heart disease. However, there was no significant difference in the prevalence of diabetes, cancer, and stroke between users and nonusers (Table 1).

The 2-Time Point Analysis and the Interaction of the Use of Web-Based Health Information

Table 2 presents the results from the negative binomial regression models predicting the number of visits to a doctor in 2012 and 2014. Model 1 and Model 3 present the adjusted models with all predisposing, enabling, and need factors. Among predisposing factors, being female was associated with higher visits to doctors in both 2012 and 2014.

The use of Web-based health information was associated with significantly more doctor visits in 2012 (*b*=0.189, SE=0.095, *P*=.048), but the association was not significant in 2014 (*b*=0.047, SE=0.091, *P*=.60), indicating that the number of visits to a doctor in the last 2 years reported concurrently is higher among those who report the use of Web-based health information, but the effect on use in the next 2 years is not significant. Model 2 and Model 4 add interaction terms with different disease conditions and the use of Web-based health

XSL•FO

RenderX

information. The inclusion of interaction terms in Model 2 and Model 4 showed the synergistic effect of the use of Web-based health information and diabetes at both time points. Having diabetes (b=0.561, SE=0.140, P<.001) was associated with a higher number of visits to a doctor, but when the respondents used Web-based health information, the predicted number of visits was lower (b=−0.575, SE=0.204, P=.005). The results were similar in the 2014 models for the main effect (b=0.493, SE=0.130, P<.001) and the interaction term (b=−0.409, SE=0.179, P=.022). The interaction between the use of Web-based health information and diabetes on doctor visits is plotted in Figures 2 and 3.

Among predisposing factors, being female was associated with higher visits to doctors across all models. Conscientiousness

was significantly associated with fewer visits to doctors in 2012 (b= −0.255, SE=0.127, P=.04), but not in 2014. Similarly, openness to experience was significantly associated with more doctor visits in 2012 (b=0.290, SE=0.100, P=.004) only. Among enabling characteristics, being able to work was associated with less number of doctor visits across all models.

To test the validity and robustness of these findings, all estimates were tested for multiplicity using the Benjamini-Yekutieli method with the corrected overall critical P value of .002. For the significant results in 2012, having the usual source of care and the main effect of diabetes were rejected with the overall corrected P values. For 2014, the main effect of having diabetes and constant were rejected. However, the interaction effect for diabetes remained significant.

Table 1. Sample characteristics: Health and Retirement Study (2012). Weighted percentages and weighted means with standard deviation in parentheses.

Variables	Total (N=991)	Users (N=476)	Nonusers (N=515)	P value[a]
Female, n (%)	557 (52.9)	290 (58.3)	267 (46.3)	.003
Married, n (%)	603 (64.0)	315 (68.1)	288 (59.0)	.021
Working, n (%)	333 (39.8)	197 (46.3)	136 (31.7)	<.001
Doctor-diagnosed health conditions, n (%)				
Hypertension	613 (55.5)	277 (49.7)	336 (62.7)	.001
Diabetes	213 (18.6)	98 (17.1)	115 (20.4)	.26
Cancer	145 (14.8)	77 (15.9)	68 (13.6)	.41
Heart disease	230 (21.1)	95 (17.8)	135 (25.3)	.019
Stroke	62 (6.0)	26 (4.5)	36 (7.7)	.10
Any insurance, n (%)	822 (80.6)	388 (78.3)	434 (83.4)	.12
Usual source care, n (%)	862 (87.6)	420 (88.0)	442 (87.2)	.78
Age, mean (SD)	65.31 (9.47)	62.79 (7.78)	68.37 (10.40)	<.001
Years of education, mean (SD)	13.64 (2.58)	14.25 (2.35)	12.90 (2.66)	<.001
Cognitive functioning[b], mean (SD)	16.29 (3.98)	17.25 (3.48)	15.12 (4.24)	<.001
Psychosocial factors[c], mean (SD)				
Neuroticism (1-4)	2.02 (0.60)	2.04 (0.62)	2.00 (0.59)	.51
Conscientiousness (1-4)	3.27 (0.41)	3.32 (0.38)	3.22 (0.43)	.001
Openness (1-4)	2.96 (0.50)	3.04 (0.49)	2.87 (0.51)	<.001
Perceived constraint (1-6)	2.00 (1.04)	1.93 (1.09)	2.07 (0.97)	.09
Perceived mastery (1-6)	4.83 (1.06)	4.88 (1.07)	4.78 (1.04)	.21
Number of visits to a doctor, mean (SD)				
2012	8.53 (13.49)	9.40 (15.76)	7.46 (10.02)	.04
2014	8.06 (10.15)	8.35 (10.08)	7.71 (10.23)	.41

[a]P value indicates significant difference between users and nonusers.

[b]Cognitive functioning indicates average summary score based on a series of tests including immediate and delayed recall tests (0-20), backward counting (0-2), and a serial 7s subtraction (0-5), with a score ranging from 0 to 27.

[c]Psychosocial factors indicate average scores on a scale.

Table 2. Negative binomial regression analysis of association between the use of Web-based health information and visits to a doctor (N=991).

Predictor	2012						2014					
	Model 1			Model 2			Model 3			Model 4		
	β	*P*	SE	β	*P*	SE	β	*P*	SE	β	*P*	SE
Predisposing factors												
Age	0.000	.96	(0.007)	0.003	.60	(0.006)	−0.005	.35	(0.005)	−0.005	.34	(0.005)
Female	0.284	.008	(0.107)	0.289	.004	(0.101)	0.213	.01	(0.085)	0.228	.008	(0.085)
Personality traits												
Neuroticism	0.080	.40	(0.096)	0.084	.35	(0.092)	0.120	.11	(0.077)	0.115	.12	(0.076)
Conscientiousness	−0.255	.04	(0.127)	−0.254	.03	(0.118)	−0.211	.10	(0.130)	−0.202	.12	(0.130)
Openness	0.290	.004	(0.100)	0.286	.004	(0.098)	0.060	.56	(0.103)	0.063	.53	(0.102)
Enabling factors												
Use Web-based health information	0.189	.048	(0.095)	0.157	.30	(0.152)	0.047	.60	(0.091)	0.008	.96	(0.157)
Education	0.018	.34	(0.019)	0.023	.20	(0.018)	0.046	.007	(0.017)	0.046	.006	(0.016)
Working	−0.265	.038	(0.127)	−0.232	.04	(0.115)	−0.270	.004	(0.092)	−0.271	.004	(0.092)
Married	−0.094	.39	(0.111)	−0.091	.37	(0.102)	0.010	.90	(0.087)	0.018	.83	(0.085)
Insurance	0.212	.12	(0.137)	0.192	.14	(0.131)	0.302	.018	(0.127)	0.314	.01	(0.122)
Regular source of care	0.514	.001	(0.158)	0.518	<.001	(0.143)	0.195	.18	(0.148)	0.196	.18	(0.148)
Cognition	0.017	.27	(0.015)	0.017	.21	(0.014)	0.008	.56	(0.013)	0.006	.66	(0.014)
Sense of control												
Perceived constraint	−0.049	.36	(0.054)	−0.052	.31	(0.052)	−0.076	.09	(0.046)	−0.079	.09	(0.047)
Perceived mastery	−0.028	.53	(0.045)	−0.027	.52	(0.042)	0.006	.89	(0.045)	0.003	.94	(0.044)
Need factors												
Hypertension	0.254	.01	(0.099)	0.272	.028	(0.123)	0.301	<.001	(0.084)	0.183	.10	(0.113)
× Use Health Information				−0.042	.81	(0.179)				0.220	.19	(0.168)
Diabetes	0.276	.007	(0.102)	0.561	<.001	(0.140)	0.298	.003	(0.100)	0.493	<.001	(0.130)
× Use Health Information				−0.575	.005	(0.204)				−0.409	.022	(0.179)
Cancer	0.304	.04	(0.146)	0.130	.42	(0.162)	0.057	.58	(0.103)	0.165	.31	(0.163)
× Use Health Information				0.256	.33	(0.267)				−0.177	.37	(0.199)
Heart Disease	0.394	.005	(0.139)	0.076	.56	(0.129)	0.212	.02	(0.093)	0.164	.21	(0.132)
× Use Health Information				0.625	.009	(0.238)				0.120	.48	(0.173)
Stroke	0.197	.40	(0.236)	0.191	.31	(0.188)	0.371	.12	(0.240)	0.396	.24	(0.338)
× Use Health Information				−0.050	.91	(0.491)				−0.181	.67	(0.430)

Figure 2. Predicted mean number of visits to doctors and the interaction of the use of Web-based health information and diabetes in 2012.
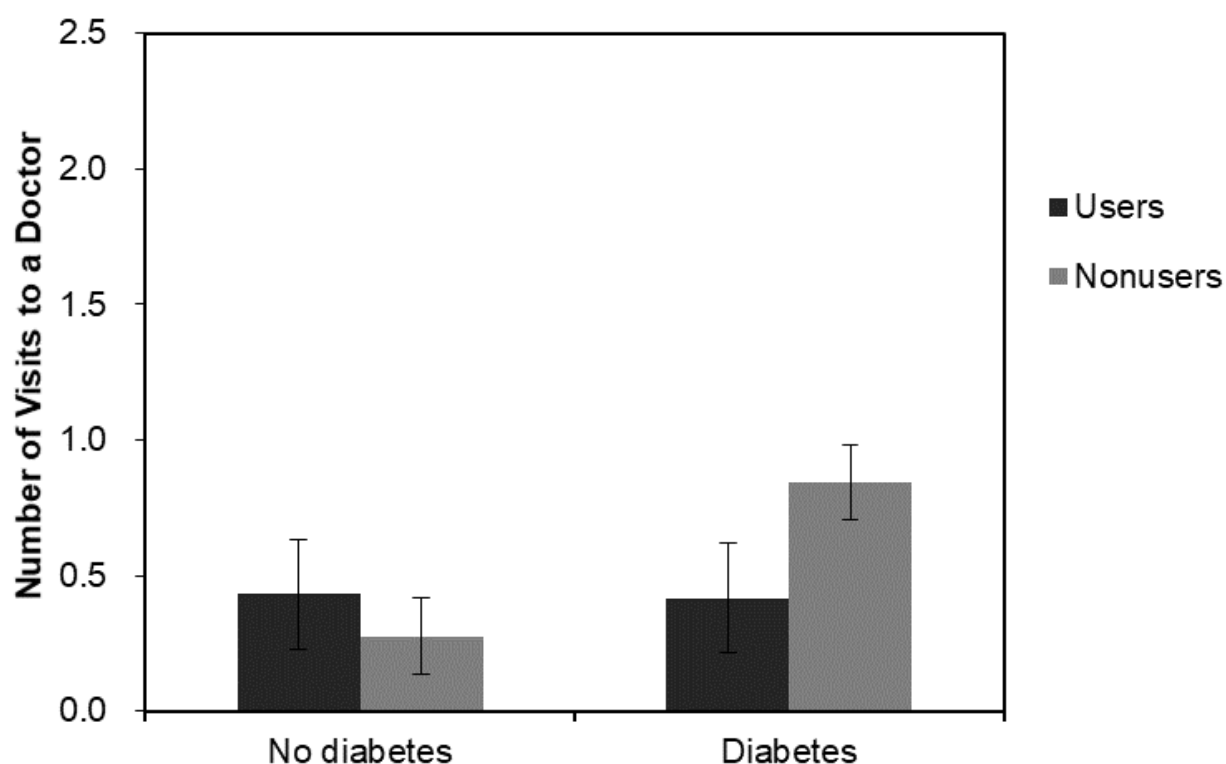
Figure 3. Predicted mean number of visits to doctors and the interaction of the use of Web-based health information and diabetes in 2014.
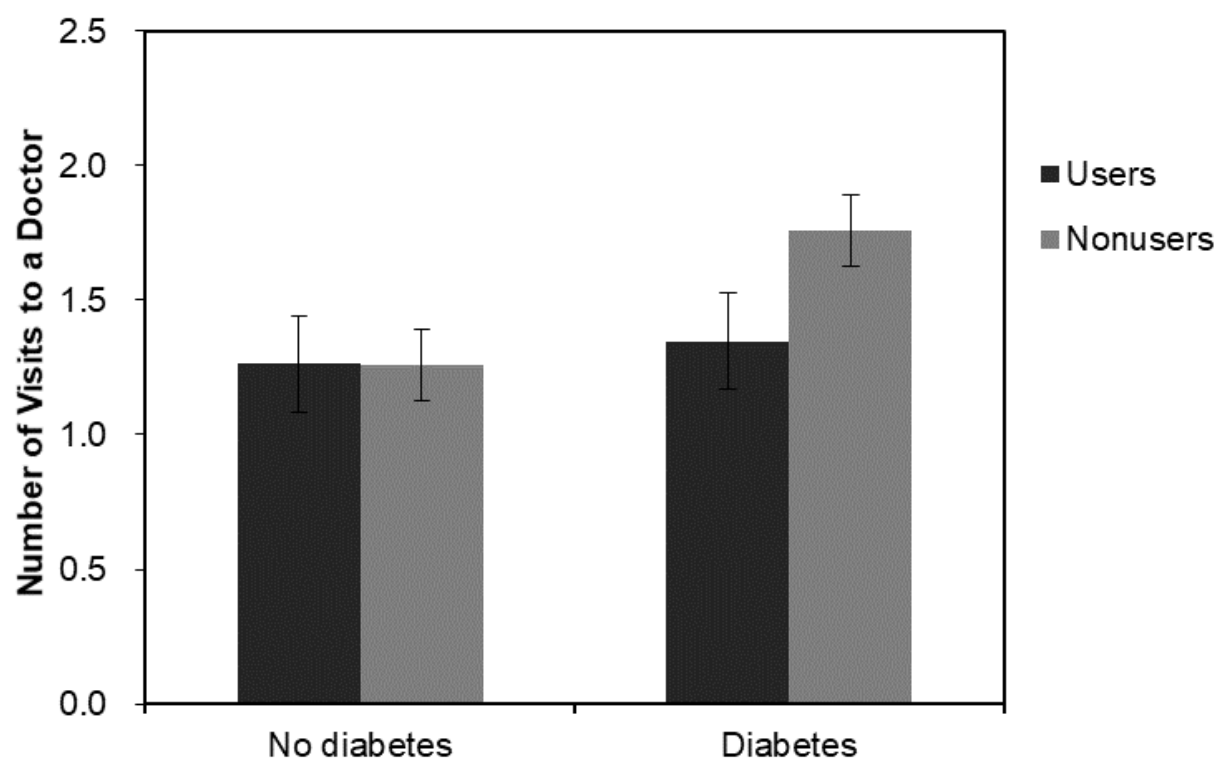
Discussion

Principal Findings

Almost half (48.0%) of the adults in this national sample of the US population aged 52 years and above used the internet for health information. This estimate is comparable to the national estimates of Web-based health information users reported in other studies, which range from 48.2% [42] to 58.0% [43] for older Americans. To our knowledge, this is the first study to examine the effect of the use of the internet for health information on the health service use among older adults at multiple time points and test the interaction effects with different diseases.

Among the predisposing factors, being female was consistently associated with a higher number of physician visits. This result is consistent with prior evidences on gender differences in the use of the internet for health information and health service use [6,7]. Prior studies found that women had higher health literacy than men, even when adjusting for sociodemographic and health factors [44,45]. Further studies will need to examine more detailed understanding of what motivates women to use health service more. Although the results on conscientiousness are counterintuitive, it is important to note that the significant association between conscientiousness and doctor visits was reported only among older men and not older women [46]. The relationship between personality and health service use remains unclear. There are more studies on diseases and personality, which found conscientious people are significantly less likely to have onset of diseases [47]. Because association between conscientiousness and doctor visits per se has received relatively little attention, further studies may focus on this relationship to clarify this association.

Regarding enabling factors, the use of Web-based health information was associated with more doctor visits at the same period, which is consistent with prior findings [6,7]. However, the association was not significant 2 years later. As most studies focused on examinations of using Web-based health information and doctor visits at one time point, using multiple time points in our study helps to understand how persistent the finding is over time. The lack of a relationship in 2014 indicates that the association is strongest when the use of the internet for health care information and health service use is coterminous, raising the possibility that the initial association was due to either reverse causation or causation by a third factor, such as need.

On the other hand, with regard to need factors, the interaction between the use of Web-based health information and chronic conditions on doctor visits was consistent for all models at both time points. Although the main effect of having diabetes was associated with more doctor visits, the association was negative when the use of the internet for health information was interacted with diabetes, suggesting a reduced number of visits to a doctor associated with being a user of Web-based health information with diabetic conditions. This result remained robust through the multiplicity tests and adds a novel finding to the existing literature.

Implications and Future Directions

Our study indicates that there may be an opportunity to further promote the development of resources for self-care and health management. Diabetes is a chronic condition that requires self-care as an essential component of management [48]. Due to this, there are many mobile phone apps and devices targeting diabetic patients [49-51], which may offer sources of health management that substitute visits to doctors. For practical reasons, some diabetic patients may prefer using Web-based health information to going to physicians.

Using the internet for health information may enable people with fewer economic resources to manage their conditions in a cost-efficient way. The national evaluation of health care costs for managing diabetes in the United States in 2012 stood at US $245 billion, out of which 9.0% was associated with physician office visits [52]. Therefore, the use of Web-based health information might be used to offset some of the burden. Additionally, online support could provide an opportunity for people with physical disabilities, who otherwise may find it difficult to visit physicians on a regular basis, to utilize health care resources. A recent meta-analysis estimating the risk of physical disability associated with diabetes reported that diabetes nearly doubles the risks of mobility disability [53]. Being able to use Web-based health information may enable older adults to manage their conditions despite physical disability. Finally, using Web-based health information could also be an enabling resource for older people who face linguistic barriers in utilizing health care services. One study found that the health-related internet use enabled Korean immigrant women in the United States to assist with self-care management, ranging from diagnosis and treatment to the recommendation of hospitals by allowing communications free of concerns about language barriers, indicating the potential empowerment by the use of the internet for health information among immigrant populations [54]. As the number of people who cannot speak fluent English is high among older immigrants, this finding has important implications for the use of Web-based health information in relation to health service use.

However, evaluating the effectiveness of the use of Web-based health information for chronic disease management remains to be addressed in future studies. One prior study reported that users of Web-based health information checked their blood glucose more regularly and practiced health behaviors better than nonusers among 344 patients with type 2 diabetes [55]. Yet, another qualitative study of 46 diabetic patients found that most participants passively received information, and although some diabetic patients collected information from Web sources, they sought medical professionals' confirmation on the reliability and applicability of information [56]. Therefore, it is not clear whether and how using Web-based health information is an alternative or a supplement to conventional health care resources.

There is also a concern regarding the quality of Web-based health information. Some prior studies reported that community-dwelling older adults aged 50 years and above showed general concerns regarding the credibility of Web-based health information and how time-consuming the Web-based

health information search can be [57]. These concerns were echoed by a study conducted on caregivers and their older care recipients, which raised questions on the reliability and effectiveness of Web-based health information [58]. Future studies should attempt to address concerns related to the reliability of Web-based health information and its implications for self-care management of chronic conditions for both older adults and their family caregivers. Additionally, the potential for the use of Web-based health information to enable socioeconomically and physically marginalized population to utilize the health care system merits further research attention.

Limitations

This study has several limitations. First, the questions on doctor visits at the first time point did not indicate a clear timing of the doctor visits relative to the use of Web-based health information. Although we examined doctor visits in both 2012 and 2014 to partly address this issue, this does not completely eliminate concerns in interpreting the findings. There is also a possibility that people may not have accurately responded to the question on visits to the doctor due to memorability reasons with regard to a 24-month time horizon. Second, the relatively simple question on the use of Web-based health information in the HRS also limited the scope of analysis. Detailed information on the purpose for seeking Web-based health information (eg, communicating with physicians, searching for information, taking care of their spouses) would have advanced our understanding on why some older adults seek health information on the internet and why women use Web-based health information more. Despite these limitations in the questions, however, this study has a unique contribution to expand our understanding on the association between the use of the internet for health information and health service utilization by using a

nationally representative sample. Finally, the study relies on self-reports of doctor-diagnosed medical conditions which may have some reporting errors. However, previous research has shown relatively high agreement between self-reports of health conditions and medical record reports [59,60].

Conclusions

This study provides the first evidence on how the use of the internet for health information and chronic health conditions are synergistically associated with doctor visits at 2 different time points. The findings indicate that, although the use of Web-based health information is independently associated with doctor visits concurrently, this effect does not remain significant 2 years later. However, for those with a certain chronic health conditions, using Web-based health information is significantly associated with fewer doctor visits at both time points, suggesting a potential avenue for future research in exploring the implications of the use of Web-based health information among older people. Examination of predisposing, enabling, and need factors, as well as their interactions, adds to our understanding of the relative importance of association between the use of Web-based health information and health service use.

Using the internet for health information may help older adults with a barrier to access conventional health care system—including those with lower socioeconomic resources, physical limitations, and linguistic barriers—to manage their conditions. However, caution needs to be exercised in extrapolating these study findings. Although accessing the health information on the Web may provide an alternative source of health care services, further research will need to evaluate quality and reliability of Web-based health information to elucidate the potential role of using this technology for health management among older people.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questions on personality traits (Health and Retirement Study, 2012).

[PDF File (Adobe PDF File), 146KB - jmir_v20i5e200_app1.pdf]

Multimedia Appendix 2

Questions on sense of control (Health and Retirement Study, 2012).

[PDF File (Adobe PDF File), 145KB - jmir_v20i5e200_app2.pdf]

References

1. US Department of Health and Human Services. 2010. Multiple Chronic Conditions: A Strategic Framework: Optimum Health and Quality of Life for Individuals with Multiple Chronic Conditions URL: https://www.hhs.gov/sites/default/files/ash/initiatives/mcc/mcc_framework.pdf [WebCite Cache ID 6yu04xYZD]

XSL•FO

RenderX

2.   Bolle S, van Weert JC, Daams JG, Loos EF, de Haes HC, Smets EM. Online health information tool effectiveness for older
     patients: a systematic review of the literature. J Health Commun 2015;20(9):1067-1083. [doi:
     10.1080/10810730.2015.1018637] [Medline: 26165846]

3.   Houston TK, Allison JJ. Users of internet health information: differences by health status. J Med Internet Res 2002
     Nov;4(2):E7 [FREE Full text] [doi: 10.2196/jmir.4.2.e7] [Medline: 12554554]

4.   Morahan-Martin JM. How internet users find, evaluate, and use online health information: a cross-cultural review.
     Cyberpsychol Behav 2004 Oct;7(5):497-510. [Medline: 15667044]

5.   Diaz JA, Griffith RA, Ng JJ, Reinert SE, Friedmann PD, Moulton AW. Patients' use of the internet for medical information.
     J Gen Intern Med 2002 Mar;17(3):180-185 [FREE Full text] [Medline: 11929503]

6.   Choi N. Relationship between health service use and health information technology use among older adults: analysis of
     the US National Health Interview Survey. J Med Internet Res 2011 Apr 20;13(2):e33 [FREE Full text] [doi:
     10.2196/jmir.1753] [Medline: 21752784]

7.   Flynn KE, Smith MA, Freese J. When do older adults turn to the internet for health information? findings from the Wisconsin
     longitudinal study. J Gen Intern Med 2006 Dec;21(12):1295-1301 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00622.x]
     [Medline: 16995892]

8.   Amante DJ, Hogan TP, Pagoto SL, English TM, Lapane KL. Access to care and use of the Internet to search for health
     information: results from the US national health interview survey. J Med Internet Res 2015;17(4):e106 [FREE Full text]
     [doi: 10.2196/jmir.4126] [Medline: 25925943]

9.   Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, et al. Factors predicting the use of technology: findings
     from the center for research and education on aging and technology enhancement (CREATE). Psychol Aging 2006
     Jun;21(2):333-352 [FREE Full text] [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]

10. Elliot AJ, Mooney CJ, Douthit KZ, Lynch MF. Predictors of older adults' technology use and its relationship to depressive
     symptoms and well-being. J Gerontol B Psychol Sci Soc Sci 2014 Sep;69(5):667-677 [FREE Full text] [doi:
     10.1093/geronb/gbt109] [Medline: 24304556]

11. Levy H, Janke AT, Langa KM. Health literacy and the digital divide among older Americans. J Gen Intern Med 2015
     Mar;30(3):284-289. [doi: 10.1007/s11606-014-3069-5] [Medline: 25387437]

12. Winstein CJ, Requejo PS, Zelinski EM, Mulroy SJ, Crimmins EM. A transformative subfield in rehabilitation science at
     the nexus of new technologies, aging, and disability. Front Psychol 2012;3:340 [FREE Full text] [doi:
     10.3389/fpsyg.2012.00340] [Medline: 23049517]

13. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the Internet and e-mail for health care information: results from a
     national survey. JAMA 2003 May 14;289(18):2400-2406. [doi: 10.1001/jama.289.18.2400] [Medline: 12746364]

14. Te PF, Baumgartner SE, Hartmann T, Tanis M. The curious case of cyberchondria: a longitudinal study on the reciprocal
     relationship between health anxiety and online health information seeking. J Anxiety Disord 2016 Oct;43:32-40. [doi:
     10.1016/j.janxdis.2016.07.009] [Medline: 27497667]

15. Andersen R. A behavioral model of families' use of health services. Chicago, IL: University of Chicago; 1968.

16. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? J Health Soc Behav 1995
     Mar;36(1):1-10. [Medline: 7738325]

17. Coulton C, Frost AK. Use of social and health services by the elderly. J Health Soc Behav 1982 Dec;23(4):330-339.
     [Medline: 7161472]

18. Kim H, Lee M. Factors associated with health services utilization between the years 2010 and 2012 in Korea: using
     Andersen's behavioral model. Osong Public Health Res Perspect 2016 Feb;7(1):18-25 [FREE Full text] [doi:
     10.1016/j.phrp.2015.11.007] [Medline: 26981338]

19. Mitchell J, Krout JA. Discretion and service use among older adults: the behavioral model revisited. Gerontologist 1998
     Apr;38(2):159-168. [Medline: 9573660]

20. Wolinsky FD, Johnson RJ. The use of health services by older adults. J Gerontol 1991 Nov;46(6):S345-S357. [Medline:
     1940101]

21. Babitsch B, Gohl D, von Lengerke T. Re-revisiting Andersen's behavioral model of health services use: a systematic review
     of studies from 1998-2011. Psychosoc Med 2012;9:Doc11 [FREE Full text] [doi: 10.3205/psm000089] [Medline: 23133505]

22. Andersen R, Davidson PS. Access to care. In: Changing the US health care system. San Francisco: Jossey Bass; 2013:33-70.

23. Schomerus G, Appel K, Meffert PJ, Luppa M, Andersen RM, Grabe HJ, et al. Personality-related factors as predictors of
     help-seeking for depression: a population-based study applying the Behavioral Model of Health Services Use. Soc Psychiatry
     Psychiatr Epidemiol 2013 Nov;48(11):1809-1817. [doi: 10.1007/s00127-012-0643-1] [Medline: 23266663]

24. Ani C, Bazargan M, Bazargan-Hejazi S, Andersen R, Hindman D, Baker R. Correlates of self-diagnosis of chronic medical
     and mental health conditions in under-served African American and Latino populations. Ethn Dis 2008;18(2 Suppl 2):S2-105.
     [Medline: 18646330]

25. Lin C, Shih H, Sher PJ. Integrating technology readiness into technology acceptance: the TRAM model. Psychol Mark
     2007 Jul;24(7):641-657. [doi: 10.1002/mar.20177]

26. Campbell R. Older women and the internet. J Women Aging 2004;16(1-2):161-174. [doi: 10.1300/J074v16n01_11] [Medline:
     15149930]

27. Campbell RJ, Harris KD, Wabby J. The internet and locus of control in older adults. 2002 Presented at: The AMIA Symposium; 2002; San Antonio, TX.

28. Roncancio AM, Berenson AB, Rahman M. Health locus of control, acculturation, and health-related internet use among Latinas. J Health Commun 2012;17(6):631-640 [FREE Full text] [doi: 10.1080/10810730.2011.635767] [Medline: 22211397]

29. Heeringa S, Connor J. Technical description of the health and retirement survey sample design. Ann Arbor: University of Michigan; 1995. URL: http://hrsonline.isr.umich.edu/sitedocs/userg/HRSSAMP.pdf [WebCite Cache ID 6yu5xSedw]

30. Gerst-Emerson K, Jayawardhana J. Loneliness as a public health issue: the impact of loneliness on health care utilization among older adults. Am J Public Health 2015 May;105(5):1013-1019. [doi: 10.2105/AJPH.2014.302427] [Medline: 25790413]

31. Landers R, Lounsbury J. An investigation of big five and narrow personality traits in relation to internet usage. Comput Human Behav 2006;22(2):283-293. [doi: 10.1016/j.chb.2004.06.001]

32. Roberts B, Jackson J, Duckworth A, Von Culin K. Personality measurement and assessment in large panel surveys. Forum Health Econ Policy 2011 Apr 18;14(3):1-23 [FREE Full text] [doi: 10.2202/1558-9544.1268] [Medline: 23503719]

33. Zimprich D, Allemand M, Lachman M. Factorial structure and age-related psychometrics of the MIDUS personality adjective items across the life span. Psychol Assess 2012 Mar;24(1):173-186 [FREE Full text] [doi: 10.1037/a0025265] [Medline: 21910548]

34. Smith J, Fisher G, Ryan L, Clarke P, House J, Weir D. Psychosocial and lifestyle questionnaire. Survey Research Center, Institute for Social Research. 2013. URL: http://hrsonline.isr.umich.edu/sitedocs/userg/HRS2006-2010SAQdoc.pdf [WebCite Cache ID 6yu6TTjfp]

35. Crimmins EM, Kim JK, Langa KM, Weir DR. Assessment of cognition using surveys and neuropsychological assessment: the health and retirement study and the aging, demographics, and memory study. J Gerontol B Psychol Sci Soc Sci 2011 Jul;66 Suppl 1:i162-i171 [FREE Full text] [doi: 10.1093/geronb/gbr048] [Medline: 21743047]

36. Xiang X, An R. The impact of cognitive impairment and comorbid depression on disability, health care utilization, and costs. Psychiatr Serv 2015 Nov;66(11):1245-1248. [doi: 10.1176/appi.ps.201400511] [Medline: 26073413]

37. McEvoy CT, Guyer H, Langa KM, Yaffe K. Neuroprotective diets are associated with better cognitive function: the health and retirement study. J Am Geriatr Soc 2017 Aug;65(8):1857-1862 [FREE Full text] [doi: 10.1111/jgs.14922] [Medline: 28440854]

38. Cameron A, Trivedi P. Microeconometrics using stata, 2nd Edition. College Station, Texas: Stata Press; 2009.

39. Kim ES, Konrath SH. Volunteering is prospectively associated with health care use among older adults. Soc Sci Med 2016 Jan;149:122-129 [FREE Full text] [doi: 10.1016/j.socscimed.2015.11.043] [Medline: 26714304]

40. Benjamini YD. The control of the false discovery rate in multiple testing under dependency. Ann Stat 2001;29(4):1165-1188.

41. Newson R. Multiple-test proceduressmile plots. Stata J 2003;3(2):109-132.

42. 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]

43. Fox S, Duggan M. Pew Research Center. Health online URL: http://www.pewinternet.org/files/old-media/Files/Reports/PIP_HealthOnline.pdf [accessed 2018-04-23] [WebCite Cache ID 6yu716JvA]

44. Lee HY, Lee J, Kim NK. Gender differences in health literacy among Korean adults: do women have a higher level of health literacy than men? Am J Mens Health 2015 Sep;9(5):370-379. [doi: 10.1177/1557988314545485] [Medline: 25122719]

45. von WC, Knight K, Steptoe A, Wardle J. Functional health literacy and health-promoting behaviour in a national sample of British adults. J Epidemiol Community Health 2007 Dec;61(12):1086-1090 [FREE Full text] [doi: 10.1136/jech.2006.053967] [Medline: 18000132]

46. Jerram K, Coleman P. The big five personality traits and reporting of health problems and health behaviour in old age. Br J Health Psychol 1999;4(2):181-192. [doi: 10.1348/135910799168560]

47. Weston S, Hill P, Jackson J. Personality traits predict the onset of disease. Soc Psychol Pers Sci 2014 Oct 13;6(3):309-317. [doi: 10.1177/1948550614553248]

48. Shrivastava SR, Shrivastava PS, Ramasamy J. Role of self-care in management of diabetes mellitus. J Diabetes Metab Disord 2013 Mar 05;12(1):14 [FREE Full text] [doi: 10.1186/2251-6581-12-14] [Medline: 23497559]

49. Demidowich AP, Lu K, Tamler R, Bloomgarden Z. An evaluation of diabetes self-management applications for android smartphones. J Telemed Telecare 2012 Jun;18(4):235-238. [doi: 10.1258/jtt.2012.111002] [Medline: 22604278]

50. Wu S. Supporting diabetes self-management among older adults through mobile technology application and intergenerational support: Imtop program design and effectiveness. 2017. URL: https://sswr.confex.com/sswr/2017/webprogram/Paper30286.html [WebCite Cache ID 6zGNOza9o]

51. Zheng L, Wang Y, Hao S, Shin AY, Jin B, Ngo AD, et al. Web-based real-time case finding for the population health management of patients with diabetes mellitus: a prospective validation of the natural language processing-based algorithm with statewide electronic medical records. JMIR Med Inform 2016 Nov 11;4(4):e37 [FREE Full text] [doi: 10.2196/medinform.6328] [Medline: 27836816]

52. American Diabetes Association. Economic costs of diabetes in the U.S. in 2012. Diabetes Care 2013 Apr;36(4):1033-1046 [FREE Full text] [doi: 10.2337/dc12-2625] [Medline: 23468086]

XSL•FO
RenderX

53. Wong E, Backholer K, Gearon E, Harding J, Freak-Poli R, Stevenson C, et al. Diabetes and risk of physical disability in adults: a systematic review and meta-analysis. Lancet Diabetes Endocrinol 2013 Oct;1(2):106-114 [FREE Full text] [doi: 10.1016/S2213-8587(13)70046-9] [Medline: 24622316]

54. Kim S, Yoon J. InformationR. The use of an online forum for health information by married Korean women in the United States URL: http://www.informationr.net/ir/17-2/paper514.html [WebCite Cache ID 6zU18HF5k]

55. Jamal A, Khan SA, AlHumud A, Al-Duhyyim A, Alrashed M, Bin SF, et al. Association of online health information-seeking behavior and self-care activities among type 2 diabetic patients in Saudi Arabia. J Med Internet Res 2015 Aug 12;17(8):e196 [FREE Full text] [doi: 10.2196/jmir.4312] [Medline: 26268425]

56. Longo DR, Schubert SL, Wright BA, LeMaster J, Williams CD, Clore JN. Health information seeking, receipt, and use in diabetes self-management. Ann Fam Med 2010;8(4):334-340 [FREE Full text] [doi: 10.1370/afm.1115] [Medline: 20644188]

57. Silver MP. Patient perspectives on online health information and communication with doctors: a qualitative study of patients 50 years old and over. J Med Internet Res 2015 Jan 13;17(1):e19 [FREE Full text] [doi: 10.2196/jmir.3588] [Medline: 25586865]

58. Nijland N, van GJ, Boer H, Steehouder MF, Seydel ER. Evaluation of internet-based technology for supporting self-care: problems encountered by patients and caregivers when using self-care applications. J Med Internet Res 2008 May 15;10(2):e13 [FREE Full text] [doi: 10.2196/jmir.957] [Medline: 18487137]

59. Bergmann MM, Byers T, Freedman DS, Mokdad A. Validity of self-reported diagnoses leading to hospitalization: a comparison of self-reports with hospital records in a prospective study of American adults. Am J Epidemiol 1998 May 15;147(10):969-977. [Medline: 9596475]

60. Bush TL, Miller SR, Golden AL, Hale WE. Self-report and medical record report agreement of selected medical conditions in the elderly. Am J Public Health 1989 Nov;79(11):1554-1556. [Medline: 2817172]

Abbreviations

FDR: False Discovery Rate
HRS: Health and Retirement Study
SAQ: self-administered leave-behind questionnaire

Original Paper

Scaling PatientsLikeMe via a "Generalized Platform" for Members with Chronic Illness: Web-Based Survey Study of Benefits Arising

Paul Wicks[1], PhD; Eileen Mack Thorley[1], MPH; Kristina Simacek[1], MA; Christopher Curran[1], BA; Cathy Emmas[2], PhD

[1]PatientsLikeMe, Cambridge, MA, United States
[2]AstraZeneca UK Ltd, Luton, United Kingdom

Corresponding Author:
Paul Wicks, PhD
PatientsLikeMe
160 Second Street
Cambridge, MA, 02142
United States
Phone: 1 6174755108
Email: pwicks@patientslikeme.com

Abstract

Background: Launched in 2006 for patients with amyotrophic lateral sclerosis, PatientsLikeMe is an online community offering patient-reported outcomes, symptom tracking, and social features. Every member of the site can see all the data reported by every other member, view aggregated reports, identify "patients like them," and learn about treatment options in order to live better with their condition. In previous studies, members reported benefits such as improved condition knowledge, increased medication adherence, and better management of side effects. However, the site evolved in 2011 from condition-specific "vertical" communities consisting only of people with the same disease to a "generalized platform," in which every patient could connect with every other patient regardless of condition and with generic, rather than condition-specific, data tools. Some, but not all, communities received further custom tracking tools.

Objective: We aimed to understand (1) whether members of PatientsLikeMe using the generalized platform still reported similar benefits and (2) assess factors associated with benefits, such as community customization, site use, and patient activation.

Methods: A cross-sectional retrospective custom survey was fielded to 377,625 members between 2016 and 2017 including the Patient Activation Measure (PAM). A benefit index was developed for comparability across conditions.

Results: The invitation was viewed by 26,048 members of whom 11,915 did not respond, 5091 opted out, 1591 provided partial data, and 17 were screened out. Complete responses were received from 7434 participants. Users perceived greatest benefit in understanding how their condition may affect them (4530/6770, 66.91% participants, excluding "does not apply" answers), understanding what might help them live better with their condition (4247/6750, 62.92%), which treatments were available (4143/6898, 60.06%), understanding treatment side effects (4182/6902, 60.59%), and important factors in making treatment decisions (3919/6813, 57.52%). The benefit index was 29% higher for the "most activated" patients (PAM level 4 vs PAM level 1; relative risk [RR]=1.29, $P<.001$), 21% higher for conditions with some community customization versus none (RR=1.21, $P<.001$), and 11% higher in those using the site most often versus least (RR=1.11, $P<.001$).

Conclusions: Members of the generalized platform reported a range of benefits related to improved knowledge and understanding of their condition and treatment management. Condition-specific customization may improve their experience still further. Future studies will explore longitudinal changes to patient activation.

KEYWORDS

personal health records; personal monitoring; technology; health care; self-help devices; personal tracking; social support; online support group; online health community

XSL•FO
RenderX

Introduction

History of Online Communities for Chronic Illness

Online communities for chronic illnesses have existed since the early 1980s through email list-servs, USENET, and online bulletin boards such as "The WELL" [1], Association of Online Cancer Resources [2], Yahoo! Groups [3], and Braintalk [4]. These emphasized lengthy and anonymous text discussions, making it hard to extract useful information as the communities grew. Later "Web 2.0" technologies and mainstream social networking sites such as Facebook and Twitter grew in popularity [5], meeting patients and caregivers where they were already engaging. However, the public nature of these newer sites may have made users more cautious about sharing health information [6,7].

PatientsLikeMe is an online community that allows members to find other patients like them, share and track their health data over time, and contribute to scientific research [8]. As a for-profit company, PatientsLikeMe works with pharmaceutical and biotechnology companies to share aggregated outcomes data, patient treatment evaluations, and to field surveys to help improve partner services and support patient centricity efforts [9]. Some question whether there is equipoise in the benefits accrued to participant members and the revenue collected by the company [10-12].

Many studies have documented the effect of online health community participation on feelings of empowerment and perceived social support [3,13]. However, systematically studying the effects of such communities on their members has been challenging [14]. The member populations are highly self-selecting, research funding is minimal [3], many users are "lurkers" [15], the most salient experiences are often qualitative, defining scope can be challenging [16], and sample sizes are small and time-limited [17]. By the time a peer-reviewed scientific study has been published, the community itself may have evolved, emptied, or shut down [18].

Early History of PatientsLikeMe

PatientsLikeMe launched in 2006 for patients with amyotrophic lateral sclerosis (ALS). Over the next 5 years, the site created distinct "vertical" communities. Members could only belong to one at a time, they could not see data from other communities, and there was no ability to add comorbid conditions. Between 2006 and 2011, nine vertical communities were developed for people with multiple sclerosis (MS), Parkinson's disease, HIV, a range of mood disorders, progressive supranuclear palsy, multiple system atrophy, fibromyalgia, myalgic encephalomyelitis / chronic fatigue syndrome, and organ transplants.

A retrospective survey among members of the first six vertical communities reported a number of benefits including: learning about symptoms they had experienced (952/1323, 71.96% rated "very helpful" or "moderately helpful"), understanding side effects of treatment (757/1323, 57.21%), and finding another patient like them on the same treatment (559/1323, 42.25%) [19]. Members reported other changes, such as deciding to start

a new medication (496/1323, 37.49%), change their medication dosage (336/1323, 25.39%), or stopping a medication (290/1323, 21.92%). A subset (151/1320, 11.44%) reported changing their physician as a result of using the site. A subsequent survey in epilepsy had similar results, with the additional finding of a dose-effect curve between the number of benefits experienced and the number of social ties made on the site [20].

However, these studies shared limitations of self-selecting populations, being conducted internally, a lack of validated instruments, and a lack of preplanned analyses. A subsequent academic collaboration with the University of California, San Francisco, and the US Department of Veterans Affairs found significant benefits for veterans with epilepsy after 6 weeks of site use. This latter study used validated measures of self-management and self-efficacy in a pre-post design with prespecified analyses [21].

The Generalized Platform

Historically, patients in the "vertical" communities each completed custom patient-reported outcome measures tailored to their condition such as the ALS Functional Rating Scale (Revised) [22] in ALS, the MS Rating Scale in MS [23], or a detailed "seizure meter" in epilepsy [20]. Additional condition-specific features included customized visualizations, symptoms, medical history, patient search features, forums, and laboratory tests. Developing and launching these tools was time-consuming, ranging from 3 to 12 months of development for each community. Given the large unmet need (and a waiting list of some 30,000 individuals requesting we build new communities), PatientsLikeMe made major changes to the platform in April 2011, described previously [24,25]

To allow any patient to join the site and to track multiple conditions, we developed a more scalable "generalized platform." This allowed any patient to use a generic quality-of-life outcome measure [26], symptom tracking, treatment tracking, and social networking tools. Members of the generalized platform would not benefit from customized visualizations or a dedicated forum; for instance, and we did not hand-curate condition-specific symptoms or treatments (see Figure 1). However, members could still opt to track their own self-selected symptoms and treatments.

Resulting in part from the generalization of the platform, the site has more than 600,000 registered members across more than 2900 conditions (as of February 2018). However, to date no research has investigated the extent to which members engaging in the general platform experience benefits.

In the years following 2011, a small number of communities benefited from additional "community upgrades" supported by pharmaceutical partnerships. These added site features that would once have only been available to "vertical" communities, such as custom questions, symptoms, laboratory tests, and patient-reported outcomes to communities that would otherwise only have had the generic site functionality. These upgraded communities included psoriasis, idiopathic pulmonary fibrosis, multiple myeloma, and lung cancer.

Figure 1. Experience of a hypothetical user before and after platform generalization. Multiple sclerosis (MS) and major depressive disorder (MDD) have community upgrades whereas type 2 diabetes (T2D) does not. MSRS: multiple sclerosis rating scale; N/A: not applicable; PLM QOL: PatientsLikeMe Quality of Life Questionnaire; PRO: patient reported outcome.
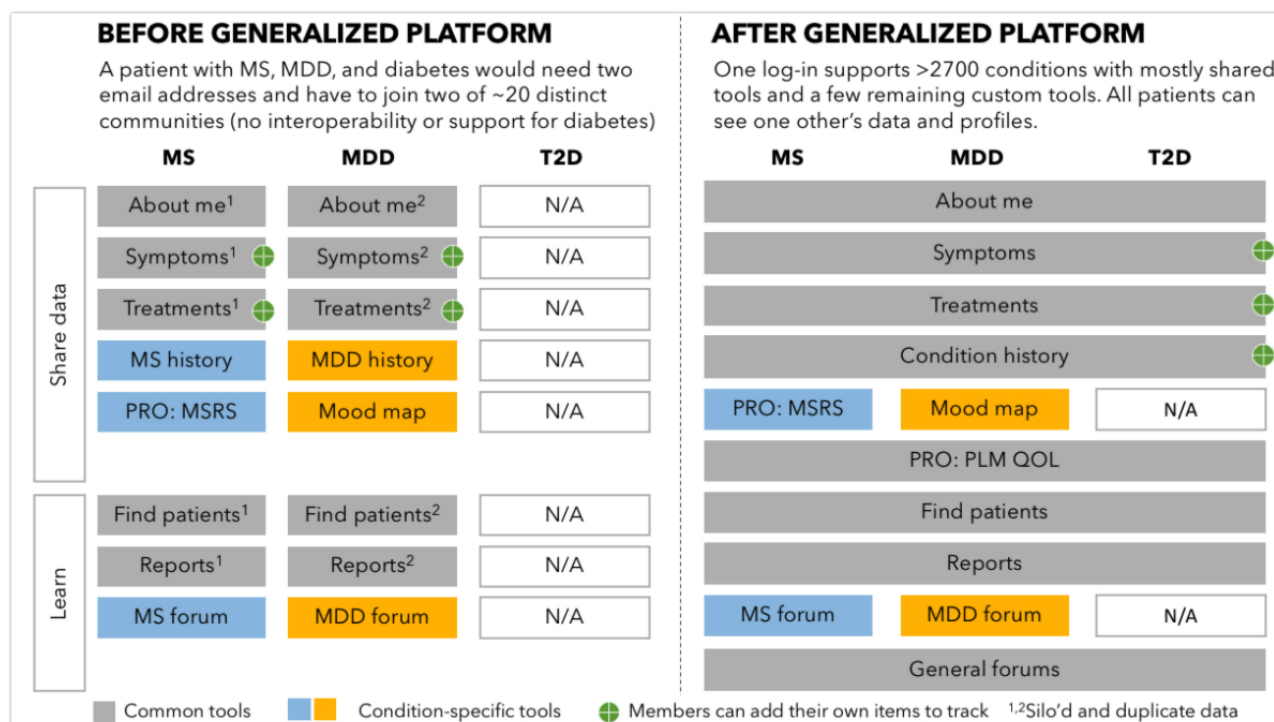
Goals of This Study

As a prespecified primary hypothesis, we sought to test whether members of the legacy vertical communities and members with community upgrades would report more benefits than generalized platform members who only had generic tools.

As a secondary prespecified hypothesis, we sought to test whether members who engaged more with the site would experience more benefits [20].

Methods

Recruitment

The invitation to participate was fielded to all registered members aged 18 years and older, except those who had opted out of research. Eligible members were invited via email or private message. Users who took no action within 3 days or who started but did not complete the survey were sent a reminder message. Patients were not compensated for participation. Presentation of response rate information including view rate, participation rate, and completion rate, are provided based on Checklist for Reporting Results of Internet E-Surveys (CHERRIES) [27]. This study was reviewed by the New England Institutional Review Board (NEIRB# 16-117) and, as a minimal risk survey, was exempted from further review.

Due to a technical error, some individuals were sent more than one invitation; these were only counted once. If they accidentally completed the survey twice, the latter was excluded. All profile data submissions were optional, so some demographic data are missing.

Patient-Perceived Benefits of PatientsLikeMe Participation

A custom cross-sectional, retrospective survey was developed to update and build on previous work [19,20]. Level of patient activation was assessed using a validated instrument, the Patient Activation Measure (PAM) [28]. The 13-item validated PAM assesses a consumer's knowledge, skills, and confidence for self-management. It groups people into one of four levels of activation (1: disengaged and overwhelmed; 2: becoming aware but struggling; 3: taking action; 4: maintaining behaviors and pushing further). Participants were also asked about their recent health utilization (eg, emergency department visits in the past year), general health perception (SF-1), and physician satisfaction. Respondents were asked to estimate how many peers were in their network before and after joining PatientsLikeMe. All survey questions are shown in Multimedia Appendix 1. Demographic data for survey responders and nonresponders were obtained from existing member profile data.

Statistical Analysis

Data Exclusion and Disease Categories

Analysis was limited to patients reporting a physician's diagnosis of a given condition. Patients were asked to designate a primary condition on their PatientsLikeMe profile; however, they were not required to do so. The median number of comorbidities is reported along with a frequency distribution of those conditions patients considered to be primary. For descriptive and summary purposes, conditions were characterized into therapeutic areas (see Multimedia Appendix 2).

A binary "community customization" variable indicated whether a community had been upgraded with additional functionality and features beyond the general platform experience, regardless of whether it had been a legacy vertical community or benefited from a community upgrade.

Reported Benefits in Relation to Site Use

Length of time patients had been registered at the time of survey ("tenure") was used as a covariate. Three site engagement metrics were generated: (1) number of days spent on the PatientsLikeMe platform ("sessions"), (2) number of days patients conducted social activity on the site ("social"), and (3) whether they had donated any structured data to the platform ("data donation"). Each metric was collapsed into highest and lowest quartiles.

Descriptive Statistics

Descriptive statistics were calculated across all survey questions and PAM scores. Summary statistics for continuous variables included n, mean, standard deviation, and range. Median and range were generated for nonnormally distributed variables. Categorical variables were summarized as frequencies and percentages. Patient-perceived benefit questions included a "does not apply" response option because some questions were not relevant to patients (eg, changing treatment if they were not receiving treatment), which were reviewed and removed from percent calculations.

The PAM scores were calculated according to recommended scoring guidelines from the scale developers, including the four levels of activation [25]. The frequency distributions of 15 custom benefit questions were reviewed among the total survey population. A "benefits score" was generated by summing each of the 15 benefits coded as 1 (present) versus 0 (absent).

Chi-square test of independence tested for associations between categorical variables, two-sample t tests compared groups on normally distributed continuous variables, with Wilcoxon-Mann-Whitney tests for nonparametric comparisons. One-sample equivalents were performed to compare benefits observed for select conditions to the overall sample. Number of benefits within the overall survey population was stratified based on the presence or absence of community customization, and P values adjusted for multiple comparisons using Bonferroni correction.

Statistical Models

Separate univariate models assessed the relationship between the dependent variable (number of benefits reported or benefits score) and each independent variable including age, gender, site engagement metrics, patient activation, tenure, and community customization. A review of the sessions and social site metric variables indicated a high correlation and the sessions variable was retained for the final model. All independent variables shown to be significant in the univariate analyses were included in the final multivariable model.

Each model run (univariate and multivariable) was specified as a log-binomial model to estimate the "benefits score" accounting for the independent variables described previously. The response variable "benefits score" was specified as the ratio of events per trials; that is, the number of benefits reported out of the total number of benefits (out of 15). For interpretation purposes, results multiplied by the inverse of this ratio (15) can be provided as the estimate (least square means) of the count of benefits. Relative risks (RRs) were estimated and presented in this model using estimate statements for the appropriate contrasts. An RR greater than 1 is interpreted as a greater chance ("risk") of an additional benefit compared to the reference category for a given independent variable. The alpha level was set to .05 and all analyses were run in SAS 9.4 by authors ET and CC.

Results

Recruitment

A total of 377,625 invitations were sent to PatientsLikeMe members between April 11, 2016 and June 20, 2017. By study close, 26,048 of the 377,625 (6.90%) sent an invitation opened the email or private message. Of 26,048 potential respondents who viewed the invitation, 9025 (34.65%) began the survey and 7434 (82.37%) of these completed the survey (Figure 2). Respondents (N=7434) were most frequently white (6106/7052, 86.59%), female (5290/7349, 71.98%), and educated (5062/6026, 84.00% at least some college). Nonresponders were younger, less likely to be white, less educated, and less likely to be on Medicare (see Table 1). Due to a technical error, some individuals were sent more than one invitation; these were only counted once. If they accidentally completed the survey twice, the latter was excluded. All profile data submission was optional, so some demographic data may be missing.

Most patients characterized themselves as being in "fair" or "good" health (5163/7434, 69.45%) and were seeing a specialist (4456/7434, 59.94%; Table 2). Most respondents had at least some level of difficulty with medication adherence (4628/7434, 62.25%). Patient activation was distributed bimodally with the largest percentage at level 3 (2639/7434, 35.50%), and a second peak at level 1 (1858/7434, 24.99%; Table 3).

Patient-Perceived Benefits of PatientsLikeMe Participation

Most respondents agreed PatientsLikeMe has furthered their understanding of how their condition could affect them (4530/6770, 66.91%) and how to live better with their condition (4247/6750, 62.92%; Table 4). A greater proportion of patients with community customization reported benefits relating to knowledge and understanding of their condition and this was consistently higher than from those in communities without customization (all P<.001; Table 4.). Benefits related to positive changes in their condition management and treatment were endorsed at lower rates and showed more variability (Table 4). However, the only benefits for which the enhanced communities did not achieve a significantly greater benefit than the generic communities were better conversations with a health provider, managing symptoms better, stopping a treatment, and changing doctors.

Figure 2. Participant flow.
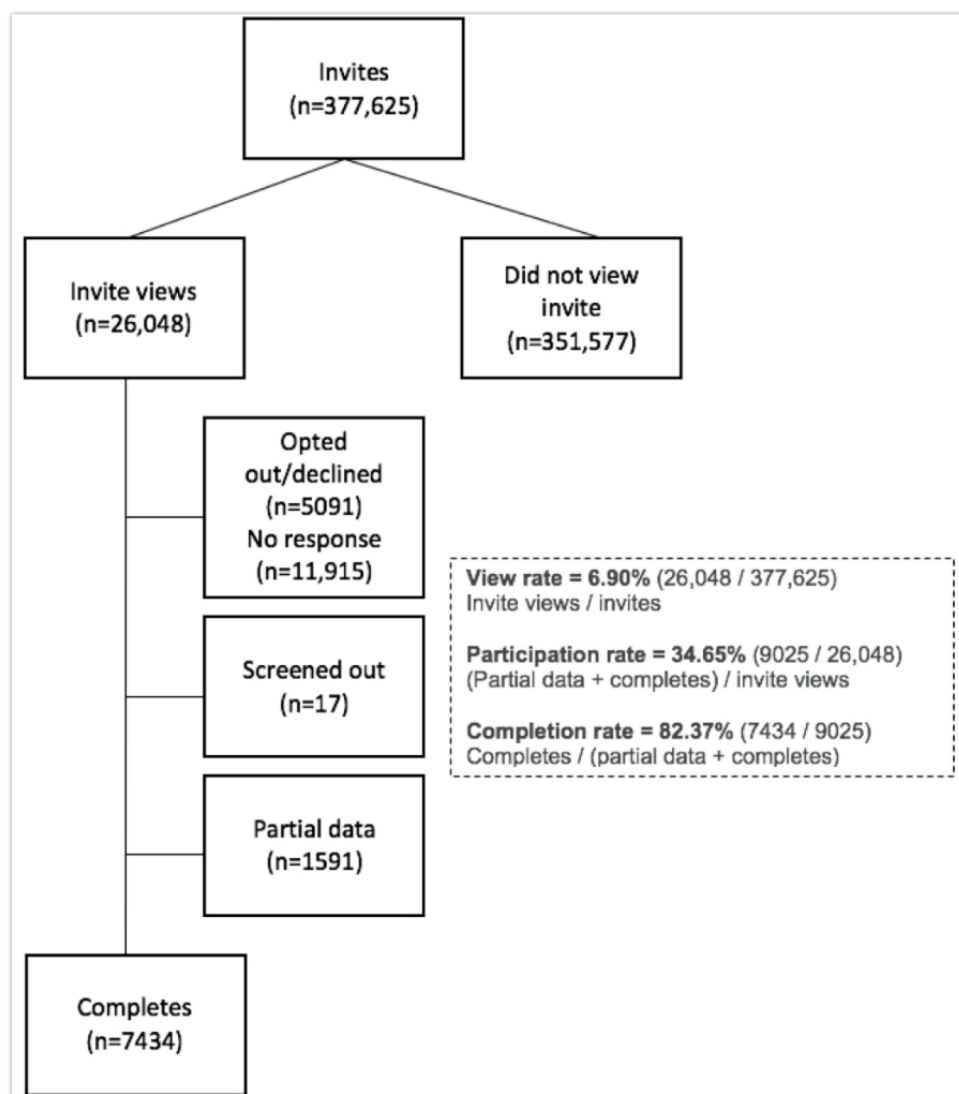
The mean number of benefits reported overall was 6.20 (SD 4.25) out of a possible total of 15 benefits, with a median of 7 (IQR 2-10). A total of 447 primary conditions were represented among survey respondents, with the top 20 conditions shown in Table 5 and across all conditions in Multimedia Appendix 3. The mean count of comorbid conditions in this survey population was 3.65 (SD 4.87) conditions, with a median of 2 (IQR 1-4). Additional analysis of the mean and median benefits for each condition (irrespective of primary condition) is presented in Multimedia Appendix 4 (n=1657 conditions represented in the Impact Survey, with 189 conditions having ≥30 patients reporting).

Reported Benefits in Relation to Site Use

The mean number of benefits reported was significantly higher among the group with community customization than those without (mean 6.57, SD 4.09 vs mean 5.60, SD 4.43; $t_{4164}=-8.02$, $P<.001$) and a similar observation was made considering the median number of benefits (median 7, IQR 3-10 vs median 6, IQR 1-9; Kruskal-Wallis $\chi^2_1=67.3$, $P<.001$).

A log-binomial model estimated the number of benefits reported, controlling for community customization, tenure, active sessions,

data donation, patient activation, and demographics (Table 6). Results indicated a significant relationship between each independent variable reviewed and the number of benefits reported with the exception of "condition groupings" (Table 6). When all independent predictors were included, the expected number of benefits was 21% higher for members with community customization compared to no community customization (RR=1.21, $P<.001$). The expected number of benefits was 29% higher for patients at the fourth level of activation, compared to the first level (RR=1.29, $P<.001$). This effect was enhanced when limiting the benefits to only those related to knowledge and understanding of one's disease (42%, RR=1.42), but had less of an effect for those benefits associated with treatment and management of one's condition (13%, RR=1.13). When a member had a higher level of engagement on the site, the expected number of benefits was 11% higher (RR=1.11, $P<.001$) than for those engaging less. It is important to note that having received benefit from the site may have encouraged respondents to return to the site for more use; however, directionality of engagement could not be directly assessed in this model.

Table 1. Demographics of respondents and nonrespondents. IQR: interquartile range.

Characteristic	Responders (n=7434)	Nonresponders (n=370,191)	P value[a]
Age (years), n	7419	321,886	<.001[b]
Mean (SD)	54 (12)	48 (14)	
Median (IQR)	54 (46-62)	48 (38-58)	
Sex, n (%)			.61
Female	5290 (71.16)	232,260 (71.60)	
Male	2059 (27.70)	92,111 (28.39)	
Prefer to skip	0	21 (<0.01)	
Race, n (%)			<.001
White	6106 (86.59)	163,857 (80.65)	
Black	302 (4.28)	9996 (4.92)	
Mixed race	291 (4.13)	7591 (3.74)	
Asian	99 (1.40)	8653 (4.26)	
Native American	90 (1.28)	2175 (1.07)	
Hawaiian	6 (<0.01)	514 (0.25)	
Prefer to skip	158 (2.24)	10,397 (5.12)	
Education, n (%)			<.001
High school grad or less	898 (14.90)	18,784 (19.51)	
Some college	2365 (39.25)	36,172 (37.58)	
College graduate	1565 (25.97)	22,881 (23.77)	
Post graduate	1132 (18.79)	15,344 (15.94)	
Prefer to skip	66 (1.10)	3080 (3.20)	
Insurance, n (%)			<.001
Medicare	1696 (28.03)	15,518 (16.27)	
Employer	1802 (29.78)	31,994 (33.54)	
Medicaid	589 (9.73)	8412 (8.82)	
National	574 (9.49)	10,691 (11.21)	
Direct	514 (8.49)	8066 (8.46)	
Veterans Affairs	153 (2.53)	1890 (1.98)	
Military	115 (1.90)	1729 (1.81)	
Indian Health Service	2 (<0.01)	108 (0.11)	
Other	97 (1.60)	1949 (2.04)	
None	351 (5.80)	8580 (8.99)	
Prefer to skip	158 (2.61)	6451 (6.76)	

[a]P values from t test for age and chi-square test for categorical variables.

[b]Satterthwaite method reported due to unequal variances.

Table 2. Health utilization and clinical characteristics of survey respondents (N=7434). IQR: interquartile range.

Characteristic	Value
General health (SF-1), n (%)	
Excellent	198 (2.66)
Very good	968 (13.02)
Good	2596 (34.92)
Fair	2567 (34.53)
Poor	1105 (14.86)
Physician type, n (%)	
A primary care physician	2325 (31.28)
An internist at a hospital	101 (1.36)
A specialist in my condition	4456 (59.94)
I don't see a physician	278 (3.74)
I don't know	183 (2.46)
Prefer to skip	91 (1.22)
Treatment satisfaction, n (%)	
Extremely dissatisfied	433 (5.82)
Very dissatisfied	527 (7.09)
Dissatisfied	1013 (13.63)
Somewhat satisfied	2326 (31.29)
Satisfied	1537 (20.68)
Very satisfied	817 (10.99)
Extremely satisfied	397 (5.34)
Does not apply	384 (5.17)
Health utilization past 12 months, mean (SD), median (IQR)	
Emergency department visits	1.26 (3.52), 0 (0-1)
Overnights in the hospital	2.41 (10.66), 0 (0-1)
Separate hospital stays	1.26 (6.97), 0 (0-1)
Adherence (how often do you have difficulty remembering to take your medications?), n (%)	
Never/rarely	2614 (35.16)
Once in a while	2211 (29.74)
Sometimes	1608 (21.63)
Usually	514 (6.91)
All the time	295 (3.97)
Does not apply	192 (2.58)

Table 3. Patient Activation Measure (PAM) distribution among survey respondents (N=7434).[a]

PAM level	Description	n (%)
Level 1	Does not feel in charge of their own health and care. Managing health is overwhelming for them with all of life's other challenges. Lacks confidence in their ability to manage health. Has few problem-solving skills and poor coping skills. They may not be very aware of own behavior.	1858 (24.99)
Level 2	May lack basic knowledge about their condition, treatment options, and/or self-care. Have little experience or success with behavior change. Look to their doctor to be the one in charge. Low confidence in their ability to manage health.	1449 (19.49)
Level 3	Have the basic facts of their conditions and treatments. Some experience and success in making behavioral changes. Some confidence in handling limited aspects of their health	2639 (35.50)
Level 4	Have made most of the necessary behavior changes, but may have difficulty maintaining behaviors over time or during times of stress	1312 (17.65)

[a]A total of 176 individuals are missing due to selection of the "I prefer to skip" option.

Table 4. Perceived benefits among all survey respondents (N=7434).

Benefit	Respondents, n (%)			P value[a]
	Total survey respondents (N=7434)[b]	With community customization (n=5344)	Without community customization (n=2045)	
Knowledge and understanding of condition (Has PatientsLikeMe improved your understanding of...)				
How your condition(s) might affect you?	4530 (66.91)	2987 (71.05)	918 (59.11)	<.001
What might help you live better with your condition(s)?	4247 (62.92)	2805 (66.77)	852 (54.97)	<.001
Treatment side effects?	4182 (60.59)	2747 (64.20)	854 (53.58)	<.001
Available treatments?	4143 (60.06)	2765 (64.69)	819 (51.32)	<.001
Important factors in making decisions about treatments?	3919 (57.52)	2538 (60.40)	808 (51.04)	<.001
What might help you get better?	3339 (50.37)	2159 (52.76)	664 (43.40)	<.001
How to deal with other problems in your life (eg, stress, work, money) that may be caused by your condition(s)?	3250 (47.86)	2102 (49.94)	679 (43.30)	<.001
Condition treatment and management (As a result of PatientsLikeMe have you...)				
Had better conversations with your health care professionals?	3592 (51.62)	2344 (54.61)	762 (47.24)	<.001
Managed your symptoms better?	3179 (45.64)	2131 (49.51)	634 (39.28)	<.001
Been better at taking your medication?	2251 (33.74)	1437 (34.94)	475 (30.55)	.14
Tried a new way to manage side effects?	2089 (29.79)	1374 (32.73)	440 (27.01)	<.001
Asked to see a specialist doctor?	1699 (24.79)	1011 (24.06)	411 (25.77)	>.99
Started a new treatment?	1046 (14.71)	692 (15.77)	196 (11.90)	.02
Stopped a treatment?	993 (13.91)	651 (14.80)	220 (13.27)	.09
Changed your doctor?	899 (12.59)	546 (12.38)	220 (13.34)	>.99

[a]P values adjusted using Bonferroni correction.

[b]Percentages calculated out of valid nonmissing data and after removal of "does not apply" responses. A range of approximately 3% to 10% were observed across benefit questions.

XSL•FO

RenderX

Table 5. Distribution of conditions most frequently represented in the survey, ordered by condition group then primary condition (N=6264).[a] IQR: interquartile range.

Patient-reported primary condition	Condition group	n (%)	Mean (SD)	Median (IQR)
Diabetes type 2	Cardiovascular and metabolic diseases	174 (2.78)	6.3 (3.9)	7 (3-9)
Fibromyalgia	Fibromyalgia (other)	668 (10.66)	6.5 (4.0)	7 (3-10)
Crohn's disease	Inflammation and autoimmunity	49 (0.78)	5.0 (4.2)	5 (1-8)
Rheumatoid arthritis	Inflammation and autoimmunity	128 (2.04)	6.4 (4.1)	7 (3-9)
Systemic lupus erythematosus	Inflammation and autoimmunity	271 (4.33)	6.0 (4.0)	6 (2-9)
Bipolar disorder	Mental and behavioral health	245 (3.91)	6.1 (4.1)	7 (2-10)
Bipolar I disorder	Mental and behavioral health	96 (1.53)	6.9 (4.5)	7 (3-10)
Bipolar II disorder	Mental and behavioral health	130 (2.08)	6.2 (4.0)	7 (3-10)
Complex post-posttraumatic stress disorder	Mental and behavioral health	54 (0.86)	6.5 (4.4)	7 (3-10)
Major depressive disorder	Mental and behavioral health	308 (4.92)	5.7 (3.9)	6 (2-8)
Posttraumatic stress disorder	Mental and behavioral health	232 (3.70)	6.4 (4.1)	7 (3-9)
Traumatic brain injury	Mental and behavioral health	122 (1.95)	6.3 (4.3)	7 (3-10)
Myalgic encephalomyelitis/chronic fatigue syndrome	Myalgic encephalomyelitis/chronic fatigue syndrome (other)	62 (0.99)	5.1 (4.0)	5 (1-8)
Amyotrophic lateral sclerosis	Neurologic	281 (4.49)	5.8 (3.6)	6 (3-9)
Epilepsy	Neurologic	109 (1.74)	7.2 (4.5)	8 (4-11)
Multiple sclerosis	Neurologic	1005 (16.04)	6.5 (3.9)	7 (3-10)
Parkinson's disease	Neurologic	468 (7.47)	6.2 (4.1)	7 (3-9)
Lung cancer	Oncology	98 (1.56)	5.8 (4.0)	6 (3-9)
Multiple myeloma	Oncology	93 (1.48)	4.9 (3.9)	5 (1-8)
Idiopathic pulmonary fibrosis	Respiratory	150 (2.39)	6.2 (3.8)	6 (3-9)

[a]Primary condition not reported and/or profile data unavailable for n=1170 patients.

XSL•FO
RenderX

Table 6. Site use-related predictors of patient-perceived health benefits in the Impact Survey: univariate and multivariate models. Ref: reference group.

Independent variables	Univariate models			Multivariate models				
	Dependent variable: total number of benefits (out of 15)			Dependent variable: total number of benefits (out of 15)			Benefits related to knowledge/understanding of one's condition (out of 7)	Benefits related to treatment/management of one's condition (out of 8)
	β	*P*	RR[a]	β	*P*	RR[a]	RR	RR
Age (years)[b]								
18-39	0.12	<.001	1.13	-0.02	0.10	0.98	1.00	0.97
40-55	0.09	<.001	1.10	-0.14	<.001	1.12	0.91	0.85
≥55	0	—	Ref	0	—	Ref	Ref	Ref
Gender								
Male	0	—	Ref	0	—	Ref	Ref	Ref
Female	0.07	<.001	1.08	0.07	<.001	1.07	1.04	1.16
Tenure at baseline[c]	0.02	<.001	1.02	0.02	<.001	1.02	1.02	1.02
Community customization								
None	0	—	Ref	0	—	Ref	Ref	Ref
Customized	0.16	<.001	1.17	0.19	<.001	1.21	1.24	1.13
Patient activation								
Level 1	0	—	Ref	0	—	Ref	Ref	Ref
Level 2	0.15	<.001	1.16	0.14	<.001	1.15	1.21	1.06
Level 3	0.21	<.001	1.24	0.21	<.001	1.23	1.32	1.12
Level 4	0.25	<.001	1.29	0.25	<.001	1.29	1.42	1.13
Sessions								
Low engagement	0	—	Ref	0	—	Ref	Ref	Ref
Mod-high engagement	0.13	<.001	1.14	0.10	<.001	1.11	1.05	1.19
Condition grouping								
Cardiovascular and metabolic diseases	0.04	.10	1.04	0.18	<.001	1.19	1.22	1.07
Infectious disease	–0.09	.08	0.91	–0.09	.08	0.91	0.87	0.98
Inflammation and autoimmunity	–0.04	.03	0.96	–0.01	.52	0.99	1.03	0.89
Mental and behavioral health	0.07	<.001	1.07	0.05	.002	1.05	1.03	1.05
Neurologic	0.07	<.001	1.07	0.01	.58	1.01	1.08	0.89
Oncology	–0.19	<.001	0.82	–0.07	.006	0.93	1.03	0.74
Respiratory	0.05	.07	1.05	0.10	.002	1.10	1.15	1.03
Other	0	—	Ref	0	—	Ref	Ref	Ref

[a]Relative risk (RR) greater than 1 represents a greater chance ("risk") of an additional benefit compared to the reference category.

[b]Representation of the survey population in the category 18-24 years was quite low (2%) and thus was combined with the category 25-39 years to create the 18-39 years category. As seen in Table 1, the mean age in the population was 54.

[c]Tenure at baseline variable converted to years (from days) for interpretation purposes.

Discussion

Principal Results

This study confirmed that members joining PatientsLikeMe in the "generalized platform" do perceive a variety of benefits from their participation on the site. The majority of members learned more about how their condition might affect them, what might help them live better with their condition, how to manage treatment side effects, to be more aware of existing treatments, and had an improved understanding of what might help them get better. Although the average member perceived around six

XSL•FO
RenderX

of a possible 15 benefits from using the site, this increased to seven benefits in conditions that had some degree of community customization, confirming our primary hypothesis. Within members of a "customized" condition, benefits were higher for the most engaged members who logged in, socialized, or entered data the most frequently, partially confirming our secondary hypothesis (the effect was absent for those without customization).

It is worth considering whether the extra design, research, testing, and coding that goes into customizing a community is worth the increase of a single reported benefit. Of the 2700 or so communities represented on the platform, the vast majority have not received site customization and yet perceive a similar number of benefits. This may reflect some "floor effect" where the basic functionality of the site, which permits any member to record their health-related quality of life, connect with others, and optionally track their own self-selected symptoms and treatments, is already doing a reasonable job at fulfilling a patient's needs. One advantage of additional community customization has been the addition of outcome measures of clinical relevance, such as oxygen use in idiopathic pulmonary fibrosis or body surface area coverage in psoriasis. Future developments on the site aim to make it much simpler (and therefore more scalable) to permit some degree of community customization through simple changes made on an "admin dashboard" rather than requiring new code. We also plan to develop and psychometrically validate a modular patient-reported outcome system that can then be validated clinically against gold standard measures in various conditions.

Limitations

There are several limitations of this study that are important to consider. Although putting the response rates of surveys in online communities into perspective is not straightforward [27], the responses provided here represent just 2% of the whole community. However, that denominator includes patients who registered a decade ago, some of whom may be deceased, too disabled to participate, or who may even have recovered to some extent from their condition. The finding that the most active ~1% of online community membership is responsible for much of its positive impact has been observed in other online health communities [15].

Given the cross-sectional design, directionality of benefits and site use characteristics could not be discerned. Although the extent of customization for communities varies, this level of granularity was not reviewed in favor of a simpler indicator. The benefits reported were self-reported with no independent validation and no control group. Furthermore, due to the potential selection bias, social desirability bias, and given that the data represent a convenience sample, the results are not generalizable to the overall population, particularly given the skew toward a more educated, female population. Another limitation is the complexity of real-world data as it relates to comorbid conditions. To minimize complexity, condition-specific data was limited to those patients who had reported a primary condition and indicated a diagnosis of this condition on the site. However, not all patients report a primary condition, this designation may change over time, and the

restriction to a primary condition does not account for the reality of complex, comorbid conditions.

Some comparisons presented in the study were not all specified a priori and are presented as exploratory results. Finally, although receiving more benefits from having participated on the site is preferable to fewer benefits, it is also possible that having received even one benefit, such as having better conversations with your health care provider or starting a new treatment, is meaningful for a patient. No weight was given to one benefit over another.

It may be interesting in future studies to evaluate combinations of benefits or to explore latent class characterizations of the latent categories or typologies of benefits.

Comparison With Prior Work

Overall, these findings align with earlier studies reported in the literature [19,20], although due to questions being asked differently it is not possible to compare benefits like-for-like. Strengths of this study include the relatively large sample size and the diversity of conditions represented, from mood disorders to infections to neurological conditions to oncology. Current literature shows that patients using other online communities benefit too, but studies have generally been limited to emotional benefits in a few communities for specific conditions or a study of one specific disease state, and often as part of an interventional education program rather than peer-to-peer communication.

A study of Reddit forums for depression found that engagement was linked to more positive emotional states and an increased use of positive words [29]. A study in the United States found that women with breast cancer who participated in a 12-week Internet-based social group reported reduced depression, stress, and cancer-related trauma [30]. Additionally, a study in Japan also focused on breast cancer patients found that online support groups provided benefits to participants, but patients who posted in the forums felt they received more benefits than the "lurkers" did [31]. This study uniquely adds to the literature in that PatientsLikeMe has been open to all patients since 2011. The main benefits reported reflect the intent of the website, which is to help patients find other patients like them, track their own personal health data over time to discover insights about their health, and to share their experiences to benefit others.

Benefits reported at lower rates, such as managing treatment side effects, asking to see a specialist doctor, or changing doctor, do not have any specific tools or functionality assigned to them on the site. Such changes in health behavior may arise from serendipitous social interactions with other members in community spaces, increased awareness of treatment options and coping methods identified on treatment reports, or increases in patient activation as members become more aware of their bodies and health state.

Findings also support PatientsLikeMe's business model. Although imperfect, other business models such as advertising or asking members to pay a subscription fee have their own challenges. Relying entirely on grant income presents a challenge to stability and continuity for an enterprise that has been operating for over a decade. The number of benefits

reported by patients identifying with conditions for which there was community customization and patient-centric research was 18% higher than for patients reporting conditions without this additional functionality and opportunity. From 2011 to the present, the majority of investment in site customization and patient-centric research has come from PatientsLikeMe partnerships with pharmaceutical companies such as AstraZeneca [9] and research grants from funders such as the Robert Wood Johnson Foundation. This study therefore demonstrates that commercial partnerships may have a positive impact on members who might not otherwise have found the site, while simultaneously enabling partners to learn about patient priorities for new therapies and implement patient-centric programs such as improving their clinical trials [32].

It is also important to note that although community customization was one factor associated with benefits, patient factors were important too, such as engagement and activation. Previous research on patient activation has indicated that 41% of US adults are characterized as being in the highest level of activation (level 4), with approximately 7% at level 1 and 15% at level 2 [33]. Lower activation has been found in those with low income, less education, Medicaid enrollees, and people with poor self-reported health [33]. Patient respondents in this survey were less activated than the general population, perhaps reflecting the higher burden they face as a result of their chronic conditions, or identifying the reason they might have sought the assistance of an online community in the first place.

Future Research

To further elucidate who benefits most from participation in the site, a study is currently being conducted to evaluate benefits over time (6 weeks) with a pre-post longitudinal study design to explore patterns in patient-perceived benefits as well as changes in patient activation overall and by conditions.

Conclusion

Online health communities offer an opportunity for patients to connect and share information that provides benefits in their health management. Generalized health communities offer benefits to users and partnering with stakeholders in patient health to enhance and customize these communities is one way to offer greater benefits to patients.

Multimedia Appendix 1

Impact survey questionnaire.

[PDF File (Adobe PDF File), 100KB - jmir_v20i5e175_app1.pdf]

Multimedia Appendix 2

List of "other" conditions.

[PDF File (Adobe PDF File), 32KB - jmir_v20i5e175_app2.pdf]

Multimedia Appendix 3

Benefits for all primary conditions.

[XLSX File (Microsoft Excel File), 49KB - jmir_v20i5e175_app3.xlsx]

Multimedia Appendix 4

Benefits for all primary and secondary conditions.

[XLSX File (Microsoft Excel File), 125KB - jmir_v20i5e175_app4.xlsx]

XSL•FO
RenderX

References

1. Rheingold H. The Virtual Community: Homesteading on the Electronic Frontier. Cambridge: MIT Press; 1993.

2. Lester J, Prady S, Finegan Y, Hoch D. Learning from e-patients at Massachusetts General Hospital. BMJ 2004 May 15;328(7449):1188-1190 [FREE Full text] [doi: 10.1136/bmj.328.7449.1188] [Medline: 15142930]

3. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. BMJ 2004 May 15;328(7449):1166 [FREE Full text] [doi: 10.1136/bmj.328.7449.1166] [Medline: 15142921]

4. Hoch D, Ferguson T. What I've learned from E-patients. PLoS Med 2005 Aug;2(8):e206 [FREE Full text] [doi: 10.1371/journal.pmed.0020206] [Medline: 16060721]

5. Grajales FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media: a review and tutorial of applications in medicine and health care. J Med Internet Res 2014;16(2):e13 [FREE Full text] [doi: 10.2196/jmir.2912] [Medline: 24518354]

6. Frost J, Vermeulen IE, Beekers N. Anonymity versus privacy: selective information sharing in online cancer communities. J Med Internet Res 2014;16(5):e126 [FREE Full text] [doi: 10.2196/jmir.2684] [Medline: 24828114]

7. van der Velden M, Emam El K. "Not all my friends need to know": a qualitative study of teenage patients, privacy, and social media. J Am Med Inform Assoc 2013 Jan 1;20(1):16-24 [FREE Full text] [doi: 10.1136/amiajnl-2012-000949] [Medline: 22771531]

8. Brownstein CA, Brownstein JS, Williams DS, Wicks P, Heywood JA. The power of social networking in medicine. Nat Biotechnol 2009 Oct;27(10):888-890. [doi: 10.1038/nbt1009-888] [Medline: 19816437]

9. Yeoman G, Furlong P, Seres M, Binder H, Chung H, Garzya V, et al. Defining patient centricity with patients for patients and caregivers: a collaborative endeavour. BMJ Innov 2017 Apr;3(2):76-83 [FREE Full text] [doi: 10.1136/bmjinnov-2016-000157] [Medline: 28890797]

10. Stock D. UWSpace. 2015. Posthumanist medicine: participatory healthcare, medical humanities, and digital media URL: https://uwspace.uwaterloo.ca/handle/10012/9568 [WebCite Cache ID 6wfCroLvP]

11. Brubaker J, Lustig C. PatientsLikeMe: empowerment and representation in a patient-centered social network. 2010 Feb 6 Presented at: 2010 ACM Conference on Computer Supported Cooperative Work; Feb 6-10, 2010; Savannah, GA URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.460.6665&rep=rep1&type=pdf

12. Whitmore L, Kempner J. Gnovis. 2013 Apr 26. Our bodies, their data: tracing biopolitical circuitry & the illness experience on PatientsLikeMe URL: http://www.gnovisjournal.org/2013/04/26/our-bodies-their-data-tracing-biopolitical-circuitry-the-illness-experience-on-patientslikeme-com/ [WebCite Cache ID 6wfDAAE6H]

13. Frost JH, Massagli MP. Social uses of personal health information within PatientsLikeMe, an online patient community: what can happen when patients have access to one another's data. J Med Internet Res 2008;10(3):e15 [FREE Full text] [doi: 10.2196/jmir.1053] [Medline: 18504244]

14. Balatsoukas P, Kennedy CM, Buchan I, Powell J, Ainsworth J. The role of social network technologies in online health promotion: a narrative review of theoretical and empirical factors influencing intervention effectiveness. J Med Internet Res 2015;17(6):e141 [FREE Full text] [doi: 10.2196/jmir.3662] [Medline: 26068087]

15. van Mierlo T. The 1% rule in four digital health social networks: an observational study. J Med Internet Res 2014;16(2):e33 [FREE Full text] [doi: 10.2196/jmir.2966] [Medline: 24496109]

16. Hamm MP, Chisholm A, Shulhan J, Milne A, Scott SD, Given LM, et al. Social media use among patients and caregivers: a scoping review. BMJ Open 2013;3(5):e002819 [FREE Full text] [doi: 10.1136/bmjopen-2013-002819] [Medline: 23667163]

17. Stellefson M, Chaney B, Barry AE, Chavarria E, Tennant B, Walsh-Childers K, et al. Web 2.0 chronic disease self-management for older adults: a systematic review. J Med Internet Res 2013 Feb;15(2):e35 [FREE Full text] [doi: 10.2196/jmir.2439] [Medline: 23410671]

18. Frick W. Harvard Business Review. 2013 Oct 31. What to do when an online community starts to fail URL: https://hbr.org/2013/10/what-to-do-when-an-online-community-starts-to-fail [accessed 2018-01-10] [WebCite Cache ID 6wfDLm7SM]

19. 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]

20. Wicks P, Keininger DL, Massagli MP, de Loge LC, Brownstein C, Isojärvi J, et al. Perceived benefits of sharing health data between people with epilepsy on an online platform. Epilepsy Behav 2012 Jan;23(1):16-23 [FREE Full text] [doi: 10.1016/j.yebeh.2011.09.026] [Medline: 22099528]

21. Hixson JD, Barnes D, Parko K, Durgin T, Van Bebber S, Graham A, et al. Patients optimizing epilepsy management via an online community: the POEM Study. Neurology 2015 Jul 14;85(2):129-136 [FREE Full text] [doi: 10.1212/WNL.0000000000001728] [Medline: 26085605]

22. Cedarbaum JM, Stambler N, Malta E, Fuller C, Hilt D, Thurmond B, et al. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group (Phase III). J Neurol Sci 1999 Oct 31;169(1-2):13-21. [Medline: 10540002]

23. Wicks P, Vaughan TE, Massagli MP. The multiple sclerosis rating scale, revised (MSRS-R): development, refinement, and psychometric validation using an online community. Health Qual Life Outcomes 2012 Jun 18;10:70 [FREE Full text] [doi: 10.1186/1477-7525-10-70] [Medline: 22709981]

24. Gupta S, Riis J. Harvard Business Review. 2011 Jan 21. PatientsLikeMe: an online community of patients URL: https:/
     /papers.ssrn.com/sol3/papers.cfm?abstract_id=2012734 [accessed 2018-01-22] [WebCite Cache ID 6wfDQ4xtP]

25. Tempini N. Till data do us part: understanding data-based value creation in data-intensive infrastructures. Inform Organ
     2017 Dec;27(4):191-210. [doi: 10.1016/j.infoandorg.2017.08.001]

26. Slawsky K, Massagli M, Wicks P. PRM30 A Comparison of the PatientsLikeMe Quality of Life Questionnaire (PLM QOL)
     with the Rand SF-36. Value Health 2011 Nov;14(7):A426. [doi: 10.1016/j.jval.2011.08.1061]

27. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
     J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]

28. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing
     and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text]
     [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]

29. Park A, Conway M. Longitudinal changes in psychological states in online health community members: understanding the
     long-term effects of participating in an online depression community. J Med Internet Res 2017 Mar 20;19(3):e71 [FREE
     Full text] [doi: 10.2196/jmir.6826] [Medline: 28320692]

30. Winzelberg AJ, Classen C, Alpers GW, Roberts H, Koopman C, Adams RE, et al. Evaluation of an internet support group
     for women with primary breast cancer. Cancer 2003 Mar 1;97(5):1164-1173 [FREE Full text] [doi: 10.1002/cncr.11174]
     [Medline: 12599221]

31. Doolittle GC, Spaulding A. Online cancer services: types of services offered and associated health outcomes. J Med Internet
     Res 2005 Jul 01;7(3):e35 [FREE Full text] [doi: 10.2196/jmir.7.3.e35] [Medline: 15998626]

32. Lowe MM, Blaser DA, Cone L, Arcona S, Ko J, Sasane R, et al. Increasing patient involvement in drug development.
     Value Health 2016;19(6):869-878 [FREE Full text] [doi: 10.1016/j.jval.2016.04.009] [Medline: 27712716]

33. Hibbard JH, Cunningham PJ. How engaged are consumers in their health and health care, and why does it matter? Res
     Brief 2008 Oct(8):1-9. [Medline: 18946947]

Abbreviations

ALS: amyotrophic lateral sclerosis
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
IQR: interquartile range
MS: multiple sclerosis
MDD: major depressive disorder
MSRS: multiple sclerosis rating scale
PAM: Patient Activation Measure
PLM QOL: PatientsLikeMe Quality of Life Questionnaire
PRO: patient reported outcome
RR: relative risk
T2D: type 2 diabetes

XSL•FO
RenderX

Original Paper

Barriers and Facilitators Affecting Patient Portal Implementation from an Organizational Perspective: Qualitative Study.

Laura Kooij[1], MSc; Wim G Groen[1], PhD; Wim H van Harten[1,2,3], MD, PhD

[1]Division of Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Amsterdam, Netherlands

[2]Department of Health Technology and Services Research, University of Twente, Enschede, Netherlands

[3]Rijnstate Hospital, Arnhem, Netherlands

Corresponding Author:
Wim H van Harten, MD, PhD
Division of Psychosocial Research and Epidemiology
Netherlands Cancer Institute
Plesmanlaan 121
Amsterdam, 1066CX
Netherlands
Phone: 31 205122861
Fax: 31 206601449
Email: w.v.harten@nki.nl

Abstract

Background: The number of patient portals is rising, and although portals can have positive effects, their implementation has major impacts on the providing health care institutions. However, little is known about the organizational factors affecting successful implementation. Knowledge of the specific barriers to and facilitators of various stakeholders is likely to be useful for future implementations.

Objective: The objective of this study was to identify the barriers to and facilitators of patient portal implementation facing various stakeholders within hospital organizations in the Netherlands.

Methods: Purposive sampling was used to select hospitals of various types. A total of 2 university medical centers, 3 teaching hospitals, and 2 general hospitals were included. For each, 3 stakeholders were interviewed: (1) medical professionals, (2) managers, and (3) information technology employees. In total, 21 semistructured interviews were conducted using the Grol and Wensing model, which describes barriers to and facilitators of change in health care practice at 6 levels: (1) innovation; (2) individual professional; (3) patient; (4) social context; (5) organizational context; and (6) economic and political context. Two researchers independently selected and coded quotes by applying this model using a (deductive) directed content approach. Additional factors related to technical and portal characteristics were added using the model of McGinn et al, developed for implementation of electronic health records.

Results: In total, we identified 376 quotes, 26 barriers, and 28 facilitators. Thirteen barriers and 12 facilitators were common for all stakeholder groups. The facilitators' *perceived usefulness* (especially less paperwork) was mentioned by all the stakeholders, followed by subjects' *positive attitude*. The main barriers were *lack of resources* (namely, lack of staff and materials), *financial difficulties* (especially complying with high costs, lack of reimbursements), and *guaranteeing privacy and security* (eg, strict regulations). Both similarities and differences were found between stakeholder groups and hospital types. For example, managers and information technology employees mainly considered *guaranteeing privacy and security* as a predominant barrier. *Financial difficulties* were particularly mentioned by medical professionals and managers.

Conclusions: Patient portal implementation is a complex process and is not only a technical process but also affects the organization and its staff. Barriers and facilitators occurred at various levels and differed among hospital types (eg, *lack of accessibility*) and stakeholder groups (eg, *sufficient resources*) in terms of several factors. Our findings underscore the importance of involving multiple stakeholders in portal implementations. We identified a set of barriers and facilitators that are likely to be useful in making strategic and efficient implementation plans.

XSL•FO
RenderX

Introduction

Patient-centeredness is an important element of high-quality care: effective communication between patients and their health care professionals, and information access can both contribute considerably to this [1]. According to the Institute of Medicine, "patients should have unfettered access to their own medical information" [2] to support them in taking control of their health (eg, using medical information to make informed health-related decisions) [2]. Information technology (IT) can play an important role in improving access to this information [3], and it also improves the participation of patients in their own care [4]. In health care, an increasingly popular way to facilitate this is by using patient portals [5]. Patient portals can be defined as "applications which are designed to give the patient secure access to health information and allow secure methods for communication and information sharing" [6], as well as for administrative purposes [7], and are mostly provided by a single health care institution [6,8]. These portals are often connected to the electronic health record (EHR) of an institution—defined as tethered patient portals [9]—to provide access to patients' medical information [3,10-12]. Some institutions allow patient portals to facilitate communication between patients and health care professionals [3,6,12], view their appointments and provide patient education [11,13], share information [12], request for repeat medication prescriptions [3], and provide tailored feedback [11,13]. Patient portals may have a range of functionalities that enable information exchange (such as having access to the EHR), which in turn may facilitate and improve the communication between the patient and the health care professional [11,14]. Previous research showed that patients are especially satisfied with access to information from the EHR and the list of their appointments [11]. Portal use can also have a positive effect on self-management of conditions [15-18], communication between patients and providers, quality of care [16,17] and participation in treatment [17]. Patient empowerment can also be improved; the accessibility of information can especially contribute to "patients' knowledge" and their "perception of autonomy and being respected" [19]. On the other hand, effects on health outcomes are reported to be mixed [6]. In summary, patient portals can be important as they provide patients with access to their own medical information, enable interaction with their health care professionals [8], and aim to involve patients in their own care processes [1].

Although patient portals can have positive effects and may develop into a standard element of care [20], their implementation has major impacts on health care institutions as it often involves a complex change in an organization [1]. This can be affected by multiple factors at the micro (eg, "individuals"), meso (eg, "resources"), and macro (eg, "sociopolitical context") levels [21]. Several implementation models are available, such as "The Consolidated Framework for Implementation Research (CFIR)," which is used in many studies as a guiding framework [22-24]. CFIR consists of 5 levels at which barriers and facilitators can occur during implementation: (1) technology-related factors (eg, "adaptability," "complexity," and "cost"); (2) outer setting (eg, "policy and incentives"); (3) inner setting (eg, "resources"); (4) process (eg, "engagement of stakeholders"); and (5) individual health professionals (eg, "individual's knowledge"). In this model, patients are part of the "outer setting," suggesting that the CFIR framework is aimed primarily at institutions [24]. Another example is the "Fit between Individuals, Tasks, and Technology" (FITT) framework, which is aimed at the adoption of IT [25]. The comprehensive model of Grol and Wensing [26] summarizes the barriers to and facilitators of change in health care practice at 6 levels: (1) innovation; (2) individual professional; (3) patient; (4) social context; (5) organizational context; and (6) economic and political context. McGinn et al [21] argue that the consideration of various stakeholder opinions can contribute to successful implementations. However, previous research mainly focused on perceptions of single stakeholder groups regarding patient portal implementation, such as physicians [27] or nurses [28]. This highlights the importance of identifying the opinions of many stakeholders during patient portal implementation. Furthermore, it remains unclear which factors are important in accomplishing change in the various groups [26].

Previous research focused on patient involvement in developing patient portals [5,14], but little is yet known about organizational factors that facilitate or hinder patient portal implementation [6]. Such knowledge is essential because the number of portals is rising. In the Netherlands, in 2017, more than 25% of hospitals provided patients with access to a patient portal, whereas this was under 10% in 2015 [29]. Comprehensive information can provide a framework for upcoming patient portal implementations, or other eHealth applications, in hospitals. The objective of this study was, therefore, to identify the barriers and facilitators among the various stakeholders within hospital organizations in the Netherlands regarding the implementation of tethered patient portals.

Methods

Sampling Procedure

Purposive sampling was used to select hospitals of the 3 different types existing in the Netherlands. In total, 2 university medical centers (UMCs), 3 teaching and 2 general hospitals (including one collaborative oncology hospital comprising 3 general hospitals) were included. Hospitals were selected by means of convenience sampling using the authors' network or by Web searching, and hospitals in various phases of implementation (contemplation, preparation, or implementation) were included. Contact persons in the hospitals were approached by phone or email. Snowball sampling was used for the selection of respondents, meaning that we informed the contact persons about the objective of the study and also asked them for contact information for 3 stakeholders, including (1) medical professionals (doctor or nurse practitioners [Advanced Practice Registered Nurses]) [30], (2) managers, and (3) IT employees.

Table 1. Barriers and facilitators at various levels of Grol and Wensing.

Levels of Grol and Wensing [26]	Examples of barriers and facilitators
Innovation: patient portal	Accessibility, attractiveness, and credibility
Individual professional	Knowledge, attitude, and motivation to change
Patient	Knowledge, skills, and attitude
Social context	Opinions of colleagues, culture of the networks, and collaboration
Organizational context	Organization of care processes, staff, and resources
Economic and political context	Financial arrangements, regulations, and policies

If the contact person belonged to one of these groups, they were also asked to participate. Once the stakeholders had agreed to participate, an interview was scheduled with each person individually. In total, 8 hospitals were approached, of which 7 agreed to participate, and 21 subjects participated in the study. No ethical review is needed for this type of study. All participants were informed about the purpose of the study, and participation was voluntary. Verbal consent for audio recording the interviews was obtained for every participant. All data were analyzed and presented anonymously.

Data Collection Procedure

The interviews were conducted by the first author (LK). A few days before the interview, each participant received a confirmation email suggesting a scheduled date and time. A document was attached describing the objectives of the study and a topic list for the interview. We also added our own definition of a typical patient portal: "a personal digital environment, facilitated by a health care institution, for example a hospital. Patients need to login to the portal to get access to, for example, their medical file (with results), patient information and appointments. Patients can also fill in questionnaires and receive personalized advice regarding, for example, quality of life and physical activity." We used a semistructured interview that was structured by applying the comprehensive model of Grol and Wensing [26] that summarizes the barriers to and facilitators of change in health care practice. This model describes 6 levels at which barriers and facilitators can occur: (1) innovation: patient portal; (2) individual professional; (3) patient; (4) social context; (5) organizational context; and (6) economic and political context. All these barriers and facilitators are described in Table 1.

All interviews were performed by telephone and lasted for, on average, 20 min. Participants were first asked for their consent to make audio recordings of the interviews. Then, the purpose of the interview was introduced, and subjects were asked if they received the introductory email. This email was then briefly discussed such that the subjects were aware of the topics to be discussed. After that, questions were asked about participants' characteristics, such as their age and work experience. To make sure an unambiguous definition of a patient portal was used, participants were asked what their definition of a patient portal was, and if necessary, it was complemented with our definition. Then, we asked them about their perceived barriers to and facilitators of patient portal implementation at all 6 levels [26]. If necessary, for example, if the question was unclear, the interviewer provided examples (and these were also sent per

email). At the end of the interview, the participants were asked to suggest additional topics or issues, if any, that had not yet been covered. The interviews were in Dutch, and the questions in Multimedia Appendix 1 are translations.

Data Analysis

The first author transcribed all interviews verbatim. Two researchers (LK and WG) independently selected text fragments that reflected a barrier to or facilitator of portal implementation and coded the transcripts in Excel according to the model of Grol and Wensing [26]. A directed content approach was used, which is mainly a deductive approach as a pre-existing model is used for coding [31]. If quotes did not fit into the Grol and Wensing model [26], we looked for categories from the McGinn model [21], which was developed for implementation of EHRs. These models have considerable overlap, but the Grol and Wensing model [26] mainly covers socio-dynamic factors, whereas the McGinn model [21] also covers technical and portal characteristics. For the remaining quotes we created new categories, which is an inductive approach. To enhance clarity and unambiguity of the categories, we renamed them to better reflect the nature of being a barrier or a facilitator. A complete overview of the categories is presented in Multimedia Appendix 2. Coding was discussed between LK and WG until consensus was reached. Saturation of the data was checked by the first author by assessing (post hoc) the percentage of new categories appearing with the analysis of every subsequent hospital.

Results

Characteristics of the Subjects

In total, we interviewed 21 stakeholders from 7 hospitals. We included 3 from each hospital including medical professionals (n=7), managers (n=7), and IT employees (n=7). The stakeholder group labeled medical professionals consisted of medical specialists (n=4) and nurse practitioners (n=3). The group of managers included a medical director (n=1), hospital division or department managers (n=5), and a project manager (n=1). IT employees were application specialists or managers (n=3), an IT manager (n=1), an IT architect and information manager (n=1), and a patient portal project manager (n=2). Mean age was 44.8 years (SD 6.7; range 25-61) and 57% (12/21) were female. We included 6 respondents (6/21, 29%) from UMCs, 9 respondents (9/21, 43%) from teaching hospitals, and 6 (6/21, 29%) from general hospitals. Participants' work experience varied from 6 years or less (10/21, 48%) to more than 21 years

(3/21, 14%). An overview of participants' characteristics is listed in Table 2.

Barriers to and Facilitators of Patient Portal Implementation

In total, we selected 376 quotes and identified 26 barriers and 28 facilitators. The results are presented according to the 6 levels of the Grol and Wensing model [26]. The full list of all barriers and facilitators—including the number of subjects for each stakeholder group—is presented in Multimedia Appendix 3. After the inclusion of 7 hospitals (using purposive sampling), we analyzed the data saturation. The data were found to be saturated, meaning that after analyzing the first 6 hospitals, no new categories emerged from the transcripts of the final hospital. We therefore did not include further hospitals.

Due to the high number of identified barriers and facilitators, only those common to all stakeholder groups (medical professionals, managers, and IT employees) are presented here. To demonstrate the similarities and differences between stakeholder groups and between hospitals types, their most mentioned barriers and facilitators are presented as well.

Barriers and Facilitators Common to All Stakeholder Groups

In total, 13 barriers and 12 facilitators (Table 3) were identified that were common to all stakeholder groups. The most relevant barriers and facilitators for each level are presented based on the number of subjects (and percentage of the total subjects) and are highlighted in *italics*. Quotes are used to illustrate the barriers and facilitators for each level that were mentioned by the majority of the subjects.

Table 2. Participants' characteristics (N=21).

Characteristics	n (%)
Gender	
Female	12 (57)
Male	9 (43)
Age (years)	
20-29	3 (14)
30-39	3 (14)
40-49	7 (33)
50-59	6 (29)
>60	2 (10)
Hospital	
University medical centers	6 (29)
Teaching hospital	9 (42)
General hospital	6 (29)
Work experience in current position in organization (years)	
≤5	10 (48)
6-10	3 (14)
11-15	1 (5)
16-20	4 (19)
≥21	3 (14)

Table 3. Barriers to and facilitators of patient portal implementation mentioned by all stakeholder groups and ranked by number of subjects.

Barriers and facilitators	Stakeholders, n (%)			
	Medical professionals (n=7)	Managers (n=7)	IT employees (n=7)	Total (n=21)
Innovation: patient portal				
Barriers				
Guaranteeing privacy and security	1 (14)	5 (71)	5 (71)	11 (52)
Lack of accessibility	2 (29)	4 (57)	3 (43)	9 (43)
Lack of perceived usefulness	4 (57)	1 (14)	2 (29)	7 (33)
Facilitators				
Perceived usefulness	7 (100)	7 (100)	7 (100)	21 (100)
Perceived ease of use	2 (29)	2 (29)	1 (14)	5 (24)
Attractiveness	1 (14)	1 (14)	2 (29)	4 (19)
Participation of end users during implementation	1 (14)	1 (14)	1 (14)	3 (14)
Individual professional				
Facilitators				
Positive attitude	3 (43)	7 (100)	3 (43)	13 (62)
Motivation to change	4 (57)	2 (29)	2 (29)	8 (38)
Having knowledge	1 (14)	2 (29)	2 (29)	5 (24)
Patient				
Barriers				
Lack of sufficient eHealth literacy	4 (57)	5 (71)	4 (57)	13 (62)
Social context				
Barriers				
Negative attitude or opinion of medical professionals	4 (57)	3 (43)	1 (14)	8 (38)
Facilitators				
Positive attitude or opinion of medical professionals	1 (14)	2 (29)	2 (29)	5 (24)
Organizational context				
Barriers				
Lack of resources	4 (57)	5 (71)	6 (86)	15 (71)
Lack of time and increased workload	4 (57)	3 (43)	1 (14)	8 (38)
Innovation-averse culture	1 (14)	4 (57)	1 (14)	6 (29)
Lack of suitable specialist staff	1 (14)	2 (29)	3 (43)	6 (29)
Adjusting organization of care processes is difficult	2 (29)	1 (14)	2 (29)	5 (24)
Structure of the organization	2 (29)	1 (14)	2 (29)	5 (24)
Change in task and new responsibilities	1 (14)	1 (14)	2 (29)	4 (19)
Facilitator				
Management support	2 (29)	3 (43)	3 (43)	8 (38)
Communication to promote the portal	1 (14)	4 (57)	1 (14)	6 (29)
Innovation-oriented culture	2 (29)	2 (29)	1 (14)	5 (24)
Economic and political context				
Barrier				
Financial difficulties	5 (71)	6 (86)	3 (43)	14 (67)
Facilitator				
Facilitating laws and regulations	1 (14)	2 (29)	1 (14)	4 (19)

XSL•FO
RenderX

Innovation: Patient Portal

Barriers

Lack of perceived usefulness, *lack of accessibility*, and *guaranteeing privacy and security* were identified as barriers for portal implementation. Important reasons related to the privacy and security were the regulations, the availability of privacy-sensitive information on the portal, and the requirements for a safe login. The login or authorization method used in the Netherlands—the so-called digital identity DigiD with additional text messaging verification—was mentioned very frequently and can therefore be considered a major barrier. This DigiD login consists of a username and password of the user's own choice and provides citizens with access to hundreds of government websites in the Netherlands [32]:

> *The security is a barrier for both the organization, and the implementation of the portal, as well for patients. The moment we secure the data according to the law and regulations, we notice that the use is not what it could be.* [Manager, university medical center]

Due to the privacy and security aspects, accessibility of the portal is increasingly becoming a limitation, and this was mainly because of the requirement for a DigiD login. Subjects mentioned lack of perceived usefulness because the portal implementation can lead to discord and practical difficulties. In addition, the portal only provides information for one health care institution, so patients do not have a complete overview of their health information.

Facilitators

Perceived usefulness, attractiveness, perceived ease of use, and *participation of end users* during implementation were seen as facilitators for implementation. All subjects (n=21) see *perceived usefulness* as a facilitator because the implementation of a patient portal could result in fewer consults, less paperwork, higher quality of care, and financial savings. Also for patients, multiple benefits were listed, including more involvement in their treatment, more transparency, and better accessibility of information:

> *It saves a lot of paperwork and hassles. It sounds ideal to me. Currently patients receive so many paper documents that they don't have an overview anymore. If we centralize this on a portal it will be more clear for them.* [Medical professional, general hospital]

A good project team and the *participation of the end users during implementation* —both patients and hospital staff—can be beneficial because their input can be used to make adjustments during the development phase. *Perceived ease of use* and specifically the design of the portal can facilitate portal use, and the *attractiveness* was widely considered to be a requirement.

Individual Professional

Facilitators

No barriers were common for all the stakeholder groups. However, all groups see *motivation to change*, *knowledge*, and their own *positive attitude* as a facilitator:

> *I am very happy that we are starting with this development and that we, I think, are taking positive steps for the healthcare in the Netherlands.* [Manager, teaching hospital]

Patient

Barriers

Only barriers were anticipated for patients (common to all stakeholder groups), especially related to patients' characteristics and patient portal use. These barriers included *lack of eHealth literacy*. This can be due to the diversity of the patient population because it will include immigrants, older patients, and people with limited literacy skills. These specific groups may experience difficulty using a portal. Patients might also fear using the portal or simply need time to get used to it:

> *We have a lot of patients with low levels of literacy [...] So a lot of people without digital access to information, and no computer. That is a barrier for the portal in this hospital.* [Manager, teaching hospital]

Social Context

Barrier and Facilitator

Negative attitude or opinion of medical professionals was seen as a barrier and a facilitator by all stakeholder groups. They stated that this is because of doctors' resistance regarding transparency of medical information, negative outcome expectancy because they think they will receive more questions and phone calls, and they are sometimes afraid to lose control:

> *...a lot of professionals are very tense about it. They are used to have the control when they get in touch with a patient or have an appointment with a patient. Now it is possible for patients to interfere with this. Doctors and other professionals are tense about that. So that is a barrier for implementation.* [Manager, university medical center]

However, *positive attitude or opinion of medical professionals* was seen as a facilitator. When medical professionals are enthusiastic, it can facilitate the implementation, and they can influence others in a positive way. It was also mentioned that medical professionals asked for IT services for patients to be improved:

> *There is also an explicit request from the medical staff to support, what they call patient IT, so that is positive.* [IT employee, general hospital]

Organizational Context

Barriers

Lack of resources, *lack of time* and *increased workload*, *innovation-averse hospital culture*, *lack of suitable specialist staff*, *difficult to adjust organization of care processes*, *structures of the organization*, and *change in task and new responsibilities* were identified as barriers. *Lack of resources* was seen as a barrier, and although material resources—such as a lack of advanced IT materials—can be a reason, mainly the lack of human resources was mentioned by stakeholders. These

XSL•FO

RenderX

resources are not only essential for implementation but also to maintain the portal and to ensure the continuity of service to patients, once the portal has been implemented. IT employees are especially important because this process requires specific knowledge. This technical knowledge is often lacking in hospitals, and it may therefore be necessary to hire suitable specialist staff. This means that there should be enough money to attract resources, which can be a problem because the budgets of hospitals are limited:

> *An organization has limited resources nowadays, so yes that is a barrier. It is not that we can open a cash box and say we will hire 20 more people to finish this together. That is not how it works.* [Medical professional, teaching hospital]

The *innovation-averse culture* in hospitals is often identified as a barrier. One reason for this is that each person wants to give his or her opinion (about the portal), and that all opinions need to be taken into account, which inevitably slows down the implementation. Health care is also seen as essentially conservative—especially by managers—meaning that health care organizations and professionals need to get used to a new medium such as a patient portal.

These new services may affect hospitals' care processes, which can be difficult to adjust. Patients usually have access to their portal 24 hours a day, 7 days a week. If they experience a problem or they ask a question, it should be addressed quickly, and this may not always be possible. *Adjusting the organization of care processes* might be necessary, for example, concerning the transparency of medical information on the portal. Adjusting these care processes can be a barrier because they are sometimes ambiguous and usually difficult to change. This may also lead to *changes in tasks and new responsibilities* for the staff. New tasks or changes in existing work processes and responsibilities may result in informing patients about the portal and answering questions that arise when reading medical information on the portal. But also *lack of time and increased workload* was noted as a barrier, and the time investment required from medical professionals was especially seen as a problem. Furthermore, *organizational structures* can also hinder implementation for the reason that each division in a hospital tends to have its own management, policy agreements, and prioritizing approach.

Facilitators

Management support, communication to promote the portal, and *innovation-oriented culture* were seen as facilitators. The support of hospitals' management can facilitate portal implementation, especially when there is a hospital-wide strategy on eHealth—and patient portals—available. On the other hand, if this is missing, then that can be a barrier to implementation. *Management support* and approval can also be a facilitator; it can help the organization to focus on the implementation instead of on the internal discussion whether or not to implement the portal:

> *...the decision of the board means everything, because then you are not going to discuss if we are going to do it and why but we are going to do this and how [...] that is an absolute must and facilitator for this*

> *kind of project to be implemented.* [IT employee, university medical center]

Clear *communication (to promote the portal)* was indicated to be facilitating and relevant for staff because it can reduce professionals' misunderstanding, for example, regarding functionalities on the portal. Sessions to inspire staff about eHealth can facilitate implementation, and hospitals can use publicity to raise awareness about the availability of the portal and thereby increase accessibility for patients.

An *innovation-oriented culture* can help for the reason that the implementation is supported by the organization, the staff are stimulated and feel motivated, and there is a positive mood.

Economic and Political Context

Barrier

Financial difficulties were seen as a barrier mainly because funding is often a problem, and technical adjustments are expensive. In addition, the reimbursement for certain applications, for example, e-consults, has not yet been arranged:

> *The barrier is that it is not directly insured care, it is a bit luxurious (to provide it to patients now). So you have to find funding for it.* [Medical professional, general hospital]

Facilitator

Facilitating laws and regulations can be beneficial, and especially the support by the government in the Netherlands for portal implementation is seen as a facilitator.

Comparison of Stakeholder Groups

We found similarities between stakeholders, for example, regarding *perceived usefulness*, but also differences (Table 4). Overall, the findings regarding *lack of resources* were fairly similar among the groups, although the majority (5/7, 71%) of the IT employees also mentioned that there are *sufficient resources* available. *Guaranteeing privacy and security* was mentioned by both managers (5/7, 71%) and IT employees (5/7, 71%) as a barrier. The majority of medical professionals (4/7, 57%) and managers (5/7, 71%) mentioned *lack of sufficient eHealth literacy* of patients as a barrier.

However, we also found differences between stakeholder groups. The *negative attitude or opinion of medical professionals* was often seen as a barrier, especially by medical professionals. They were most often negative about providing patients with medical information via the patient portal because they were afraid it would lead to more work (such as more questions from patients), and they were worried about losing control. A remarkable finding is that all the managers (7/7, 100%) see their own *positive attitude* as a facilitator; however, this is true for only less than the half (3/7, 43%) of the other groups. All the medical professionals mentioned the *perceived usefulness* of the portal, but they (4/7, 57%) also indicated a *lack of perceived usefulness* because they think that the portal can lead to practical problems. However, the majority of this group is *motivated to change* (4/7, 57%) compared with only a minority in the other 2 stakeholder groups (both 2/7, 29%).

Table 4. Top 3 barriers and facilitators for each stakeholder group and ranked by number of subjects.

Barriers and facilitators by stakeholder group	n (%)
Medical professionals (n=7)	
Perceived usefulness (+[a])	7 (100)
Financial difficulties (−[b])	5 (71)
Lack of perceived usefulness (−)	4 (57)
Motivation to change (+)	4 (57)
Lack of sufficient eHealth literacy (−)	4 (57)
Negative attitude or opinion of medical professionals (−)	4 (57)
Lack of resources (−)	4 (57)
Lack of time and increased workload (−)	4 (57)
Managers (n=7)	
Perceived usefulness (+)	7 (100)
Positive attitude (+)	7 (100)
Financial difficulties (−)	6 (86)
Guaranteeing privacy and security (−)	5 (71)
Lack of sufficient eHealth literacy (−)	5 (71)
Lack of resources (−)	5 (71)
IT[c] employees (n=7)	
Perceived usefulness (+)	7 (100)
Lack of resources (−)	6 (86)
Guaranteeing privacy and security (−)	5 (71)
Sufficient resources (+)	5 (71)

[a]"+" indicates facilitator.

[b]"−" indicates barrier.

[c]IT: information technology.

Comparison of Hospital Types

In Table 5, the top 3 barriers and facilitators for each hospital type are listed. A complete overview of all barriers and facilitators—including the number of subjects for each hospital type—is presented in Multimedia Appendix 4. Differences were found in the barriers mentioned by subjects from different hospital types. The majority (5/6, 80%) of subjects from UMCs mentioned *lack of accessibility* as a barrier, and the difficult login method was especially seen as a barrier in these hospitals. In general hospitals, most subjects think that the *positive attitude or opinion of medical professionals* will facilitate implementation because medical professionals are enthusiastic. *Lack of time and increased workload* is also an important barrier in general hospitals because everybody is already always busy. Along with the differences, we also found similarities between the 3 hospital types. *Perceived usefulness* was mentioned by all subjects (21/21, 100%), but also *lack of resources* was seen in every hospital type as an important barrier. The UMCs and general hospitals see that the *lack of sufficient eHealth literacy* can hinder patient portal use. The most similarities were found between the teaching and general hospitals. *Positive attitude, guaranteeing privacy and security,* and *financial difficulties* were mentioned by the majority of subjects in both teaching and general hospitals. This is an important difference from the UMCs, which can perhaps be explained by differences in the financing of these hospital types.

Table 5. Barriers and facilitators—top 3 for each hospital type and ranked by number of subjects.

Barriers and facilitators by hospital type	n (%)
UMCs[a] (n=6)	
Perceived usefulness (+[b])	6 (100)
Lack of accessibility (–[c])	5 (83)
Lack of sufficient eHealth literacy (–)	4 (67)
Lack of resources (–)	4 (67)
Teaching hospitals (n=9)	
Perceived usefulness (+)	9 (100)
Lack of resources (–)	7 (78)
Financial difficulties (–)	7 (78)
Guaranteeing privacy and security (–)	6 (67)
Positive attitude (+)	6 (67)
General hospitals (n=6)	
Perceived usefulness (+)	6 (100)
Positive attitude (+)	5 (83)
Guaranteeing privacy and security (–)	4 (67)
Lack of sufficient eHealth literacy (–)	4 (67)
Positive attitude or opinion of medical professionals (+)	4 (67)
Lack of resources (–)	4 (67)
Lack of time and increased workload (–)	4 (67)
Financial difficulties (–)	4 (67)

[a]UMC: university medical center.

[b]"+" indicates facilitator.

[c]"–" indicates barrier.

Comparison of Hospitals With and Without an Implemented Patient Portal

Although we did not explicitly ask the included hospitals in which phase of implementation they were, we could deduce this from the interviews. In total, we included 7 hospitals. Two of these hospitals had no patient portal but were planning implementation. Three hospitals had minimal experience with portals—small pilots with limited functionalities or a classic portal version—but were also in the implementation phase. Only 2 hospitals had an active patient portal; however, stakeholders of one hospital mentioned they were still implementing to extend their current functionalities. In Table 6, we list the barriers and facilitators that were mentioned by (at least one stakeholder) all the included hospitals both with a patient portal (n=2) and without a patient portal (n=5). A complete overview is presented in Multimedia Appendix 5. Although there were similarities (eg, *financial difficulties, lack of sufficient eHealth literacy)*, we also found differences. All

hospitals without a patient portal mentioned *negative attitude or opinion of medical professionals* and *lack of specialist staff* as barriers. These factors could negatively influence implementation. Although the hospitals with a patient portal see barriers for the implementation of their patient portals, they also mentioned multiple facilitators, for example, *perceived ease of use, motivation to change,* and *sufficient resources*. The barriers *lack of a generic guideline* (n=1) and *participation of end users during implementation* (n=1) were only mentioned by hospitals with a patient portal. *Lack of a generic guideline* was a barrier expressed by a manager (n=1), meaning that it could have been beneficial for implementation if there would have been coordination or a standard format. All stakeholders of one hospital that had implemented a portal noticed *participation of end users during implementation*. In that case, they referred back to the implementation and stated that it was useful to involve end users—both patients and health care professionals—during implementation and for each hospital division to be well represented in the project organization.

Table 6. Barriers and facilitators mentioned by all hospitals (at least one subject per hospital) with and without a patient portal and ranked by total number of subjects.

Barriers to and facilitators of hospitals with and without a patient portal	Hospitals with a patient portal[a], n (%)	Hospitals without a patient portal[b], n (%)
Barriers and facilitators common for hospitals with and without a patient portal (ie, unanimously reported by hospitals of both groups)		
Barriers		
Financial difficulties	4 (67)	10 (67)
Lack of sufficient eHealth literacy	4 (67)	9 (60)
Lack of resources	2 (33)	12 (80)
Negative attitude or opinion of colleagues in general	3 (50)	9 (60)
Facilitators		
Perceived usefulness	6 (100)	15 (100)
Positive attitude	3 (50)	10 (67)
Barriers and facilitators only reported unanimously by hospitals with a patient portal		
Barriers		
Lack of time and increased workload	4 (67)	
Innovation-averse culture	3 (50)	
Adjusting organization of care processes	3 (50)	
Structures of the organization	3 (50)	
Change in task and new responsibilities	2 (33)	
Facilitators		
Perceived ease of use	3 (50)	
Motivation to change	2 (33)	
Having knowledge	2 (33)	
Positive attitude or opinion of medical professionals	2 (33)	
Good collaboration with colleagues	2 (33)	
Sufficient resources	2 (33)	
Conducive financial arrangements	2 (33)	
Barriers only reported unanimously by hospitals without a patient portal		
Negative attitude or opinion of medical professionals		7 (47)
Lack of suitable specialist staff		5 (33)

[a]n=2 hospitals; n=6 subjects.

[b]n=5 hospitals; n=15 subjects.

Discussion

Summary of Main Findings

In this study, we have presented an overview of the barriers and facilitators related to patient portal implementation among various stakeholders within the hospital organization. In total, we identified 26 barriers and 28 facilitators. Positive factors related to *perceived usefulness* (eg, cost savings, accessibility for patients to their information) were mentioned by all subjects. The facilitators individuals' *positive attitude* and *management support* (eg, strategy plan for eHealth and patient portals) were also mentioned by majority of the subjects. The main barriers reported were *lack of resources* (especially lack of staff), *financial difficulties* (high costs, lack of reimbursement), and

guaranteeing privacy and security (eg, strict regulations). We want to emphasize that no inferences can be drawn about the prevalence of phenomena observed beyond the current sample.

We found several similarities between stakeholders (eg, regarding *perceived usefulness*) but also remarkable differences that highlight the importance of involving multiple stakeholders. One interesting finding is that approximately half the medical professionals see their own *positive attitude* and *motivation to change* as facilitators. Although medical professionals' motivation to change is the highest of all stakeholder groups, *lack of time and increased workload* was perceived by them as a barrier. Apparently, they are willing to change, but at the same time, they assume that they do not have enough time to achieve implementation and portal use. The barriers *guaranteeing privacy and security* and *lack of resources* were mentioned by

the majority of IT employees. This shows the challenges this group is dealing with when implementing a secure portal. Managers were the only group of which all (7/7, 100%) stated that they had a *positive attitude.* This is in clear contrast with the proportion of medical professionals and IT employees (both 3/7, 43%). Managers also stand out in their statements about the culture with more than the half of the managers (4/7, 57%) thinking the culture is hindering implementation, whereas only a minority of both the medical professionals (1/7, 14%) and IT employees (1/7, 14%) stated this. Managers mentioned that hospital culture is conservative and slow to change.

Comparison With Previous Research

Koivunen et al [28] identified nurses' barriers and facilitators regarding portal implementation. Their findings were comparable with ours; for example, concerning the barriers *lack of resources* and *lack of time.* However, in their study, nurses were included and were mainly negative because they had doubts about the benefits of the portal; moreover, they were unwilling to use a new technical tool because they believed that their primary tasks are to be more important. This differs from our findings as we found *positive attitudes* among all included stakeholders (medical professionals, managers, and IT employees), and all our subjects mentioned *perceived usefulness* as a facilitator for patient portal implementation. One reason for these differences may be the selection of stakeholders, as we focused on those directly involved and did not include nurses, only medical doctors and nurse practitioners ("Advanced Practice Registered Nurses") [30]. Keplinger et al [27] also considered physicians' attitudes regarding patient portal implementation. Some of their findings are in line with ours, for example, the expected increase in workload and positive attitudes regarding the patient portal. However, they also found differences in attitudes both before and after implementation. For example, before implementation, more than half of the physicians assumed that their workload would increase, whereas only one-third actually experienced such an increase in workload.

McGinn et al [21] showed the relevance of including the perspectives of various stakeholders regarding EHR implementation. Their results are both similar and different from our results. They found that the main factors common to all stakeholder groups were found at various levels and included "perceived ease of use," "costs," "motivation to use EHR," and "privacy and security concerns." These findings are similar to ours perhaps because *financial difficulties, guaranteeing privacy and security,* and *positive attitude* were mentioned by the majority of our subjects. The use of the internet and other electronic applications is becoming increasingly common in health care [33], and patients' eHealth literacy needs to be taken into account. *Participation of end users during implementation* was mentioned as a facilitator and can be used to focus on the eHealth literacy of the users.

McGinn et al [21] argue that the consideration of various stakeholder opinions may contribute to successful EHR implementations. Similarities with and differences from our results were found. The main factors common to all stakeholder groups were found at various levels and included "design and

technical concerns," "costs," "lack of time and workload," and "privacy and security." The findings are similar to ours, and this can be the case because both EHRs as well as patient portals are complex technologies that affect multiple levels of an organization. However, we also found differences because in our study, *perceived usefulness* and *lack of sufficient eHealth literacy (patients)* were mentioned by the majority of the subjects. *Lack of accessibility* (because of login methods perceived as difficult) was mentioned by almost half of the subjects. This difference can be due to an EHR being primarily aimed at professionals and a patient portal being primarily intended as a service for patients. The differences found among these implementation studies highlight the importance of identifying barriers and facilitators for each technology separately taking into account the perspectives of the several stakeholder groups that are involved.

Implementation Frameworks and Models

There are many implementation models, and they have considerable overlap [34]. A combination of 2 models was used for categorization of the selected quotes, that is, the model of Grol and Wensing [26] for socio-dynamic factors and by McGinn [21] mainly for portal characteristic and technical factors. Although this combination of frameworks appeared to be a feasible approach, we also added categories and renamed existing ones, so they better match with our findings. An essential difference between our approach and, the CFIR framework is that in our study, patients are included as a separate factor, whereas in the CFIR framework, they are part of the "outer setting" [24]. In the FITT framework, separate categories such as "social context" and "organizational context" are missing, and the aspects related to social interaction, for example, are categorized under "individual" within the FITT model. We found these categories to be relevant as a separate level because many subjects reported on them [25]. In the McGinn model [21], a subcategory is "participation of end users during the design," which does not cover all the input we received, particularly because it is not aimed at the complete implementation process. One of the added categories is *participation of end users during implementation.* Another new category is *sufficient eHealth literacy,* which encompasses the skills and knowledge necessary to use electronic applications [33]. The models we used only address patients' skills and knowledge [26] and applicability—of EHR implementation—to patients' characteristics [21]. Patients' lack of eHealth literacy was identified as a barrier by the majority of the subjects.

Practical Suggestions and Insights for Portal Implementations

Our findings suggest that implementation is affected by barriers and facilitators at various levels. McGinn et al [21] describe 3 key levels: the macro, meso, and micro levels. We present some suggestions and insights for organizations that intend to implement a patient portal.

Micro Level: Individual and Social Factors

Our findings suggest that stakeholders' *positive attitudes* can contribute to implementation. They greatly value their colleagues' opinions, so apparently this can play a crucial role

in the implementation process. Clear communication with all stakeholders during the implementation process and about the patient portal functionalities can increase stakeholders' understanding and can help to avoid misunderstandings.

Meso Level: Organizational and Operational Developments

The implementation can be affected by operational factors in the organization [21]; for example, *lack of resources*, *management support*, and *lack of suitable specialist staff*. To successfully implement a patient portal, a project team is essential that includes resources and staff with technical knowledge about patient portals and implementation processes. Management support is important; for example, by including the plan for portal implementation in their organizational strategy. Organizations should also be aware that the implementation of a patient portal is not only a technical implementation but also involves a change in the organizational socio-dynamics, including changes in employees' tasks, new responsibilities, and a shift in control from health care professionals to patients.

Macro Level: Sociopolitical Influences

Governments in Western countries are increasingly promoting and supporting portal implementation and use. In the United States, financial support is generated by the Health Information Technology for Economic and Clinical Health Act and arranged by the Centers for Medicare and Medicaid Services. The goal of these incentive programs is to support the implementation [35], adoption, and "meaningful use" of the EHRs [6,35,36]. This includes, for example, providing patients with access to or acquiring an electronic copy of their health data [36]. In the Netherlands, the Ministry of Health and the Dutch Hospital Association developed a funding program to support information exchange for both patients and professionals. The ultimate goal of this program is that in 2020, all Dutch people will have access to their own medical information. Therefore, all institutions must have a patient portal by the end of 2019 or a link to a Personal Health Record (PHR) to which the institution can upload medical information [37]. Government commitment thus can be beneficial for hospitals, especially in view of the opportunities for funding. Hospitals can exploit governments' ambitions and policies and patient representatives demands, for example, to make EHR data accessible for every patient, as a motivation to facilitate implementation.

Limitations

This study has several limitations. First, we used semistructured interviews in which we provided participants with prompts/examples for each level. Providing subjects with examples may have restricted participants in their answers about new barriers and facilitators or to "think outside the box" on these topics, so we might have missed factors. However, we used the combined models of Grol and Wensing [26] and McGinn et al [21], and many stakeholders mentioned barriers and facilitators that fell outside our scope. Although we have confidence in the richness of the current data, we already reached data saturation after 6 hospitals, limiting the total number of hospitals and subjects. There were also differences

in the included hospitals with regard to the phase of patient portal implementation. Some had already provided a portal, whereas others were in the middle of the implementation process or had no portal at all. Although we found only limited differences between the hospitals with and without an implemented patient portal, this could still have introduced bias into the responses because of the recall or the imagination of information. This means that the results might have been influenced by the current state of hospitals because participants sometimes had to recall information from the time of implementation or had to imagine an implementation process (if there is no portal or no implementation).

Although we presented many different types of barriers and facilitators, we acknowledge that quantity should not be taken as a proxy for importance. We therefore added quotes to the results so as to highlight the specific nature of specific barriers and facilitators. For data analysis, we used a directed content analysis (deductive) approach. This can be a possible limitation because we started with an already existing model with defined categories. However, as the methods allows, we did not completely hold on to the categories in the models as we added additional categories ourselves and renamed the existing (generic) categories to barriers and facilitators that better fit our findings. Despite these limitations this is, to the best of our knowledge, the first qualitative study to identify barriers and facilitators for patient portal implementation involving multiple stakeholder groups.

Future Perspectives and Research Directions

Instead of organizing health care around professionals and institutions, some contend that it should increasingly be arranged around patients [2]. In a recent review, we found little evidence for the efficacy of IT-supported shared care [38]; however, many initiatives exist that may facilitate patient-centered or shared care. We already see movement in this direction as information systems are evolving from purely organizational to regional and even international systems [39]. For instance, a PHR is an example of an application in which patients can access their health information that has been collected from various health care institutions but is controlled by the patients [40]. In several European countries, these national systems have already been introduced. For example, in France, there is a national initiative called "Dossier Médical Personnel," which is accessible over the internet. The information is uploaded by the involved clinicians; however, patients are in charge about what is included in the portal and who is authorized to access it. In Estonia, health professionals transfer information into a system called the "Estonian Health Information System," providing patients with information via a patient portal [41]. These initiatives show a shift from hospital-financed, -owned, and -managed health records for which access is granted through portals, toward PHRs in which providers upload the data and ownership by patients is facilitated. The present uptake/compliance rates of portals are however still rather low (seldom above 50%), so this is an aspect that should receive attention if widespread use is foreseen.

Future research is necessary to confirm the practical utility of our proposed model when used among various stakeholder

groups and to test whether it is useful to tailor implementation strategies to these various stakeholders, and organizations, taking possible development routes into account. In addition, there is a lack of knowledge regarding the association between patient portal implementation and patient portal adoption (ie, actual uptake and use by patients). One important element we identified is eHealth literacy, and this should ideally be included in the implementation and evaluation strategies for health technology tools. Moreover, the expectations before implementations and the experiences afterward can vary among health care professionals [27] and patients [11]. Further research into "satisfiers" determining the attitude of professionals toward using these technologies is recommended because evidence of the effectiveness of technology-related aspects on patient empowerment and on health outcomes is a strong facilitator.

Conclusions

Patient portal implementation is a complex process that is not just a technical process, but it also affects an organization and its staff. We found barriers and facilitators at various levels that differed depending on hospital types (eg, *lack of accessibility*) and stakeholder groups (eg, *sufficient resources*) in terms of several factors. Our findings underscore the importance of involving multiple stakeholders in portal implementation projects. We identified a set of barriers and facilitators, which are likely to be useful in making strategic and efficient portal implementation plans.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Interview questions.

[PDF File (Adobe PDF File), 89KB - jmir_v20i5e183_app1.pdf]

Multimedia Appendix 2

Barriers and facilitators categorized according to the model of Grol & Wensing and the model of McGinn et al.

[PDF File (Adobe PDF File), 90KB - jmir_v20i5e183_app2.pdf]

Multimedia Appendix 3

Barriers and facilitators for each stakeholder group.

[PDF File (Adobe PDF File), 58KB - jmir_v20i5e183_app3.pdf]

Multimedia Appendix 4

Barriers and facilitators for each hospital type.

[PDF File (Adobe PDF File), 58KB - jmir_v20i5e183_app4.pdf]

Multimedia Appendix 5

Barriers and facilitators for hospitals with and without a patient portal.

[PDF File (Adobe PDF File), 32KB - jmir_v20i5e183_app5.pdf]

References

1. Reid PP, Compton WD, Grossman JH, Fanjiang G. Building a Better Delivery System: A New Engineering/Health Care Partnership. Washington, DC: The National Academies Press; 2005.
2. Corrigan J. Crossing the quality chasm. In: Reid PP, Compton WD, Grossman JH, Fanjiang G, editors. Building a Better Delivery System: A New Engineering/Health Care Partnership. Washington, DC: The National Academies Press; 2005.
3. Snyder CF, Wu AW, Miller RS, Jensen RE, Bantug ET, Wolff AC. The role of informatics in promoting patient-centered care. Cancer J 2011;17(4):211-218 [FREE Full text] [doi: 10.1097/PPO.0b013e318225ff89] [Medline: 21799327]
4. 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. [doi: 10.1093/jamia/ocx043] [Medline: 28460042]

5. Kuijpers W, Groen WG, Loos R, Oldenburg HS, Wouters MW, Aaronson NK, et al. An interactive portal to empower cancer survivors: a qualitative study on user expectations. Support Care Cancer 2015 Sep;23(9):2535-2542 [FREE Full text] [doi: 10.1007/s00520-015-2605-0] [Medline: 25876157]

6. 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]

7. Bourgeois FC, Mandl KD, Shaw D, Flemming D, Nigrin DJ. Mychildren's: integration of a personally controlled health record with a tethered patient portal for a pediatric and adolescent population. AMIA Annu Symp Proc 2009 Nov 14;2009:65-69 [FREE Full text] [Medline: 20351824]

8. Amante DJ, Hogan TP, Pagoto SL, English TM. A systematic review of electronic portal usage among patients with diabetes. Diabetes Technol Ther 2014 Nov;16(11):784-793. [doi: 10.1089/dia.2014.0078] [Medline: 24999599]

9. Bates DW, Wells S. Personal health records and health care utilization. J Am Med Assoc 2012 Nov 21;308(19):2034-2036. [doi: 10.1001/jama.2012.68169] [Medline: 23168828]

10. 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]

11. Kuijpers W, Groen WG, Oldenburg HS, Wouters MW, Aaronson NK, van Harten WH. eHealth for breast cancer survivors: use, feasibility and impact of an interactive portal. JMIR Cancer 2016 May 10;2(1):e3 [FREE Full text] [doi: 10.2196/cancer.5456] [Medline: 28410178]

12. 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]

13. Groen WG, Kuijpers W, Oldenburg HS, Wouters MW, Aaronson NK, van Harten WH. Supporting lung cancer patients with an interactive patient portal: feasibility study. JMIR Cancer 2017 Aug 08;3(2):e10 [FREE Full text] [doi: 10.2196/cancer.7443] [Medline: 28790025]

14. Kuijpers W, Groen WG, Oldenburg HS, Wouters MW, Aaronson NK, van Harten WH. Development of MijnAVL, an interactive portal to empower breast and lung cancer survivors: an iterative, multi-stakeholder approach. JMIR Res Protoc 2015 Jan 22;4(1):e14 [FREE Full text] [doi: 10.2196/resprot.3796] [Medline: 25614924]

15. Calvillo J, Román I, Roa LM. How technology is empowering patients? A literature review. Health Expect 2015 Oct;18(5):643-652 [FREE Full text] [doi: 10.1111/hex.12089] [Medline: 23711169]

16. Earnest MA, Ross SE, Wittevrongel L, Moore LA, Lin CT. Use of a patient-accessible electronic medical record in a practice for congestive heart failure: patient and physician experiences. J Am Med Inform Assoc 2004;11(5):410-417 [FREE Full text] [doi: 10.1197/jamia.M1479] [Medline: 15187074]

17. Gagnon MP, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012 Feb;36(1):241-277 [FREE Full text] [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]

18. 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:i8-12 [FREE Full text] [doi: 10.1136/amiajnl-2011-000473] [Medline: 22071530]

19. Groen WG, Kuijpers W, Oldenburg HS, Wouters MW, Aaronson NK, van Harten WH. Empowerment of cancer survivors through information technology: an integrative review. J Med Internet Res 2015 Nov 27;17(11):e270 [FREE Full text] [doi: 10.2196/jmir.4818] [Medline: 26614438]

20. Sarkar U, Bates DW. Care partners and online patient portals. J Am Med Assoc 2014;311(4):357-358. [doi: 10.1001/jama.2013.285825] [Medline: 24394945]

21. McGinn CA, Grenier S, Duplantie J, Shaw N, Sicotte C, Mathieu L, et al. Comparison of user groups' perspectives of barriers and facilitators to implementing electronic health records: a systematic review. BMC Med 2011 Apr 28;9:46 [FREE Full text] [doi: 10.1186/1741-7015-9-46] [Medline: 21524315]

22. 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 Dec 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]

23. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]

24. Varsi C, Ekstedt M, Gammon D, Ruland CM. Using the Consolidated Framework for Implementation Research to identify barriers and facilitators for the implementation of an Internet-based patient-provider communication service in five settings: a qualitative study. J Med Internet Res 2015 Nov 18;17(11):e262 [FREE Full text] [doi: 10.2196/jmir.5091] [Medline: 26582138]

XSL•FO
RenderX

25. Ammenwerth E, Iller C, Mahler C. IT-adoption and the interaction of task, technology and individuals: a fit framework and a case study. BMC Med Inform Decis Mak 2006 Jan 09;6:3 [FREE Full text] [doi: 10.1186/1472-6947-6-3] [Medline: 16401336]

26. Grol R, Wensing M. What drives change? Barriers to and incentives for achieving evidence-based practice. Med J Aust 2004 Mar 15;180(6 Suppl):S57-S60. [Medline: 15012583]

27. Keplinger LE, Koopman RJ, Mehr DR, Kruse RL, Wakefield DS, Wakefield BJ, et al. Patient portal implementation: resident and attending physician attitudes. Fam Med 2013 May;45(5):335-340 [FREE Full text] [Medline: 23681685]

28. Koivunen M, Hätönen H, Välimäki M. Barriers and facilitators influencing the implementation of an interactive Internet-portal application for patient education in psychiatric hospitals. Patient Educ Couns 2008 Mar;70(3):412-419. [doi: 10.1016/j.pec.2007.11.002] [Medline: 18079085]

29. M&I/Partners. : ; 2017. Patientportalen: de opstart voorbij URL: https://www.zorgvisie.nl/PageFiles/103493/patientportalenonderzoek2017%20mxipartnersDEF.pdf [accessed 2018-04-13] [WebCite Cache ID 6yeAlMlWE]

30. Nurse Journal. 2017. Requirements to Become A Nurse Practitioner URL: https://nursejournal.org/nurse-practitioner/what-to-know-to-become-a-nurse-practitioner/ [accessed 2018-04-13] [WebCite Cache ID 6yeAw9Pyr]

31. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]

32. Digid. 2017. About DigiD URL: https://www.digid.nl/over-digid/ [accessed 2017-09-15] [WebCite Cache ID 6tUkBVc3k]

33. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]

34. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci 2015 Apr 21;10:53 [FREE Full text] [doi: 10.1186/s13012-015-0242-0] [Medline: 25895742]

35. Powell KR. Patient-perceived facilitators of and barriers to electronic portal use: a systematic review. Comput Inform Nurs 2017 Nov;35(11):565-573. [doi: 10.1097/CIN.0000000000000377] [Medline: 28723832]

36. Blumenthal D, Tavenner M. The "meaningful use"; regulation for electronic health records. N Engl J Med 2010 Aug 05;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]

37. Nederlandse Vereniging van Ziekenhuizen. VIPP-programma URL: https://www.nvz-ziekenhuizen.nl/onderwerpen [accessed 2017-09-15] [WebCite Cache ID 6tUk27L6B]

38. Kooij L, Groen WG, van Harten WH. The effectiveness of information technology-supported shared care for patients with chronic disease: a systematic review. J Med Internet Res 2017 Jun 22;19(6):e221 [FREE Full text] [doi: 10.2196/jmir.7405] [Medline: 28642218]

39. Haux R. Health information systems - past, present, future. Int J Med Inform 2006;75(3-4):268-281. [doi: 10.1016/j.ijmedinf.2005.08.002] [Medline: 16169771]

40. 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;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]

41. de Lusignan S, Ross P, Shifrin M, Hercigonja-Szekeres M, Seroussi B. A comparison of approaches to providing patients access to summary care records across old and new europe: an exploration of facilitators and barriers to implementation. Stud Health Technol Inform 2013;192:397-401. [Medline: 23920584]

Abbreviations

CFIR: Consolidated Framework for Implementation Research
EHR: electronic health record
FITT: Fit between Individuals, Tasks, and Technology
IT: information technology
PHR: personal health record
UMC: university medical centers

XSL•FO
RenderX

XSL•FO

RenderX

Original Paper

The Importance of Visit Notes on Patient Portals for Engaging Less Educated or Nonwhite Patients: Survey Study

Macda Gerard[1], BA; Hannah Chimowitz[2], BA; Alan Fossa[2], MPH; Fabienne Bourgeois[3], MPH, MD; Leonor Fernandez[2], MD; Sigall K Bell[2], MD

[1]Wayne State University School of Medicine, Detroit, MI, United States

[2]Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, United States

[3]Department of Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Sigall K Bell, MD
Department of Medicine
Beth Israel Deaconess Medical Center
330 Brookline Ave
Boston, MA, 02215
United States
Phone: 1 6174700771
Email: sbell1@bidmc.harvard.edu

Abstract

Background: OpenNotes, a national initiative to share clinicians' visit notes with patients, can improve patient engagement, but effects on vulnerable populations are not known very well.

Objective: Our aim is to examine the importance of visit notes to nonwhite and less educated patients.

Methods: Patients at an urban academic medical center with an active patient portal account and ≥1 available ambulatory visit note over the prior year were surveyed during June 2016 until September 2016. The survey was designed with patients and families and assessed importance of reading notes (scale 0-10) for (1) understanding health conditions, (2) feeling informed about care, (3) understanding the provider's thought process, (4) remembering the plan of care, and (5) making decisions about care. We compared the proportion of patients reporting 9-10 (extremely important) for each item stratified by education level, race/ethnicity, and self-reported health. Principal component analysis and correlation measures supported a summary score for the 5 items (Cronbach alpha=.93). We examined factors associated with rating notes as extremely important to engage in care using logistic regression.

Results: Of 24,722 patients, 6913 (27.96%) completed the survey. The majority (6736/6913, 97.44%) read at least one note. Among note readers, 74.0% (727/982) of patients with ≤high school education, 70.7% (130/184) of black patients, and 69.9% (153/219) of Hispanic/Latino patients reported that notes are extremely important to feel informed about their care. The majority of less educated and nonwhite patients reported notes as extremely important to remember the care plan (62.4%, 613/982 ≤high school education; 62.0%, 114/184 black patients; and 61.6%, 135/219 Hispanic/Latino patients) and to make care decisions (62.3%, 612/982; 59.8%, 110/184; and 58.5%, 128/219, respectively, and *P*<.003 for all comparisons to more educated and white patients, respectively). Among patients with the poorest self-reported health, 65.9% (499/757) found notes extremely important to be informed and to understand the provider. On multivariable modeling, less educated patients were nearly three times as likely to report notes were extremely important to engage in care compared with the most educated patients (odds ratio [OR] 2.9, 95% CI 2.4-3.3). Nonwhite patients were twice as likely to report the same compared with white patients (OR 2.0, 95% CI 1.5-2.7 [black] and OR 2.2, 95% CI 1.6-2.9 [Hispanic/Latino and Asian], *P*<.001 for each comparison). Healthier patients, women, older patients, and those who read more notes were more likely to find notes extremely important to engage in care.

Conclusions: Less educated and nonwhite patients using the portal each assigned higher importance to reading notes for several health behaviors than highly educated and white patients, and may find transparent notes especially valuable for understanding their health and engaging in their care. Facilitating access to notes may improve engagement in health care for some vulnerable populations who have historically been more challenging to reach.

(J Med Internet Res 2018;20(5):e191)   doi:10.2196/jmir.9196

XSL·FO

RenderX

KEYWORDS

Introduction

As patients seek access to their health information, electronic health records (EHRs) have become increasingly central to health care delivery [1]. Centers for Medicare and Medicaid Services meaningful use incentives have prompted a rapid rise in the number of health care organizations that have implemented EHRs and accompanying Web-based patient portals to increase patient engagement [2]. Among patients engaging with their electronic health data, portal use has been associated with clinical benefits, such as increased rates of diabetes-related medication adherence [3,4]. In addition, patients who access visit notes on the portal report better understanding of their health conditions, feeling more prepared for visits, feeling more in control of health care, better remembering the care plan, and better taking medications as prescribed [5]. Similar results have been reported among US veterans, a population that may represent older, low income, or chronically ill Americans [6,7]. However, potential benefits derived from portal use may be influenced by many factors, including education, age, race/ethnicity, health literacy, and health status [8]. Such factors may, therefore, prevent some traditionally more vulnerable patient populations from realizing the potential benefits of electronic health information transparency [9-11].

The effects of health technology on the digital divide are actively debated. Studies report that nonwhite patients and those with fewer formal years of education are less likely to register for patient portals compared with less vulnerable patient populations [12-19]. However, other studies have found that once patients are enrolled on the patient portal, actual portal use may not differ by race, ethnicity, or socioeconomic status [15]. Although digital disparities exist, some data suggest that the gap is narrowing over time [20,21]. For instance, in 2016, 68% of Internet users had less than a high school education, compared with 41% in 2010 [20]. Between 2010 and 2016, the gap in Internet utilization in general among black users compared with white users nearly closed (85% vs 88%, respectively) [20]. In addition, at least some patients in safety net hospitals are interested in using electronic communication with their providers [22]. Disadvantaged patients may lack access to information needed to make informed care decisions [5], and at the same time, may be at the highest risk of negative health outcomes. Although some experts warn that patient portals and other electronic health (eHealth) innovations may widen disparities in care [12,19], others point out that with literacy-appropriate, user-centered design and better support, such health technology can play an important role to help bridge the divide [4,23].

OpenNotes, a national movement dedicated to making health care more open and transparent by encouraging health care providers to share their visit notes with patients through the patient portal, is giving patients easier access to their medical information [5,24]. As access to visit notes through the patient portal spreads across the country [24,25], we are often asked about whether OpenNotes can benefit vulnerable patients. The question is complicated because the term vulnerable has been broadly interpreted to include any patients who are at risk of health disparities, with respect to race and ethnicity, income, education level, chronic illness, disability, English as a second language, and limited health literacy, among other definitions [26,27]. Little is known about whether and how each of these patient groups may be affected by electronic access to their notes, and whether they view note access as important to their health.

We aimed to better understand how a subset of vulnerable patients engages with their health information. We asked portal-registered patients how important notes are to them for several patient engagement and health behaviors. We hypothesized that patients of different racial and ethnic backgrounds, education levels, and health status may report variable degrees of note importance, and that this information could help guide organizational strategies to engage patients through the patient portal.

Methods

Survey Development

We conducted a cross-sectional survey of an adult patient population using the patient portal at an urban US academic medical center. To design the survey, we convened a multidisciplinary team of patients and family members (members of the Patient and Family Advisory Council), health care delivery researchers, nurses, doctors, social workers, and patient engagement and safety experts. This group met regularly to design the survey for 1 year. Survey item structure was adapted from prior published OpenNotes surveys, and new items were developed to focus more specifically on patient engagement and safety [5]. This analysis focused on demographic data and on a subset of survey items assessing patient-perceived importance of notes for various patient engagement activities.

The survey introduction explained open notes and included a screenshot to remind patients where their notes are accessed on the portal. Questions focused on notes (rather than portal use more generally). We asked participants who reported reading at least one note:

How important is reading your notes for:

1. Understanding your health and medical conditions
2. Feeling informed about your care
3. Understanding how your provider(s) are thinking about your medical conditions
4. Remembering the plan for your care (what the provider(s) suggests you do next)
5. Helping you make decisions about your care.

Response options for each item ranged from 0-10, displayed horizontally after each item and anchored with the words not at all important on one end and extremely important at the other. The response scale was selected to allow for greater granularity in assessing the importance of notes to patients of varying

demographic backgrounds, as prior data suggested that, overall, the majority of patients supported the idea of OpenNotes [5]. We used administrative data to determine age, sex, number of notes available on the portal, and number of notes accessed.

The survey was reviewed for face validity by members of the Patient and Family Advisory Council and also underwent an external review by a survey scientist with expertise in development of national validated instruments assessing patient and family experience, and it was revised based on their feedback. We then performed formal cognitive testing with 3 additional patients of varying sociodemographic backgrounds. Final survey items relevant to this analysis are shown in Multimedia Appendix 1.

Participants

Our survey sample consisted of a simple random sample of 31,049 patients at 1 urban US academic health center with active portal accounts and at least one available out-patient visit note available during the prior year. Participants were invited to complete the questionnaire through the patient portal between June 2016 and September 2016 and received up to 2 subsequent reminders. Ten raffle prizes (iPads) were used as incentives for survey participation.

Analysis

Patient Characteristics

We compared demographics, note availability, and note reading between respondents and nonrespondents using the chi-square test for categorical data and the Wilcoxon sign rank test or t test for continuous data.

Importance of Notes: Bivariate Analysis

We compared the proportion of patients reporting notes were extremely important (designated as 9-10 on a scale of 0-10), by bivariate analysis using the chi-squared test for each of the five health behaviors across the three sociodemographic factors of interest: (1) education, (2) race/ethnicity, and (3) self-reported health.

Multivariable Modeling

We used principal component analysis to assess the psychometric properties of the 5 survey items addressing the importance of notes for patient engagement activities. The 5 items showed good internal consistency and represented 1 domain (correlation analysis revealed a Cronbach alpha of .93, see Multimedia Appendix 2). We used a logistic regression model to assess independent demographic and portal use factors associated with reporting that notes are extremely important to engage in care. As our psychometric analysis supported a summary measure, we calculated the mean score of all 5 items, and used a mean summary score of 9-10 as the outcome of interest. We also ran the model for each of the 5 study questions independently to confirm the results. All statistical analyses were performed in SAS version 9.4 (SAS Institute Inc., Cary, NC, USA). The study was approved by our Institutional Review Board.

Results

Patient Characteristics

A total of 79.62% (24,722/31,049) invited patients logged on to the portal during the study period and among these, 27.96% (6913/24,722) of patients completed the survey (Figure 1).

The mean age of respondents was 56 years, 62.82% (4343/6913) were women and 82.41% (5697/6913) were white (Table 1). Among participants, 70.40% (4867/6913) reported a bachelor's degree and 72.21% (4992/6913) held private insurance. Respondents and nonrespondents did not differ by gender and number of hospitalizations. However, compared with nonrespondents, patients who completed the survey were slightly older (51 vs 56 years) and more likely to use Medicare as their primary insurance (17.22%, 4156/24,136 vs 23.65%, 1635/6913). Both groups had a median of 7 notes available, but respondents accessed more notes than nonrespondents (median 4 vs 2). The majority of patients invited to participate in the study were white (77.03%, 23,917/31,049) and college graduates (63.76%, 19,797/31,049), as reflective of the patient population at our academic medical center. Compared with nonrespondents, participants were somewhat more likely to be white (75.49%, 18,220/24,136 vs 82.41%, 5697/6913) and a college graduate (61.86%, 14,930/24,136 vs 70.40%, 4867/6913). Additional participant characteristics and a comparison of respondents versus nonrespondents are shown in Table 1.

Importance of Notes: Bivariate Analysis

Among all 6913 respondents, 94.68% (6545/6913) reported reading at least one visit note during the prior 12 months, and an additional 2.76% (191/6913) patients read at least one note at some point in the past (126 patients reported never reading a note in the past, and 51 patients reported Don't Know and were excluded from the importance of note-reading analysis; Figure 1). Among patients who reported reading at least one visit note, 6391 completed all 5 important items and were included in the analysis. The majority of these respondents reported that notes were important for engaging in their care across all five items (Multimedia Appendix 2).

Education

Overall, the majority of patients assigned high importance to reading notes, but there were significant differences between patients with varying formal education levels in all 5 survey items ($P<.001$ for all comparisons; Multimedia Appendix 3). Nearly three-fourths of less educated patients reported that notes are extremely important to feel informed about their care (74.0%, 727/982), and to understand how their provider(s) are thinking about their medical conditions (73.3%, 720/982). Compared with respondents with masters or doctorate education, patients with a high school education or less were twice as likely to report that notes are extremely important to remember the plan of care (36.15%, 1069/2957 vs 62.4%, 613/982), and help them make decisions about their care (37.37%, 1105/2957 vs 62.3%, 612/982).

Figure 1. Flowchart of patient recruitment, enrollment, and sampling.
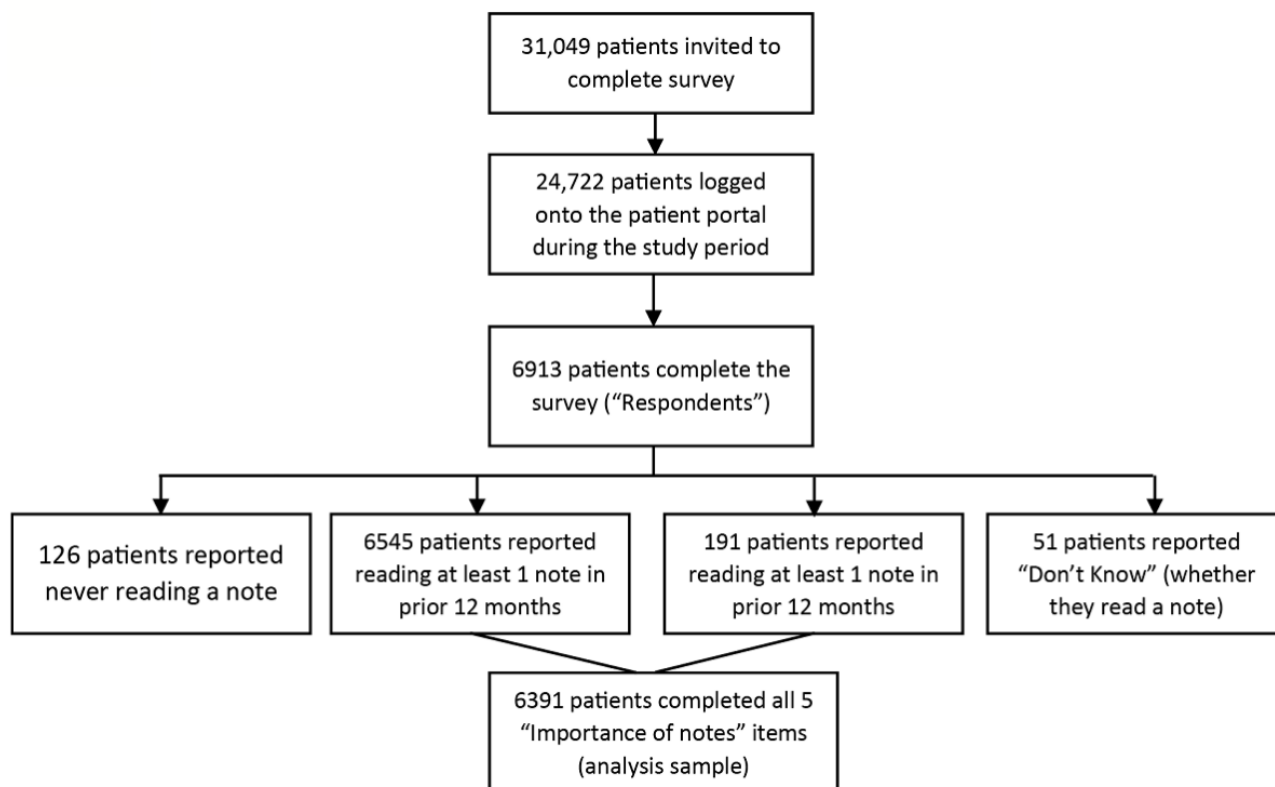
Race/Ethnicity

We found significant differences among patients of nonwhite race/ethnicity reporting on the importance of notes for all 5 items ($P \leq .003$ for all comparisons). For example, 64.1% (118/184) of black respondents felt notes were extremely important to understand their health and medical conditions compared with 51.15% (2607/5097) of white respondents. Similarly, 62.0% (114/184) of black patients compared with 41.83% (2132/5097) of white patients reported notes were extremely important to remember the care plan, and 70.7% (130/184) of black patients rated notes as extremely important to understand how the provider is thinking about their medical conditions compared with 59.17% (3016/5097) of white patients. Hispanic/Latino patients and Asian patients also consistently reported higher importance of notes in all categories compared with white patients (Multimedia Appendix 3).

Self-Reported Health

Compared with healthier patients, patients with poorer self-reported health were as or more likely to ascribe high importance of notes to understand health conditions, understand how the provider is thinking, remember the care plan, and make decisions about their care ($P \leq .02$ for all comparisons except informed about care; Multimedia Appendix 3). Across all health strata, 62.4% to 65.9% of patients reported notes were extremely important to feel informed about care. Overall, the distribution of responses was U-shaped, with patients reporting excellent health and those reporting fair or poor health being most likely to indicate that notes were extremely important to engage in care, compared with those with very good or good self-reported health. The sickest patients were as, or more, likely to find notes

extremely important for engaging in care than the healthiest patients.

Multivariable Modeling

Each of the variables described in the bivariate analysis remained significant in multivariable modeling (Table 2). Controlling for other demographic characteristics, patients with a high school education or less were more likely than those with more than a college education to report that notes are extremely important to engage in care (odds ratio [OR] 2.9, CI 2.4-3.3, $P<.001$). Similarly, controlling for other characteristics, nonwhite patients were more likely than white patients to report that notes are extremely important to engage in care (OR 2.0, CI 1.5-2.7, $P<.001$ for black patients, and OR 2.2, CI 1.7-2.9, $P<.001$ for Asian patients, and OR 2.2, CI 1.6-2.9, $P<.001$ for Hispanic or Latino patients). As in bivariate analysis, patients with excellent health remained more likely than those with very good or good health to report notes were extremely important. After accounting for race, education, and the other model factors, the sickest patients were as likely to report notes were extremely important as patients with good or very good health, but less likely to do so than those with excellent health (OR 0.7, CI 0.6-0.9, $P=.001$). Females compared with males, older patients compared with younger patients (by each incremental year of age), and patients who accessed more notes compared with those who accessed fewer notes (by each additional note accessed) were also more likely to report that notes were extremely important to engage in care, albeit with smaller effect sizes. Individual models for each of the 5 engagement items and a sensitivity analysis restricting the outcome to a mean of 10 on the summary score (rather than 9-10) revealed similar results (data not shown).

Table 1. Comparison of demographics and health care use across safety survey response status.

Demographics[a] and health care use	Total (n=31,049)	Nonrespondents (n=24,136)	Respondents (n=6913)	P value
Race/ethnicity				<.001
Asian, n (%)	1737 (5.59)	1440 (5.97)	297 (4.30)	
Black, n (%)	1458 (4.70)	1187 (4.92)	271 (3.92)	
Hispanic/Latino, n (%)	765 (2.46)	653 (2.71)	112 (1.62)	
White, n (%)	23917 (77.03)	18220 (75.49)	5697 (82.41)	
Other, n (%)	1531 (4.93)	1217 (5.04)	314 (4.54)	
Unknown, n (%)	1641 (5.29)	1419 (5.88)	222 (3.21)	
Age in years				<.001
Age, mean (range)	52 (19-101)	51 (19-101)	56 (19-96)	
Gender				.40
Female, n (%)	19371 (62.39)	15028 (62.26)	4343 (62.82)	
Male, n (%)	11678 (37.61)	9108 (37.74)	2570 (37.18)	
Education				<.001
Less than high school, n (%)	332 (1.07)	300 (1.24)	32 (0.46)	
High school graduate, n (%)	6566 (21.15)	5233 (21.68)	1333 (19.28)	
College graduate, n (%)	19797 (63.76)	14930 (61.86)	4867 (70.40)	
Unknown, n (%)	4354 (14.02)	3673 (15.22)	681 (9.85)	
Health insurance				<.001
Medicaid, n (%)	1663 (5.36)	1390 (5.76)	273 (3.95)	
Medicare, n (%)	5791 (18.65)	4156 (17.22)	1635 (23.65)	
Private, n (%)	23,517 (75.74)	18,525 (76.75)	4992 (72.21)	
Self-pay, n (%)	78 (0.25)	65 (0.27)	13 (0.19)	
Hospitalization				.28[b]
Hospitalizations over prior year (median, range)	0 (0-17)	0 (0-17)	0 (0-12)	
Notes available				<.001[b]
Number of notes available (median, IQR[c])	7 (4-13)	7 (4-13)	7 (4-14)	
Notes read				<.001[b]
Number of notes read (median, IQR)	3 (1-6)	2 (1-5)	4 (2-8)	

[a]Demographic data taken from administrative sources.

[b]Wilcoxon sign-rank test.

[c]IQR: interquartile range.

Table 2. Odds ratios estimated using logistic regression of factors associated with reporting that notes are extremely important for five health behaviors.

Covariates[a]	Odds ratio (95% CI)	*P* value
Education (reference: masters or doctoral)		
Associates or bachelors	1.4 (1.3-1.6)	<.001
High school graduate or less	2.9 (2.4-3.3)	<.001
Race/ethnicity (reference: white)		
Asian	2.2 (1.7-2.9)	<.001
Black	2.0 (1.5-2.7)	<.001
Hispanic or Latino	2.2 (1.6-2.9)	<.001
Other/multiple races	1.4 (1.1-2.0)	.02
Self-reported health (reference: excellent)		
Fair or poor	0.7 (0.6-0.9)	.001
Good	0.7 (0.6-0.8)	<.001
Very good	0.7 (0.6-0.9)	<.001
Gender (reference: male)		
Female	1.3 (1.1-1.4)	<.001
Number of notes read	1.03 (1.02-1.04)	<.001
Age	1.01 (1.01-1.01)	<.001

[a]Observations (n=464) excluded because of missing data.

Discussion

Principal Findings and Comparison With Prior Work

Our study of nearly 7000 patient portal users reveals that less educated and nonwhite patients are each independently more likely to report that reading visit notes is extremely important to engage in their care than more educated and white patients. Our study reveals several insights that can help guide future research.

First, although clinicians and health care leaders may expect OpenNotes to most benefit tech-savvy, highly educated patients, our findings suggest that even after controlling for other demographic factors, less educated patients using the portal are nearly three times as likely to report that reading visit notes is extremely important to understand and engage in their care. Nearly three-fourths of patients with a high school education or less rated reading notes as extremely important for being informed about their care and for understanding the doctor's thought process, and two-thirds reported the same for understanding their health conditions. Our findings resonate with other studies suggesting that although health literacy and access to technology are critical issues [21,28], portal registration may be a key actionable barrier to engagement for at least some patients with fewer years of formal education [14,17]. Among portal users, those with less formal education may find shared notes particularly valuable, perhaps because patients can return to their notes and review information at their own leisure and pace or share them with family or other sources of support after the visit is complete [29,30].

Irrespective of patient's educational backgrounds, experts estimate that 40%-80% of health visit information is forgotten or misremembered by patients [31]. Information decay is even more pronounced when it is not written or when it is complex. In our study, differences between patients of varying educational backgrounds were particularly stark with respect to the importance of OpenNotes to help patients remember the care plan and make health decisions. These results suggest that sharing notes with patients (and less educated patients in particular) can be important first steps to enhancing adherence and shared decision making; future research focused on these areas is needed.

Second, our findings demonstrate that nonwhite patients were twice as likely to assign extremely high importance to OpenNotes for engaging in their care when compared with white respondents, suggesting that patients of different races and ethnicities may find transparent notes helpful. The health care professionals caring for such patients may also make use of open notes as a way to engage patients of varying backgrounds. Research shows that black patients have more distrust of the health care system compared with white patients, and that this distrust may stem from perceived differences between health care professionals' values and their own, rather than from their perception of the provider's competence per se [32,33]. Other ethnic groups such as Asian and Latino populations may also experience distrust [34] or feel that health care providers do not understand their background and values [35]. In our study, over 70% of nonwhite patients reported that reading notes was extremely important to understand how the provider thinks. As greater transparency can lead to greater levels of trust, the invitation to read visit notes may itself strengthen patient-clinician relationships [11]. OpenNotes may provide an opportunity for providers to mitigate distrust by spelling out their thought process. Health care providers who also document

a clear understanding of the patient's concerns and values may take steps toward earning more trust, although further research is needed. However, lack of solicitation or understanding of patient values or use of judgmental language in notes could potentially exacerbate distrust of clinicians and may reinforce patient concerns about divergent priorities. Effects of clinician tone, language, and literacy writing levels in notes requires further study.

Finally, the relationship between self-reported health and importance of notes was more complex. On bivariate analysis, we noted a U-shaped distribution in the data, whereby patients with the highest and lowest self-reported health were most likely to report that notes were extremely important to engage in care across each of the five health behaviors. Overall, roughly two-thirds of patients with a range of poor to excellent health rated shared visit notes as extremely important to feel informed about their care. However, although prior studies suggest that chronically ill patients are more likely to report benefits from personalized health records [21,36], in our study, patients with poorest self-reported health were as likely as those with good or very good health to find notes extremely important for engaging in care, but patients with excellent health were most likely to do so after controlling for race, education, and other demographic factors. This finding may be attributable to several possible factors. For example, our assessment of health did not designate chronically ill patients from others, but rather relied on patient self-reported health, which may not correlate directly with chronic illness. Our population was healthy overall, and we did not have a large enough sample size to distinguish between patients with fair versus poor health, potentially diluting effects by our groupings. In addition, nonwhite patients and those with lower levels of education may have been disproportionately represented among patients with fair or poor health, and race and education demonstrated larger effect sizes. Finally, sicker patients may find notes are not as important because they are too ill or frail to consistently read notes. Instead, informal family and friend caregivers may be the individuals to benefit most from reading notes of sicker patients to remain informed about the patient's care, as suggested in other studies [37,38]. Future research with larger patient populations who have poor health may help better explore these effects.

About two-thirds (66%) of patients with poor/fair health and 62% of patients with excellent health reported notes were extremely important to understand how the provider is thinking. Better understanding the provider's thought process may help patients across the health spectrum see the rationale for health recommendations, potentially influencing adherence to treatment plans [39]. As OpenNotes is centered on transparency, and transparency improves trust, it may also be a first step toward improving trust between clinicians and patients, which is itself associated with greater adherence [3,40]. Taken together, our findings reinforce prior studies demonstrating that among patients who use their personalized health record, those with fewer formal years of education and lower income are more likely to feel they have learned about their health care, ask their doctor a question they may not have asked before, or do something specific to improve their health [21,36].

We were intrigued to find that older female patients were more likely to view notes as extremely important, perhaps because women and the elderly may be less assertive, and may find answers to their questions in notes (rather than having to ask the doctor), although the effect was small and this hypothesis requires formal testing. Similarly, patients who read more notes were more likely to report they were extremely important, suggesting that greater use of notes is associated with greater value (perhaps not surprisingly), although the effect size was small. Further research may better differentiate the effects of reading notes for patients who have more visits (and therefore more notes), patients who read notes more frequently, and those who read notes repeatedly.

There is ongoing debate about whether technologic innovations in eHealth will increase or decrease health disparities [3]. Hospitals, health care systems, and clinicians may not prioritize portal registration or other electronic health information engagement for less educated or nonwhite or patients, operating under the assumption that these groups are less likely to benefit. Our findings add to a growing evidence base suggesting that challenging these assumptions may prove important, particularly as health care provider endorsement remains a key predictor of portal use [41,42]. Our findings highlight that at least some patients who are less educated and nonwhite are very interested in accessing their notes on the portal and are two to three times as likely to find them valuable for various patient engagement and health care activities. Better understanding their health conditions may help patients feel better prepared for their visits and more informed for making decisions about their care [5,29]. As trust has been linked with shared decision making in minority populations [43], greater transparency may support patient engagement through stronger relationships with clinicians, provided clinicians use this tool to demonstrate an understanding and appreciation for patients' values and beliefs.

Limitations

The study has several important limitations. The survey was conducted at a single US institution, thus limiting generalizability. Although it was a large Internet survey, the majority of respondents were white and educated. However, our data from about 1000 patients with high school education or less, nonwhite race/ethnicity, and or fair or poor self-reported health can help inform future research questions for larger studies of diverse patient populations. The study also had a limited response rate, considering that incentives may drive participation. The responses are likely biased by patients who are more activated (as respondents were already registered on the portal and using OpenNotes)—a limitation that is intrinsic to the study question. Assessing how important notes are to patients for various health engagement behaviors necessitates that patients have read at least one note because responses from patients who never read notes would be hypothetical. As patients had to have an active portal account to receive a study invitation, the perspectives of patients without portal access are not necessarily represented in our data, and more research is needed to better understand and overcome barriers facing vulnerable patient populations who do not register for portals, such as rural patient populations lacking broadband access or those with Internet access who do not enroll on the portal.

Our study focused on patient perceptions of how important OpenNotes are for various health behaviors including understanding health conditions, remembering the care plan, understanding how the provider thinks, feeling informed, and making health decisions, but did not directly assess patient behaviors themselves. Whether access to health information improves adherence and empowers patients to ask questions, voice concerns, or create stronger partnerships with clinicians that enhance shared decision making needs further research. Finally, our study did not assess health literacy. Data show that patients with limited health literacy are less likely to use the portal and will require more active support, outreach, and user-centered approaches [36,44,45], and further research is needed to best engage this population. Moreover, most portals and almost all notes are in English only, although many centers are actively working on translations. Although many patient portals are largely inaccessible for direct use by patients with very limited English proficiency, they may be an important tool for friend or family care givers with whom the patient chooses to share notes (particularly if patients struggle to retain information from the visit or need more time to review and digest it), until reliable translations are routinely available through the portal. Factors affecting vulnerable populations are complex and likely do not operate in isolation. Although we modeled the relative contributions of age, sex race, education, health status, and note access in this first exploration of how some vulnerable patient groups feel about the importance of notes, more nuanced modeling to tease out the contributions of additional factors such as low health literacy, non-English preference, rural versus urban populations, and other cultural or economic influences would be helpful in studies with larger groups of vulnerable patients, including patients in other countries and those with other health insurance systems and access to health care. Such studies may help further to design most effective interventions.

Conclusions

Sharing health information with at least some vulnerable patient populations through OpenNotes may help engage them. Patients who are less educated and nonwhite are two to three times as likely to assign highest importance to reading their notes for various health behaviors including understanding their health, remembering the care plan, and making informed decisions compared with more educated and white patients. Realizing the possible benefits of OpenNotes for more vulnerable patients may require broad-scale social outreach and portal registration, patient/family and clinician education, and user-friendly portals that are designed in collaboration with diverse end users. Despite these challenges, our study suggests that at least some vulnerable patients are interested in access to their notes and perceive important benefits that may lead to improved engagement and enhanced patient-provider understanding.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Relevant survey items.

[PDF File (Adobe PDF File), 24KB - jmir_v20i5e191_app1.pdf]

Multimedia Appendix 2

Item distribution characteristics and principal component analysis results.

[PDF File (Adobe PDF File), 179KB - jmir_v20i5e191_app2.pdf]

Multimedia Appendix 3

Percent of patients reporting notes are extremely important for each of five health behaviors by patient demographics.

[PDF File (Adobe PDF File), 582KB - jmir_v20i5e191_app3.pdf]

References

1. Lohr S. New York Times. 2015 Mar 31. The Healing Power of Your Own Medical Records URL: https://www.nytimes.com/2015/04/01/technology/the-healing-power-of-your-own-medical-data.html?mcubz=0 [WebCite Cache ID 6Xmf660nc]

2. Wright A, Feblowitz J, Samal L, McCoy AB, Sittig DF. The Medicare Electronic Health Record Incentive Program: provider performance on core and menu measures. Health Serv Res 2014 Feb;49(1 Pt 2):325-346 [FREE Full text] [doi: 10.1111/1475-6773.12134] [Medline: 24359554]

3. Lyles CR, Sarkar U, Schillinger D, Ralston JD, Allen JY, Nguyen R, et al. Refilling medications through an online patient portal: consistent improvements in adherence across racial/ethnic groups. J Am Med Inform Assoc 2016 Apr;23(e1):e28-e33 [FREE Full text] [doi: 10.1093/jamia/ocv126] [Medline: 26335983]

4. Czaja SJ, Zarcadoolas C, Vaughon WL, Lee CC, Rockoff ML, Levy J. The usability of electronic personal health record systems for an underserved adult population. Hum Factors 2015 May;57(3):491-506 [FREE Full text] [doi: 10.1177/0018720814549238] [Medline: 25875437]

5. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med 2012 Oct 02;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]

6. Woods SS, Forsberg CW, Schwartz EC, Nazi KM, Hibbard JH, Houston TK, et al. The association of patient factors, digital access, and online behavior on sustained patient portal use: a prospective cohort of enrolled users. J Med Internet Res 2017 Oct 17;19(10):e345 [FREE Full text] [doi: 10.2196/jmir.7895] [Medline: 29042345]

7. Nazi KM, Turvey CL, Klein DM, Hogan TP, Woods SS. VA OpenNotes: exploring the experiences of early patient adopters with access to clinical notes. J Am Med Inform Assoc 2015 Mar;22(2):380-389. [doi: 10.1136/amiajnl-2014-003144] [Medline: 25352570]

8. Irizarry T, Shoemake J, Nilsen ML, Czaja S, Beach S, DeVito Dabbs A. Patient portals as a tool for health care engagement: a mixed-method study of older adults with varying levels of health literacy and prior patient portal use. J Med Internet Res 2017 Mar 30;19(3):e99 [FREE Full text] [doi: 10.2196/jmir.7099] [Medline: 28360022]

9. Lavizzo-Mourey R. LinkedIn. 2017 Jan 18. Where Health Conversations Should Start and End: the Patient URL: https://www.linkedin.com/pulse/where-health-conversations-should-start-end-patient-lavizzo-mourey [accessed 2018-04-10] [WebCite Cache ID 6ya4IQ9Td]

10. Oliver Wyman. 2016. Right place, right time: marketplace perspectives on providing health information to vulnerable populations URL: http://www.oliverwyman.com/our-expertise/insights/2016/jun/oliver-wyman-_-altarum-institute-study.html [accessed 2018-03-06] [WebCite Cache ID 6xiw3kLEb]

11. Bell SK, Mejilla R, Anselmo M, Darer JD, Elmore JG, Leveille S, et al. When doctors share visit notes with patients: a study of patient and doctor perceptions of documentation errors, safety opportunities and the patient-doctor relationship. Br Med J Qual Saf 2017 Apr;26(4):262-270 [FREE Full text] [doi: 10.1136/bmjqs-2015-004697] [Medline: 27193032]

12. Jhamb M, Cavanaugh KL, Bian A, Chen G, Ikizler TA, Unruh ML, et al. Disparities in electronic health record patient portal use in nephrology clinics. Clin J Am Soc Nephrol 2015 Nov 06;10(11):2013-2022 [FREE Full text] [doi: 10.2215/CJN.01640215] [Medline: 26493242]

13. Ketterer T, West DW, Sanders VP, Hossain J, Kondo MC, Sharif I. Correlates of patient portal enrollment and activation in primary care pediatrics. Acad Pediatr 2013;13(3):264-271. [doi: 10.1016/j.acap.2013.02.002] [Medline: 23680344]

14. Yamin CK, Emani S, Williams DH, Lipsitz SR, Karson AS, Wald JS, et al. The digital divide in adoption and use of a personal health record. Arch Intern Med 2011 Mar 28;171(6):568-574 [FREE Full text] [doi: 10.1001/archinternmed.2011.34] [Medline: 21444847]

15. Goel MS, Brown TL, Williams A, Hasnain-Wynia R, Thompson JA, Baker DW. Disparities in enrollment and use of an electronic patient portal. J Gen Intern Med 2011 Oct;26(10):1112-1116 [FREE Full text] [doi: 10.1007/s11606-011-1728-3] [Medline: 21538166]

16. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, López A, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc 2011 May 01;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]

17. Ancker JS, Barrón Y, Rockoff ML, Hauser D, Pichardo M, Szerencsy A, et al. Use of an electronic patient portal among disadvantaged populations. J Gen Intern Med 2011 Oct;26(10):1117-1123 [FREE Full text] [doi: 10.1007/s11606-011-1749-y] [Medline: 21647748]

18. Gordon NP, Hornbrook MC. Differences in access to and preferences for using patient portals and other eHealth technologies based on race, ethnicity, and age: a database and survey study of seniors in a large health plan. J Med Internet Res 2016 Mar 04;18(3):e50 [FREE Full text] [doi: 10.2196/jmir.5105] [Medline: 26944212]

19. Roblin DW, Houston 2nd TK, Allison JJ, Joski PJ, Becker ER. Disparities in use of a personal health record in a managed care organization. J Am Med Inform Assoc 2009;16(5):683-689 [FREE Full text] [doi: 10.1197/jamia.M3169] [Medline: 19567790]

20. Pew Research Center. 2017. Internet/Broadband Fact Sheet URL: http://www.pewinternet.org/fact-sheet/internet-broadband/ [accessed 2017-02-19] [WebCite Cache ID 6oNny1FMh]

21. Christopher Gibbons M. Use of health information technology among racial and ethnic underserved communities. Perspect Health Inf Manag 2011 Jan 01;8:1f [FREE Full text] [Medline: 21307989]

22. Schickedanz A, Huang D, Lopez A, Cheung E, Lyles CR, Bodenheimer T, et al. Access, interest, and attitudes toward electronic communication for health care among patients in the medical safety net. J Gen Intern Med 2013 Jul;28(7):914-920 [FREE Full text] [doi: 10.1007/s11606-012-2329-5] [Medline: 23423453]

23. Sarkar U, Gourley GI, Lyles CR, Tieu L, Clarity C, Newmark L, et al. Usability of commercially available mobile applications for diverse patients. J Gen Intern Med 2016 Dec;31(12):1417-1426 [FREE Full text] [doi: 10.1007/s11606-016-3771-6] [Medline: 27418347]

24. Walker J, Darer JD, Elmore JG, Delbanco T. The road toward fully transparent medical records. N Engl J Med 2014 Jan 02;370(1):6-8 [FREE Full text] [doi: 10.1056/NEJMp1310132] [Medline: 24304001]

25. Tang PC, Smith MD. Democratization of health care. J Am Med Assoc 2016 Oct 25;316(16):1663-1664 [FREE Full text] [doi: 10.1001/jama.2016.14060] [Medline: 27669345]

26. Patient-Centered Outcomes Research Institute. PCORI. 2017. Healthcare Delivery and Disparities Research URL: https://www.pcori.org/about-us/our-programs/healthcare-delivery-and-disparities-research [accessed 2018-03-06] [WebCite Cache ID 6xiufVOBf]

27. Agency for Healthcare Research and Quality. AHRQ. 2016 Apr. Agency for Healthcare Research and Quality: Division of Priority Populations URL: https://www.ahrq.gov/sites/default/files/wysiwyg/research/findings/factsheets/priority-populations/prioritypopulations_factsheet.pdf [accessed 2018-04-10] [WebCite Cache ID 6ya5sItex]

28. Patel VN, Dhopeshwarkar RV, Edwards A, Barron Y, Likourezos A, Burd L, et al. Low-income, ethnically diverse consumers' perspective on health information exchange and personal health records. Inform Health Soc Care 2011 Dec;36(4):233-252. [doi: 10.3109/17538157.2011.554930] [Medline: 21851182]

29. Bell SK, Delbanco T, Walker J. 2017 Oct 12. OpenNotes: How the power of knowing can change health care URL: https://catalyst.nejm.org/opennotes-knowing-change-health-care/ [accessed 2018-04-10] [WebCite Cache ID 6ya6cg2sh]

30. Gerard M, Fossa A, Folcarelli PH, Walker J, Bell SK. What patients value about reading visit notes: a qualitative inquiry of patient experiences with their health information. J Med Internet Res 2017 Jul 14;19(7):e237 [FREE Full text] [doi: 10.2196/jmir.7212] [Medline: 28710055]

31. Kessels RP. Patients' memory for medical information. J R Soc Med 2003 May;96(5):219-222 [FREE Full text] [Medline: 12724430]

32. Armstrong K, Putt M, Halbert CH, Grande D, Schwartz JS, Liao K, et al. Prior experiences of racial discrimination and racial differences in health care system distrust. Med Care 2013 Feb;51(2):144-150 [FREE Full text] [doi: 10.1097/MLR.0b013e31827310a1] [Medline: 23222499]

33. Armstrong K, McMurphy S, Dean LT, Micco E, Putt M, Halbert CH, et al. Differences in the patterns of health care system distrust between blacks and whites. J Gen Intern Med 2008 Jun;23(6):827-833 [FREE Full text] [doi: 10.1007/s11606-008-0561-9] [Medline: 18299939]

34. Schwei RJ, Kadunc K, Nguyen AL, Jacobs EA. Impact of sociodemographic factors and previous interactions with the health care system on institutional trust in three racial/ethnic groups. Patient Educ Couns 2014 Sep;96(3):333-338 [FREE Full text] [doi: 10.1016/j.pec.2014.06.003] [Medline: 24973901]

35. Ngo-Metzger Q, Legedza AT, Phillips RS. Asian Americans' reports of their health care experiences. Results of a national survey. J Gen Intern Med 2004 Feb;19(2):111-119 [FREE Full text] [Medline: 15009790]

36. Undem T. CHCF. National Health IT Consumer Survey URL: https://www.chcf.org/publication/consumers-and-health-information-technology-a-national-survey/ [accessed 2018-03-06] [WebCite Cache ID 6xiuOImnX]

37. Chimowitz H, Gerard M, Fossa A, Bourgeois F, Bell S. Empowering informal caregivers with health information: OpenNotes as a safety strategy. Jt Comm J Qual Patient Saf 2018 Mar;44(3):130-136. [doi: 10.1016/j.jcjq.2017.09.004]

38. Wolff JL, Darer JD, Berger A, Clarke D, Green JA, Stametz RA, et al. Inviting patients and care partners to read doctors' notes: OpenNotes and shared access to electronic medical records. J Am Med Inform Assoc 2017 Apr 01;24(e1):e166-e172. [doi: 10.1093/jamia/ocw108] [Medline: 27497795]

39. Náfrádi L, Galimberti E, Nakamoto K, Schulz PJ. Intentional and unintentional medication non-adherence in hypertension: the role of health literacy, empowerment and medication beliefs. J Public Health Res 2016 Dec 09;5(3):762 [FREE Full text] [doi: 10.4081/jphr.2016.762] [Medline: 28083523]

40. Bauer AM, Parker MM, Schillinger D, Katon W, Adler N, Adams AS, et al. Associations between antidepressant adherence and shared decision-making, patient-provider trust, and communication among adults with diabetes: diabetes study of Northern California (DISTANCE). J Gen Intern Med 2014 Aug;29(8):1139-1147 [FREE Full text] [doi: 10.1007/s11606-014-2845-6] [Medline: 24706097]

41. Vydra TP, Cuaresma E, Kretovics M, Bose-Brill S. Diffusion and use of tethered personal health records in primary care. Perspect Health Inf Manag 2015;12:1c [FREE Full text] [Medline: 26755897]

42. 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]

43. Peek ME, Gorawara-Bhat R, Quinn MT, Odoms-Young A, Wilson SC, Chin MH. Patient trust in physicians and shared decision-making among African-Americans with diabetes. Health Commun 2013;28(6):616-623 [FREE Full text] [doi: 10.1080/10410236.2012.710873] [Medline: 23050731]

44.   Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, Lopez A, et al. The literacy divide: health literacy and the use of an internet-based patient portal in an integrated health system-results from the diabetes study of northern California (DISTANCE). J Health Commun 2010;15 Suppl 2:183-196 [FREE Full text] [doi: 10.1080/10810730.2010.499988] [Medline: 20845203]

45.   Lyles CR, Ratanawongsa N, Bolen SD, Samal L. mHealth and health information technology tools for diverse patients with diabetes. J Diabetes Res 2017;2017:1704917 [FREE Full text] [doi: 10.1155/2017/1704917] [Medline: 28326330]

Abbreviations

eHealth: electronic health
EHR: electronic health record
IQR: interquartile range
OR: odds ratio

XSL·FO
RenderX

Original Paper

Computer-Based Driving in Dementia Decision Tool With Mail Support: Cluster Randomized Controlled Trial

Mark J Rapoport[1,2], MD, FRCPC; Carla Zucchero Sarracini[1], BA (Hons); Alex Kiss[3,4], PhD; Linda Lee[5], MD, MClSc(FM), CCFP(COE), FCFP; Anna Byszewski[6,7], MEd, MD, FRCPC; Dallas P Seitz[8,9], MD, PhD, FRCPC; Brenda Vrkljan[10], PhD, OT Reg (Ont); Frank Molnar[6,7], MSc, MDCM, FRCPC; Nathan Herrmann[1,2], MD, FRCPC; David F Tang-Wai[11,12], MDCM, FRCPC; Christopher Frank[13,14], MD, FCFP; Blair Henry[15,16], D Bioethics; Nicholas Pimlott[16,17], MD, PhD, CCFP; Mario Masellis[18,19], MD, PhD, FRCPC; Gary Naglie[20,21], MD, FRCPC

[1]Department of Psychiatry, Sunnybrook Health Sciences Centre, Toronto, ON, Canada

[2]Department of Psychiatry, University of Toronto, Toronto, ON, Canada

[3]Department of Research Design and Biostatistics, Sunnybrook Research Institute, Toronto, ON, Canada

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

[5]Department of Family Medicine, McMaster University, Hamilton, ON, Canada

[6]Division of Geriatric Medicine, The Ottawa Hospital, Ottawa, ON, Canada

[7]Division of Geriatric Medicine, University of Ottawa, Ottawa, ON, Canada

[8]Seniors Mental Health Program, Providence Care, Kingston, ON, Canada

[9]Department of Psychiatry, Queen's University, Kingston, ON, Canada

[10]School of Rehabilitation Science, McMaster University, Hamilton, ON, Canada

[11]Memory Clinic, University Health Network, Toronto, ON, Canada

[12]Division of Neurology, Division of Geriatric Medicine, University of Toronto, Toronto, ON, Canada

[13]Specialized Geriatric Services, Providence Care, Kingston, ON, Canada

[14]Division of Geriatric Medicine, Queen's University, Kingston, ON, Canada

[15]Clinical Ethics Centre, Sunnybrook Health Sciences Centre, Toronto, ON, Canada

[16]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada

[17]Department of Family and Community Medicine, Women's College Hospital, Toronto, ON, Canada

[18]Department of Medicine, Sunnybrook Health Sciences Centre, Toronto, ON, Canada

[19]Division of Neurology, University of Toronto, Toronto, ON, Canada

[20]Department of Medicine, Baycrest Health Sciences, Toronto, ON, Canada

[21]Division of Geriatric Medicine, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Mark J Rapoport, MD, FRCPC
Department of Psychiatry
Sunnybrook Health Sciences Centre
2075 Bayview Avenue
Toronto, ON, M4N 3M5
Canada
Phone: 1 416 480 4085
Fax: 1 416 480 5318
Email: mark.rapoport@sunnybrook.ca

Abstract

Background: Physicians often find significant challenges in assessing automobile driving in persons with mild cognitive impairment and mild dementia and deciding when to report to transportation administrators. Care must be taken to balance the safety of patients and other road users with potential negative effects of issuing such reports.

Objective: The aim of this study was to assess whether a computer-based Driving in Dementia Decision Tool (DD-DT) increased appropriate reporting of patients with mild dementia or mild cognitive impairment to transportation administrators.

Methods: The study used a parallel-group cluster nonblinded randomized controlled trial design to test a multifaceted knowledge translation intervention. The intervention included a computer-based decision support system activated by the physician-user, which provides a recommendation about whether to report patients with mild dementia or mild cognitive impairment to transportation administrators, based on an algorithm derived from earlier work. The intervention also included a mailed educational package and Web-based specialized reporting forms. Specialists and family physicians with expertise in dementia or care of the elderly were stratified by sex and randomized to either use the DD-DT or a control version of the tool that required identical data input as the intervention group, but instead generated a generic reminder about the reporting legislation in Ontario, Canada. The trial ran from September 9, 2014 to January 29, 2016, and the primary outcome was the number of reports made to the transportation administrators concordant with the algorithm.

Results: A total of 69 participating physicians were randomized, and 36 of these used the DD-DT; 20 of the 35 randomized to the intervention group used DD-DT with 114 patients, and 16 of the 34 randomized to the control group used it with 103 patients. The proportion of all assessed patients reported to the transportation administrators concordant with recommendation did not differ between the intervention and the control groups (50% vs 49%; $Z=-0.19$, $P=.85$). Two variables predicted algorithm-based reporting—caregiver concern (odds ratio [OR]=5.8, 95% CI 2.5-13.6, $P<.001$) and abnormal clock drawing (OR 6.1, 95% CI 3.1-11.8, $P<.001$).

Conclusions: On the basis of this quantitative analysis, in-office abnormal clock drawing and expressions of concern about driving from caregivers substantially influenced physicians to report patients with mild dementia or mild cognitive impairment to transportation administrators, but the DD-DT tool itself did not increase such reports among these expert physicians.

Trial Registration: ClinicalTrials.gov NCT02036099; https://clinicaltrials.gov/ct2/show/NCT02036099 (Archived by WebCite at http://www.webcitation.org/6zGMF1ky8)

KEYWORDS

dementia; mild cognitive impairment; automobile driving; decision support systems, clinical

Introduction

Motor Vehicle Collisions

In 2010, there were 1.24 million fatalities from motor vehicle collisions (MVCs) internationally, representing the eighth leading cause of death, and this is predicted to rise to fifth place by 2030 [1]. It has also been estimated that at least 20 people sustain nonfatal injuries for every MVC fatality [2]. The crash rate per mile driven begins to increase at 65 years [3], and older drivers have the highest fatality rate per mile driven among drivers over the age of 25 years [3]. Although most of the older drivers are safe drivers, various medical conditions may impact their driving ability [3]. The risk of collisions increases with age, and although this increased risk may be largely attributable to those with low mileage, collisions in older adults are more likely to be lethal than in younger adults [4]. In clinical practice, predicting driving safety in this population is very challenging [5].

Dementia

Dementia refers to a syndrome of chronic and usually progressive cognitive decline caused by changes in the structure and function of the brain. Alzheimer disease (AD) is implicated, either alone or in combination with other causes in more than two-thirds of the cases of dementia in epidemiological and autopsy samples [6,7]. Patients with AD show an inevitable decline in cognition, which ultimately will affect driving abilities over time [8]. Age is the biggest risk factor for AD, with individuals between the ages of 80 to 89 years being 7 times as likely to have AD, and those aged 90 years and older being 38 times as likely to have AD, relative to those between the ages of 70 to 79 years in a community study [9]. It is estimated that

there will be 6.7 million older adults with dementia in the United States by 2031 [10] and 1.1 million in Canada by 2038 [11].

Driving Safety and Dementia

Data from the Canadian province of Ontario also indicate that in 2000, an estimated 34,105 people with AD and related dementias were driving, with the number projected to climb to 98,032 in 2028 [12]. Many patients with mild AD may be safe to drive for some time [8,13,14], and driving cessation in dementia is associated with depression and social isolation as well as mortality [15]. On the other hand, there are significant safety concerns associated with driving in this population. Crash rates may be increased by 2 to 8 times [16,17] in dementia, although several studies have been negative with a failure to control for driving exposure [18], and our most recent systematic review update yielded inconclusive results about this increased crash risk [19]. Patients with dementia have more consistently been demonstrated to have a significantly increased rate of failure when given on-road tests of driving abilities [20-24], with a risk ratio (RR) of 10.77 (95% CI 3.00-38.62) for on-road failure rates among patients with very mild and mild dementia in our recent meta-analysis [19]. Other types of dementia, beyond AD, for example, dementia from cerebrovascular disease, Parkinson disease dementia, Lewy body dementia, and frontotemporal dementia have less predictable impacts on driving ability [8,25-28]. In-office tests have limited ability to predict crashes and on-road test failures in dementia [29-31]. Composite measures of attention, visuospatial skills, global cognition, and especially executive dysfunction are associated with crashes and on-road test failures in part [29,31], but misclassification rates are high and cutoff scores are lacking, limiting their clinical utility [30]. Mild cognitive impairment (MCI) is a condition in which there is concern about a change

XSL•FO

RenderX

of cognition, with objective evidence of cognitive impairment, but with preserved independent functioning [32]. MCI is felt to be a risk factor for dementia, but few studies have explored its association with driving safety, and different classification systems for MCI make this a particularly challenging area for physicians [33].

Medical Reporting on Driving Safety and Dementia

Seven US states and most Canadian provinces have legislation mandating the reporting of medically impaired drivers to transportation administrators [34], and clinical guidelines issued by the Canadian [35] and American Medical Associations [36] emphasize individualized assessments of drivers with dementia. However, these guidelines do not provide concrete suggestions about issues pertaining to reporting in cases of either mild dementia or MCI. Many physicians avoid discussing driving concerns, do not report their patients to transportation administrators, nor advise them on the issue of driving cessation [37-39], at least in part because of concerns over negative impacts on the doctor-patient relationship [37,38].

We conducted an earlier modified Delphi study, Driving and Dementia in Ontario (DADIO) [40,41], in which physician experts in dementia were asked whether or not they would report a patient with mild dementia or MCI to transportation administrators based on 26 hypothetical case scenarios. After 5 iterations, consensus was ultimately obtained for the majority of scenarios, and an algorithm was created to reflect the case scenarios and corresponding expert-derived reporting decisions. In that study, caregiver concern and abnormal performance on the clock drawing test (CDT) [42] accounted for 62% of the variance in reporting such patients. We also found that male physicians were 14% more likely to report than their female counterparts [41].

A multifaceted computer-based knowledge translation intervention was developed using the algorithm developed from the DADIO study as well as qualitative interviews with physicians, caregivers of former drivers with dementia, and transportation administrators. The interviews focused on facilitators and barriers to mandatory reporting and on the algorithm. The intervention, called the Driving in Dementia Decision Tool (DD-DT), also incorporated an updated review of the literature and international guidelines. The DD-DT aims to increase consistency in physician decision-making related to reporting drivers with mild dementia or MCI to transportation administrators [43].

The objective of this study was to evaluate the effects of DD-DT on physicians' reporting of patients with mild dementia and MCI to transportation administrators, to evaluate its effect on physician recommendations to patients to undergo specialized on-road testing, and to examine its effect on the physicians' perceptions of the doctor-patient and doctor-caregiver relationship.

Methods

Design

We conducted a cluster randomized controlled trial (RCT), in which physicians (the clusters) were randomized to either the

DD-DT intervention or a control group. The trial ran from September 9, 2014 to January 29, 2016.

Intervention

DD-DT and its development are described elsewhere in detail [43]. Briefly, a computer-based clinical decision support system (CCDSS) was created to guide decisions for reporting patients with mild dementia or MCI to transportation administrators (see Multimedia Appendix 1). A training video was embedded in the DD-DT website. When using DD-DT, physicians were asked to input the following variables—patient's cognitive diagnosis (MCI or mild dementia), history of MVCs in the last 2 years ("driving history"), caregiver or informant concern about the patient's driving, behavioral or neuropsychiatric disturbances in the patient, level of independence in the performance of activities of daily living and instrumental activities of daily living, and results of in-office cognitive assessment, including the speed of performance on these tests, the patient's performance on CDT [42], the Mini-Mental State Exam (MMSE) [44] score, or the Montreal Cognitive Assessment (MoCA) [45] score. Input of findings on the Trail Making Test (versions A and B) [46] was considered optional.

Depending on the data input, participants received a recommendation of "Report" to transportation administrators (see Multimedia Appendix 2), "Don't Report" (see Multimedia Appendix 3), or "No Consensus" (see Multimedia Appendix 4), with the latter recommendation indicating that the data input does not lead to a definitive recommendation, as determined by the DADIO study and algorithm [40]. Additionally, if physicians chose to report the patient, they were directed to prepopulated Ministry of Transportation of Ontario reporting forms within the computer-based DD-DT intervention (see Multimedia Appendix 5) to facilitate the reporting process. If they chose not to issue a report despite an algorithmic recommendation to do so, physicians were asked to document their rationale. Information packages for patients and caregivers were also mailed to participants, so physicians had them available to give to the patients regardless of the decision to report to transportation administrators. The package included a copy of the driving and dementia toolkit for patients and caregivers [47], which includes information on coping with grief, how to contact the local branch of the Alzheimer society, suggestions about how to initiate conversations regarding driving safety with persons with dementia, a list of local Ministry-approved specialized driving assessment centers, and a list of alternative sources of transportation in the participant's region. Thus, the DD-DT intervention encompassed the computer-based decision support system, a specialized reporting form, and a mailed information package.

Control

Physicians in the control group were instructed to input the same data onto the website as the DD-DT intervention group. However, the control group version of the DD-DT did not generate an algorithm-based reporting decision. Neither a patient/caregiver information package nor a Ministry of Transportation reporting form was provided for the control group. Instead, the physician received a generic prompt

reminding them of the mandatory reporting legislation for patients in Ontario with conditions that may affect driving.

Participants (Physicians)

Information about the study was distributed by email using national physician organizations that represent Geriatric Medicine and Geriatric Psychiatry, as well as groups of family physicians specializing in Care of the Elderly or dementia care. The investigators also contacted members of their respective disciplines to further facilitate recruitment. We also advertised in 3 continuing medical education journals and at a primary care conference on May 7 to May 9, 2015. To be included, physicians had to have access to a computer and printer in the clinical area where they saw patients. They also needed to confirm they saw at least 12 new patients per year with mild dementia or MCI. Participants in the DADIO project were eligible to participate, provided that they did not attend a June 2012 study meeting with investigators, where they would have been exposed to the algorithm, which informed the DD-DT tool. All participants were in Ontario, Canada, a jurisdiction with mandatory requirements on medical practitioners to report individuals suffering from any conditions that may make it dangerous to operate a motor vehicle [48].

Assessments (Patients)

Physicians were instructed to use the tool only with patients aged 60 years and over, who had mild dementia or MCI, and who continued to drive. Participating physicians were instructed not to use the tool for patients with moderate or severe dementia or those for whom the most responsible cause of the cognitive impairment was a primary psychiatric disorder, delirium, or alcohol or substance use.

Outcomes

After using the tool, physicians in both the intervention and control groups indicated whether or not they decided to report the patient to the Registrar of the Ministry of Transportation of Ontario (referred to as transportation administrators in this manuscript). They also indicated if a specialized road test was recommended for the patient in question. As the main goal of the tool was to enhance the physicians' reporting of patients deemed to be at significant risk of unsafe driving, the primary outcome of this study was the number of reports made to the local transportation administrators concordant with recommendations of the DD-DT algorithm (true positives). The primary outcome was the proportion of all assessments in which participants made an algorithm-concordant report (true positives/all assessments). Secondary outcomes included "any reports" filed to the transportation administrators, whether or not concordant with the DD-DT algorithm (true positives+false positives/all assessments), and recommendations for a specialized on-road test. Finally, we explored the participants' perception of the doctor-patient relationship and the doctor-caregiver relationship after each assessment. A 5-point Likert scale was used with scores ranging from −2 ("Extremely negative") to +2 ("Extremely positive"), with 0 representing "Neither negative nor positive." We also asked about their perception of how any pressure they felt from family members

played into their decision, with scores ranging from 0 to +4 and 0 representing "Not at all" and 4 representing "A great deal."

Mediators

We also measured physician-related factors that may have predicted reporting, including physician years in practice, the sex of the physician, the type of practice (family physician or specialist), whether the physician practiced in an urban community (based on their postal code), and whether their practice was hospital-based or community-based. Participants also completed the Risk-Taking Scale (RTS) that assessed their level of comfort with risk-taking versus risk-aversion, along with the Stress from Medical Uncertainty Scale (SUS), both of which are described by Pines [49] as potential physician-related mediating factors.

Patient-related mediating factors that were analyzed included the patient's age, whether the caregiver was concerned about the patient's driving (coded as yes, no, or uncertain), and whether the CDT was abnormal, based on the Mini-Cog scoring system [50]. Other potential mediators, including patient diagnosis and scores on other cognitive tests, were measured but not included in the multivariable analysis because of the limitations posed by the sample size.

Randomization

We used a cluster RCT design, in which each physician participant was considered as a "cluster." A statistician-generated randomization sequence was used to ensure equal probability of each physician participant being assigned to the intervention or control group, in a 1:1 ratio. Randomization was stratified based on sex in permuted blocks of 4 and 8 to ensure equal numbers of males and females in the intervention and control groups. Physicians who agreed to participate and provided informed consent were provided with a link to access the Web-based decision tool. Given the nature of the intervention, participants were aware of group assignment and, as such, blinding was impossible.

Analysis

Univariate analysis of variance, chi-square, and Fisher exact tests were used to compare physician and patient demographic and clinical variables between the intervention and control groups, and between those randomized who used the tool versus those who did not use the tool. We also used these univariate tests to compare the same variables between patients reported in concordance with the DD-DT algorithm (true positives) and those for whom a per-algorithm report was not issued (true negatives and false negatives). For these analyses, a tool recommendation of "No Consensus" led to a classification of concordance with the algorithm, regardless of the physician's subsequent decision and action. Then, a 2-sample, 2-sided test of proportions compared the primary outcome, reporting percentage of those deemed to be of concern (ie, "reporting concordant with algorithm"), between groups. A generalized estimating equation model with a logit link function was run to compare reporting between groups as well as physician and patient mediating factors, which were adjusted for the clustered nature of the data, assuming an exchangeable correlation structure. Before analysis, multicollinearity among the predictor

variables was assessed using tolerance statistics, with a value of less than 0.4 as the cut point. If multicollinearity was found, then only 1 member of a correlated set of variables was retained for the final model. Bivariate analyses were carried out on the remaining set of mediators in relation to the outcome. The final model included those with P values <.20 on these analyses. A similar analysis was conducted for the secondary outcomes of "any reports to transportation administrators" and "any road-test recommendations" (ie, regardless of whether these were concordant with the DD-DT algorithm). Ordinal regression models were run to examine doctor-patient relationship and doctor-caregiver relationship in relation to the predictors of interest.

Sample Size Calculation

We anticipated a base rate of reporting to transportation administrators of 13.0% (13/100) in the control group based on data from academic family practice [39]. We conducted an informal poll with 6 of the knowledge-users involved in the design of this study, enquiring about what they would view as the "minimal clinically important difference" as a trade-off for the main cost of time spent on assessment. This yielded a mean absolute difference of 12.2% (SD 3). We based the sample size on a more conservative absolute difference of 10% (ie, 13% vs 23%), in line with a recent comprehensive review of multifaceted knowledge translation interventions [51]. We planned for a sample of 36 clusters (ie, physicians) in both the intervention and control groups, and assumed an average of 7 patient assessments per physician. Using an unpooled 2-sided Z test, an alpha of .05, and an intracluster correlation coefficient of .02, a sample size of 252 assessments in each group would yield 80% power to detect an absolute difference of 10% between groups. We anticipated a 33.0% (33/100) attrition rate and, as such, planned to recruit 54 participants per arm. A maximum of 12 assessments were allowed per participant, and data were censored after that number had been reached.

Changes to the Study Protocol

We initially included physicians who indicated they typically saw at least 12 new patients per year with mild dementia or MCI; however, we quickly realized that many potential participants did not have this volume of patients with MCI or mild dementia who were still driving. We, therefore, subsequently removed this minimum volume requirement.

Several steps were taken to keep study participants engaged and to encourage them to remember to use the online tool when seeing patients with mild dementia or MCI. We sent participants in both groups up to 4 newsletters via email over the course of the study (with later recruited participants receiving fewer than 4 letters), which provided updates on research about driving and dementia, but deemed to be unlikely to influence the participants' decisions regarding whether or not to report their patients to driving administrators, or to recommend a road test (see Multimedia Appendix 6). We also invited participants to implement a chart flag system, which provided a visual reminder to use the online tool when seeing patients with mild dementia or MCI. In addition, we increased the honorarium to participants from Can $20 to $40 per use midway through the course of the study in an effort to encourage use of the tool. Finally, we gave

control group participants an opportunity to use the intervention version of the tool following the conclusion of the study.

Ethics

The study was approved by the Research Ethics Office of Sunnybrook Health Sciences Centre, #269-2013, and written informed consent was obtained for all participants.

Results

Participants and Evaluations

A total of 191 physicians expressed an interest in participating in this study. Of these, 77 did not reply to follow-up emails after initial contact, 22 were ineligible because they did not meet patient population eligibility criteria, 17 were not interested, and 6 were ineligible for other reasons (3 had been exposed to the algorithm that formed the basis of the decision tool at the DADIO study meeting, 2 did not have a printer in the area where they saw patients, and 1 was a resident physician still in training). A total of 69 physicians were enrolled in the study and 35 were randomized to the intervention group and 34 to the control group. Of the 35 physicians in the intervention group, only 20 used DD-DT at least once, with a total of 114 eligible individual patient assessments in this group. Of the 34 physicians in the control group, only 16 used the control tool at least once, with a total of 103 eligible patient assessments in that group (see Figure 1).

Participants who were randomized but did not use the tool were no different from those who did use the tool in terms of group assignment, gender, type or location of practice, or years in practice. However, those who were randomized but did not use the tool scored higher on RTS than those who did use the tool, indicating a higher tolerance of risk ($F_{1,60}=4.702$, $P=.03$, and a nonsignificant tendency to score lower on SUS ($F_{1,60}=3.765$, $P=.057$), indicating less stress from uncertainty (see Multimedia Appendix 7).

There were no significant differences in the intervention and control groups in the physician sex, years in practice, type of practice or location of practice, or on the RTS or SUS scales (Table 1). The patients assessed by the physicians in each group were similar in age, sex, history of collisions, and other measured clinical features (Table 2). Overall, 117 out of 217 patients assessed had MCI and 100 had dementia; in 16 of the latter cases, the physician was uncertain if the patient had moderate dementia. Table 3 shows etiology of MCI or dementia for patients assessed.

Reporting to Transportation Administrators Concordant With the Tool Recommendation

In the univariate analysis, the proportion of patients reported to transportation administrators concordant with the tool recommendation did not differ statistically between the intervention and the control groups, with a report issued in 50.0% (57/114) of the assessments in the intervention group and 48.5% (50/103) of the assessments in the control group. Raw data on the concordance between tool recommendation and participant action are provided in Multimedia Appendix 8. There were no differences in the measured physician-related

variables between patients reported concordant with the recommendation versus other assessments in which a recommendation-concordant report was not issued (Tables 4 and 5); however, all of the measured patient clinical variables, except patient age and rural or urban status, were associated with reports issued concordant with the recommendation (Tables

6 and 7). Overall, there was concordance between the tool recommendation and subsequent physician decision in 199 out of the 217 cases, including 59 intervention group cases and 37 control group cases in which the concordance was due to a tool recommendation of "No Consensus."

Figure 1. Consolidated Standards of Reporting Trials (CONSORT) diagram of recruitment, randomization, allocation, follow-up, and analysis.
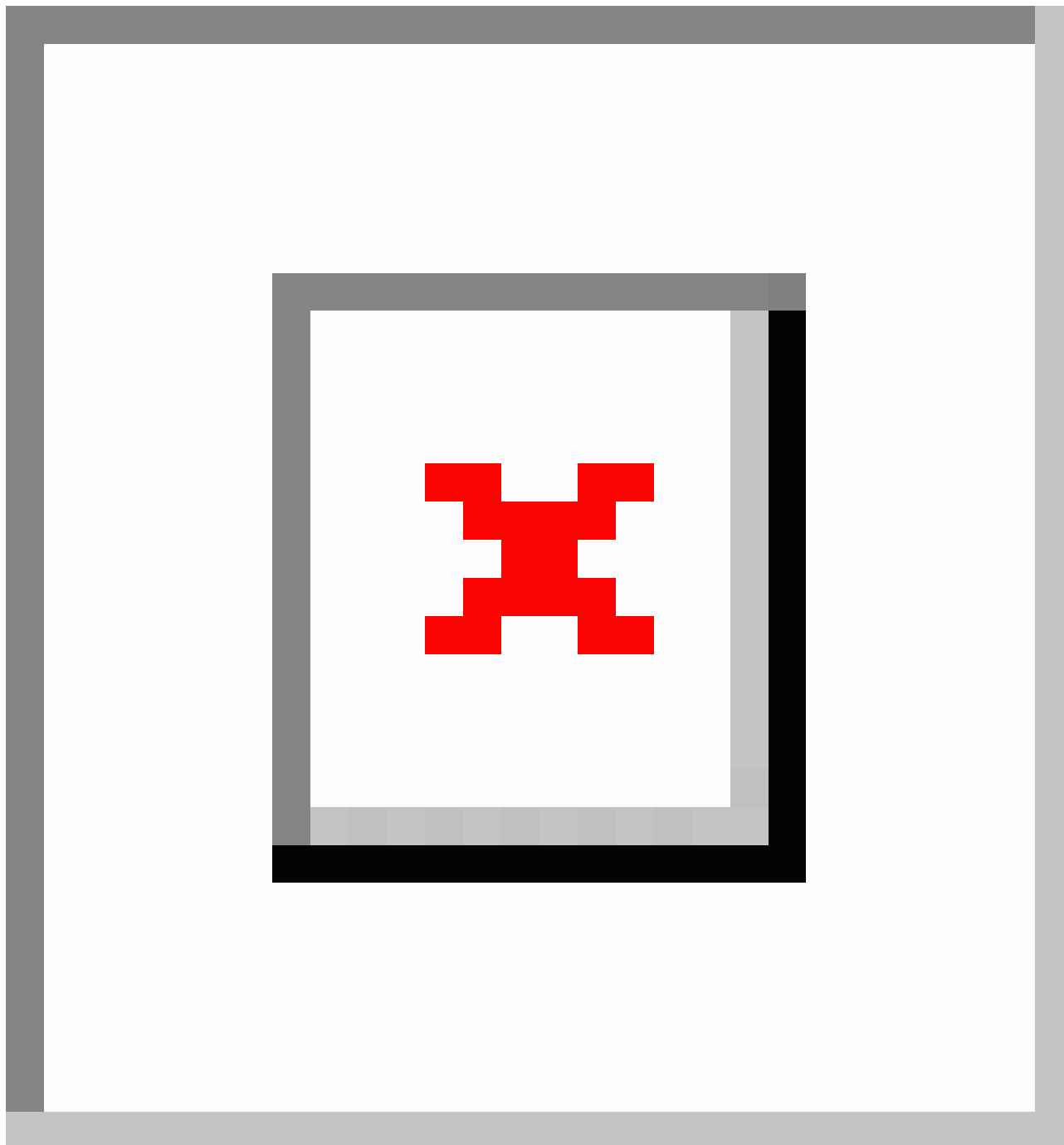
Table 1. Between-group comparisons for the participants.

Variable	Control group (n=16)	Intervention group (n=20)	Statistic		
			Chi-square (degrees of freedom)	F test (degrees of freedom)	*P* value
Physician sex (male), n (%)	8 (50)	5 (25)	2.4 (1)	—	.12
Physician years in practice, mean (SD)	13.44 (10.33)	14.58 (9.51)	—	0.118 (1,35)	.73
Type of practice (family medicine rather than specialty), n (%)	6 (38)	10 (50)	0.6 (1)	—	.45
Geographic location of practice (rural rather than urban), n (%)	0 (0)	2 (10)	—	—	.49[a]
Hospital-based practice (hospital-based rather than community-based practice), n (%)	11 (69)	12 (60)	0.3 (1)	—	.59
Risk taking scale, mean (SD)	17.56 (4.15)	16.50 (5.17)	—	0.446 (1,35)	.51
Stress from medical uncertainty scale, mean (SD)	42.13 (9.86)	42.15 (11.37)	—	0.0001 (1,35)	.99

[a]Fisher exact test.

Table 2. Between-group comparisons for the patients assessed by the participants.

Variable	Control group	Intervention group	Statistic		
			t test (degrees of freedom)	*Z* value	*P* value
Patient age in years[a], mean (SD)	78.12 (7.62)	77.73 (7.16)	0.24 (34)	—	.81
Patient sex[a] (male), n (%)	62 (60.2)	62 (54.4)	—	0.74	.46
Patient location[a] (rural), n (%)	15 (14.6)	30 (26.3)	—	0.96	.34
Mild cognitive impairment[a] (ie, not mild dementia), n (%)	54 (52.4)	63 (55.3)	—	−0.28	.78
Caregiver concern[a,b], n (%)	37 (35.9)	36 (31.6)	—	−0.66	.51
Motor vehicle collisions[a,b], n (%)	11 (10.7)	14 (12.3)	—	0.41	.68
Abnormal clock drawing test[a,c], n (%)	62 (60.2)	61 (53.5)	—	−0.65	.51
MMSE[d], mean (SD)	24.07 (3.70)	25.74 (3.56)	−1.67 (18)	—	.11
MoCA[e], mean (SD)	20.77 (3.84)	20.42 (3.87)	0.44 (32)	—	.66
Behavioral or neuropsychiatric disturbances[a,b], n (%)	25 (24.3)	21 (18.4)	—	−0.85	.39
Cognitive slowing[a], n (%)	32 (31.1)	33 (29.0)	—	−0.08	.94
Abnormal Trails B result[f,g], n (%)	33 (62.3)	51 (76.1)	—	1.60	.11

[a]Based on 217 observations.

[b]For caregiver concern, motor vehicle collisions, and behavioral or neuropsychiatric disturbances, a response of "unknown" was treated as "no" and combined with "no" for analyses.

[c]We used the Mini-Cog algorithm for scoring abnormal performance on the CDT. Specifically, "The CDT is considered normal if all the numbers are present in the correct sequence and position, and the hands readably display the requested time." [52].

[d]MMSE: Mini-Mental State Exam. Based on 75 observations.

[e]MoCA: Montreal Cognitive Assessment. Based on 182 observations.

[f]Abnormal Trails B result defined as completion time of >3 min or ≥3 errors [53].

[g]Based on 120 observations.

XSL•FO
RenderX

Table 3. Etiology of mild cognitive impairment (MCI) or dementia for patients assessed.

Type of disorder	Patients[a], n (%)
MCI[a]	86 (39.6)
Vascular cognitive impairment	22 (10.1)
Other MCI	9 (4.1)
Alzheimer disease	47 (21.7)
Vascular dementia	5 (2.3)
Mixed dementia	39 (18.0)
Lewy body dementia	1 (0.5)
Frontotemporal dementia	1 (0.5)
Dementia due to brain injury	1 (0.5)
Dementia type not yet diagnosed	5 (2.3)
Other dementia	1 (0.5)

[a]Based on 217 observations.

[b]MCI: mild cognitive impairment.

Table 4. Physician predictors of reports concordant with the tool recommendation (dichotomous variables).

Variable	Number of reports concordant with the tool recommendation, n (%)	Statistic Chi-square (degrees of freedom)	P value
Intervention group	57 (50.0)	—	—
Control group	50 (48.5)	0.05 (1)	.83
Male physicians	37 (52)	—	—
Female physicians	70 (47.9)	0.35 (1)	.57
Rural physicians	5 (38)	—	—
Urban physicians	102 (50.0)	0.7 (1)	.42
Hospital-based practice	83 (49.7)	—	—
Community-based practice	24 (48)	0.05 (1)	.83
Family physicians	35 (45)	—	—
Specialists	72/140 (51.4)	0.7 (1)	.40

Table 5. Physician predictors of reports concordant with the tool recommendation (continuous variables).

Variable	Reported concordant with algorithm (n=107), mean (SD)	Other assessments[a] (n=110), mean (SD)	Statistic F test (degrees of freedom)	P value
Physician years in practice	13.813 (9.31)	13.368 (8.49)	0.135 (1,216)	.71
Risk taking scale	16.75 (6.06)	16.77 (4.57)	0.001 (1,216)	.97
Stress from medical uncertainty scale	41.78 (11.04)	41.06 (11.16)	0.223 (1,216)	.64

[a]For this category, in 110 assessments, there were 92 cases in which physicians followed the algorithm recommendation not to report to transportation administrators, 2 cases in which physicians reported when the tool recommended not to do so, and 16 cases in which physicians did not report when the tool recommended a report.

Table 6. Patient predictors of reports concordant with the tool recommendation (dichotomous variables).

Variable	Number of reports concordant with the tool recommendation, n (%)	Statistic[a]	
		Chi-square (degrees of freedom)	P value
Male patients	54 (43.5)	—	—
Female patients	53 (57)	3.8 (1)	.05
Rural patients	20 (44)	—	—
Urban patients	87 (50.6)	0.6 (1)	.46
Mild dementia	74 (74.0)	—	—
Mild cognitive impairment	33 (28.2)	45.2 (1)	<.001
Caregiver concern	56 (77)	—	—
No caregiver concern	51 (35.4)	33.1 (1)	<.001
History of collisions	21 (84)	—	—
No history of collisions	86 (44.79)	13.6 (1)	<.001
Abnormal clock drawing test[a]	82 (66.7)	—	—
Normal clock drawing test	25 (27)	34.2 (1)	<.001
Behavioral or neuropsychiatric disturbances	37 (80)	—	—
No behavioral or neuropsychiatric disturbances	70 (40.9)	22.6 (1)	<.001
Cognitive slowing	49 (75)	—	—
No cognitive slowing	58 (38.1)	25.2 (1)	<.001
Abnormal Trails B result[b]	58 (69)	—	—
Normal Trails B result	2 (6)	40.6 (1)	<.001

[a]We used the Mini-Cog algorithm for scoring abnormal performance on clock drawing test (CDT). Specifically, "The CDT is considered normal if all the numbers are present in the correct sequence and position, and the hands readably display the requested time." [52].

[b]Abnormal Trails B result defined as completion time of >3 min or ≥3 errors [53]. n=120 cases included the Trails B data.

Table 7. Patient predictors of reports concordant with the tool recommendation (continuous variables). MMSE: Mini-Mental State Exam; MoCA: Montreal Cognitive Assessment.

Variable	Reported per algorithm (n=107), mean (SD)	Other assessments[a] (n=110), mean (SD)	Statistic	
			F test (degrees of freedom)	P value
Patient age in years	78.60 (6.98)[b]	77.25 (7.70)[c]	1.835 (1,216)	.18
MMSE[d]	24.14 (4.05)[e]	25.97 (3.10)[f]	4.901 (1,74)	.03
MoCA[g]	19.03 (3.67)[h]	22.17 (3.36)[i]	36.015 (1,181)	<.001

[a]For other assessments, there were 92 cases in which physicians followed the algorithm recommendation not to report to transportation administrators, 2 cases in which physicians reported when the tool recommended not to do so, and 16 cases in which physicians did not report when the tool recommended a report.

[b]n=107 observations.

[c]n=110 observations.

[d]MMSE: Mini-Mental State Exam.

[e]n=36 observations.

[f]n=39 observations.

[g]MoCA: Montreal Cognitive Assessment.

[h]n=92 observations.

[i]n=90 observations.

There were 8 intervention group cases in which the tool recommended reporting to transportation administrators, but the physician did not issue a report. The rationales for disregarding the tool recommendation included a perceived need for further assessment before reporting educational or medical factors playing a role in low cognitive scores, among others (see Multimedia Appendix 9).

As all tolerance values for the multicollinearity assessment were greater than 0.4, we did not exclude any variables. In the multivariable analysis, the effect of group was not significant (odds ratio [OR]=1.1, 95% CI 0.4-3.0, P=.85). Controlling for the effects of group, patient age, participant sex, participant rural or urban status, participant type of practice, participant years in practice, participant RTS scores, participant SUS scores, and clock-drawing abnormalities, the presence of caregiver concern about driving safety was associated with an increase of physician reporting in accordance with the DD-DT algorithm (OR 5.8, 95% CI 2.5-13.6, P<.001). Similarly, the presence of clock-drawing abnormalities, controlling for caregiver concern and the variables mentioned in the immediately preceding analysis, was associated with an increased likelihood of such reports (OR 6.1, 95% CI 3.1-11.8, P<.001). In the same model, SUS fell just short of statistical significance such that for each 1 unit increase, the odds of issuing such reports were increased (OR 1.04, 95% CI 1.00-1.09, P=.06).

Any Reports to the Transportation Administrators

Similar results were found for the analysis of "any reports" issued to transportation administrators, regardless of whether or not these were concordant with the tool recommendation. Physicians in the intervention group reported 50.9% (58/114) patients assessed, compared with those in the control group who reported 49.5% (51/103) patients. Although group was not significant (OR 1.1, 95% CI 0.4-2.8, P=.90), a multivariable analysis controlling for the same variables described for the primary analysis, filing "any report" was significantly associated with the presence of caregiver concern about driving (OR 5.2, 95% CI 2.3-12.0, P<.001) and clock-drawing abnormalities (OR 5.4, 95% CI 3.0-9.9, P<.001).

Recommendations for Specialized On-Road Testing

Recommendations for specialized on-road testing were issued by physicians in the intervention group for 32.4% (37/114) patients assessed compared with 33.0% (34/103) patients assessed by physicians in the control group (Z=0.70, P=.48). In this case, recommendations for testing were associated with the presence of caregiver concern about driving (OR 2.24, 95% CI 1.17-4.28, P=.01) and clock-drawing abnormalities (OR 2.26, 95% CI 1.12-4.53, P=.02) using the same model as the prior analyses.

Impact on the Physician's Relationship With the Patient and Caregiver

After using the tool, the physicians indicated negative relationships (ie, scores of −2 or −1) with patients in 21.0% (45/215) of assessments, and with caregivers in only 7.0% (15/216) of assessments. There was no significant difference between the intervention or control groups in the physicians' perceptions of their relationship with the patient or caregiver, controlling for the decision to report, the presence of caregiver concern about driving, or the presence of clock-drawing abnormalities. However, the filing of a report (controlling for group, caregiver concern, and clock-drawing abnormalities) was significantly associated with a lower likelihood of a perceived good relationship with the caregiver (OR 0.34, 95%

CI 0.19-0.62, P=.01) and with the patient (OR 0.23, 95% CI 0.12-0.43, P<.001).

Discussion

Principal Findings

We found that use of a multifaceted DD-DT in a Canadian province with mandatory reporting legislation did not increase the likelihood of physician reporting of patients with MCI or mild dementia to transportation administrators, as compared with a legislation reminder of the legislation. Other researchers have assembled algorithms, pathways, or educational approaches to guide clinicians in assessing and reporting to authority drivers with dementia [54-56]. Such approaches have been found to improve physician knowledge and confidence [54,56], and a decision aid geared at patients with dementia, rather than their physicians, reduced decisional conflict in an uncontrolled pilot study [57]. However, this is the first study to assess the impact of these interventions on actual reporting of patients to transportation administrators. Our between-group differences were not clinically meaningful, and a post-hoc sample size calculation indicates that 39,240 assessments in each group would be required to find the difference that we observed with 80% power and a 2-sided alpha of .05.

It is unclear why DD-DT did not increase reporting rates to transportation administrators, but there are several possibilities. We found a much higher reporting rate in the control group than we had anticipated based on earlier work in academic family practice [39] or in our prior study with hypothetical cases in which there was consensus to report just over one-quarter of patients with MCI or mild dementia [40]. It is likely that the effect of being observed in a research study (ie, the "Hawthorne" effect) increased the reporting rate substantially beyond what we had anticipated.

We initially planned to recruit more family physicians than specialists. Specialists, and indeed highly specialized family physicians such as those who participated in this study, may intuitively and implicitly use reporting-related heuristics by virtue of their training and extensive experience with such patients [58], and thereby be less likely to incrementally benefit from DD-DT. Indeed, there is some RCT evidence that more robust clinical outcomes are seen when CCDSSs are used by generalists [59-61], and observational evidence that generalists may be more likely to use best-practice algorithms than specialists [62]. Physician behavior is difficult to change, and to some extent, fears of malpractice suits may drive behavior [63]. Malpractice-related concerns may be particularly salient with regard to not reporting potentially medically-impaired drivers to transportation administrators, and specialists may feel that it is important for them to make such reports to preserve patients' relationship with their primary care providers [34,37,64]. Previously documented physician concerns about the impact of reporting on the doctor-patient relationship [34,37,64] were confirmed empirically in this study. We did not, however, confirm earlier findings pertaining to physician-related predictors of reporting, such as years in practice [65], physician sex, or self-rated perceived stress from medical uncertainty [41].

Clinical predictors were robustly associated with reporting to transportation administrators. Specifically, concern by caregivers and abnormality on a CDT were found to sway physician behavior, above and beyond the effect of randomized group assignment in our multivariable analysis. A lower MMSE was predictive of reporting to the licensing authority in a large-scale Swedish registry study in which only 9% of 5113 patients with dementia were reported [66]. In our earlier Delphi study [40] in which we presented hypothetical cases, and in this study with real patients, physician reporting was highly tied to cognitive findings and caregiver concerns. Although studies have documented low agreement between physicians' predictions and on-road results [67,68], when faced with uncertainty about patients with mild dementia in jurisdictions with mandatory reporting, physicians appear to use their judgment about the clinical picture in deciding whether to file these reports. Caregiver concern about driving ability, in particular, is a challenging area for clinicians. When present, it has been recommended as an important consideration and marker for the need for driving assessment, but the absence of caregiver concern is considered less predictive or helpful [69]. Caregiver concerns were associated with road-test outcomes in 2 recent studies of cognitively impaired drivers [70,71]. However, in one of those studies, the caregiver concerns were only correlated with on-road and naturalistic driving outcomes when the caregiver was an adult child (mostly female in that study), but when the caregiver was a spouse (mostly male), better ratings of driving ability were counter-intuitively associated with worse driving performance [70]. Although our univariate analysis showed that abnormal performance on the CDT led physicians to issue a report to transportation administrators, a prior study showed that various scoring systems of the CDT had limited predictive ability of impairment of on-road test performance among those with mild AD and healthy controls [72]. Similarly, Trails B, which was associated with reporting in the univariate analysis, has been shown in a number of studies to be associated with on-road driving performance in dementia, but with limited predictive ability [73], and limitations in the evidence base.

Strengths of This Study

A review of RCTs of CCDSSs and Knowledge Management systems in 2012 found that less than 20% of 148 studies assessed clinical outcomes [74]. Our study had some important strengths when considering the evaluation of CCDSSs, building in features that have been found to be associated with more successful outcomes in RCTs of CCDSSs [75]. We used a cluster randomization design of physicians to the intervention and control conditions. Physicians in both of these groups were required to enter clinical data about the patients assessed, equalizing the Hawthorne effect by adding a control. We also circumvented the known effects of using checklists [76,77] on physician behavior by requiring clinical data entry in both groups. We randomized the physician rather than the patient to circumvent contamination bias, in which the physician learns the tool with an intervention patient and applies it via "contamination" to a control patient [76]. We stratified our randomization by physician sex, which was found to be an important predictor in our earlier study. We required physicians

to provide a reason for their decision if it was discordant with the recommendation of the DD-DT.

Limitations of This Study

There are some limitations to our work that should be considered. First, there were high levels of nonuse or low use of the tool by the participants enrolled in our study, similar to a naturalistic observation study of CCDSS in primary care for heart failure [78]. We may have found more use of the tool and more between-group differences had we embedded the DD-DT into the physicians' existing electronic medical records and work-flow procedures of each clinic [74,79], and explicitly provided the justification of the decision support with research evidence. Although there were few differences between those physicians who did and did not use the tool, those who did not use the tool tended to have a higher tolerance of risk and lower stress from medical uncertainty. Second, because we had a large number of cases in which the DD-DT produced a No Consensus recommendation, the number of cases counted as concordant with the algorithm was higher than expected. Third, our results may not be representative of doctors and patients in rural settings, as we included very few of these. Fourth, the patients assessed were heterogeneous with respect to etiology of MCI or dementia, and in a minority of cases, physicians were not completely sure that the dementia was of mild severity. Fifth, we did not incorporate the Trail Making B test into the multivariable analysis, even though it was predictive in the univariate analysis. The rationale was that the Trail Making B test was optional for participants, and only 56% of participants used this as part of their assessment. Furthermore, the sample size precluded adding more than our a priori variables to the multivariable model. Sixth, the results may not generalize to jurisdictions in which there is discretionary rather than mandatory reporting legislation for medically impaired drivers, or to jurisdictions with mandatory reporting legislation that specifically requires reporting of individuals with dementia. Finally, the study results may not apply to physicians who see small numbers of patients with MCI and mild dementia, and have limited expertise in assessing driving risk in patients with cognitive impairment. The tool may be more useful to such physicians than it was in our sample with more expert physicians.

We confirmed that in a jurisdiction with mandatory reporting of medically impaired drivers, physicians base their decision to report concerns about the driving of patients with mild dementia or mild cognitive impairment on caregiver concerns and abnormal clock drawing. However, we did not find that DD-DT increased these reports beyond a simple reminder about the legislation. We also confirmed a negative impact of reporting on the doctor-patient relationship, as perceived by the physician. A preliminary analysis of the qualitative data shows that in general, family physicians had more positive views of the tool than specialist physicians, and some highlighted barriers were identified, including the lack of integration with electronic medical records and the fact that the DD-DT could not incorporate certain contextual nuances. It will be important to understand further the barriers and facilitators of using DD-DT, and a more extensive qualitative analysis of interviews conducted with those physicians who participated in the

intervention group, as well as a group of other health care   professionals, is under way.

Authors' Contributions

MJR was responsible for the conceptual framework of the study, oversaw all phases of the design and implementation of the research, data analysis, and preparation of the manuscript. CZS was the research coordinator, who helped create and test the online tool, and contributed to data analysis and all drafts of the manuscript. AK designed and carried out the data analysis plan and contributed to the manuscript. LL provided extensive guidance and input on the development of the tool and assisted with recruitment of participants and editing of the manuscript. AB contributed to the development of the tool, oversaw the group which created the caregiver resource package, and assisted with editing the manuscript. DPS, BV, FM, NH, DFTW, CF, NP, MM, and GN provided extensive guidance and input throughout the tool development process and assisted with recruitment of participants and editing of the manuscript. BH provided feedback on the conceptual framework and drafts of the manuscript.

Conflicts of Interest

DPS has participated in clinical trials sponsored by Roche Pharmaceuticals.

Multimedia Appendix 1

Screenshot of the Driving in Dementia Decision Tool study website home page.

[PNG File, 838KB - jmir_v20i5e194_app1.png]

Multimedia Appendix 2

Screenshot of recommendation to report patient to transportation administrators.

[PNG File, 161KB - jmir_v20i5e194_app2.png]

Multimedia Appendix 3

Screenshot of recommendation not to report patient to transportation administrators.

[PNG File, 127KB - jmir_v20i5e194_app3.png]

Multimedia Appendix 4

Screenshot of no consensus recommendation.

[PNG File, 704KB - jmir_v20i5e194_app4.png]

Multimedia Appendix 5

Sample Ministry of Transportation of Ontario reporting form.

[JPG File, 366KB - jmir_v20i5e194_app5.jpg]

Multimedia Appendix 6

References contained in newsletters to DD-DT study participants.

[PDF File (Adobe PDF File), 19KB - jmir_v20i5e194_app6.pdf]

XSL•FO

RenderX

Multimedia Appendix 7

Comparisons between enrolled participants who used the tool and those who did not.

[[PDF File (Adobe PDF File), 54KB](#) - [jmir_v20i5e194_app7.pdf](#)]

Multimedia Appendix 8

Tool recommendations compared with participant action.

[[PDF File (Adobe PDF File), 13KB](#) - [jmir_v20i5e194_app8.pdf](#)]

Multimedia Appendix 9

Rationale for "Do not Report" by 6 intervention group participants when tool recommended that the patient be reported.

[[PDF File (Adobe PDF File), 17KB](#) - [jmir_v20i5e194_app9.pdf](#)]

Multimedia Appendix 10

CONSORT-EHEALTH checklist (V 1.6.1).

[[PDF File (Adobe PDF File), 539KB](#) - [jmir_v20i5e194_app10.pdf](#)]

References

1. World Health Organization. Geneva: World Health Organization; 2013. Global status report on road safety 2013: supporting a decade of action URL: http://www.who.int/violence_injury_prevention/road_safety_status/2013/en/ [accessed 2017-06-23] [WebCite Cache ID 6uByKmEmh]
2. Peden M, Scurfield R, Sleet D, Mohan D, Hyder AEA. World Health Organization. Geneva: World Health Organization; 2004. World report on road traffic injury prevention URL: http://www.who.int/violence_injury_prevention/publications/road_traffic/world_report/en/ [accessed 2017-06-23] [WebCite Cache ID 6uBzXO7uK]
3. Li G, Braver ER, Chen LH. Fragility versus excessive crash involvement as determinants of high death rates per vehicle-mile of travel among older drivers. Accid Anal Prev 2003 Mar;35(2):227-235. [Medline: 12504143]
4. Langford J, Charlton JL, Koppel S, Myers A, Tuokko H, Marshall S, et al. Findings from the Candrive/Ozcandrive study: low mileage older drivers, crash risk and reduced fitness to drive. Accid Anal Prev 2013 Dec;61:304-310. [doi: 10.1016/j.aap.2013.02.006] [Medline: 23477415]
5. Dobbs BM, Carr DB, Morris JC. Evaluation and management of the driver with dementia. Neurologist 2002 Mar;8(2):61-70. [Medline: 12803692]
6. Schneider JA, Arvanitakis Z, Bang W, Bennett DA. Mixed brain pathologies account for most dementia cases in community-dwelling older persons. Neurology 2007 Dec 11;69(24):2197-2204. [doi: 10.1212/01.wnl.0000271090.28148.24] [Medline: 17568013]
7. Feldman H, Levy AR, Hsiung G, Peters KR, Donald A, Black SE, ACCORD Study Group. A Canadian cohort study of cognitive impairment and related dementias (ACCORD): study methods and baseline results. Neuroepidemiology 2003;22(5):265-274. [doi: 10.1159/000071189] [Medline: 12902621]
8. Duchek JM, Carr DB, Hunt L, Roe CM, Xiong C, Shah K, et al. Longitudinal driving performance in early-stage dementia of the Alzheimer type. J Am Geriatr Soc 2003 Oct;51(10):1342-1347. [Medline: 14511152]
9. Lindsay J, Laurin D, Verreault R, Hébert R, Helliwell B, Hill GB, et al. Risk factors for Alzheimer's disease: a prospective analysis from the Canadian Study of Health and Aging. Am J Epidemiol 2002 Sep 01;156(5):445-453. [Medline: 12196314]
10. Hebert LE, Scherr PA, Bienias JL, Bennett DA, Evans DA. Alzheimer disease in the US population: prevalence estimates using the 2000 census. Arch Neurol 2003 Aug;60(8):1119-1122. [doi: 10.1001/archneur.60.8.1119] [Medline: 12925369]
11. Alzheimer Society of Canada. Toronto: Alzheimer Society of Canada; 2010. Rising tide: The impact of dementia on Canadian society URL: http://alzheimer.ca/sites/default/files/Files/national/Advocacy/ASC_Rising_Tide_Full_Report_e.pdf [accessed 2018-04-19] [WebCite Cache ID 6ynw29jef]
12. Hopkins RW, Kilik L, Day DJ, Rows C, Tseng H. Driving and dementia in Ontario: a quantitative assessment of the problem. Can J Psychiatry 2004 Jul;49(7):434-438. [doi: 10.1177/070674370404900704] [Medline: 15362247]
13. Hunt L, Morris JC, Edwards D, Wilson BS. Driving performance in persons with mild senile dementia of the Alzheimer type. J Am Geriatr Soc 1993 Jul;41(7):747-752. [Medline: 8315186]
14. Molnar FJ, Patel A, Marshall SC, Man-Son-Hing M, Wilson KG. Systematic review of the optimal frequency of follow-up in persons with mild dementia who continue to drive. Alzheimer Dis Assoc Disord 2006;20(4):295-297. [doi: 10.1097/01.wad.0000213843.43871.c7] [Medline: 17132976]
15. Chihuri S, Mielenz TJ, DiMaggio CJ, Betz ME, DiGuiseppi C, Jones VC, et al. Driving cessation and health outcomes in older adults. J Am Geriatr Soc 2016 Feb;64(2):332-341 [FREE Full text] [doi: 10.1111/jgs.13931] [Medline: 26780879]

XSL•FO
RenderX

16. Drachman DA, Swearer JM. Driving and Alzheimer's disease: the risk of crashes. Neurology 1993 Dec;43(12):2448-2456. [Medline: 8255438]

17. Friedland RP, Koss E, Kumar A, Gaine S, Metzler D, Haxby JV, et al. Motor vehicle crashes in dementia of the Alzheimer type. Ann Neurol 1988 Dec;24(6):782-786. [doi: 10.1002/ana.410240613] [Medline: 3207361]

18. Man-Son-Hing M, Marshall SC, Molnar FJ, Wilson KG. Systematic review of driving risk and the efficacy of compensatory strategies in persons with dementia. J Am Geriatr Soc 2007 Jun;55(6):878-884. [doi: 10.1111/j.1532-5415.2007.01177.x] [Medline: 17537088]

19. Chee JN, Rapoport MJ, Molnar F, Herrmann N, O'Neill D, Marottoli R, et al. Update on the risk of motor vehicle collision or driving impairment with dementia: a collaborative international systematic review and meta-analysis. Am J Geriatr Psychiatry 2017 Dec;25(12):1376-1390. [doi: 10.1016/j.jagp.2017.05.007] [Medline: 28917504]

20. Barco PP, Baum CM, Ott BR, Ice S, Johnson A, Wallendorf M, et al. Driving errors in persons with dementia. J Am Geriatr Soc 2015 Jul;63(7):1373-1380. [doi: 10.1111/jgs.13508] [Medline: 26140521]

21. Davis JD, Papandonatos GD, Miller LA, Hewitt SD, Festa EK, Heindel WC, et al. Road test and naturalistic driving performance in healthy and cognitively impaired older adults: does environment matter? J Am Geriatr Soc 2012 Nov;60(11):2056-2062 [FREE Full text] [doi: 10.1111/j.1532-5415.2012.04206.x] [Medline: 23110378]

22. Lincoln NB, Radford KA, Lee E, Reay AC. The assessment of fitness to drive in people with dementia. Int J Geriatr Psychiatry 2006 Nov;21(11):1044-1051. [doi: 10.1002/gps.1604] [Medline: 16955424]

23. Ott BR, Heindel WC, Papandonatos GD, Festa EK, Davis JD, Daiello LA, et al. A longitudinal study of drivers with Alzheimer disease. Neurology 2008 Apr 01;70(14):1171-1178 [FREE Full text] [doi: 10.1212/01.wnl.0000294469.27156.30] [Medline: 18216302]

24. Hird MA, Egeto P, Fischer CE, Naglie G, Schweizer TA. A systematic review and meta-analysis of on-road simulator and cognitive driving assessment in Alzheimer's disease and mild cognitive impairment. J Alzheimers Dis 2016 May 11;53(2):713-729. [doi: 10.3233/JAD-160276] [Medline: 27176076]

25. Emre M, Ford PJ, Bilgiç B, Uç EY. Cognitive impairment and dementia in Parkinson's disease: practical issues and management. Mov Disord 2014 Apr 15;29(5):663-672. [doi: 10.1002/mds.25870] [Medline: 24757114]

26. Roh JH, Lee JH. Recent updates on subcortical ischemic vascular dementia. J Stroke 2014 Jan;16(1):18-26 [FREE Full text] [doi: 10.5853/jos.2014.16.1.18] [Medline: 24741561]

27. Turk K, Dugan E. Research brief: a literature review of frontotemporal dementia and driving. Am J Alzheimers Dis Other Demen 2014 Aug;29(5):404-408. [doi: 10.1177/1533317513518656] [Medline: 24413539]

28. Yamin S, Stinchcombe A, Gagnon S. Driving competence in mild dementia with Lewy bodies: in search of cognitive predictors using driving simulation. Int J Alzheimers Dis 2015;2015:806024 [FREE Full text] [doi: 10.1155/2015/806024] [Medline: 26713169]

29. Stern RA, Abularach LM, Seichepine DR, Alosco ML, Gavett BE, Tripodis Y. Office-based assessment of at-risk driving in older adults with and without cognitive impairment. J Geriatr Psychiatry Neurol 2016 Sep 19:-. [doi: 10.1177/0891988716666378] [Medline: 27647790]

30. Molnar FJ, Patel A, Marshall SC, Man-Son-Hing M, Wilson KG. Clinical utility of office-based cognitive predictors of fitness to drive in persons with dementia: a systematic review. J Am Geriatr Soc 2006 Dec;54(12):1809-1824. [doi: 10.1111/j.1532-5415.2006.00967.x] [Medline: 17198485]

31. Bennett JM, Chekaluk E, Batchelor J. Cognitive tests and determining fitness to drive in dementia: a systematic review. J Am Geriatr Soc 2016 Sep;64(9):1904-1917. [doi: 10.1111/jgs.14180] [Medline: 27253511]

32. Langa KM, Levine DA. The diagnosis and management of mild cognitive impairment: a clinical review. J Am Med Assoc 2014 Dec 17;312(23):2551-2561 [FREE Full text] [doi: 10.1001/jama.2014.13806] [Medline: 25514304]

33. Olsen K, Taylor JP, Thomas A. Mild cognitive impairment: safe to drive? Maturitas 2014 Jun;78(2):82-85. [doi: 10.1016/j.maturitas.2014.03.004] [Medline: 24709165]

34. Rapoport MJ, Herrmann N, Molnar FJ, Man-Son-Hing M, Marshall SC, Shulman K, et al. Sharing the responsibility for assessing the risk of the driver with dementia. CMAJ 2007 Sep 11;177(6):599-601 [FREE Full text] [doi: 10.1503/cmaj.070342] [Medline: 17846442]

35. Canadian Medical Association. CMA Driver's Guide: Determining Medical Fitness to Operate Motor Vehicles. Ottawa: Canadian Medical Association; 2012.

36. Carr D, Schwartzberg JG, Manning L, Sempek J. NHTSA. Washington: National Highway Traffic Safety Administration; 2010. Physician's Guide to Assessing and Counseling Older Drivers, 2nd edition URL: https://www.nhtsa.gov/staticfiles/nti/older_drivers/pdf/811298.pdf [accessed 2018-04-19] [WebCite Cache ID 6ynxDjNCp]

37. Jang RW, Man-Son-Hing M, Molnar FJ, Hogan DB, Marshall SC, Auger J, et al. Family physicians' attitudes and practices regarding assessments of medical fitness to drive in older persons. J Gen Intern Med 2007 Apr;22(4):531-543 [FREE Full text] [doi: 10.1007/s11606-006-0043-x] [Medline: 17372806]

38. Marshall SC, Gilbert N. Saskatchewan physicians' attitudes and knowledge regarding assessment of medical fitness to drive. CMAJ 1999 Jun 15;160(12):1701-1704 [FREE Full text] [Medline: 10410630]

39. Pimlott NJ, Siegel K, Persaud M, Slaughter S, Cohen C, Hollingworth G, et al. Management of dementia by family physicians in academic settings. Can Fam Physician 2006 Sep;52(9):1108-1109 [FREE Full text] [Medline: 17279222]

XSL•FO
RenderX

40.   Rapoport MJ, Naglie G, Herrmann N, Zucchero Sarracini C, Mulsant BH, Frank C, et al. Developing physician consensus on the reporting of patients with mild cognitive impairment and mild dementia to transportation authorities in a region with mandatory reporting legislation. Am J Geriatr Psychiatry 2014 Dec;22(12):1530-1543. [doi: 10.1016/j.jagp.2013.12.002] [Medline: 24406250]

41.   Rapoport MJ, Herrmann N, Haider S, Sarracini CZ, Molnar F, Frank C, et al. Physician sex is a predictor of reporting drivers with mild cognitive impairment and mild dementia to transportation authorities. J Am Geriatr Soc 2014 Jan;62(1):201-203. [Medline: 25180390]

42.   Sunderland T, Hill JL, Mellow AM, Lawlor BA, Gundersheimer J, Newhouse PA, et al. Clock drawing in Alzheimer's disease. A novel measure of dementia severity. J Am Geriatr Soc 1989 Aug;37(8):725-729. [Medline: 2754157]

43.   Cameron DH, Zucchero Sarracini C, Rozmovits L, Naglie G, Herrmann N, Molnar F, et al. Development of a decision-making tool for reporting drivers with mild dementia and mild cognitive impairment to transportation administrators. Int Psychogeriatr 2017 Sep;29(9):1551-1563. [doi: 10.1017/S1041610217000242] [Medline: 28325164]

44.   Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [Medline: 1202204]

45.   Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]

46.   Spreen O, Strauss E. A Compendium of Neuropsychological Tests: Administration, Norms and Commentary. New York: Oxford University Press; 1998.

47.   Byszewski AM, Aminzadeh F, Robinson K, Molnar F, Dalziel WEA. Regional Geriatric Program of Eastern Ontario. Ottawa: Regional Geriatric Program of Eastern Ontario; 2011. The driving and dementia toolkit: For Patients and Caregivers URL: http://www.rgpeo.com/media/30422/d%20%20d%20toolkit%20pt%20crgvr%20eng%20with%20hyperlinks.pdf [accessed 2018-04-19] [WebCite Cache ID 6ynyFRsiU]

48.   Government of Ontario. 1990. Highway Traffic Act R.S.O. Chapter H.8, Section 203(1) URL: https://www.ontario.ca/laws/statute/90h08 [accessed 2018-04-19] [WebCite Cache ID 6ynyXYw8J]

49.   Pines JM, Hollander JE, Isserman JA, Chen EH, Dean AJ, Shofer FS, et al. The association between physician risk tolerance and imaging use in abdominal pain. Am J Emerg Med 2009 Jun;27(5):552-557. [doi: 10.1016/j.ajem.2008.04.031] [Medline: 19497460]

50.   Borson S, Scanlan JM, Chen P, Ganguli M. The Mini-Cog as a screen for dementia: validation in a population-based sample. J Am Geriatr Soc 2003 Oct;51(10):1451-1454. [Medline: 14511167]

51.   Grimshaw JM, Eccles MP, Lavis JN, Hill SJ, Squires JE. Knowledge translation of research findings. Implement Sci 2012 May 31;7:50 [FREE Full text] [doi: 10.1186/1748-5908-7-50] [Medline: 22651257]

52.   Borson S, Scanlan J, Brush M, Vitaliano P, Dokmak A. The mini-cog: a cognitive 'vital signs' measure for dementia screening in multi-lingual elderly. Int J Geriatr Psychiatry 2000 Nov;15(11):1021-1027. [Medline: 11113982]

53.   Roy M, Molnar F. Systematic review of the evidence for Trails B cut-off scores in assessing fitness-to-drive. Can Geriatr J 2013;16(3):120-142 [FREE Full text] [doi: 10.5770/cgj.16.76] [Medline: 23983828]

54.   Byszewski AM, Graham ID, Amos S, Man-Son-Hing M, Dalziel WB, Marshall S, et al. A continuing medical education initiative for Canadian primary care physicians: the driving and dementia toolkit: a pre- and postevaluation of knowledge, confidence gained, and satisfaction. J Am Geriatr Soc 2003 Oct;51(10):1484-1489. [Medline: 14511173]

55.   Carter K, Monaghan S, O'Brien J, Teodorczuk A, Mosimann U, Taylor J. Driving and dementia: a clinical decision pathway. Int J Geriatr Psychiatry 2015 Feb;30(2):210-216 [FREE Full text] [doi: 10.1002/gps.4132] [Medline: 24865643]

56.   Meuser TM, Carr DB, Berg-Weger M, Niewoehner P, Morris JC. Driving and dementia in older adults: implementation and evaluation of a continuing education project. Gerontologist 2006 Oct;46(5):680-687. [Medline: 17050760]

57.   Carmody J, Potter J, Lewis K, Bhargava S, Traynor V, Iverson D. Development and pilot testing of a decision aid for drivers with dementia. BMC Med Inform Decis Mak 2014 Mar 19;14:19 [FREE Full text] [doi: 10.1186/1472-6947-14-19] [Medline: 24642051]

58.   Bate L, Hutchinson A, Underhill J, Maskrey N. How clinical decisions are made. Br J Clin Pharmacol 2012 Oct;74(4):614-620 [FREE Full text] [doi: 10.1111/j.1365-2125.2012.04366.x] [Medline: 22738381]

59.   Anchala R, Kaptoge S, Pant H, Di Angelantonio E, Franco OH, Prabhakaran D. Evaluation of effectiveness and cost-effectiveness of a clinical decision support system in managing hypertension in resource constrained primary health care settings: results from a cluster randomized trial. J Am Heart Assoc 2015 Jan 05;4(1):e001213 [FREE Full text] [doi: 10.1161/JAHA.114.001213] [Medline: 25559011]

60.   Carroll AE, Bauer NS, Dugan TM, Anand V, Saha C, Downs SM. Use of a computerized decision aid for ADHD diagnosis: a randomized controlled trial. Pediatrics 2013 Sep;132(3):e623-e629 [FREE Full text] [doi: 10.1542/peds.2013-0933] [Medline: 23958768]

61.   Ranta A, Dovey S, Weatherall M, O'Dea D, Gommans J, Tilyard M. Cluster randomized controlled trial of TIA electronic decision support in primary care. Neurology 2015 Apr 14;84(15):1545-1551. [doi: 10.1212/WNL.0000000000001472] [Medline: 25795645]

XSL•FO
RenderX

62. Cook DA, Sorensen KJ, Linderbaum JA, Pencille LJ, Rhodes DJ. Information needs of generalists and specialists using online best-practice algorithms to answer clinical questions. J Am Med Inform Assoc 2017 Jul 01;24(4):754-761. [doi: 10.1093/jamia/ocx002] [Medline: 28339685]

63. Bishop TF, Federman AD, Keyhani S. Physicians' views on defensive medicine: a national survey. Arch Intern Med 2010 Jun 28;170(12):1081-1083. [doi: 10.1001/archinternmed.2010.155] [Medline: 20585077]

64. Redelmeier DA, Yarnell CJ, Thiruchelvam D, Tibshirani RJ. Physicians' warnings for unfit drivers and the risk of trauma from road crashes. N Engl J Med 2012 Sep 27;367(13):1228-1236. [doi: 10.1056/NEJMsa1114310] [Medline: 23013074]

65. Adler G, Rottunda SJ. The driver with dementia: a survey of physician attitudes, knowledge, and practice. Am J Alzheimers Dis Other Demen 2011 Feb;26(1):58-64. [doi: 10.1177/1533317510390350] [Medline: 21282279]

66. Lovas J, Fereshtehnejad S, Cermakova P, Lundberg C, Johansson B, Johansson K, et al. Assessment and reporting of driving fitness in patients with dementia in clinical practice: data from SveDem, the Swedish Dementia Registry. J Alzheimers Dis 2016 May 05;53(2):631-638 [FREE Full text] [doi: 10.3233/JAD-160254] [Medline: 27163829]

67. Ranchet M, Tant M, Akinwuntan AE, Morgan JC, Devos H. Fitness-to-drive disagreements in individuals with dementia. Gerontologist 2017 Oct 01;57(5):833-837. [doi: 10.1093/geront/gnw119] [Medline: 27496014]

68. Brown LB, Ott BR, Papandonatos GD, Sui Y, Ready RE, Morris JC. Prediction of on-road driving performance in patients with early Alzheimer's disease. J Am Geriatr Soc 2005 Jan;53(1):94-98 [FREE Full text] [doi: 10.1111/j.1532-5415.2005.53017.x] [Medline: 15667383]

69. Iverson DJ, Gronseth GS, Reger MA, Classen S, Dubinsky RM, Rizzo M, Quality Standards Subcomitee of the American Academy of Neurology. Practice parameter update: evaluation and management of driving risk in dementia: report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2010 Apr 20;74(16):1316-1324 [FREE Full text] [doi: 10.1212/WNL.0b013e3181da3b0f] [Medline: 20385882]

70. Bixby K, Davis JD, Ott BR. Comparing caregiver and clinician predictions of fitness to drive in people with Alzheimer's disease. Am J Occup Ther 2015;69(3):6903270030p1-6903270030p7 [FREE Full text] [doi: 10.5014/ajot.2015.013631] [Medline: 25871601]

71. Hemmy L, Rottunda S, Adler G. The older driver with cognitive impairment: perceptions of driving ability and results of a behind the wheel test. Geriatrics (Basel) 2016 Mar;1(1):- [FREE Full text] [doi: 10.3390/geriatrics1010006] [Medline: 29354644]

72. Manning KJ, Davis JD, Papandonatos GD, Ott BR. Clock drawing as a screen for impaired driving in aging and dementia: is it worth the time? Arch Clin Neuropsychol 2014 Feb;29(1):1-6 [FREE Full text] [doi: 10.1093/arclin/act088] [Medline: 24296110]

73. Papandonatos GD, Ott BR, Davis JD, Barco PP, Carr DB. Clinical utility of the trail-making test as a predictor of driving performance in older adults. J Am Geriatr Soc 2015 Nov;63(11):2358-2364 [FREE Full text] [doi: 10.1111/jgs.13776] [Medline: 26503623]

74. Lobach D, Sanders GD, Bright TJ, Wong A, Dhurjati R, Bristow E, et al. Enabling health care decisionmaking through clinical decision support and knowledge management. Evid Rep Technol Assess (Full Rep) 2012 Apr(203):1-784. [Medline: 23126650]

75. Roshanov PS, Fernandes N, Wilczynski JM, Hemens BJ, You JJ, Handler SM, et al. Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials. Br Med J 2013;346:f657 [FREE Full text] [Medline: 23412440]

76. Wyatt J. Quantitative evaluation of clinical software, exemplified by decision support systems. Int J Med Inform 1997 Dec;47(3):165-173. [Medline: 9513005]

77. Haugen AS, Søfteland E, Almeland SK, Sevdalis N, Vonen B, Eide GE, et al. Effect of the World Health Organization checklist on patient outcomes: a stepped wedge cluster randomized controlled trial. Ann Surg 2015 May;261(5):821-828. [doi: 10.1097/SLA.0000000000000716] [Medline: 24824415]

78. Lugtenberg M, Pasveer D, van der Weijden T, Westert GP, Kool RB. Exposure to and experiences with a computerized decision support intervention in primary care: results from a process evaluation. BMC Fam Pract 2015 Oct 16;16:141 [FREE Full text] [doi: 10.1186/s12875-015-0364-0] [Medline: 26474603]

79. Nurek M, Kostopoulou O, Delaney BC, Esmail A. Reducing diagnostic errors in primary care. A systematic meta-review of computerized diagnostic decision support systems by the LINNEAUS collaboration on patient safety in primary care. Eur J Gen Pract 2015 Sep;21 Suppl:8-13 [FREE Full text] [doi: 10.3109/13814788.2015.1043123] [Medline: 26339829]

Abbreviations

AD: Alzheimer disease
CCDSS: computer-based clinical decision support system
CDT: Clock Drawing Test
DADIO: Driving and Dementia in Ontario study
DD-DT: Driving in Dementia Decision Tool
MCI: mild cognitive impairment

MMSE: Mini-Mental State Exam
MoCA: Montreal Cognitive Assessment
MVC: motor vehicle collisions
OR: odds ratio
RCT: randomized controlled trial
RTS: Risk-Taking Scale
SUS: Stress from Medical Uncertainty Scale

XSL•FO
RenderX

Original Paper

Online Self-Administered Cognitive Testing Using the Amsterdam Cognition Scan: Establishing Psychometric Properties and Normative Data

Heleen EM Feenstra[1,2], MSc; Ivar E Vermeulen[3], PhD; Jaap MJ Murre[2], PhD; Sanne B Schagen[1,2], PhD

[1]Department of Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Amsterdam, Netherlands

[2]Department of Psychology, University of Amsterdam, Amsterdam, Netherlands

[3]Department of Communication Science, VU University Amsterdam, Amsterdam, Netherlands

Corresponding Author:
Heleen EM Feenstra, MSc
Department of Psychosocial Research and Epidemiology
Netherlands Cancer Institute
Plesmanlaan 121
Amsterdam, 1066 CX
Netherlands
Phone: 31 205129077
Email: h.feenstra@nki.nl

Abstract

Background: Online tests enable efficient self-administered assessments and consequently facilitate large-scale data collection for many fields of research. The Amsterdam Cognition Scan is a new online neuropsychological test battery that measures a broad variety of cognitive functions.

Objective: The aims of this study were to evaluate the psychometric properties of the Amsterdam Cognition Scan and to establish regression-based normative data.

Methods: The Amsterdam Cognition Scan was self-administrated twice from home—with an interval of 6 weeks—by 248 healthy Dutch-speaking adults aged 18 to 81 years.

Results: Test-retest reliability was moderate to high and comparable with that of equivalent traditional tests (intraclass correlation coefficients: .45 to .80; .83 for the Amsterdam Cognition Scan total score). Multiple regression analyses indicated that (1) participants' age negatively influenced all (12) cognitive measures, (2) gender was associated with performance on six measures, and (3) education level was positively associated with performance on four measures. In addition, we observed influences of tested computer skills and of self-reported amount of computer use on cognitive performance. Demographic characteristics that proved to influence Amsterdam Cognition Scan test performance were included in regression-based predictive formulas to establish demographically adjusted normative data.

Conclusions: Initial results from a healthy adult sample indicate that the Amsterdam Cognition Scan has high usability and can give reliable measures of various generic cognitive ability areas. For future use, the influence of computer skills and experience should be further studied, and for repeated measurements, computer configuration should be consistent. The reported normative data allow for initial interpretation of Amsterdam Cognition Scan performances.

Introduction

Online Cognitive Assessments

Following the rise of computerized cognitive testing over the past decades [1,2], online cognitive testing is now also increasingly applied in both research and clinical practice [3]. Online testing shares its main advantages with computerized testing: standardization and precise (multiple) response measurements [4-6] but adds to that the advantages associated with self-administered testing: flexibility (in time and location) and cost-efficiency [1,7-11]. Importantly, this allows online

assessments to take place from home. Furthermore, central management of online test platforms allows for continuous software updates and opens ways for gathering normative data during testing. In sum, online testing may greatly facilitate large-scale cognitive data collection, which is needed in many fields of research [12-19]. One of these fields is oncology. Many cancer patients develop cognitive problems during the course of their disease [20]. With the growing community of cancer survivors and the increasingly chronic nature of many common cancers, the management of symptoms related to the disease and its treatments has become an important part of long-term survivorship care [21].

Development of the Amsterdam Cognition Scan

To advance our understanding of cognitive decline in cancer patients, we developed a self-administered online neuropsychological test battery: the Amsterdam Cognition Scan (ACS). The goal of the ACS is to measure broad cognitive functioning for research purposes. Although this new test battery was designed for the oncology setting, it measures various cognitive abilities and is therefore equally suitable for cognitive studies in other settings. The ACS measures attention, information processing, learning and memory, executive functioning, and psychomotor speed. Psychometric properties have been studied in Dutch adult noncentral nervous system cancer patients. Overall, adequate test-retest reliability was observed (n=96; 59% [57/96] female; mean age 51.8 years, SD 11.9; 57% [54/96] high education level), with correlations comparable with those of equivalent traditional tests (r / ρ=.29-.79) [22]. A second study (n=201; 55.7% [112/201] female; mean age 53.5 years, SD 12.3; 61.2% [123/201] high education level) showed concurrent validity, ie, consistency with scores from equivalent traditional tests, to be medium to large (r / ρ=.42-.70; total score r=.78), except for a visuospatial memory test (ρ=.36) [22]. Correlations were affected—as expected—by design differences between online tests and their offline counterparts. Furthermore, usability proved to be high as almost all participants could successfully complete the test battery from home—unsupervised and without technical problems. Most of these participants (90%) indicated to prefer online home assessments over online or traditional assessments from the hospital.

Objectives

To further facilitate use of the ACS, we collected reference data in a sample of 248 healthy adults. This enables indicating sensitivity to demographic characteristics, which is relevant as age, education, and—to a lesser extent—gender are often found to be associated with (various types of) cognitive performance [23,24]. Even though every new test requires representative normative data [25,26], online neuropsychological tests that are currently available often lack normative data or use data from offline assessments [1,27].

In sum, the aims of this study were to evaluate the psychometric properties of the ACS in a healthy sample and to establish regression-based demographically corrected normative data.

Methods

Participants

Reference data were collected from 248 healthy Dutch-speaking adults. All participants were recruited (October 2013-November 2014) via cancer patients of the Netherlands Cancer Institute who were participating in the ACS validation study [22]. First, patients of the validation study provided contact details of friends or family members suitable for participation in the study. Subsequently, the research team sent out invitation letters and contacted invitees by telephone 2 weeks later. Those who were interested in participation were asked several questions to verify eligibility. All participants were required to have sufficient proficiency of the Dutch language, basic computer skills (ie, being able to operate the mouse and send emails independently), and access to a computer with an Internet connection. Exclusion criteria were history of cancer and self-reported neurological or psychiatric conditions that could influence cognitive functioning (eg, schizophrenia, psychosis, clinical depression, substance dependence, or brain pathology). Because one of the seven neuropsychological tests—Place the Beads—was still under development during data collection, data for this test were obtained later in a different sample of 421 healthy Dutch adults that were recruited through an online respondent panel using similar inclusion and exclusion criteria as for the main reference sample.

Procedure

The study was approved by the review board of the Netherlands Cancer Institute conform ethical guidelines for human experimentation stated in the Declaration of Helsinki. Before the start of the assessments, informed consent was obtained from all participants. Assessments were completed in an unmonitored setting, either from home or from other private locations. Participants were instructed to find a quiet environment without distractions and to complete the test in one sitting. First, the ACS was administered: seven neuropsychological tests and two questionnaires—the Hospital Anxiety and Depression Scale (HADS) [28] on symptoms of depression and anxiety and the Multidimensional Fatigue Inventory-20 (MFI-20) [29] on fatigue—were presented. The complete battery—including two fixed, standardized breaks of 2 min each—took about 1 hour (on average 56 min for the first assessment) to complete. The ACS was followed by an online debriefing on test conditions during the assessment (eg, disruptions and technical issues). Around 6 weeks later (median=45 days; SD=13.8), the ACS was repeated using the same order and versions of neuropsychological tests.

For the Place the Beads reference sample, subscribers to an online respondent panel were stratified to match demographic characteristics of the main reference sample and subsequently invited by email. Participants from this sample received similar instructions but were presented with the Place the Beads test only. Both participants and nonparticipants received a second invitation, either as a reminder or as an invitation to repeat the test after 6 to 8 weeks (median=79 days; SD=29.5).

Computer Skills and Neuropsychological Assessment

Before the start of the first neuropsychological test, computer skills were assessed via tests of keyboard type skills (Type Skills), mouse click skills (Click Skills), and mouse drag skills (Drag Skills) that were newly developed by the research team (see Multimedia Appendix 1 for a description of these tests and Figure 1 for screenshots of these tests). To create a compound score that reflects overall computer skills, standardized scores for these three tests were averaged, such that higher scores indicate better overall computer skills [22].

The neuropsychological assessment consisted of seven tests that provide 12 measurements on generic cognitive ability areas, including a composite score. All tests were based on conventional traditional neuropsychological tests and developed to measure similar cognitive measurement constructs. An overview of these neuropsychological tests, their measurement constructs, and the traditional test versions are shown in Table 1; a more detailed description of these tests is provided in Multimedia Appendix 1; and Figure 2 and Multimedia Appendix

2 provide screenshots of the different elements of the ACS. In chronological order, the following tests were assessed: Connect the Dots I and II, Wordlist Learning, Reaction Speed, Place the Beads (analyses on this test are based on the Place the Beads reference sample), Box Tapping, Fill the Grid, Wordlist Delayed Recall and Recognition, and Digit Sequences I and II. Tests were developed for online self-administration. For participants, all that was required was a computer with sound and an Internet connection. One could use a mouse or touchpad to respond, as preferred. The ACS was programmed to run on all major Internet browsers and on all common variations in operating systems, which means that no downloads were required to access or complete the battery. Every neuropsychological test started with an instruction video in which screen captures of the test, combined with voiceover instructions, were presented. Subsequent practice with feedback ensured that participants understood the instructions and were ready to take the actual test. The ACS was presented in Dutch but is available in English as well. A German, French, and Spanish version is currently being developed.

Figure 1. Computer skill test from the Amsterdam Cognition Scan.
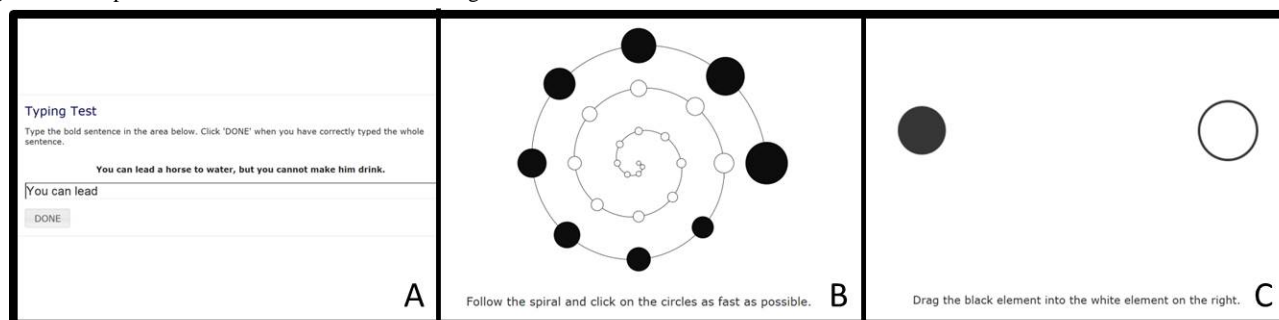
Table 1. Tests of the Amsterdam Cognition Scan and their equivalent traditional tests.

Online tests[a]	Test domains	Main outcome measure	Traditional equivalent
1. Connect the Dots I; Connect the Dots II	Visuomotor tracking, planning, cognitive flexibility, divided attention	Completion time (I and II)	Trail Making Test A Trail Making Test B [30]
2.a. Wordlist Learning	Verbal learning	Total number of correct words (trial 1 to 5)	15 Words test (Dutch version of Rey Auditory Verbal Learning test) [31]
3. Reaction Speed	Information processing speed and attention	Mean reaction time	Visual Reaction Time (subtest FePsy) [32]
4. Place the Beads	Planning, response inhibition, visuospatial memory	Total number of extra moves	Tower of London, Drexel University (TOL-dx) [33]
5. Box Tapping	Visuospatial short-term memory	Total number of correctly repeated sequences	Corsi Block-tapping Test [34]
6. Fill the Grid	Fine motor skills	Completion time	Grooved Pegboard [35]
2.b. Wordlist Delayed Recall & Recognition	Retention of information: free recall and recognition	Total number of correct words; free recall and recognition	15 Words test
7. Digit Sequences I; Digit Sequences II	I: attention II: working memory	Total number of correctly repeated sequences (I and II)	WAIS III Digit Span (forward and backward) [36]

[a]Online tests displayed in chronological order

Figure 2. Neuropsychological tests of the Amsterdam Cognition Scan, in chronological order.
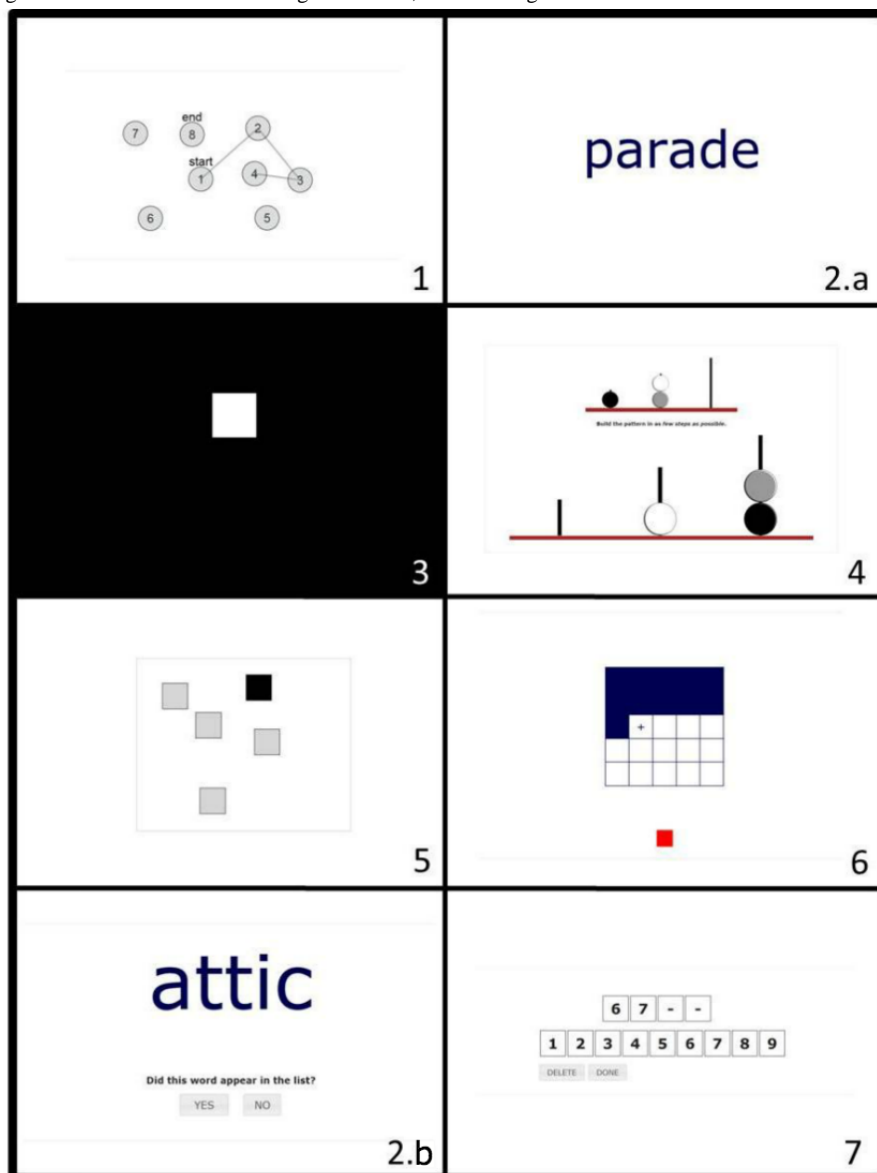
Data Analysis

Outliers on the neuropsychological test scores were identified and excluded from data analyses to limit the influence of extreme scores, possibly reflecting periods of participant distraction or technical glitches. For reaction time outcomes (Connect the Dots, Reaction Speed, and Fill the Grid), we used the median absolute deviation (MAD) method [37] to detect outliers. MADs were calculated and applied times 3.5 separately per age group: ≤40 years, 41-59 years, and ≥60 years to indicate upper and lower data limits. For tests that rely on the number of correct responses and for which zero scores are more likely to reflect usability issues than true (floor) performance (Wordlist Learning, Box Tapping, and Digit Sequences), zero scores were considered outliers and excluded from analyses.

A composite score for the total online neuropsychological test battery (total score) was calculated by averaging the standardized scores of the main outcome measures Connect the Dots I and II, Wordlist Learning, Reaction Speed, Box Tapping, Fill the Grid, Wordlist Delayed Recall, and Digit Sequences I

and II. Standardized scores of Connect the Dots, Reaction Speed, and Fill the Grid were reversed scored (z-score times −1) first. The Place the Beads test was not included in the composite score because this test was completed by a different group of participants. Wordlist Recognition was not included in the composite score as there was very little variance in the data.

To assess practice effects, paired sample t tests on the differences between performance on the first and the second assessments were performed. A significant ($P<.05$) improvement over time was interpreted as indicating practice effects [38].

Test-retest reliability was assessed using the intraclass correlation coefficient (ICC): two-way mixed effects model with measures of absolute agreement (95% CI). Similar to other correlation measures, higher ICC values indicate less error variance and better test-retest reliability, but unlike other measures, ICCs take both random and systemic error (eg, practice effects) into account when calculating reliability [39,40]. Generally, as a criterion for acceptable test-retest reliability, ICC values of .60 or .70 and higher are recommended

[39,41-43]. In agreement with these recommendations, we used a criterion of .60 to indicate which tests have acceptable reliability results and which do not. To enable interpretation of reliability results and comparisons with reported test-retest reliabilities in the literature, Pearson *r* and Spearman rho (depending on the distribution of scores on the particular measurements) were also calculated.

An absolute measure (in the unit of the measurement instrument) of reliability was calculated by the SE of measurement (SEM), using the mean squared error (MS; residual) term from the ICC formula as shown in equation (1):

(1) $SEM = \sqrt{MS}_{residual}$

In addition, the smallest detectable change (SDC), indicating—similar to the reliable change index—an interval for change beyond measurement error, was calculated as shown in equation (2) [39]:

(2) $SDC = 1.96 \times \sqrt{2} \times SEM$

When applying the SDC to group scores (eg, for research purposes), averaged scores make the measurement error smaller. Therefore, to indicate real change of group mean scores, SDC was divided by the square-root of the sample size, as shown in equation (3) [44]:

(3) $SDC_{group} = \sqrt{n}_{sample}$

Multiple regression analyses (MRA) were performed on Connect the Dots I and II, Wordlist Learning and Delayed Recall, Reaction Speed, Place the Beads, Box Tapping, Fill the Grid, Digit Sequences I and II, and the ACS total score to explore sensitivity to demographic variables. To do so, first, reference group raw scores were converted into normalized and standardized scores (mean 0, SD 1). Reverse scoring was applied for Connect the Dots, Reaction Speed, and Fill the Grid. For Connect the Dots I, Reaction Speed, and Fill the Grid, inverse transformations were applied (1 / [reversed] test score), whereas for Connect the Dots II a log10 transformation and for Place the Beads, a squared root transformation was applied. Next, for all measures, we regressed standardized scores on the predictive variables age, age-squared, gender (0=female, 1=male), and education (Verhage education scores [45]) entered blockwise. These variables were selected based on literature [25,26,46] and on previous findings of influencing factors on ACS performance in a sample of cancer patients [22]. Age was centered (age – mean group age) to avoid multicollinearity with the quadratic age term [47], which was added to model nonlinear relationships between age and test performance [26,48]. The Verhage education score, which ranges from 1 to 7, was transformed to a high-low score: 0=high (Verhage 6 and 7) and 1=low or medium (Verhage 1, 2, 3, 4, and 5). Low and medium levels were merged to one level because the lowest level was represented by only one participant with Verhage 3. In addition, for explorative analyses on the influence of computer skills and experience, we ran all models with tested computer skills as supplementary predictor and performed correlational analyses on self-reported mean number of hours of computer use per week (computer use indicated as [1] 0-5, [2] 5-15, [3] 15-35, or [4] >35 hours a week). Model parameters—consisting of standardized and unstandardized beta weights for each predictor

and a predictive constant—and the SD of residuals of the participant's observed scores were estimated for each outcome measure.

Next, nonsignificant predictors ($P>.05$) were excluded from the models, and MRA were rerun. The resulting model parameters provided the basis for regression-based formulas that can be used to calculate demographically adjusted scores.

To check assumptions for MRA, multicollinearity was assessed by calculating the variance inflation factors (VIFs) and checking whether any values were >10 [49]. In addition, for potential future interpretation of CIs and statistical significance, homoscedasticity was evaluated by visual inspection of residual-predicted values scatter plots, and normal distribution of the residuals was evaluated by visual inspection of residual histograms and p-p plots.

All statistical analyses were performed using SPSS Statistics for Windows, version 22.0 (IBM Corp). For reliability analyses, probabilities of $P<.01$ (two-tailed) were considered statistically significant to reduce chance of type one error. For multiple regressions and analyses on practice effects, probabilities of $P<.05$ (two-tailed) were considered statistically significant to be more conservative.

Results

Main Findings

For the main reference sample, letters were sent to 353 people (friends and family of patients from the ACS validation study), of whom 294 (83.3%) agreed to participate. Of these 294 participants, a total of 250 (70.8% of all 353 invitees) completed both assessments. A total of 11 participants dropped out before the start of the study and 33 during the study. The data of 2 participants were excluded from analyses because of missing demographic information (age and level of education). Therefore, analyses are based on 248 participants. In addition, 6 participants were excluded from analyses on the Wordlist Learning test, as they indicated to have used a notepad during one or both assessments. One participant was excluded from analyses on Digit Sequences II, as this participant used the entry field to memorize the digits before entering them in reverse order. Table 2 shows demographic and medical characteristics (history of brain pathology and medication use) of the dropout group and of the reference sample (n=248). All participants of the main reference sample were in the age range of 19 to 81 years (mean 49.1, SD 12.9), and 63.3% (157/248) were female. Seventy-one percent of the participants had high education levels (Verhage scale 6 or 7). For the Place the Beads reference sample, about 600 people (subscribers to an online respondent panel) were invited by email, of which 541 (90.2%) opened the link to the online test. Of these 541 participants, 433 (72.2% from all 600 invitees) completed the first assessment. A total of 12 participants were excluded from analyses as there was no information available on their gender. The 421 participants of the final Place the Beads sample were in the age range of 18 to 78 years (mean 47.9, SD 13.8), and 251 59.6% (251/421) were female. A subgroup of 143 Place the Beads participants also completed a second assessment for test-retest analyses. See

Table 2 for an overview of demographic characteristics of these subgroups of the Place the Beads samples. Figure 3 illustrates participation and completion rates of the main reference sample and the Place the Beads reference sample.

Conditions During the Self-Administered Assessments

The vast majority (75.4%, 187/248) of the participants used a standard mouse, 19% (47/248) used a trackpad, and 5.6% (14/248) used something else (eg, joystick or pen mouse) at the first assessment. Furthermore, most participants used the same device type at the second assessment (92.7%, 230/248). The results of the online debriefing on the first assessment with the ACS (n=235) are displayed in Table 3. Main findings are that 13.6% (32/235) participants were disrupted, and 11.5% (27/235) participants experienced some type of technical problem during the assessment; mostly problems with the Internet connection (23/235, 9.8%) and a few cases of problems with hardware (sound system or failed attempts having tried to use a tablet;

1.7%, 4/235). Technical problems were in all cases resolvable and did not prevent participants from completing the assessment. By far, most participants preferred online assessments from home over online or traditional assessments from the hospital (97.3% out of n=225 answering this question).

Test-Retest Reliability and Practice Effects

Test-retest reliability results indicated statistically significant correlations between the first and the second assessment for all 12 outcome measures (all $P<.001$; see Table 4). HADS and MFI-20 questionnaire results are presented separately (see Multimedia Appendix 3). The ICCs of the individual tests ranged from .45 (Place the Beads) to .80 (Fill the Grid). The total score had an ICC of .83. The majority of measures had ICCs above our .60 criterion (Connect the Dots I and II, Wordlist Recognition, Reaction Speed, Fill the Grid, Digit Sequences II, and the total score), except for Wordlist Learning and Delayed Recall, Place the Beads, Box Tapping, and Digit Sequences I.

Table 2. Demographics and clinical characteristics of the main reference sample and the Place the Beads reference sample.

Characteristics	Dropouts[a] (n=44)	Main reference sample (n=248)	Place the Beads reference sample (n=421)	Place the Beads test-retest subsample (n=143)
Demographics				
Gender, n (%)				
Female	25 (57)	157 (63.3)	251 (59.6)	65 (45.5)
Male	19 (43)	91 (36.7)	170 (40.4)	78 (54.5)
Age (years), mean (SD)	44.4 (14.1)	49.2 (13)	47.9 (13.8)	48.3 (12.7)
Education level[b], n (%)				
Low	—	—	8 (1.9)	2 (1.4)
Medium	11 (25)	73 (29.4)	196 (46.6)	64 (44.8)
High	33 (75)	175 (70.6)	217 (51.5)	77 (53.8)
Computer experience (years), mean (SD)	19.9 (7.2)	19.5 (6.3)	—	—
Clinical characteristics, n (%)				
Dyslexic	3 (7)	4 (1.6)	—	—
Concussion (history)	11 (25)	59 (23.8)	—	—
Whiplash (history)	1 (2)	10 (4)	—	—
Stroke (history)	0	1 (0.4)	—	—
Transient ischemic attacks (history)	0	1 (0.4)	—	—
Heart disease (current treatment)	1 (2)	5 (2)	—	—
High blood pressure (current treatment)	0	28 (11.3)	—	—
Diabetes mellitus (current treatment)	1 (2)	6 (2.4)	—	—
Sleep medication	0	5 (2)	—	—
Antidepressants	2 (5)	8 (3.2)	—	—
Painkillers	0	5 (2)	—	—

[a]Dropouts from the group of 296 people who agreed to participate as part of the main reference sample: 11 dropped out before starting the assessments, 15 dropped out after starting—but not completing—the first assessment, and 18 dropped out after completing the first assessment. Of other nonparticipants (n=59), no characteristics were available because of the recruitment procedure.

[b]Education is based on Verhage education scores 1 to 7 [45], corresponding with the following US years of education: 1: 1 to 5 years, 2: 6 years, 3: 7 to 8 years, 4: 7 to 9 years, 5: 7 to 10 years, 6: 7 to 16 years, and 7: 17 to 20 years. Low=Verhage 1 or 2; medium=Verhage 3, 4, or 5; and high=Verhage 6 or 7.

Figure 3. Participation and completion rates of the main reference sample and the Place the Beads reference sample.
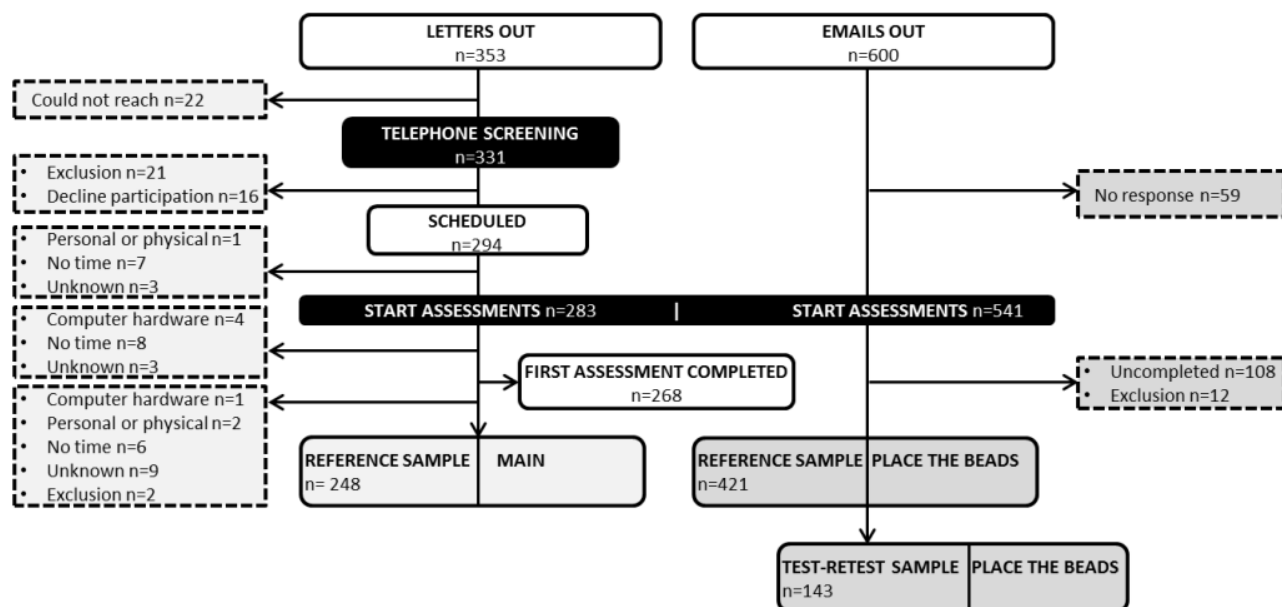
Table 3. Online debriefing on conditions during first assessments with the Amsterdam Cognition Scan (n=235).

Question	Frequency (%)
Others present during assessment	40 (17)
Task unclear after practice session	0 (0)
Received help from others	1 (0.4)
Used aid[a]	4 (1.7)
Were disrupted	32 (13.6)
By telephone or doorbell	14 (6)
By house members	17 (7.2)
Experienced technical problems	27 (11.5)
Internet connection	23 (9.8)
Hardware problem	4 (1.7)
Computer use a week (hours)	
0-5	49 (20.9)
5-15	62 (26.4)
15-35	84 (35.7)
>35	40 (17)

[a]In all cases, this concerned paper and pencil use.

In addition, we found Pearson and Spearman (depending on score distributions) correlations highly comparable with the highest test-retest correlations of the traditional tests reported in literature on studies with similar retest intervals (2 weeks-6 months). For Wordlist Delayed Recall, we found the reliability coefficient to be somewhat lower than in the literature but still relatively high (ρ=.64 vs r=.80) [50].

On the basis of the SEM, the SDC was calculated for use with mean group results for all neuropsychological outcome measures

(see Table 4). Paired samples t tests showed significant practice effects for the following tests: Connect the Dots I and II (both P<.001), Wordlist Learning and Delayed Recall (both P<.001), Place the Beads (P=.002), Box tapping (P<.001), and Fill the Grid (P<.001). A ceiling effect was observed for the Wordlist Recognition measure, indicated by small SDs and no improvement over time. For the seven measures that showed significant performance change over time, the difference scores (mean 1 – mean 2) were added to the group SDC to account for practice effects.

Table 4. Test-retest reliability and practice effects (significant *t* values in italics) for the Amsterdam Cognition Scale (Total participants [N]=248). ICC: intraclass correlation coefficient; N/A: not applicable; SDC: smallest detectable change; SEM: standard error of measurement.

Test	n (%)	Mean 1 (SD)	Mean 2 (SD)	*t* value (degrees of freedom), *P* value	SEM	SDC group[a,b]	ICC[c]	Correlation coefficient[c]	Literature[d]
Connect the Dots I	240 (96.8)	34.83 (9.39)	32.48 (7.79)	*5.43* (239), <.001	4.76	*1.00*	.67	.75[e]	.55-.73
Connect the Dots II	246 (99.2)	58.05 (18.45)	54.95 (17.62)	*3.6* (245), <.001	9.55	*1.89*	.71	.74[e]	.56-.79
Wordlist Learning	241 (97.2)	52.02 (10.13)	58.11 (9.80)	*−12.31* (240), <.001	5.44	*1.36*	.59	.75[e]	.80
Wordlist Delayed recall	241 (97.2)	11.15 (2.71)	12.09 (2.88)	*−5.34* (240), <.001	3.69	*0.72*	.50	.64[e]	.83
Wordlist Recognition	242 (97.6)	44.17 (1.57)	44.31 (1.72)	−1.76 (241), .08	0.90	0.16	.70	.54[e]	.48
Reaction Speed	241 (97.2)	308.88 (44.57)	310.49 (44.85)	−0.69 (240), .49	25.60	4.60	.67	.74[e]	.20-.82
Place the Beads[f]	143 (100)	25.63 (14.03)	21.54 (15.81)	*3.17* (142), .002	11.15	*2.93*	.45	.50[e]	.38-.70
Box Tapping	232 (93.5)	9.17 (2.26)	9.70 (1.91)	*−3.91* (231), <.001	1.47	*0.30*	.49	.46[e]	.42-.79
Fill the Grid	241 (97.2)	62.59 (12.69)	60.73 (11.42)	*3.85* (240), <.001	5.30	*0.95*	.80	.81[e]	.72-.86
Digit Sequences I	245 (98.8)	10.40 (2.22)	10.61 (2.17)	−1.52 (244), .13	1.49	0.28	.54	.54[g]	.61-.78
Digit Sequences II	242 (97.6)	8.50 (2.72)	8.78 (2.87)	−1.81 (241), .07	1.68	0.32	.64	.64[g]	.46-.71
Total score[h]	206 (83.1)	0.06 (.53)	0.07(.55)	−0.23 (205), .82	0.05	0.01	.83	.83[g]	N/A

[a]SDC group: SDC / √n.

[b]For measures with a significant practice effect (*t* value in italics), Mean1 – Mean2 difference scores were added.

[c]All correlation coefficients were significant at *P*<.001.

[d]Range of Pearson r correlation coefficients as found in the literature on studies with retest intervals between 2 weeks and 6 months on adults without progressive disease is presented: Trail Making Test A [51-54]; Trail Making Test B [23,53,54]; 15 Word Test Learning and Delayed Recall [50]; 15 Word Test Recognition [55]; Visual Reaction Time FePsy [55,56]; Tower of London [23,56-58]; Corsi Block-tapping [52,57,59,60]; Grooved Pegboard [61,62]; WAIS Digit Span [36,63].

[e]Spearman ρ.

[f]Analyses performed on data from the Place the Beads sample; participants that completed both assessments only (n=143).

[g]Pearson *r*.

[h]Total score is based on mean standardized scores from Connect the Dots I and II, Wordlist Learning and Delayed Recall, Reaction Speed, Box Tapping, Fill the Grid, and Digit Sequences I and II.

Influence of Age, Gender, Education, and Computer Skills on Neuropsychological Test Scores

The multiple regression models—all including the factors age, age-squared, gender, and/or education if significant (*P*<.05)—are presented in Table 5. MRA, first, showed that higher age significantly deteriorated scores on all neuropsychological tests, as well as on the total score (see betas in Table 5). In addition to this linear age effect, a negative quadratic age effect was found for Connect the Dots II and the total score, indicating that age deteriorated scores on these measures increasingly with older age. Second, men outperformed women on Connect the Dots I, Reaction Speed, Box Tapping, Fill the grid, and the total score. Women outperformed men on Wordlist Delayed Recall. Finally, participants with low or medium education showed lower test scores than participants with high education on Place the Beads, Digit Sequences I and II, and the total score. Mean scores and SDs for all outcome measures stratified by gender and age are provided in Table 6. MRA performed only on participants who used a standard mouse (thus excluding participants who used a trackpad or "other" device) yielded highly similar results; the effect of demographic characteristics were similar, except for age no longer being significant for performance on Digits I and gender no longer being significant for performance on Wordlist Learning and Reaction Speed.

Table 5. Multiple linear regression models including the factors age, age-squared, gender, and/or education if significant ($P<.05$) for Amsterdam Cognition Scan outcome measures. All multiple regression analyses (MRA) are performed with normalized and standardized (mean 0, SD 1) scores. Education: 0=high, 1=low or medium; Gender: 0=female, 1=male.

Test	Variable	Beta	SE beta	Standard beta	t value	P value	R^2	SD (residual)
Connect the Dots I[a,b]	Constant	−.149	.066		2.26			
	Age	−.042	.004	.549	10.44	<.001		
	Gender	.401	.109	-.193	-3.68	<.001	0.324	0.819
Connect the Dots II[a,c]	Constant	.101	.067		1.51			
	Age	−.045	.004	−.580	-11.06	<.001		
	Age-squared	−.001	.000	−.126	-2.40	.02	0.33	0.815
Wordlist Learning	Constant	−.001	.060		-.02			
	Age	−.027	.005	−.346	-5.78	<.001	0.116	0.938
Wordlist Delayed Recall	Constant	.102	.077		1.32			
	Age	−.018	.005	−.234	-3.79	<.001		
	Gender	−.282	.128	−.136	-2.21	.03	0.068	0.961
Reaction Speed[a,b]	Constant	−.106	.078		1.35			
	Age	−.019	.005	.248	3.98	<.001		
	Gender	.272	.128	−.132	-2.12	.04	0.067	0.962
Place the Beads[a,d,e]	Constant	.109	.067		1.63			
	Age	−.014	.004	−.195	-3.96	<.001		
	Education[f]	−.226	.098	−.113	-2.30	.02	0.058	0.968
Box Tapping	Constant	−.145	.077		-1.90			
	Age	−.023	.005	−.298	-4.86	<.001		
	Gender	.372	.128	.179	2.91	.004	0.11	0.939
Fill the Grid[a,b]	Constant	−.144	.069		-2.08			
	Age	−.040	.004	−.505	-9.18	<.001		
	Gender	.358	.114	.173	3.15	.002	0.267	0.852
Digit Sequences I	Constant	.103	.074		1.39			
	Age	−.010	.005	−.130	-2.08	.04		
	Education	−.350	.137	−.160	-2.56	.01	0.036	0.978
Digit Sequences II	Constant	.100	.074		1.35			
	Age	−.015	.005	−.196	-3.15	<.001		
	Education	−.357	.137	−.163	-2.61	.02	0.06	0.966
Total score	Constant	.068	.044		1.56			
	Age	−.025	.002	−.597	-11.04	<.001		
	Age-squared	.000	.000	−.117	-2.16	.03		
	Gender	.120	.057	.111	2.10	.02		
	Education	−.129	.061	−.112	-2.11	.03	0.351	0.420

[a]Reverse scoring was applied before MRA.

[b]Inverse transformations were applied (eg, 1/Connect the Dots I).

[c]Log10 transformation was applied.

[d]Squared root transformation was applied.

[e]Analyses performed on data from the Place the Beads sample.

[f]Note that education levels were more balanced in the Place the Beads sample by including more participants with middle or lower education than in

XSL•FO
RenderX

the main sample.

Table 6. Mean raw scores and SDs for the Amsterdam Cognition Scan measures stratified by gender and age.

Test	Age (years)	Gender	
		Female, mean (SD)	Male, mean (SD)
Connect the Dots I	≤40	31.28 (6.8)	26.15 (4.5)
	41-59	34.86 (8.6)	33.16 (6.9)
	≥60	46.55 (13.8)	39.84 (10.8)
Connect the Dots II	≤40	47.56 (12.9)	48.2 (12.7)
	41-59	55.6 (14.9)	54.41 (11.55)
	≥60	79.17 (22.8)	68.4 (18.07)
Wordlist Learning	≤40	56.76 (8.1)	55.45 (9.2)
	41-59	53.96 (8.8)	50.09 (9.8)
	≥60	48.03 (13)	46.26 (9.3)
Wordlist Delayed Recall	≤40	12.14 (1.8)	11.95 (2.6)
	41-59	11.54 (2.5)	10.44 (3.2)
	≥ 60	10.53 (2.9)	9.57 (3.4)
Wordlist Recognition	≤40	44.65 (0.7)	44.59 (0.96)
	41-59	44.31 (1.5)	43.93 (1.9)
	≥60	43.67 (2)	43.78 (1.5)
Reaction Speed	≤40	304.4 (35)	289.7 (33)
	41-59	309.2 (47)	297.9 (38)
	≥60	336.1 (57)	321.5 (41)
Place the Beads[a]	≤40	23.49 (14.1)	21.68 (15.5)
	41-59	27.3 (13.3)	27.07 (15.4)
	≥60	35.47 (17.7)	28.24 (15.3)
Box Tapping	≤40	9.38 (1.8)	10.24 (1.7)
	41-59	8.94 (2.6)	9.67 (2)
	≥60	7.85 (2.1)	8.78 (2.7)
Fill the Grid	≤40	60 (10.3)	50.21 (8.3)
	41-59	62 (11.7)	60.7 (10.7)
	≥60	73.1 (12.2)	72.36 (14.8)
Digit Sequences I	≤40	10.73 (2.1)	10.45 (2.6)
	41-59	10.14 (2.5)	11.02 (2)
	≥60	10.14 (1.9)	9.96 (2.3)
Digit Sequences II	≤40	9.22 (2.9)	9.32 (2.9)
	41-59	8.45 (2.7)	8.98 (2.4)
	≥60	7.23 (2.7)	8.17 (2.6)
Total score[b]	≤40	2.8 (3.7)	3.88 (3.9)
	41-59	0.64 (4.2)	1.18 (3.38)
	≥60	-4.22 (4.8)	-3.5 (4.2)

[a]Analyses performed on data from the Place the Beads sample.

[b]Total score is based on standardized score of all main neuropsychological outcome measures, except for Wordlist Recognition and Place the Beads.

XSL•FO
RenderX

To further investigate the widespread association between cognitive performance and gender, explorative analyses were performed. These analyses showed no general age difference between women and men (t_{246}=−0.79, *P*=.43). Level of education was distributed somewhat differently, with more very highly educated (Verhage 7) men than women (χ^2_4[n=248]=9.62, *P*=.047). Moreover, on average, men had more years of self-reported computer experience (t_{246}=−3.07, *P*=.002) and higher scores on the computer skills compound score ($t_{232,1}$=−2.07, *P*=.04). Concerning tested computer skills, when looking into gender differences between young (≤40 years), middle aged (41-59 years), and older aged (≥60 years) participants separately, only in the younger age group, men significantly outperformed women (t_{57}=−2.07, *P*=.04). We entered both self-reported computer experience and tested computer skill scores as additional predictors in the MRA but removed self-reported computer experience as it was not associated with any of the neuropsychological outcome measures. Tested computer skills, on the other hand, proved to be associated with performance on all time-based measures (Connect the Dots I and II, Reaction Speed, and Fill the Grid) and the total score (see Multimedia Appendix 4). For most of these measures (Connect the Dots I, Reaction Speed, Fill the Grid, and the total score), a beneficial effect of being male was found in the main MRA. After entering computer skills as a predictor, gender was no longer significantly associated with Connect the Dots II, Reaction speed, and the total score, but it remained significantly associated—albeit less so—with Connect the Dots I and Fill the Grid. Additionally, a second self-reported measure of computer experience—"mean number of hours computer use per week"—was positively correlated with several ACS measures: Connect the Dots I (ρ=.37; *P*<.001), Connect the Dots II (ρ=.27; *P*<.001), Wordlist Learning (ρ=.27; *P*<.001), Reaction Speed (ρ=.16; *P*=.02), Box Tapping (ρ=.33; *P*<.001), Fill the Grid (ρ=.36; *P*<.001), and the total score (ρ=.33; *P*<.001), as well as with tested computer skills (ρ=.4; *P*<.001).

There was no proof of multicollinearity (the highest VIF value was 1.06), indicating statistical independence of the proposed predictor variables. Evaluation of the p-p normality plots of the final models showed that standardized residuals were not normally distributed for Wordlist Delayed Recall and Box Tapping, and evaluation of residual-predicted values scatter plots indicated heteroscedasticity for Connect the Dots II. For the current analyses—focusing on estimation of model parameters—lack of normality and homoscedasticity does not invalidate results [64].

Regression-Based Normative Data

The regression models, as presented in Table 5, provide regression-based normative data. First, for each measure, demographically based predicted scores can be calculated using the constant value and the unstandardized beta weights. Age is centered by calculating the difference between the age of the patient and the mean age of the reference sample (49.19 years). For example, on Connect the Dots I, a woman aged 55 years would have a predicted score of −.149 + (55-49.19)(−.042) + 0(.401)=−.39. As we used normalized and standardized scores for the initial multiple regression analyses, actual scores need to be normalized and standardized using the same transformations as well. Subsequently, by calculating the standardized difference between actual and predicted score, a demographically corrected norm score is established. If the woman of our example has an actual score of 60 seconds, this corresponds to a normalized (in the case of Connect the Dots I reverse transformations were applied) and standardized score of ([1/60]-.0304)/.00793=−1.73, resulting in a difference score of −.39-(−1.73)=−1.34 and a norm score of −1.34/.819=−1.64.

Discussion

On the basis of assessments from 248 healthy Dutch adults, we studied psychometric properties and established regression-based, demographically corrected, normative data for the ACS.

Test-Retest Reliability and Practice Effects

Test-retest reliability was found to be adequate for most tests. ICCs (>.60) indicated sufficient consistency over time for 7 out of 12 outcome measures. Consistency between overall ACS performance (total score) was especially high. This could be expected as averaged scores generally generate higher correlations than single measurements [38,39]. Test-retest correlations were highly comparable with the highest test-retest correlations for equivalent traditional tests as reported in the literature. Current reliability results were also comparable with the test-retest results on the ACS as observed in a sample of cancer patients (6 week interval; n=96) [22], although we currently find higher reliability for Reaction Speed, Box tapping, and Fill the Grid. An important difference between this study with healthy subjects and our previous study with cancer patients was the degree of variability in computer hardware. In the patient study, repeated testing was performed from two different settings (home and hospital, in a counterbalanced design) and thus, on two different computers; in this study, the majority of participants performed both assessments on the same computer. As Reaction Speed, Box tapping, and Fill the Grid are highly dependent on mouse input, it is likely that consistent hardware over assessments has improved test-retest reliability. Therefore, in future applications with repeated testing, it will be important to pursue consistency of computer configuration; preferably using one computer and one browser type.

We found significant differences between the first and the second assessment on most of our tests (all measures except for Digit Sequences I and II, Reaction Speed, and Wordlist Recognition). For this last test, a ceiling effect was found, as is commonly found in the literature for its traditional equivalent [38,65]. Practice effects are commonly observed with repeated cognitive testing, especially for infrequently practiced modes of response and for memory tests [64]. Therefore, in case of repeated testing (eg, before and after treatment), it is important to take such practice effects into account. Our future research will focus on establishing parallel versions of tests that are most susceptible to practice effects to limit overall occurrence of practice effects. This means that after developing parallel versions, psychometric properties will be reevaluated. In this study, we took practice effects into account by adjusting the group SDCs for observed practice effects and by using ICC

measures that take systematic error into account for our test-retest correlational analyses.

Influence of Age, Gender, Education, and Computer Skills on Neuropsychological Test Scores

Considering sensitivity to demographic characteristics, multiple regression-based parameters indicated distinct associations of age, gender, and education with test performance. Age was, as expected, the strongest predictor of cognitive performance [66]. In addition to linear age effects, we found quadratic age effects for two outcome measures (Connect the Dots II and the total score), indicating acceleration of age-related decline with advancing age. Previously, these effects were reported for functions such as verbal memory [25], spatial visualization, processing speed, and reasoning [66,67]. However, in these previous studies and in this study, quadratic age effects are found to be small compared with linear age effects.

A high level of education was predictive for better test performance on only a few outcome measures (Place the Beads, Digit Sequences I and II, and the total score). This suggests that there is limited influence of education level on ACS performance. Note that our main reference sample did not include participants with a low education level, which hampers the assessment of the influence of level of education on cognitive performance. Therefore, MRA should be repeated after collecting reference data from a more heterogeneous reference sample, and information on level of education should preferably be collected in future studies using the ACS.

Gender appeared to be a generic predictor of cognitive performance on the ACS, with men outperforming women on 5 out of 12 measures and women outperforming men on one of the measures. Previous studies have shown more specific effects of gender on cognitive performance; it is commonly found that women outperform men on verbal learning and memory tests, whereas for many other cognitive functions, gender differences are less common [64,68]. To better understand our generic finding of men outperforming women, additional analyses were performed, indicating that computer skills, rather than gender, predicted performance on several gender-sensitive outcome measures. Sensitivity to computer skills may also be influenced by age, as young men in particular were found to outperform women of similar age on computer skills.

We have not yet included our measure of computer skills in the MRA and regression models, as we first want to learn more about its measurement construct and its relation with both online tests and traditional tests. Preliminary data on correlations between tested computer skills and the traditional equivalents of our online neuropsychological tests in a different sample of healthy adults (n=40; 72.5% [29/40] female; mean age 40 years, SD 15.4; 80% [32/40] high education level) indicated associations between our computer skills measure and several traditional neuropsychological tests, mainly in the domains of processing speed and motor coordination. Positive correlations were found for the Trail Making Test (TMT) A (ρ=.5; P=.001), 15 Words test (Dutch version of Rey Auditory Verbal Learning test), Learning (ρ=.74; P<.001) and Delayed Recall (ρ=.58; P=.004), Reaction Speed (ρ=.47; P=.002), Grooved Pegboard (ρ=.5; P=.001), and the battery total score (ρ=.77; P<.001).

Moreover, self-reported average number of hours of computer use a week was not associated with performance on traditional neuropsychological tests (except for TMT B: ρ=.34; P<.03), whereas in analyses on the online ACS data, this measure was associated with tests on the domains of processing speed and motor coordination (tests that require relatively demanding use of computer input devices). These additional analyses indicate that number of hours computer use is potentially a more valid measure of computer skills than our tests of computer skills, as it did not correlate with offline measures or online measures for which influence of computer skills was not to be expected. Further research on these measures should result in determining an optimal measure of computer skills, which subsequently can be included in the regression-based norms.

Regression-Based Normative Data

Normative data are established by providing MRA parameters on sensitivity for age, gender, and/or education for all main neuropsychological measures. The MRA parameters constitute formulas that allow for calculating regression-based, demographically corrected norm scores. In addition, to illustrate the influence of the main demographic predictors on cognitive performance, we presented mean test scores stratified by age and gender. Normative data, as presented, are suitable for interpreting ACS results from future studies with Dutch adult populations. American normative data for the English version of the ACS will be available in the near future as well. Concerning future use of the ACS, note that the Wordlist Recognition measure tends to result in a ceiling performance and that the reference data on this test had insufficient variance to enable interpretation of performance. Therefore, currently, Wordlist Learning and Wordlist Delayed Recall are advised to be used as the main measures of verbal learning and memory. Furthermore, as in general response time latencies can differ substantially depending on input device type [69], norm data should ideally be device specific. However, since most people used a standard mouse, we were at present not able to differentiate between devices. In future studies, after collection of additional reference data with a variety of device types, MRA should be repeated.

Usability

In addition to psychometric properties and normative data, this study showed that the ACS has high usability for cognitive testing with healthy Dutch adults, as was previously found in application with Dutch cancer patients. Virtually all participants completed the test battery from home without additional help. Temporary Internet disconnection occasionally occurred, but this did not prevent participants from completing the ACS successfully. On the basis of the debriefing, we suggest explicitly mentioning in pretest instructions to have a functional sound system, not to use paper-and-pencil, and to strive not to be disrupted during the assessment. This was done in our own instructions but could have been emphasized even more.

Conclusion

In conclusion, this study with a healthy adult sample shows that the ACS can give reliable measures of various generic cognitive ability areas. For repeated measurements, computer

configuration should be consistent. Combined with our normative data that describe which demographic characteristics influence performance, these results allow for initial interpretation of ACS performances. However, to improve the interpretation of the test scores, we will continue to collect reference data, including in a lower educational group and repeat multiple regression analyses after acquiring a more heterogeneous reference sample. We also aim to improve our MRA parameters by performing input device specific calculations and by studying the influence of computer skills and experience and related factors more in depth. The Dutch version of the ACS is estimated to be available for research purposes by the end of 2018. By that time, scoring software based on our normative data will be available as well.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The Amsterdam Cognition Scan—online neuropsychological and computer skill tests.

[PDF File (Adobe PDF File), 47KB - jmir_v20i5e192_app1.pdf]

Multimedia Appendix 2

Screenshots from the Amsterdam Cognition Scan—including in chronological order (1) the general instruction video, (2) computer skill tests, and (3) seven neuropsychological tests with in-between the standardized break animation video.

[PPTX File, 2MB - jmir_v20i5e192_app2.pptx]

Multimedia Appendix 3

Questionnaire results on test-retest reliability and practice effects(significant t values in italics).

[PDF File (Adobe PDF File), 27KB - jmir_v20i5e192_app3.pdf]

Multimedia Appendix 4

Multiple linear regression models including the factors age, age-squared, gender, education, and computer skills (tested) for Amsterdam Cognition Scan outcome measures.

[PDF File (Adobe PDF File), 38KB - jmir_v20i5e192_app4.pdf]

References

1. Bauer RM, Iverson GL, Cernich AN, Binder LM, Ruff RM, Naugle RI. Computerized neuropsychological assessment devices: joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. Clin Neuropsychol 2012;26(2):177-196 [FREE Full text] [doi: 10.1080/13854046.2012.663001] [Medline: 22394228]

2. Iverson GL, Brooks BL, Ashton VL, Johnson LG, Gualtieri CT. Does familiarity with computers affect computerized neuropsychological test performance? J Clin Exp Neuropsychol 2009 Jul;31(5):594-604. [doi: 10.1080/13803390802372125] [Medline: 18972312]

3. Hansen TI, Lehn H, Evensmoen H, Håberg A. Initial assessment of reliability of a self-administered web-based neuropsychological test battery. Comput Human Behav 2016 Oct;63(6):91-97 [FREE Full text] [doi: 10.1016/j.chb.2016.05.025] [Medline: 26009791]

4. Barak A, English N. Prospects and limitations of psychological testing on the internet. J Technol Hum Serv 2002;19(2-3):65-89. [doi: 10.1300/J017v19n02_06]

5. Bilder RM. Neuropsychology 3.0: evidence-based science and practice. J Int Neuropsychol Soc 2011;17(1):7-13 [FREE Full text] [doi: 10.1017/S1355617710001396] [Medline: 21092355]

6. Parsey CM, Schmitter-Edgecombe M. Applications of technology in neuropsychological assessment. Clin Neuropsychol 2013 Sep;27(8):1328-1361 [FREE Full text] [doi: 10.1080/13854046.2013.834971] [Medline: 24041037]

7. Birnbaum MH. Human research and data collection via the internet. Annu Rev Psychol 2004;55(8):803-832 [FREE Full text] [doi: 10.1146/annurev.psych.55.090902.141601] [Medline: 14744235]

8. Naglieri JA, Drasgow F, Schmit M, Handler L, Prifitera A, Margolis A, et al. Psychological testing on the Internet: new problems, old issues. Am Psychol 2004 Apr;59(3):150-162. [doi: 10.1037/0003-066X.59.3.150] [Medline: 15222858]

9. Reips UD. Internet-based psychological experimenting: five dos and five don'ts. Soc Sci Comput Rev 2002 Aug;20(3):241-249. [doi: 10.1177/089443930202000302]

XSL•FO

RenderX

10.   Buchanan T, Smith JL. Using the internet for psychological research: personality testing on the World Wide Web. Br J Psychol 1999 Feb;90(1):125-144. [Medline: 10085550]

11.   Caine C, Mehta MP, Laack NN, Gondi V. Cognitive function testing in adult brain tumor trials: lessons from a comprehensive review. Expert Rev Anticancer Ther 2012 May;12(5):655-667. [doi: 10.1586/era.12.34] [Medline: 22594900]

12.   Assmann KE, Bailet M, Lecoffre AC, Galan P, Hercberg S, Amieva H, et al. Comparison between a self-administered and supervised version of a web-based cognitive test battery: results from the NutriNet-Sante cohort study. J Med Internet Res 2016 Apr 05;18(4):e68 [FREE Full text] [doi: 10.2196/jmir.4862] [Medline: 27049114]

13.   Van de Weijer-Bergsma E, Kroesbergen EH, Prast EJ, Van Luit JE. Validity and reliability of an online visual-spatial working memory task for self-reliant administration in school-aged children. Behav Res Methods 2015 Sep;47(3):708-719. [doi: 10.3758/s13428-014-0469-8] [Medline: 24771322]

14.   Haworth CM, Harlaar N, Kovas Y, Davis OS, Oliver BR, Hayiou-Thomas ME, et al. Internet cognitive testing of large samples needed in genetic research. Twin Res Hum Genet 2007 Aug;10(4):554-563. [doi: 10.1375/twin.10.4.554] [Medline: 17708696]

15.   Wild K, Howieson D, Webbe F, Seelye A, Kaye J. Status of computerized cognitive testing in aging: a systematic review. Alzheimers Dement 2008 Nov;4(6):428-437 [FREE Full text] [doi: 10.1016/j.jalz.2008.07.003] [Medline: 19012868]

16.   Gates NJ, Kochan NA. Computerized and on-line neuropsychological testing for late-life cognition and neurocognitive disorders: are we there yet? Curr Opin Psychiatry 2015 Mar;28(2):165-172. [doi: 10.1097/YCO.0000000000000141] [Medline: 25602241]

17.   Hansen TI, Haferstrom EC, Brunner JF, Lehn H, Håberg AK. Initial validation of a web-based self-administered neuropsychological test battery for older adults and seniors. J Clin Exp Neuropsychol 2015;37(6):581-594 [FREE Full text] [doi: 10.1080/13803395.2015.1038220] [Medline: 26009791]

18.   Troyer AK, Rowe G, Murphy KJ, Levine B, Leach L, Hasher L. Development and evaluation of a self-administered on-line test of memory and attention for middle-aged and older adults. Front Aging Neurosci 2014;6:335 [FREE Full text] [doi: 10.3389/fnagi.2014.00335] [Medline: 25540620]

19.   Aalbers T, Baars MA, Olde RM, Kessels RP. Puzzling with online games (BAM-COG): reliability, validity, and feasibility of an online self-monitor for cognitive performance in aging adults. J Med Internet Res 2013;15(12):e270 [FREE Full text] [doi: 10.2196/jmir.2860] [Medline: 24300212]

20.   Schagen SB, Klein M, Reijneveld JC, Brain E, Deprez S, Joly F, et al. Monitoring and optimising cognitive function in cancer patients: present knowledge and future directions. EJC Suppl 2014 Jun;12(1):29-40 [FREE Full text] [doi: 10.1016/j.ejcsup.2014.03.003] [Medline: 26217164]

21.   Janelsins MC, Kesler SR, Ahles TA, Morrow GR. Prevalence, mechanisms, and management of cancer-related cognitive impairment. Int Rev Psychiatry 2014 Feb;26(1):102-113 [FREE Full text] [doi: 10.3109/09540261.2013.864260] [Medline: 24716504]

22.   Feenstra HE, Murre JM, Vermeulen IE, Kieffer JM, Schagen SB. Reliability and validity of a self-administered tool for online neuropsychological testing: the Amsterdam Cognition Scan. J Clin Exp Neuropsychol 2017 Jul 03;40(3):253-273. [doi: 10.1080/13803395.2017.1339017] [Medline: 28671504]

23.   Strauss E, Shermann EM, Spreen O. A compendium of neuropsychological tests: Administration, norms, and commentary. Oxford: Oxford University Press; 2006.

24.   Van der Elst W, Molenberghs G, Van Tetering M, Jolles J. Establishing normative data for multi-trial memory tests: the multivariate regression-based approach. Clin Neuropsychol 2017;31(6-7):1173-1187. [doi: 10.1080/13854046.2017.1294202] [Medline: 28276864]

25.   Van der Elst W, Van Boxtel MP, Van Breukelen GJ, Jolles J. Rey's verbal learning test: normative data for 1855 healthy participants aged 24-81 years and the influence of age, sex, education, and mode of presentation. J Int Neuropsychol Soc 2005 May;11(3):290-302. [doi: 10.1017/S1355617705050344] [Medline: 15892905]

26.   Parmenter BA, Testa SM, Schretlen DJ, Weinstock-Guttman B, Benedict RH. The utility of regression-based norms in interpreting the minimal assessment of cognitive function in multiple sclerosis (MACFIMS). J Int Neuropsychol Soc 2010 Jan;16(1):6-16. [doi: 10.1017/S1355617709990750] [Medline: 19796441]

27.   Germine L, Nakayama K, Duchaine BC, Chabris CF, Chatterjee G, Wilmer JB. Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. Psychon Bull Rev 2012 Oct;19(5):847-857. [doi: 10.3758/s13423-012-0296-9] [Medline: 22829343]

28.   Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [Medline: 6880820]

29.   Smets EM, Garssen B, Bonke B, De Haes JC. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. J Psychosom Res 1995 Apr;39(3):315-325. [Medline: 7636775]

30.   Reitan RM. Validity of the Trail Making Test as an indicator of organic brain damage. Percept Mot Skills 1958;8:271-276.

31.   Van den Burg W, Saan RJ, Deelman BG. 15-woordentest - Provisional Manual. Groningen: University Hospital, Department of Neuropsychology; 1985.

32.   Alpherts W, Aldenkamp AP. FePsy: The Iron Psyche. Heemstede: Instituut voor epilepsiebestrijding; 1994:A.

33. Culbertson W, Zillmer E. Tower of london-drexel (TOL-DX) technical manual. North Tonawanda, NY: Multi-Health Systems; 2001.

34. Kessels RP, Van Zandvoort MJ, Postma A, Kappelle LJ, De Haan EH. The Corsi Block-Tapping Task: standardization and normative data. Appl Neuropsychol 2010;7(4):252-258. [doi: 10.1207/S15324826AN0704_8] [Medline: 11296689]

35. Kløve H. Grooved Pegboard Test. Lafayette: Lafayette Instruments; 1963.

36. Tulsky D, Zhu J, Ledbetter MF. WAIS-III/WMS-III Technical Manual. San Antonio, TX: Psychological Corporation; 1997.

37. Leys C, Ley C, Klein O, Bernard P, Licata L. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. J Exp Soc Psychol 2013 Jul;49(4):764-766. [doi: 10.1016/j.jesp.2013.03.013]

38. Andreotti C, Root JC, Schagen SB, McDonald BC, Saykin AJ, Atkinson TM, et al. Reliable change in neuropsychological assessment of breast cancer survivors. Psychooncology 2016 Jan;25(1):43-50 [FREE Full text] [doi: 10.1002/pon.3799] [Medline: 25808921]

39. De Vet HC, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine: A practical guide. Cambridge: Cambridge University Pre; 2011.

40. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 2005 Feb;19(1):231-240. [doi: 10.1519/15184.1] [Medline: 15705040]

41. Anastasi A. Psychological testing. New York: Macmillan; 1998.

42. Cole WR, Arrieux JP, Schwab K, Ivins BJ, Qashu FM, Lewis SC. Test-retest reliability of four computerized neurocognitive assessment tools in an active duty military population. Arch Clin Neuropsychol 2013 Nov;28(7):732-742. [doi: 10.1093/arclin/act040] [Medline: 23819991]

43. Ruano L, Sousa A, Severo M, Alves I, Colunas M, Barreto R, et al. Development of a self-administered web-based test for longitudinal cognitive assessment. Sci Rep 2016 Jan 08;6:19114 [FREE Full text] [doi: 10.1038/srep19114] [Medline: 26743329]

44. Van Kampen DA, Willems WJ, Van Beers LW, Castelein RM, Scholtes VA, Terwee CB. Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). J Orthop Surg Res 2013 Nov 14;8:40 [FREE Full text] [doi: 10.1186/1749-799X-8-40] [Medline: 24225254]

45. Verhage F. Intelligentie en leeftijd: Onderzoek bij Nederlanders van twaalf tot zevenenzeventig jaar. Assen: Van Gorcum; 1964.

46. Vaughan RM, Coen RF, Kenny R, Lawlor BA. Preservation of the semantic verbal fluency advantage in a large population-based sample: normative data from the TILDA study. J Int Neuropsychol Soc 2016 May;22(5):570-576. [doi: 10.1017/S1355617716000291] [Medline: 27055803]

47. Van der Elst W, Hoogenhout EM, Dixon RA, De Groot RH, Jolles J. The Dutch Memory Compensation Questionnaire: psychometric properties and regression-based norms. Assessment 2011 Dec;18(4):517-529 [FREE Full text] [doi: 10.1177/1073191110370116] [Medline: 20519736]

48. Van der Elst W, Van Boxtel MP, Van Breukelen GJ, Jolles J. The Stroop color-word test: influence of age, sex, and education; and normative data for a large sample across the adult age range. Assessment 2006 Mar;13(1):62-79. [doi: 10.1177/1073191105283427] [Medline: 16443719]

49. Belsley D, Kuh E, Welsch RE. Regression diagnostics: Identifying influential data and sources of collinearity. Hoboken, New Jersey: John Wiley & Sons; 2005.

50. Saan RJ, Deelman BG. The 15-Words Test A and B. Groningen: University of Groningen, Department of Neuropsychology; 1986.

51. Spreen O, Strauss E. A compendium of neuropsychological tests. New York: Oxford University Press; 1998.

52. Silverstein SM, Jaeger J, Donovan-Lepore A, Wilkniss SM, Savitz A, Malinovsky I, et al. A comparative study of the MATRICS and IntegNeuro cognitive assessment batteries. J Clin Exp Neuropsychol 2010 Nov;32(9):937-952. [doi: 10.1080/13803391003596496] [Medline: 20455131]

53. Bruggemans EF, Van de Vijver FJ, Huysmans HA. Assessment of cognitive deterioration in individual patients following cardiac surgery: correcting for measurement error and practice effects. J Clin Exp Neuropsychol 1997 Aug;19(4):543-559. [doi: 10.1080/01688639708403743] [Medline: 9342689]

54. Bornstein RA, Baker GB, Douglass AB. Short-term retest reliability of the Halstead-Reitan Battery in a normal sample. J Nerv Ment Dis 1987 Apr;175(4):229-232. [Medline: 3559534]

55. Andreotti C, Hawkins KA. RBANS Norms based on the relationship of age, gender, education, and WRAT-3 reading to performance within an older african american sample. Clin Neuropsychol 2015;29(4):442-465. [doi: 10.1080/13854046.2015.1039589] [Medline: 26035646]

56. Lemay S, Bédard M, Rouleau I, Tremblay PG. Practice effect and test-retest reliability of attentional and executive tests in middle-aged to elderly subjects. Clin Neuropsychol 2004 May;18(2):284-302. [doi: 10.1080/13854040490501718] [Medline: 15587675]

XSL•FO
RenderX

57. Lowe C, Rabbitt P. Test/re-test reliability of the CANTAB and ISPOCD neuropsychological batteries: theoretical and practical issues. Cambridge Neuropsychological Test Automated Battery. International Study of Post-Operative Cognitive Dysfunction. Neuropsychologia 1998 Sep;36(9):915-923. [Medline: 9740364]

58. Schnirman GM, Welsh MC, Retzlaff PD. Development of the Tower of London-revised. Assessment 1998 Dec;5(4):355-360. [doi: 10.1177/107319119800500404] [Medline: 9835658]

59. Nys GM, Van Zandvoort MJ, De Kort PL, Jansen BP, Van der Worp HB, Kappelle LJ, et al. Domain-specific cognitive recovery after first-ever stroke: a follow-up study of 111 cases. J Int Neuropsychol Soc 2005 Nov;11(7):795-806. [Medline: 16519259]

60. California Computerized Assessment Package. 1995. CaLCAP psychometric properties URL: http://www.calcaprt.com/metrics.htm [accessed 2018-04-25] [WebCite Cache ID 6ywUQPOja]

61. Dikmen SS, Heaton RK, Grant I, Temkin NR. Test-retest reliability and practice effects of expanded Halstead-Reitan Neuropsychological Test Battery. J Int Neuropsychol Soc 1999 May;5(4):346-356. [Medline: 10349297]

62. Ruff RM, Parker SB. Gender- and age-specific changes in motor speed and eye-hand coordination in adults: normative values for the Finger Tapping and Grooved Pegboard Tests. Percept Mot Skills 1993 Jun;76(3 Pt 2):1219-1230. [doi: 10.2466/pms.1993.76.3c.1219] [Medline: 8337069]

63. Ryan JJ, Lopez SJ, Paolo AM. Temporal stability of digit span forward, backward, and forward minus backward in persons aged 75-87 years. Cogn Behav Neurol 1996;9(3):-.

64. Lezak M, Howieson DB. Neuropsychological assessment. Oxford: Oxford University Press; 2004.

65. Mathews M, Abner E, Caban-Holt A, Kryscio R, Schmitt F. CERAD practice effects and attrition bias in a dementia prevention trial. Int Psychogeriatr 2013 Jul;25(7):1115-1123 [FREE Full text] [doi: 10.1017/S1041610213000367] [Medline: 23570673]

66. Salthouse TA. When does age-related cognitive decline begin? Neurobiol Aging 2009 Apr;30(4):507-514 [FREE Full text] [doi: 10.1016/j.neurobiolaging.2008.09.023] [Medline: 19231028]

67. Verhaeghen P, Salthouse TA. Meta-analyses of age-cognition relations in adulthood: estimates of linear and nonlinear age effects and structural models. Psychol Bull 1997 Nov;122(3):231-249. [Medline: 9354147]

68. Robbins TW, James M, Owen AM, Sahakian BJ, McInnes L, Rabbitt P. Cambridge Neuropsychological Test Automated Battery (CANTAB): a factor analytic study of a large sample of normal elderly volunteers. Dementia 1994;5(5):266-281. [Medline: 7951684]

69. Cernich AN, Brennana DM, Barker LM, Bleiberg J. Sources of error in computerized neuropsychological assessment. Arch Clin Neuropsychol 2007 Feb;22(Suppl 1):S39-S48 [FREE Full text] [doi: 10.1016/j.acn.2006.10.004] [Medline: 17097851]

Abbreviations

ACS: Amsterdam Cognition Scan
HADS: Hospital Anxiety and Depression Scale
ICC: intraclass correlation coefficient
MFI-20: Multidimensional Fatigue Inventory-20
MRA: multiple regression analyses
SDC: smallest detectible change
VIF: variance inflation factor

XSL•FO
RenderX

Original Paper

An eHealth Capabilities Framework for Graduates and Health Professionals: Mixed-Methods Study

Melissa Brunner[1,2], BAppSc, MHlthSc; Deborah McGregor[1], BAppSc, M Health Care; Melanie Keep[3], PhD(Psych); Anna Janssen[1], BA, MPub; Heiko Spallek[4], PhD; Deleana Quinn[1], BPsych; Aaron Jones[5], MClinEd; Emma Tseris[6], PhD; Wilson Yeung[7], MHA, MPA; Leanne Togher[3], PhD; Annette Solman[8], M Nursing Research; Tim Shaw[1], PhD

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

[2]Speech Pathology, Graduate School of Health, The University of Technology Sydney, Ultimo, Australia

[3]Faculty of Health Sciences, The University of Sydney, Camperdown, Australia

[4]School of Dentistry, The University of Sydney, Westmead, Australia

[5]Sydney Local Health District, Camperdown, Australia

[6]Faculty of Arts and Social Sciences, The University of Sydney, Camperdown, Australia

[7]eHealth NSW, Chatswood, Australia

[8]NSW Health Education Training Institute, Gladesville, Australia

Corresponding Author:
Melissa Brunner, BAppSc, MHlthSc
Research in Implementation Science and eHealth
Faculty of Health Sciences
The University of Sydney
Charles Perkins Centre
Camperdown, 2006
Australia
Phone: 61 401190686
Email: melissa.brunner@sydney.edu.au

Abstract

Background: The demand for an eHealth-ready and adaptable workforce is placing increasing pressure on universities to deliver eHealth education. At present, eHealth education is largely focused on components of eHealth rather than considering a curriculum-wide approach.

Objective: This study aimed to develop a framework that could be used to guide health curriculum design based on current evidence, and stakeholder perceptions of eHealth capabilities expected of tertiary health graduates.

Methods: A 3-phase, mixed-methods approach incorporated the results of a literature review, focus groups, and a Delphi process to develop a framework of eHealth capability statements.

Results: Participants (N=39) with expertise or experience in eHealth education, practice, or policy provided feedback on the proposed framework, and following the fourth iteration of this process, consensus was achieved. The final framework consisted of 4 higher-level capability statements that describe the learning outcomes expected of university graduates across the domains of (1) digital health technologies, systems, and policies; (2) clinical practice; (3) data analysis and knowledge creation; and (4) technology implementation and codesign. Across the capability statements are 40 performance cues that provide examples of how these capabilities might be demonstrated.

Conclusions: The results of this study inform a cross-faculty eHealth curriculum that aligns with workforce expectations. There is a need for educational curriculum to reinforce existing eHealth capabilities, adapt existing capabilities to make them transferable to novel eHealth contexts, and introduce new learning opportunities for interactions with technologies within education and practice encounters. As such, the capability framework developed may assist in the application of eHealth by emerging and existing health care professionals. Future research needs to explore the potential for integration of findings into workforce development programs.

Introduction

Background

Developing an eHealth-ready workforce [1,2] is becoming a key priority for addressing the complex challenges in health care globally [3,4]. Modern health care now incorporates a multitude of eHealth technologies that can be used for (1) monitoring, tracking, and informing health; (2) interacting for health such as using digital technologies to enable health communication among practitioners and between health professionals and clients or patients; and (3) data that enable health via collecting, managing, and using health data to improve outcomes [5]. eHealth has been recognized as pivotal in recent health reforms, with the potential to provide more efficient, cost-effective care with better outcomes [4,6,7]. As such, health services increasingly expect health professionals, new graduates, and experienced workforce alike, to be eHealth ready within practice across diverse digital health environments [4,8]. In turn, this is directing efforts to integrate eHealth education into clinical health degrees, specifically allied health, nursing, pharmacy, dentistry, and medical programs [1,2,9].

Significant work has been conducted by Gray et al (2014) to identify the educational needs of health professionals necessary for an eHealth-capable workforce [9]. A key factor limiting the eHealth readiness of current and future health professionals is the lack of coordinated, formal education in the use of digital technologies in health [9]. The health practice context has become vastly more complex, mirroring changes in patients, the health system, and medical science, making it increasingly challenging for clinicians to safely and efficiently navigate health care [10]. As digital health becomes more widespread, so does the requirement for health professionals to be well versed in navigating and using such technologies [3]. eHealth education, therefore, needs to be coordinated to explain and explore eHealth in both current and future health care contexts, and to incorporate clear specifications as to the eHealth capabilities expected of the current and future health care workforce [9]. Research into, and the development of, coordinated approaches to eHealth curriculum design and implementation are necessary to effectively integrate eHealth capabilities among health graduates and to support the health care workforce in using eHealth technologies [3]. Consequently, specific eHealth education is required, even if students report to be competent and confident in using technology [11]. Such skills must not to be equated with technology proficiency, and most certainly not with information fluency [12]. Information literacy is defined in essence as "the ability to access, evaluate, and use information from a variety of sources" [13] and has become a critical skill for the present generation of students, and indeed for the 21st-century citizen [14,15]. For this to be achieved, a systematic approach to curriculum design and collaborative efforts from stakeholders are needed [9,16].

Multiple issues need to be addressed when designing strategies for embedding eHealth into undergraduate curricula [4,9,16].

Drawing from research into teaching evidence-based medicine [17], eHealth would be more effectively taught (ie, demonstrate improved knowledge, skills, and attitudes) when it is integrated into clinical subjects rather than as an adjunct [16]. Education needs to improve awareness and understanding of the purpose of eHealth in practice contexts, as well as training in the use of technologies [4,9]. Competency frameworks in health informatics [18,19] provide valuable reference points in technical and informatics literacy for workforce development. eHealth-enabled health care, however, extends beyond informatics and requires professionals to integrate digital technologies into the health care and management processes [5]. As such, although health informatics competencies provide a strong foundation for the technical requirements of eHealth, the current and future health workforce must also receive education in the use of eHealth in practice to perform a range of professional functions, such as clinical decision making, patient empowerment, promotion of health and wellness, critical reflection and ethical decision making, and enabling new models of care [3,20]. Although such workforce eHealth competency frameworks are emerging for specific professions [21], limited understanding and agreement remain regarding the core eHealth competencies expected of tertiary graduates.

Although the tertiary education of health professionals is traditionally framed around a set of specific competency standards decreed by professional associations that allow measurable behaviors to be observed and assessed [22], competency frameworks are most beneficial when the skills and practice are consistent across contexts, eg, in the assessment of a specific injury or illness [23]. eHealth practice, on the other hand, might look different in different contexts. For example, effective integration of mobile technologies in one context might involve using mobile devices to enhance the measurement and assessment of injury. In another context, clinicians could be using apps to remotely prescribe exercises and monitor patients' physical activity. The fluidity in how eHealth skills are demonstrated across different contexts and the speed with which eHealth technologies evolve both clash with the rigidity and speed at which competencies and educational frameworks can be developed. As such, this suggests that competency standards may not be the most appropriate framing of eHealth skills and education [5,22,23].

In contrast to competency, capability has been defined as "a holistic concept that describes how an individual or organization applies their ability in a confident manner to problems in new and unfamiliar circumstances as well as in familiar situations" [24]. As such, a capable workforce includes lifelong learners who are able to identify the need for change, adapt to familiar and novel situations and environments, and work collaboratively with other stakeholders to provide and potentially transform care [22,23]. Although capability has been described as being similar to competence, it in fact encompasses competence and extends beyond the technical skills implied by competence to emphasize the components of adaptability to change, lifelong learning, and self-efficacy [23-25]. As such, capability-informed

frameworks address wider aspects of professionalism, focusing on supporting continuous development rather than assessment of a skill at a specific point in time [25].

Objectives

A capabilities approach is, therefore, able to inform and modify competency-based frameworks that better reflect the complexity of real-life environments as is found in the field of using eHealth for health care and wellness [22,25]. With this in mind, this study aimed to develop an eHealth capabilities framework based on current evidence and stakeholder perceptions of eHealth capabilities that are expected of workforce-ready tertiary health graduates. This study formed one component of a larger project, eHealthMap, which aimed to align the University's approach to eHealth education with the best evidence and national expectations of workforce-ready graduates. A key focus of the study was to ensure alignment between the required capabilities of new graduates with workforce capability requirements. This led to a collaborative design approach between the University and the New South Wales Ministry of Health.

Methods

Triphasic Approach

A 3-phase approach was used. This consisted of a literature review in the first phase, followed by qualitative studies to identify relevant themes, gain expert opinions, and raise consensus with key stakeholders in the second stage. In the third phase, data from the 2 previous stages were reviewed using a Delphi process to develop a framework of capability statements for guiding health curricula and measurement of workforce readiness in the use of eHealth technologies for health care. The project occurred between October 2016 and November 2017.

Literature Review

A systematic search for literature was undertaken during October 2016 to December 2016 to identify papers related to workforce readiness in the use of eHealth technologies. The databases and search engines used included CINAHL, Medline, ERIC, PsychINFO, Google Scholar, and Google. Databases and search engines were systematically searched for literature on eHealth competencies using combinations and variations of the following key search terms: *eHealth*, *digital health*, *health professional*, *workforce readiness*, *graduate*, *student*, *higher education*, *capability statement*, and *competency*. The fields of health, medicine, nursing, public health, allied health, pharmacy, psychology, physiotherapy, occupational therapy, speech language pathology, dentistry, paramedicine, social work, dietetics, nutrition, radiography, audiology, exercise and sports science, optometry, orthoptics, ophthalmology, and podiatry were included. Researchers also reached out to health and academic communities and reviewed Web pages from a number of professional societies and health organizations to identify the existence of other eHealth readiness competencies and capabilities.

The search was limited to papers published after 2000 to maximize relevance to current clinical contexts and papers written in English. Research studies and reports were included if they (1) referred to readiness competencies or capabilities for students, graduates, or workforce; (2) referred to the use of digital technologies in health care; and (3) provided empirical support for the capability statements or framework. Included peer-reviewed and gray literature were then reviewed to identify existing core competencies or capability statements relating to eHealth readiness. Data were extracted by 1 author (author 2), analysis and synthesis [26] conducted independently by 2 authors (authors 1 and 2), with consensus provided by a third author as required (author 3).

Focus Groups

Focus groups with key informants from the health workforce and higher education were used to ascertain the perceived eHealth capabilities requirements of new graduates [27]. Purposeful sampling recruited 23 participants with significant expertise or vested interests in eHealth education, practice, or policy. Focus groups were conducted during a workshop held in February 2017 at the University of Sydney. Ethical approval was obtained from the University of Sydney Human Research Ethics Committee (Protocol No. 2016/811) before participant recruitment.

The workshop included 2 focus group sessions: (1) a large focus group (n=23, 1 hour and 14 min) and (2) 4 small breakout focus groups (n=23, with 5-6 participants per group, 45 min). A semistructured script encouraged topic exploration and included the following questions: (1) What eHealth competencies do you expect health graduates to be able to demonstrate? and (2) Could you provide some examples of how these competencies are taught or applied in your organization? Focus group discussions were facilitated by the researcher team experienced in qualitative research methods. To stimulate discussions, participants were presented with evidence of core eHealth readiness competencies identified via the literature review. Of the focus group participants, 65% (15/23) were female, and 48% (11/23) were University of Sydney faculty representatives. Faculty representatives encompassed the fields of physiotherapy, speech pathology, psychology, nursing, dentistry, pharmacy, medicine, information technology and engineering, and mathematics and statistics. The balance of the participants (12/23, 52%) included broad representation from health services (9/23, 39%) and state and national government health agencies (3/23, 13%), including senior executives, clinicians, and senior health administrators. Participants also included recent health professional graduates (2/23, 9%) and 1 enrolled student from the University of Sydney. Interviews were transcribed verbatim and thematically analyzed [26] independently by 2 authors (authors 1 and 2). Codes were applied to the text of the transcripts, with themes systematically refined until saturation was achieved.

Development of an eHealth Capability Framework

The third phase involved identifying core capability statements by using the Delphi method [28] to refine and establish core eHealth capability statements. This was conducted over 4 iterations [29] during 2017, which allowed participants recruited from the focus groups (n=23) and via invitation to a wider group of key stakeholders (n=16) to systematically consider the capability statements as they evolved. Each round consisted of a meeting and follow-up email correspondence, promoting discussion and consensus between participants. A predetermined

XSL•FO

RenderX

quorum of participants (n=12, ie, participation quorum of 50% based on the total number of participants recruited during the focus group phase, where N=23) was present for each round required to reach consensus, and each participant provided feedback throughout the process during meetings or via email. At the end of each round, feedback was incorporated into the capability framework and the revised statements presented for deliberation at the next meeting. Incorporation of feedback resulted in improved clarity and limited redundancy where similar capabilities were collapsed or condensed. Iterations continued until such time as consensus among the participants was achieved [29]. Following consensus agreement, endorsement from key academic and industry organizations was sought.

Results

Literature Review

After duplicates were removed, the search revealed 92 relevant papers. Papers were excluded if they were opinion pieces, provided no empirical support, or reported no specific details regarding eHealth capabilities. This left a total of 30 papers included in this review. Emergent themes from the literature were identified (by authors 1 and 2) and refined through research team discussion at regular meetings.

The health, activity, and participation issues identified within the literature were initially arranged into the following 6 categories: (1) health information management (n=24); (2) communication (n=19); (3) professionalism (n=24); (4) information systems and technologies (n=25); (5) patient focus (n=13); and (6) health analytics (n=14). Categorization of the

6 core domains as identified in the literature is presented in Table 1.

Focus Groups

Three overarching themes, each with multiple subthemes, emerged from the thematic analyses: (1) reinforce fundamental clinical capabilities, (2) acknowledge and adapt existing capabilities, and (3) introduce and provide opportunities for new learning. Themes and subthemes are presented in Table 2 and have been reported previously [20].

Development of the eHealth Capability Framework

At the completion of the fourth round, consensus was achieved. The final capability framework consisted of 4 overarching domains and 40 performance cues (Multimedia Appendix 1). The 4 upper-level eHealth capability statements describe what an eHealth-ready health graduate should be able to demonstrate (Table 3). The nature of these statements is such that more specific examples of what they might look like in different contexts are needed. These specific examples are provided in the form of performance cues in the framework. As these statements reflect expected capabilities among entry-level clinical positions, levels of mastery are not specified. This would be more appropriate at the curriculum level where educators consider how these capabilities might be developed and assessed in their specific disciplinary and educational context. Feedback from the focus groups indicated that, ideally, the eHealth capability framework would be integrated with relevant professional competency-based occupational standards. Throughout the development process, key stakeholders consistently recommended that the framework needed to incorporate the essential components of both safety and effective communication across all the 4 high-level capability statements.

Table 1. Categorization of core domains of eHealth capability as identified in the literature review.

Domain	Summary of capabilities included in the domain	References
Health information management	Recording and storing health information in electronic systems, data quality, and information governance	[2,9,30-46]
Communication	Using digital technologies to support interprofessional relationships, consumer-provider relationships, and multiprofessional care coordination	[2,9,30-46]
Professionalism	Critical appraisal, evidence-based practice, eHealth literacy, continued professional development, ethical use of information, and management and leadership	[2,9,30-42,44-50]
Information systems and technologies	Using information systems and technologies to support routine clinical care, business processes, and patient-centered service provision	[2,9,30-35,37-39,41-48,50]
Patient focus	Patient empowerment, use of technology for self-management and wellness, patient eHealth literacy, and education	[2,9,31-33,35,37,39-43,46]
Health analytics	Use of data analytics in practice for informed decision making, quality improvement, service planning, and delivery	[2,9,31-33,37,39-44,46-47]

Table 2. Fundamental themes of eHealth capability identified from focus groups.

Fundamental themes	Subthemes
Reinforce fundamental capabilities	Quality and safety, Communication, problem solving, critical analysis, patient-centeredness, professionalism, lifelong learning skills
Acknowledge and adapt existing capabilities	Understand purpose of systems, advanced digital literacy, adaptive behavior, active participation in codesign
Introduce and provide opportunities for new learning	Working with health data, integration of health information sources, eHealth-enabled new models of care, data analytics, data governance, data privacy, data security, shifting role of the health care professional

Table 3. Overarching domains of eHealth capabilities.

Number and domain		Capability statement
1	Digital technologies, systems, and policies	Understand the purpose and function of digital health technologies and systems implemented at a local, state, or national level, including consideration of legal, policy, and ethical implications
2	Clinical practice and applications	Integrate digital health into clinical practice to deliver safe and quality care, including provision of best practice models of care
3	Data analysis and knowledge creation	Use data and data analysis to inform, deliver, and improve health and health care practice at an individual, team, or systems level
4	System and technology implementation	Participate in digital health implementation, evaluation, and codesign processes to drive improvement and stimulate change

Discussion

Principal Findings

This study aimed to develop a framework of eHealth capabilities that could be used to inform development of higher education curricula for health students and professional development opportunities for the current health workforce. The framework resulting from our literature review, focus groups, and Delphi process describes 4 overarching domains of eHealth capabilities: (1) digital technologies, systems, and policies; (2) clinical practice and applications; (3) data analysis and knowledge creation; and (4) system and technology implementation. Each domain is supported by a capability statement that articulates what learners need to be able to know and do to demonstrate achievement of this capability. Examples of these knowledge and behaviors are provided in the form of performance cues. The key components of the eHealth capability framework are provided in Multimedia Appendix 1.

The systematic process for developing the framework was necessary for the successful integration of eHealth and continuous quality improvement at all levels in the health system [9]. Critical to long-term implementation and adoption of the framework is the capability approach, which contrasts with a competency approach by recognizing the complexity and ever-changing nature of eHealth in practice. The idea of capability not only addresses complexity and evolving contexts but also incorporates the aspects of critical analysis and ethical practice in health care [25]. Furthermore, this research identified that eHealth capabilities should be integrated with clinical competencies and be sufficiently flexible so that they can be adapted to specific contexts.

Given the inherent nature of eHealth as being innovative and transformative, it is critical that we enable our health workforce to be suitably prepared and adaptable. Tertiary education of future health graduates needs to move from a focus on technical skills to encompass broader eHealth capabilities, such as professional competencies and attributes of an adaptable, improvement-minded, and innovative workforce. The eHealth capabilities framework extends beyond this technical proficiency to include the integration of technology into current practice and demonstrating a strong ethos around lifelong learning and transforming care. By re-evaluating traditional resources used in education and incorporating knowledge around eHealth capabilities, educational curriculum will provide greater opportunities to build capability in working with systems undergoing digital transformation [3]. Consistent with previous research [4,9], we identified the need for educational curriculum to reinforce existing eHealth capabilities, adapt existing capabilities to make them transferable to novel eHealth contexts, and introduce new learning opportunities for interactions with technologies within education and practice encounters. In essence, the capability framework developed assists in bridging the gap between academia and the application of digital health by emerging and existing health care professionals.

This framework is not a curriculum; it is a starting point to guide curriculum development and redesign. Furthermore, the framework is intended to stimulate discussion with industry stakeholders regarding workforce capability with regard to eHealth in practice. As a guide, it will enable further development of both curriculum and competency evaluation [48]. The eHealth capabilities framework is not intended to set a rigid curriculum for eHealth education, but rather to provide a key resource and common standards for the review, development, and alignment of profession-specific curricula to ensure high-quality and consistent student learning experiences. Attaining capability will ideally involve embedding eHealth within problem-based, case-based, and practice-based learning

XSL·FO

RenderX

experiences and incorporate digital simulations and codesign projects. This work emphasizes core professional practice principles that underpin all activities involving eHealth, including quality and safety, consumer-centeredness, critical thinking, and evidence-based practice. As technologies evolve and practices involving eHealth grow, it is critical to maintain attention to these core principles. The inclusion of performance cues within the framework provides a starting point to describe how successful learning might be demonstrated. These are intended to guide the development of assessments that support the achievement of specific learning outcomes and activities of the discipline-specific curriculum.

It was clear that there is a need to create an eHealth capability framework that accurately reflected the roles and work contexts in the digital age. In addition to moving beyond a focus on the technical skills, this included offering theoretical and practical opportunities for health graduates to integrate eHealth into practice, eg, using technologies to provide new models of care and facilitate consumer empowerment, or to use routinely collected digital health data to inform practice. One of the key challenges for education in this space is the ability for curriculum to adequately address the divide between digitally capable individuals and their capacity to apply digital skills to clinical situations that often use archaic systems [49]. Although current students may now use digital technologies in their everyday lives, they might have limited knowledge of the opportunities and issues that technology can bring to the health care landscape [49]. As a result, students require tailored opportunities to ensure that they develop or translate skills and knowledge to effectively practice in evolving digital workplaces.

Limitations

In this study, we did not aim to achieve data saturation as this was beyond the scope of this initial work. The participants in this study brought with them a breadth of experience and representation of multiple health professions, but the lack of consumer representation was a study limitation. It is likely that future research incorporating additional participants working across a wider range of contexts, and is inclusive of diverse health care consumers, may provide additional insights into eHealth-enabled interdisciplinary practice. Further research exploring the implications for the existing health care workforce is also warranted, with a focus on identifying the potential relevance and impact of the capability statements on policy and practice, including recruitment, professional development, performance management, and systems improvement activities. Building on the results of this study, the University is undertaking a curriculum mapping process with a commitment

to the development of high-quality teaching and learning activities and resources. This will include an analysis of clinical placement experiences, work-based training, and workplace orientation programs to further identify and address gaps in the preparation of our health graduates and workforce for eHealth contexts. We further excluded from the discussion implications of automation and computerization on the actual workforce and their predicted displacement [50]. Thus, the impact artificial intelligence might have on the health care delivery system was not in the scope of this research.

Practical Implications

This framework has direct implications for curriculum redevelopment in health education and professional development opportunities for the current health workforce. The framework could be used to assess which eHealth capabilities are currently being taught in health profession degrees, and how. Exemplars of effective eHealth education could be collated to form resources and professional development for health educators across the sector. Importantly, in using the framework to map current curricular, educators and course coordinators can identify the capabilities not addressed in their programs at present. This provides opportunity for revising the curriculum to better prepare graduates.

For the current health workforce, the framework requires further development. At present, the capabilities reflect the knowledge and skills required of a graduate. Further research and development are required to articulate intermediate and advanced levels of capability across the 4 domains of the framework. Once established, this more comprehensive framework could guide professional development opportunities and self-reflection or self-assessment for practicing health professionals.

Conclusions

This paper describes the foundational level of eHealth capability expected of tertiary health students at graduation and as they enter the health workforce. Ideally, the eHealth capability framework will inform how tertiary health programs deliver and assess essential eHealth education. The results of this study will inform a cross-faculty eHealth curriculum that aligns with workforce expectations and will be of interest to professional associations, health services, and organizations. Future research needs to explore the potential for integration of findings into workforce development programs, particularly with consideration of intermediate and advanced levels of capability in collaboration with workforce and industry stakeholders.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Brunner 2018_eHealth Capability Framework_Appendix.
[PDF File (Adobe PDF File), 199 KB - jmir_v20i5e10229_app1.pdf]

References

1. Deaprtment of Health and Aging, Australian Government. 2011. The eHealth Readiness of Australia's Allied Health Sector URL: http://www.health.gov.au/internet/publications/publishing.nsf/Content/ehealth-readiness-allied-toc/$FILE/Allied%20Health%20ehealth%20readiness%20survey%20report.pdf [accessed 2018-02-07] [WebCite Cache ID 6x2lIFsSA]

2. Health Workforce Australia. 2013 Oct. Health Information Workforce Report URL: https://www.aims.org.au/documents/item/401 [accessed 2018-02-07] [WebCite Cache ID 6x2lD5I7z]

3. Gray K. Public Health Platforms: An Emerging Informatics Approach to Health Professional Learning and Development. J Public Health Res 2016 Apr 26;5(1):665 [FREE Full text] [doi: 10.4081/jphr.2016.665] [Medline: 27190977]

4. Li S, Bamidis PD, Konstantinidis ST, Traver V, Car J, Zary N. Setting priorities for EU healthcare workforce IT skills competence improvement. Health Informatics J 2017 Apr 01:1460458217704257. [doi: 10.1177/1460458217704257] [Medline: 28441906]

5. 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]

6. Jolly R. Parliament of Australia. 2011. The e health revolution - easier said than done URL: https://www.aph.gov.au/about_parliament/parliamentary_departments/parliamentary_library/pubs/rp/rp1112/12rp03 [accessed 2018-02-07] [WebCite Cache ID 6x2kbZ0Or]

7. New South Wales Health. 2016. eHealth Strategy for NSW Health 2016-2026 URL: http://www.health.nsw.gov.au/eHealth/Documents/eHealth-Strategy-for-NSW-Health-2016-2026.pdf [accessed 2018-02-07] [WebCite Cache ID 6x2kTAYiF]

8. Queensland Government Metro South Health. 2016. Digital Hospital URL: https://metrosouth.health.qld.gov.au/princess-alexandra-hospital/digital-hospital [accessed 2018-02-07] [WebCite Cache ID 6x2lNS3BE]

9. Gray K, Dattakumar A, Maeder A, Butler-Henderson K, Chenery H. clinicalinformaticseducation.pbworks. 2014. Advancing Ehealth education for the clinical health professions URL: http://clinicalinformaticseducation.pbworks.com/w/file/fetch/74500403/PP10_1806_Gray_report_2014.pdf [accessed 2018-02-07] [WebCite Cache ID 6x2lg2mTP]

10. Obermeyer Z, Lee TH. Lost in Thought - The Limits of the Human Mind and the Future of Medicine. N Engl J Med 2017 Sep 28;377(13):1209-1211 [FREE Full text] [doi: 10.1056/NEJMp1705348] [Medline: 28953443]

11. Lam M, Hines M, Lowe R, Nagarajan S, Keep M, Penman M, et al. Preparedness for eHealth: Health Sciences Students' Knowledge, Skills, and Confidence. JITE:Research 2016;15:305-334. [doi: 10.28945/3523]

12. Oblinger DG. Growing up with Google: What it means to education. Emerging technologies for learning 2008 Mar:11-29 [FREE Full text]

13. Doyle CS. US Department of Education. 1992. Outcome Measures for Information Literacy within the National Education Goals of 1990. Final Report to National Forum on Information Literacy URL: https://files.eric.ed.gov/fulltext/ED351033.pdf [accessed 2018-02-07] [WebCite Cache ID 6x2mR4xKw]

14. Hancock VE. US Department of Education. 1993. Information Literacy for Lifelong Learning URL: https://files.eric.ed.gov/fulltext/ED358870.pdf [accessed 2018-02-07] [WebCite Cache ID 6x2mV55Dw]

15. Duggan F. The Guardian. 2013. Why universities should acquire - and teach - digital literacy URL: https://www.theguardian.com/education/2013/apr/23/universities-should-teach-digital-literacy [accessed 2018-02-07] [WebCite Cache ID 6x2mYhvmq]

16. Clark J, Baker B, Baker D. Getting eHealth into basic nursing education: report of the RCN information in nursing project. Stud Health Technol Inform 2009;146:534-539. [Medline: 19592900]

17. Khan KS, Coomarasamy A. A hierarchy of effective teaching and learning to acquire competence in evidenced-based medicine. BMC Med Educ 2006 Dec 15;6:59 [FREE Full text] [doi: 10.1186/1472-6920-6-59] [Medline: 17173690]

18. Australian Nursing and Midwifery Federation. 2015. National informatics standards for nurses and midwives URL: http://anmf.org.au/documents/National_Informatics_Standards_For_Nurses_And_Midwives.pdf [accessed 2018-02-07] [WebCite Cache ID 6x2mtnonb]

19. Health Informatics Society of Australia. 2013 Dec. Certified Health Informatician Australasia (CHIA) Health Informatics Competencies Framework Edition 1 URL: http://www.healthinformaticscertification.com/wp-content/uploads/2016/02/CHIA-competencies-Framework_FINAL.pdf [accessed 2018-02-07] [WebCite Cache ID 6x2mymlVX]

20. McGregor D, Keep M, Brunner M, Janssen A, Quinn D, Avery J, et al. Preparing E-Health Ready Graduates: A Qualitative Focus Group Study. Stud Health Technol Inform 2017;239:91-96. [Medline: 28756442]

21. Academy of Medical Royal Colleges. 2011 Jun. eHealth Competency Framework: Defining the role of the expert clinician URL: http://www.aomrc.org.uk/wp-content/uploads/2016/05/EHealth_Competency_Framework_0611.pdf [accessed 2018-02-07] [WebCite Cache ID 6x2n81o58]

22. O'Connell J, Gardner G, Coyer F. Beyond competencies: using a capability framework in developing practice standards for advanced practice nursing. J Adv Nurs 2014 Dec;70(12):2728-2735. [doi: 10.1111/jan.12475] [Medline: 25109608]

23. Phelps R, Hase S, Ellis A. Competency, capability, complexity and computers: exploring a new model for conceptualising end-user computer education. Br J Educ Technol 2005 Jan;36(1):67-84. [doi: 10.1111/j.1467-8535.2005.00439.x]

24. Townsend P, Cairns L. Developing the Global Manager Using a Capability Framework. Management Learning 2016 Aug 19;34(3):313-327. [doi: 10.1177/13505076030343002]

25. Lester S. Professional standards, competence and capability. HESWBL 2014 Feb 11;4(1):31-43. [doi: 10.1108/HESWBL-04-2013-0005]

26. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

27. Creswell JW. Research design: Qualitative, quantitative, and mixed methods approaches. Thousand Oaks, California: Sage Publications; 2013.

28. Dalkey N, Helmer O. An Experimental Application of the DELPHI Method to the Use of Experts. Manag Sci 1963 Apr;9(3):458-467. [doi: 10.1287/mnsc.9.3.458]

29. Skulmoski GJ, Hartman FT, Krahn J. The Delphi method for graduate research. JITE 2007;6:1-21 [FREE Full text]

30. Archer N. Project Management Characteristics and Education for eHealth Students. EJHI 2011;6(4):e30 [FREE Full text]

31. Barakat A, Woolrych RD, Sixsmith A, Kearns WD, Kort HS. eHealth Technology Competencies for Health Professionals Working in Home Care to Support Older Adults to Age in Place: Outcomes of a Two-Day Collaborative Workshop. Med 2 0 2013 Sep 05;2(2):e10 [FREE Full text] [doi: 10.2196/med20.2711] [Medline: 25075233]

32. Blake A, Kaye R. A Digital Solution to the EU/US Skilled eHealth/Health IT Worker Shortage: The HITCOMP Tool and Repository. : EU-US Workforce Development Workgroup; 2015 Presented at: eHealth Week; May 11-13, 2015; Riga, Latvia URL: https://www.eiseverywhere.com/file_uploads/ ace6e27c257b052431e3e7d7b16a23a3_BLAKE_ADigitalSolutiontotheEU_USSkilledeHealthHealthITWorkerShortage. pdf

33. Blak R, Shaw T, Blake A, Hübner U, Kaye R, Schug S, et al. Healthcare Information and Management Systems Society Foundation. 2016. EU-US eHealth Works to Improve Global Workforce Development URL: http://www.himss.org/sites/ himssorg/files/eu-us-ehealth-workforce-development.pdf [accessed 2018-02-07] [WebCite Cache ID 6x2p0Hq2k]

34. Britt RK, Hatten KN. The Development and Validation of the eHealth Competency Scale: A Measurement of Self-Efficacy, Knowledge, Usage, and Motivation. Tech Commun Q 2016 Apr 07;25(2):137-150 [FREE Full text] [doi: 10.1080/10572252.2016.1149621]

35. Connoly R. Ethical leadership: Training, education, and eLearning to promote ethics as a core competency within the eHealth Program. Ottawa, Canada: Royal Roads University; 2008.

36. Crawford A, Sunderji N, López J, Soklaridis S. Defining competencies for the practice of telepsychiatry through an assessment of resident learning needs. BMC Med Educ 2016 Jan 26;16:28 [FREE Full text] [doi: 10.1186/s12909-016-0529-0] [Medline: 26813286]

37. Department of Health and Human Services, State of Victoria. 2016. Allied health: credentialling, competency and capability framework (revised edition): Driving effective workforce practice in a changing health environment URL: https://www2. health.vic.gov.au/Api/downloadmedia/%7B426782F8-57EB-4DD9-B9CF-25CB92256A02%7D [accessed 2018-02-07] [WebCite Cache ID 6x2pTW5rN]

38. Gilstad H. Toward a comprehensive model of eHealth literacy. 2014 Presented at: Practical Aspects of Health Informatics 2014; May 19, 2014; Trondheim, Norway URL: http://ceur-ws.org/

39. Gray K, Choo D, Butler-Henderson K, Whetton S, Maeder A. Clinical Informatics Education. 2015 Aug. An online health informatics subject for clinical health profession students in Australian universities: an extension of the OLT Priority Project Coordinated interprofessional curriculum renewal for ehealth capability in clinical health profession degrees. Final report URL: http://clinicalinformaticseducation.pbworks.com/w/file/fetch/103478171/ Clinical%20Informatics%20and%20eHealth%20OLT%20Final%20Report%202015.pdf [accessed 2018-02-07] [WebCite Cache ID 6x2rm5NIl]

40. Gray K, Choo D, Butler-Henderson K, Whetton S, Maeder A. Clinical Informatics Education. 2015. Clinical Informatics and eHealth: Open Curriculum URL: http://clinicalinformaticseducation.pbworks.com/w/file/fetch/97197057/ Clinical%20Informatics%20and%20eHealth%20OLT%20Curriculum%20V1.1.pdf [accessed 2018-02-07] [WebCite Cache ID 6x2sP5DiZ]

41. Honey ML, Skiba DJ, Procter P, Foster J, Kouri P, Nagle LM. Nursing Informatics Competencies for Entry to Practice: The Perspective of Six Countries. Stud Health Technol Inform 2017;232:51-61. [Medline: 28106582]

42. Klecun E, Lichtner V, Cornford T. e-Literacy in health care. Stud Health Technol Inform 2014;205:843-847. [Medline: 25160306]

43. Procter PM. Ubiquitous Adoption of Innovative and Supportive Information and Communications Technology Across Health and Social Care Needs Education for Clinicians. Stud Health Technol Inform 2017;235:358-362. [Medline: 28423814]

44. Röhrig R, Stausberg J, Dugas M, GMDS project group, Medical Informatics Education in Medicine. Development of national competency-based learning objectives "Medical Informatics" for undergraduate medical education. Methods Inf Med 2013;52(3):184-188. [doi: 10.3414/ME13-04-0001] [Medline: 23615960]

XSL·FO
RenderX

45.   van Houwelingen CT, Moerman AH, Ettema RG, Kort HS, Ten Cate O. Competencies required for nursing telehealth activities: A Delphi-study. Nurse Educ Today 2016 Apr;39:50-62 [FREE Full text] [doi: 10.1016/j.nedt.2015.12.025] [Medline: 27006033]

46.   Värri A, Blake R, Roberts J, Fenton S, Cleary M, Zacks S, et al. Transatlantic collection of health informatics competencies. FinJeHeW 2016 May 23;8(2-3):127-136 [FREE Full text]

47.   Friedman CP, Donaldson KM, Vantsevich AV. Educating medical students in the era of ubiquitous information. Med Teach 2016 May;38(5):504-509 [FREE Full text] [doi: 10.3109/0142159X.2016.1150990] [Medline: 27027546]

48.   Jogerst K, Callender B, Adams V, Evert J, Fields E, Hall T, et al. Identifying interprofessional global health competencies for 21st-century health professionals. Ann Glob Health 2015;81(2):239-247 [FREE Full text] [doi: 10.1016/j.aogh.2015.03.006] [Medline: 26088089]

49.   Biggins D, Holley D, Zezulkova M. Digital Competence and Capability Frameworks in Higher Education: Importance of Life-long Learning, Self-Development and Well-being. EAI Endorsed Transactions on e-Learning 2017;4(13):e1 [FREE Full text]

50.   Susskind R, Susskind D. Harvard Business Review. 2016. Technology Will Replace Many Doctors, Lawyers, and Other Professionals URL: https://hbr.org/2016/10/robots-will-replace-doctors-lawyers-and-other-professionals [accessed 2018-02-07] [WebCite Cache ID 6x2uFGNkp]

Abbreviations

eHealth: electronic health

Original Paper

The Effectiveness of a Computer-Tailored E-Learning Program for Practice Nurses to Improve Their Adherence to Smoking Cessation Counseling Guidelines: Randomized Controlled Trial

Dennis de Ruijter[1], MSc; Math Candel[2], PhD; Eline Suzanne Smit[3], PhD; Hein de Vries[1], PhD; Ciska Hoving[1], PhD

[1]Care and Public Health Research Institute, Department of Health Promotion, Maastricht University, Maastricht, Netherlands

[2]Care and Public Health Research Institute, Department of Methodology & Statistics, Maastricht University, Maastricht, Netherlands

[3]Amsterdam School of Communication Research, Department of Communication Science, University of Amsterdam, Amsterdam, Netherlands

Corresponding Author:
Dennis de Ruijter, MSc
Care and Public Health Research Institute
Department of Health Promotion
Maastricht University
Maastricht,
Netherlands
Phone: 31 43 388 24 06
Email: d.deruijter@maastrichtuniversity.nl

Abstract

Background: Improving practice nurses' (PN) adherence to smoking cessation counseling guidelines will benefit the quality of smoking cessation care and will potentially lead to higher smoking abstinence rates. However, support programs to aid PNs in improving their guideline uptake and adherence do not exist yet.

Objective: The aim of this study was to assess the effects of a novel computer-tailored electronic learning (e-learning) program on PNs' smoking cessation guideline adherence.

Methods: A Web-based randomized controlled trial (RCT) was conducted in which an intervention group (N=147) with full access to the e-learning program for 6 months was compared with a control group (N=122) without access. Data collection was fully automated at baseline and 6-month follow-up via online questionnaires, assessing PNs' demographics, work-related factors, potential behavioral predictors based on the I-Change model, and guideline adherence. PNs also completed counseling checklists to retrieve self-reported counseling activities for each consultation with a smoker (N=1175). To assess the program's effectiveness in improving PNs' guideline adherence (ie, overall adherence and adherence to individual counseling guideline steps), mixed linear and logistic regression analyses were conducted, thus accommodating for the smokers being nested within PNs. Potential effect moderation by work-related factors and behavioral predictors was also examined.

Results: After 6 months, 121 PNs in the intervention group (82.3%, 121/147) and 103 in the control group (84.4%, 103/122) completed the follow-up questionnaire. Mixed linear regression analysis revealed that counseling experience moderated the program's effect on PNs' overall guideline adherence (beta=.589; 95% CI 0.111-1.068; $P_{\text{Holm-Bonferroni}}$ =.048), indicating a positive program effect on adherence for PNs with a more than average level of counseling experience. Mixed logistic regression analyses regarding adherence to individual guideline steps revealed a trend toward moderating effects of baseline levels of behavioral predictors and counseling experience. More specifically, for PNs with less favorable scores on behavioral predictors (eg, low baseline self-efficacy) and high levels of counseling experience, the program significantly increased adherence.

Conclusions: Results from our RCT showed that among PNs with more than average counseling experience, the e-learning program resulted in significantly better smoking cessation guideline adherence. Experienced PNs might have been better able to translate the content of our e-learning program into practically applicable counseling strategies compared with less experienced colleagues. Less favorable baseline levels of behavioral predictors among PNs possibly contributed to this effect, as there was more room for improvement by consulting the tailored content of the e-learning program. To further substantiate the effectiveness of e-learning programs on guideline adherence by health care professionals (HCPs), it is important to assess how to support a wider range of HCPs.

Trial Registration: Netherlands Trial Register NTR4436; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=4436 (Archived by WebCite at http://www.webcitation.org/6zJQuSRq0)

XSL•FO
RenderX

Introduction

Smoking is the most preventable cause of illness and premature death worldwide [1,2]. In the Netherlands, 24.1% of adults still smoked in 2016 [3], illustrating the persistent need for effective smoking cessation strategies. For example, the general practice setting has great potential for cessation support, as over 75% of Dutch smokers visit their general practice at least once a year [4]. General practice health care professionals (HCPs) such as practice nurses (PN) and general practitioners (GPs) are trained to use evidence-based smoking cessation guidelines (ie, the STIMEDIC guideline [5] being the most recent one) to counsel their smoking patients. Applying such guidelines in structured cessation treatment, which combines behavioral and pharmacological support [5], is known to have beneficial effects on smokers' abstinence rates [6]. However, only in 25% to 33% of consultations do smokers receive a quit smoking advice in their general practice [4]. Moreover, once a quit advice has been given, more extensive smoking cessation support should be provided, which is most often the responsibility of PNs [7]. Yet, also concerning subsequent steps of evidence-based smoking cessation guidelines, PNs' adherence is suboptimal [8]. Consequently, improving PNs' guideline adherence would benefit the quality of the smoking cessation care in the general practice and could therefore lead to higher smoking abstinence rates [9,10].

An earlier study investigating PNs' needs for guideline adherence support found that they were interested in an individually relevant, easy-to-use, and practically applicable program or intervention [11]. Moreover, research showed that PNs' guideline adherence is positively related to their level of self-efficacy for using a guideline and perceiving advantages of using a guideline [8,12-14]. Such behavioral predictors could be targeted through intervention programs aimed to improve PNs' guideline adherence. More specifically, providing PNs with content tailored to behavioral predictors fulfills their need for an individually relevant program, and therefore, tailored content is more likely to be read and remembered, compared with nontailored program content [15]. For instance, tailored content can be matched with PNs' individual level of self-efficacy and their perceived advantages of guideline usage: information can be provided regarding potentially difficult counseling situations (eg, when limited time is available) or regarding specific benefits of using a smoking cessation guideline during consultations (eg, increasing counseling quality), as identified by each individual PN in an earlier evaluation [16].

Additionally, providing PNs with online access to tailored content (ie, computer-tailored, CT) enables them to consult it time-efficiently and whenever and wherever they desire [17,18]. Previously tested CT programs proved to be effective in changing various (determinants of) health behaviors, including smoking cessation [19,20]. Therefore, by targeting PNs'

behavioral predictors via a Web-based CT support program, positive behavior change can be achieved among PNs, meaning that they improve their smoking cessation guideline adherence. Moreover, despite PNs' interest in tailored adherence support [11], such (Web-based) CT programs do not yet exist with the aim to increase PNs' smoking cessation guideline adherence. Therefore, we developed and tested a novel Web-based CT electronic learning (e-learning) program for PNs to support them to improve their smoking cessation counseling guideline adherence [21].

The aim of the study described here was to assess the effects of the CT e-learning program on PNs' smoking cessation guideline adherence in a randomized controlled trial (RCT). We hypothesized that PNs' guideline adherence would significantly improve as a result of exposure to the CT e-learning program.

Methods

Study Design

We conducted an RCT to investigate the effectiveness of the CT e-learning program on PNs' smoking cessation guideline adherence, compared with no intervention. A full description of the design of the RCT can be found elsewhere [21]. Evaluation by the Medical Ethics Committee Atrium-Orbis-Zuyd (14-N-17) revealed that no medical ethical clearance for this study was needed according to the rules of the Medical Research Involving Human Subjects Act (WMO). The study is registered with the Dutch Trial Register (NTR4436).

The Computer-Tailored E-Learning Program

The CT e-learning program was structurally based on previously developed CT programs [22,23] and consisted of (1) Several e-learning modules in which PNs had access to individually tailored advice, a forum, and smoking cessation counseling materials (both to inform themselves and to provide to smokers) and (2) Three general modules with project information, frequently asked questions about the RCT, and a counseling checklist to monitor self-reported counseling activities during the trial [21]. The content of advice modules was tailored to several respondent characteristics theoretically grounded in the I-Change Model (ICM [24]), which were previously demonstrated to be effective in achieving behavior change [25-28]: demographics (eg gender), premotivational factors (eg, knowledge), motivational factors (eg, self-efficacy), postmotivational factors (eg, coping planning), intention (to use a smoking cessation guideline), and behavior (ie, self-reported application of smoking cessation guideline steps).

Participants and Procedure

PNs across the Netherlands were contacted through email, newsletters, and website messages via national organizations for PNs or primary care professionals in general, as well as via a project website [29] and social media platforms (ie, Twitter,

XSL•FO

RenderX

LinkedIn, and Facebook). Additionally, individual PNs were contacted by the research team via telephone through their general practice. Eligible PNs were actively engaged in smoking cessation counseling in a Dutch general practice, had Internet access and an active email account, and were sufficiently proficient in Dutch. Upon interest and obtaining important project information via telephone and email, PNs were prompted to visit the CT e-learning program to complete an online informed consent form, were randomized (ie, allocation by a computer software randomization device), and were asked to fill out the Web-based baseline questionnaire. As PN enrollment in the trial was spread over a period of 6 months, randomization was conducted at respondent level at the time of enrollment of an individual PN.

Individual PNs who were randomly allocated to the intervention group of the trial had access to all e-learning and general modules described above and received a tailored feedback letter based on their answers to the baseline questionnaire; this letter provided individual PNs with a summary of various pieces of tailored advice (ie, on different motivational factors and behavior) and instructions on where to find more elaborate advice in the e-learning modules. PNs in the control group only had access to the general modules. During a 6-month time period (ie, upon completion of the baseline questionnaire), PNs in the intervention and control group were free to visit the modules of the CT e-learning program that were available to them based on their group allocation as many times as they wanted. PNs could directly print content from the modules and save this content on their computer.

During the trial, PNs in both the intervention and control group were asked to engage in smoking cessation counseling with their smoking patients when the opportunity arose. All PNs were asked to recruit these smokers to partake in the trial (Figure 1). When smokers agreed, PNs were instructed to record smokers' date of birth and email address. Smokers then directly received an email invitation to participate in the trial and to fill out an online questionnaire.

Data Collection Among Practice Nurses

Baseline and 6-month follow-up questionnaires for PNs were informed by the ICM [24] and were based on questionnaires previously used among HCPs to assess smoking cessation activities [13,23,30]. Questionnaires were identical for intervention and control group PNs and administered in a Web-based format. The baseline questionnaire for PNs consisted of questions concerning demographic characteristics, potential behavioral predictors of adherence, and their guideline adherence. The follow-up questionnaire included the same questions about potential behavioral predictors and PNs' guideline adherence. Additional data on PNs' smoking cessation guideline adherence were collected via the counseling checklist that PNs filled out after each consultation with a smoker throughout the 6-month intervention period and during a 6-month follow-up period.

Demographics

After providing online informed consent, every PN was requested to fill out their first and last name, gender, date of birth, and smoking status (smoker, ex-smoker, nonsmoker). Subsequently, they filled out in how many general practices they worked, how many hours they worked per week, and whether or not they were listed in the Dutch Stop Smoking Quality Register (ie, a register with qualified smoking cessation professionals). The final questions concerned the practice in which a PN worked most hours per week; they filled out practice name, experience in smoking cessation counseling in years, the presence of designated smoking cessation consulting hours (yes or no), and whether patients' smoking status was systematically registered in their patient files (yes or no).

Behavioral Predictors

Several socio-cognitive factors were assessed as potential predictors of guideline adherence, informed by the ICM [24]: intention, knowledge, attitude, self-efficacy, social influence, action planning, and coping planning. Items were based on previously used questionnaires about behavioral predictors related to nurses' smoking cessation counseling [13,19,30].

PNs' intention to use a smoking cessation guideline was assessed by two questions addressing the intention to use (1) Any evidence-based smoking cessation guideline and (2) The most recent Dutch counseling protocol, specifically (ie, STIMEDIC guideline [5]: *Do you intend to use the STIMEDIC quit smoking guideline?*), using the same answering scale (1=definitely not, 4=do not know, 7=definitely).

PNs' knowledge of evidence-based smoking cessation guidelines was assessed by 18 true-false items concerning the content of the STIMEDIC guideline (eg, *the first consultation of a smoking cessation trajectory starts with providing a quit advice)*. As such, PNs scored points for every statement they appropriately identified to correctly reflect the content of the counseling protocol (range 0-18).

PNs' attitude was assessed by seven items about perceived advantages (eg, *using an evidence-based guideline improves the quality of my smoking cessation counseling*) and seven items about perceived disadvantages (eg, *using an evidence-based guideline is time-consuming for me*) of using an evidence-based smoking cessation guideline (1=completely disagree, 5=completely agree). These items were subsequently combined into separate scales for perceived advantages (Cronbach alpha=0.82; Ω=0.82) and perceived disadvantages (Cronbach alpha=0.74; Ω=0.74).

PNs' level of self-efficacy was assessed by ten items describing potentially difficult situations when trying to adhere to an evidence-based smoking cessation guideline (eg, *when it is very busy at the general practice*) and asking PNs how difficult they would find it to follow a guideline in each of these situations (1=very difficult, 5=very easy). All items were combined into a self-efficacy scale (Cronbach alpha=0.84; Ω=0.84).

Figure 1. Flow of counseled smokers recruited by practice nurses.
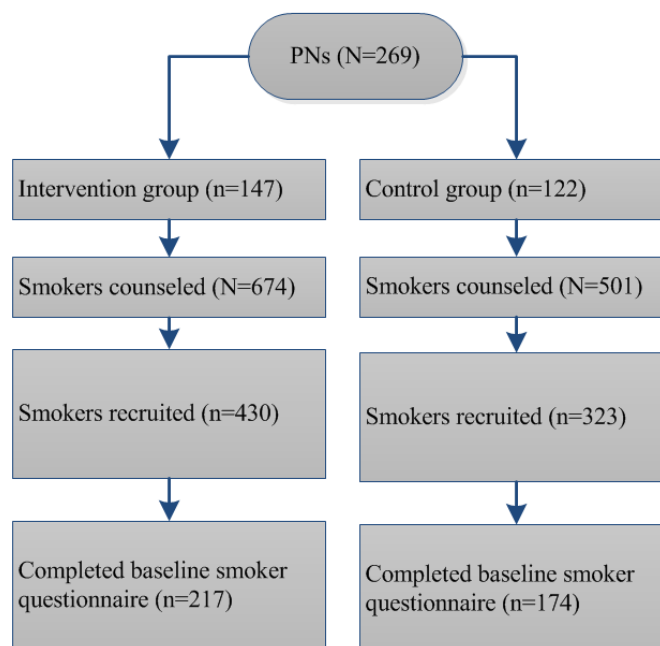
PNs' perceived social influence was assessed by three items about social modeling (eg, *the GP works with an evidence-based smoking cessation guideline*), five items about social support (eg, *colleagues in other practices support the use of an evidence-based smoking cessation guideline*), and five items about social norms (eg, *the practice manager thinks using an evidence-based smoking cessation guideline is important*). All items assessed the potential influence of important others within and outside the general practice (1=completely disagree, 3=neutral, 5=completely agree) and were subsequently combined into separate scales for social modeling (Cronbach alpha=0.62; Ω=0.63), social support (Cronbach alpha=0.73; Ω=0.74), and social norms (Cronbach alpha=0.71; Ω=0.72). Scores that represented a *not applicable* answering category were assigned a neutral score.

PNs' intention to make action plans and coping plans was assessed by eight and ten items, respectively (yes or no). Action plans addressed specific activities for preparing a smoking cessation consultation with a patient (eg, *discussing a patient's smoking status with the GP*), whereas coping plans addressed their aspiration to develop a concrete plan for dealing with potentially difficult situations (ie, plans for dealing with the same potentially difficult situations as assessed in the self-efficacy questions). Subsequently, sum scores for both action plans (range 0-8) and coping plans (range 0-10) were computed.

Guideline Adherence

Questions on guideline adherence concerned the nine evidence-based counseling steps, as described in the STIMEDIC guideline [5]: (1) advising to quit smoking, (2) assessing smoking profile and smoking history, (3) assessing motivation to quit, (4) increasing motivation, (5) assessing barriers to quitting, (6) discussing barriers, (7) informing about cessation aids, (8) making a quit plan and setting a quit date, and (9) arranging follow-up after the quit date. PNs' adherence at baseline was assessed by asking PNs to self-report their adherence to each guideline step (eg, *I advised my patient to quit smoking; step 1*) during complete smoking cessation trajectories (ie, intake and follow-up consultations) of their last ten patients (range 0-10). These data on PNs' adherence were used to create CT advice for PNs in the intervention group regarding their behavior. In the effect analyses, PNs' baseline guideline adherence score was used as a covariate. Additionally, during the trial period, guideline adherence was assessed by asking PNs to self-report their adherence to each guideline step (ie, *Please select which subjects were addressed during the consultation with your smoking patient*) after every consultation with a smoking patient (yes or no) using the counseling checklist (ie, one of the general modules in the CT e-learning program available for PNs in both the intervention and control group). This resulted in a score from 0 (none of the steps were adhered to) to 9 (all steps were adhered to) for each individual consultation with a smoker. Checklists of consultations with the same smoker were combined into a single score for guideline adherence, reflecting a PN's adherence during a complete counseling trajectory of a smoking patient, which was the primary outcome measure in the effect analyses.

Sample Size Calculation

We calculated the required sample size based on the possibility to detect a difference of medium effect size (ie, adherence to two additional guideline steps) between intervention and control group PNs (alpha=5%; beta=10%). As a result, at least 95 PNs per condition at the end of the trial would be sufficient [21]. However, to detect a medium effect size for an interaction with the intervention factor when assuming an intraclass correlation of .25, at least 105 PNs per condition are needed to ensure a statistical power of 80%. Considering 30% attrition, we aimed to include 300 PNs at baseline.

Statistical Analyses

Reliability analyses (ie, Cronbach alpha and Ω) were conducted using R version 3.4.0 (R Foundation for Statistical Computing), and other statistical analyses were conducted using SPSS version 23.0 (IBM Corp). Descriptive analyses were conducted to summarize PNs' characteristics, whereas independent-samples *t* tests and chi-square tests determined significant (*P*<.05) baseline differences between intervention and control group PNs. Logistic regression was used to determine selective dropout of PNs after baseline, including variables potentially related to PNs' guideline adherence (ie, specific work-related variables, behavioral predictors, and baseline guideline adherence). On the basis of analyses for baseline differences and selective dropout, statistically significant variables were identified and included as covariates in further analyses.

As smokers were nested within PNs participating in the trial, mixed regression analyses were conducted to assess the effects of exposure to the CT e-learning program on PNs' smoking cessation guideline adherence. Both PNs' overall adherence score (range 0-9) was used as outcome measure and their adherence score for each guideline step separately (ie, step-based adherence; 0=nonadherent, 1=adherent). Therefore, both linear and logistic mixed models were run, including the same covariates. Effect moderators were tested by including interaction effects with PNs' group allocation (ie, intervention or control) to the regression models tested. On the basis of literature, several work-related factors (ie, counseling experience and presence of consulting hours [31,32]) and behavioral predictors (ie, intention, attitude, self-efficacy, and social influence [8,33,34]), potentially moderating the program's effect on PNs' adherence, were tested. First, nonsignificant interaction effects were stepwise deleted using a backward deletion procedure, meaning that at each step the least significant interaction effect was removed. Second, nonsignificant covariates were deleted from the model following the same procedure, with the restriction that these covariates remained in the model if they were also part of a significant interaction term. Upon finding a significant interaction effect, subsequent subgroup analyses were conducted to determine the nature of the moderation using adjusted alpha levels (Holm-Bonferroni method) to correct for multiple testing. For subgroup analyses, the final mixed regression model was repeated, while replacing the original moderator with three centered versions of the moderator, centered by subtracting the mean – 1 SD, the mean, and the mean + 1 SD from the original scores on the moderating variable. This allows for testing the effects of the e-learning program for three subgroups: one group corresponding to a score of the mean – 1 SD on the moderating variable, a second group with a score at the average on the moderating variable, and a third group with a score at mean + 1 SD on the moderating variable.

As 211 PNs (78.4%, 211/269) completed at least one checklist (which was needed to calculate the primary outcome measure), it meant that 58 PNs were excluded from effect analyses. For this reason, sensitivity analyses were conducted by replacing missing values on the primary outcome measure (guideline adherence) with scores assuming some dependency between the score being missing and the adherence score itself, either following an optimistic or a pessimistic scenario. In both scenarios, missing data were imputed for these 58 PNs based on the average number of patients counseled during the trial per PN. Furthermore, PNs' dropout status was taken into account, as some PNs, who did not complete any checklists, also did not fill out the follow-up questionnaire (ie, PN dropouts). One might expect PNs who did complete the follow-up questionnaire (ie, retained PNs) to be more motivated and to be more adherent if they would have completed the checklists during the intervention period. This was taken into account when imputing data. In the optimistic imputation scenario, retained PNs were assumed to be adherent in 90% of the consultations with smokers, whereas PN dropouts were assumed to be adherent in only 80% of their consultations. In the pessimistic scenario, a 50% probability of adherence was assumed for retained PNs and a 20% probability of adherence for PN dropouts. The datasets obtained under these two imputation scenarios were analyzed with the mixed regression models, as obtained after backward deletion of nonsignificant (interaction) effects in the analysis of only the complete cases.

Results

Sample Characteristics

Figure 2 shows the flow of PNs included in the trial from initial assessment of eligibility to randomization and completion of baseline and follow-up questionnaires. Of the 346 PNs assessed for eligibility, 49 (14.2%) did not meet inclusion criteria, and 18 (5.2%) refrained from participation. After randomization, 147 (49.5%, 147/297) and 122 (41.1%, 122/297) PNs were allocated to the intervention and control group, respectively, and completed the baseline questionnaire. Unequal group sizes are the result of chance, as randomization took place at respondent level, and each PN had a 50% probability of being allocated to either group.

The baseline sample of PNs (Table 1) had a mean age of 47.3 years (range 23-66), the vast majority was female (97.8%, 263/269), and very few PNs (1.1%, 3/269) were current smokers. Nearly half (47.2%, 127/269) worked in more than one general practice, and PNs worked on average almost 26 hours a week (range 3-42). Many PNs (66.9%, 180/269) were listed in the Dutch Stop Smoking Quality Register, and the mean-reported PN counseling experience was 5.6 years (range 0-20). Finally, almost half of the PNs (47.6%, 128/269) worked in a general practice with designated smoking cessation consulting hours, and nearly all (92.6%, 249/269) reported to systematically register their patients' smoking status in their patient files. During the trial, PNs engaged in smoking cessation counseling with 5.6 different patients on average (range 1-26).

Baseline characteristics of PNs were comparable between intervention and control group, except for the presence of designated smoking cessation consulting hours, which was significantly more often reported by PNs in the control group (χ^2_1=10.1; *P*=.001). Therefore, the presence of designated consulting-hours was included as a covariate in all effect analyses.

Attrition Analyses

After 6 months, 254 PNs remained in the trial and were invited for the follow-up measurement, which was completed by 88.9% (121/136) of intervention group and 87.3% (103/118) of control group PNs, respectively (Figure 2). Attrition analyses revealed that PNs who completed the follow-up measurement had a higher baseline intention to use the STIMEDIC guideline (odds ratio, OR 1.41, 95% CI 1.00-1.98) and had more baseline knowledge about the STIMEDIC guideline content (OR 1.39, 95% CI 1.07-1.82) compared with PNs who dropped out before the follow-up measurement (Table 2). Therefore, these two variables were also included as covariates in all effect analyses.

Effect Analyses

Overall Adherence

Table 3 shows the mixed regression results on PNs' overall guideline adherence; mean-centered values are reported for variables included in interaction effects to enable meaningful interpretation of the effect of *group* (eg, in Table 3, the effect of group allocation illustrates the effect of the intervention for PNs that score average on counseling experience). The results reveal a significant interaction effect of group allocation with counseling experience ($P=.045$) and a main effect of perceived advantages of guideline use ($P=.03$). Subgroup analyses showed that for PNs with more than average counseling experience (ie, mean + 1SD = 9.4 years of experience), allocation to the intervention group (ie, access to the CT e-learning program) resulted in a significantly higher overall adherence compared with the control group (beta=.589; 95% CI 0.111-1.068; $P_{\text{Holm-Bonferroni}}=.048$). The subgroup analysis for less experienced PNs (ie, mean 5.6 years of experience and mean − 1SD = 1.9 years of experience) revealed no significant intervention effect.

Figure 2. Flow and randomization of practice nurses that were recruited from January 2016 to June 2016.
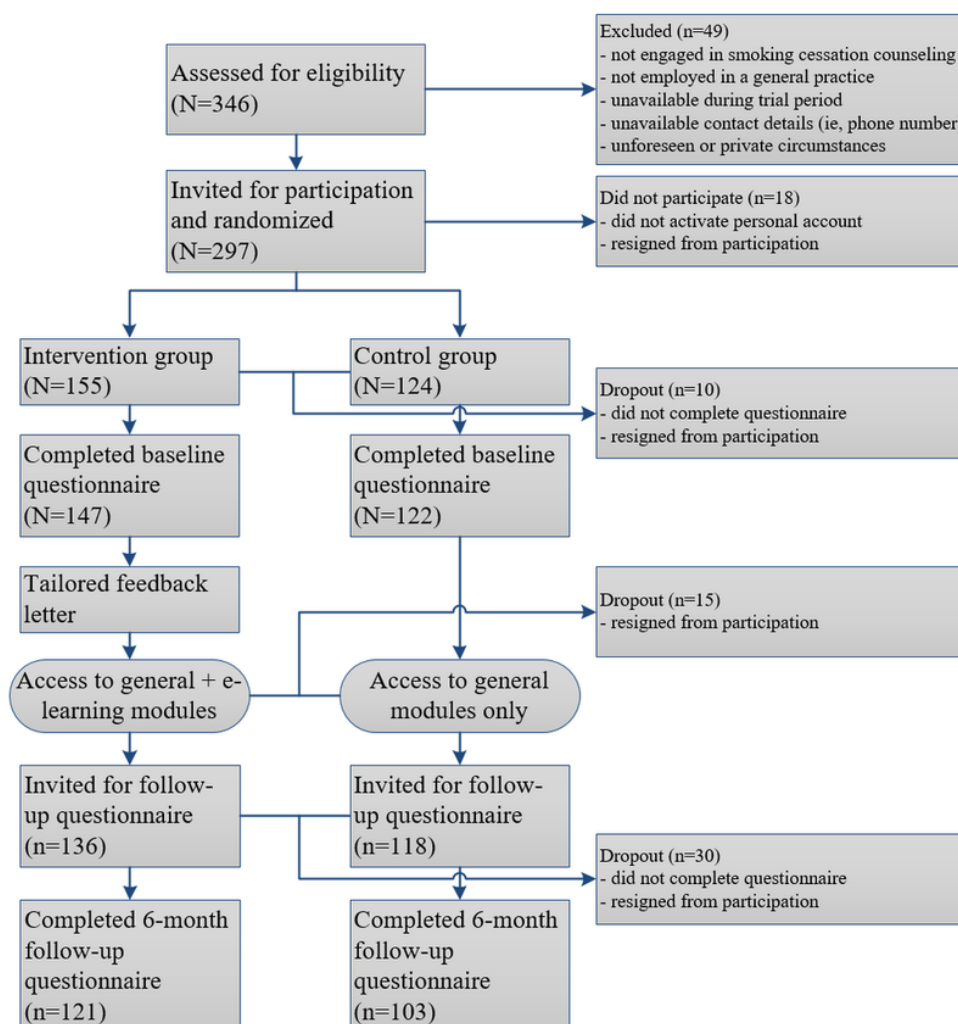
Table 1. Characteristics of practice nurses and comparison of characteristics between intervention and control groups.

Characteristics	Overall sample (N=269)	Intervention group (n=147)	Control group (n=122)	Chi-square	T (degrees of freedom)	P value
Age in years, mean (SD)	47.3 (9.5)	48.0 (9.6)	46.5 (9.4)	N/A[a]	−1.345 (267)	.18
Female, n (%)	263 (97.8)	143 (97.3)	120 (98.4)	X^2_1=0.4	N/A	.55
Smoking status, n (%)				X^2_2=0.4	N/A	.84
Nonsmoker	151 (56.1)	84 (57.1)	67 (54.9)			
Ex-smoker	115 (42.8)	61 (41.5)	54 (44.3)			
Smoker	3 (1.1)	2 (1.4)	1 (0.8)			
Employed in >1 practice, n (%)	127 (47.2)	70 (47.6)	57 (46.7)	X^2_1=0.0	N/A	.88
Working hours, mean (SD)	25.7 (7.4)	25.7 (7.5)	25.7 (7.4)	N/A	−0.007 (267)	.99
Registration in Stop Smoking Quality Register, n (%)	180 (66.9)	95 (64.6)	85 (69.7)	X^2_1=0.8	N/A	.38
Counseling experience in years, mean (SD)	5.6 (3.7)	5.5 (3.6)	5.8 (3.8)	N/A	-0.602 (267)	.55
Consulting hours present, n (%)	128 (47.6)	57 (38.8)	71 (58.2)	X^2_1=10.1	N/A	.001
Registration of smoking status in patient files, n (%)	249 (92.6)	135 (91.8)	114 (93.4)	X^2_1=0.3	N/A	.62
Number of counseled patients during the trial, mean (SD)	5.6 (4.4)	5.8 (4.1)	5.3 (4.8)	N/A	-0.700 (209)	.49

[a]N/A: not applicable.

Table 2. Baseline characteristics of practice nurses (PN) and their odds to predict PN retention.

Characteristics	Mean (SD)	Range	Odds ratio (95% CI)	P value
Working hours	25.7 (7.4)	3-42	0.95 (0.91-1.00)	.06
Counseling experience in years	5.6 (3.7)	0-20	0.96 (0.88-1.06)	.42
Intention to use any evidence-based guideline (1-7)	6.3 (0.8)	4-7	0.86 (0.51-1.45)	.57
Intention to use STIMEDIC (1-7)	5.5 (1.1)	2-7	1.41 (1.00-1.98)	.049
STIMEDIC knowledge (0-18)	14.6 (1.3)	11-18	1.39 (1.07-1.82)	.02
Perceived advantages (1-5)	4.2 (0.6)	2-5	0.80 (0.41-1.56)	.52
Perceived disadvantages (1-5)	1.9 (0.6)	1-4	0.56 (0.28-1.13)	.10
Self-efficacy (1-5)	2.8 (0.6)	1.3-5	0.62 (0.34-1.13)	.12
Social modeling (1-5)	3.2 (0.9)	1-5	0.70 (0.38-1.29)	.25
Social support (1-5)	3.3 (0.8)	1-5	1.61 (0.60-4.28)	.35
Social norms (1-5)	1.4 (0.7)	1.4-5	1.34 (0.47-3.83)	.58
Action planning (0-8)	6.4 (1.7)	0-8	0.97 (0.78-1.22)	.82
Coping planning (0-10)	6.7 (2.7)	0-10	1.01 (0.89-1.16)	.86
Baseline guideline adherence (0-9)	8.5 (1.6)	0-10	1.06 (0.86-1.31)	.59

Table 3. Results of backward linear mixed regression analysis on practical nurses' overall guideline adherence.

Final model	Coefficient	95% CI	P value
Group (control=0; intervention=1)	.245	−0.087 to 0.577	.19
Counseling experience	−.046	−0.113 to 0.021	.18
Perceived advantages	.319	0.031-0.608	.03
Group* counseling experience	.092	0.002-0.183	.045

Step-Based Adherence

Multimedia Appendix 1 shows the results on PNs' step-based adherence for each step separately; again mean-centered values are reported to enable meaningful interpretation of the effect of *group* in the presence of a significant interaction. Regarding adherence to *step 1* (ie, advising to quit smoking), a significant main effect was found of PNs' baseline adherence to step 1 (P=.002). Regarding adherence to *step 2* (ie, assessing smoking profile and smoking history), a significant interaction effect of group allocation with perceived advantages of guideline use (P=.001) was identified. Regarding adherence to *step 3* (ie, assessing motivation to quit), an interaction effect with self-efficacy (P=.02) was found. Regarding adherence to *step 4* (ie, increasing motivation), a borderline significant interaction effect of group allocation with counseling experience (P=.06) was found, as well as main effects of baseline adherence to step 4 (P=.01) and of perceived advantages of guideline use (P=.03). Regarding adherence to *step 5* (ie, assessing barriers to quitting), significant interaction effects of group allocation with social modeling (P=.009) and with social support were found (P=.046). Regarding adherence to *step 6* (ie, discussing barriers), a significant main effect was found of perceived advantages of guideline use (P=.045). Regarding adherence to *step 7* (ie, informing about cessation aids), significant interaction effects of group allocation with perceived disadvantages (P=.01) and with self-efficacy (P=.001) and a borderline significant interaction effect with counseling experience (P=.05) were found. Furthermore, a significant main effect of social modeling (P=.01) was found. Regarding adherence to *step 8* (ie, making a quit plan and setting a quit date), significant interaction effects of group allocation with counseling experience (P=.01), with social support (P=.001), and with social norms (P=.005) were found. Concerning adherence to *step 9* (ie, arranging follow-up after the quit date), an interaction with social support (P=.04) was found.

Detailed results on subgroup analyses for each significant interaction effect are presented in Multimedia Appendix 2. Overall, the results show that higher adherence scores on individual guideline steps of intervention group PNs, compared with control group PNs, occur for (1) High levels of counseling experience and perceived social support and (2) Low levels of perceived advantages, perceived disadvantages, self-efficacy and social modeling, and little social norms.

Sensitivity Analyses

After replicating the final mixed regression models following an *optimistic imputation scenario*, results for overall guideline adherence indicated that only the main effect of perceived advantages remained significant. Regarding step-based adherence, similar results were found for adherence to guideline steps 1, 2, and 8. Concerning adherence to steps 5 and 7 only, the interaction effects with social modeling and self-efficacy could be replicated, and concerning adherence to step 4, only the main effect of perceived advantages remained significant. Main and interaction effects concerning guideline steps 3, 6, and 9 were no longer or only marginally significant after conducting these optimistic sensitivity analyses.

After replicating the final mixed regression models following a *pessimistic imputation scenario*, main and interaction effects for overall adherence were no longer or only marginally significant. Regarding step-based adherence, similar results were found for adherence to guideline steps 1, 6, 8, and 9, whereas regarding step 4, only main effects of baseline adherence to step 4 and perceived advantages remained significant. Main and interaction effects concerning guideline steps 2, 3, 5, and 7 were no longer or only marginally significant after conducting these pessimistic sensitivity analyses.

Discussion

Principal Findings

The present RCT tested the effectiveness of a novel CT e-learning program for PNs to improve their adherence to evidence-based smoking cessation guidelines in Dutch general practice. Our results suggest that among PNs with more than average smoking cessation counseling experience, access to the CT e-learning program resulted in significantly better guideline adherence. A more detailed inspection of PNs' guideline adherence revealed comparable results for several specific guideline steps (ie, increasing motivation, discussing cessation aids, and making a quit plan). Additionally, a trend was observed that for PNs reporting less favorable baseline levels for behavioral predictors such as self-efficacy and favorable levels of perceived social support, the e-learning program was effective in improving guideline adherence.

Regarding PNs' counseling experience, subgroup analyses illustrated that more experienced PNs in the intervention group adhered significantly better to the guideline (ie, difference of adherence to 0.6 steps) at follow-up compared with equally experienced PNs in the control group. This finding means that the CT e-learning program successfully promoted guideline adherence among PNs with more counseling experience, which is in line with the fact that a significant positive association between counseling experience and PNs' baseline guideline adherence was found (Multimedia Appendix 3). Experienced PNs might have been better able to practically apply the content of the e-learning program during their counseling, as they have likely dealt with many different types of smokers and difficult counseling situations in the past. Due to these past coping experiences, it could have been easier for them, compared with less experienced PNs, to translate the theoretically grounded e-learning content to counseling situations encountered in practice, leading to improved guideline adherence. Yet, evidence about the influence of HCPs' counseling experience on their application of evidence-based guidelines is scarce. One study among physicians investigated the association between work experience and guideline adherence after taking an e-learning course but could not establish such an association [35]. Another study focused on the relation between work experience and knowledge and found that more experienced professionals scored higher on knowledge after using a nontailored interactive video intervention [36]. However, an association of experience with knowledge was not established among PNs in the present trial (Multimedia Appendix 3), and neither was it found in a study about the effects of an e-learning course on the knowledge

level of nursing staff [37]. Hence, an alternative explanation could be considered; perhaps experienced PNs were able to spend their counseling time more efficiently, giving them opportunity to visit the CT e-learning program more often (eg, in between consultations) and—as a consequence—benefit more from its content. However, although more frequent users were more adherent in the present trial, program usage (mean number of module visits 3.2, SD 6.2, range 0-48) was not significantly associated with PNs' guideline adherence, nor did we identify that more experienced PNs were more frequent users (data not shown). This is comparable with a study among GPs, in which factors such as working experience were not predictive of usage of an e-learning program to promote dementia guideline adherence [38]. It is important to further investigate if and why HCPs, including PNs, with more counseling experience benefit more from CT e-learning programs. For example, process evaluation of users' interaction with a program can aid our understanding of the working mechanisms (eg, targeted behavioral predictors) of exposure to tailored program content [39]. Such insights could subsequently be used to improve a program's effectiveness for individuals with varying levels of counseling experience (eg, by additionally tailoring program content on an individual's level of experience).

Results regarding PNs' step-based guideline adherence illustrated that, besides counseling experience, several behavioral predictors could explain better adherence scores of PNs in the intervention group compared with the control group. PNs' perceived advantages and disadvantages of guideline application, level of self-efficacy, social modeling, social norms, and social support at baseline all moderated the e-learning program's effect on PNs' step-based adherence. Although moderation effects could not be replicated precisely for each separate guideline step, a trend was observed that especially a high level of baseline social support and lower baseline levels of the other predictors were related to better effects of the program on adherence. As the content of the e-learning program was designed to especially target such behavioral predictors of guideline adherence [21], it is likely that intervention group PNs, initially scoring less favorable on these predictors, were more able to significantly improve their adherence through progress made on these behavioral predictors. An analysis of PNs' change scores concerning these behavioral predictors, indeed found that intervention group PNs had more favorable scores compared with control group PNs, indicating more progress in terms of their levels of perceived advantages, disadvantages, self-efficacy, social modeling, social norms, and social support (data not shown). Although these change scores were not statistically significant, it is likely that the small improvements on all these predictors together contributed to better step-based adherence scores among intervention group PNs [24].

Strengths and Limitations

A strength of the present trial is that we enrolled a substantial number of PNs and managed to retain 83.3% (224/269) of them at 6-month follow-up. In comparison, in other studies among nurses, retention rates were considerably lower, eg, studies among hospital nurses that reported 3-month and 6-month follow-up retention rates of 68% and 58% [40], or 56% and

48%, respectively [41]. A high retention rate is essential to obtain adequate power for conducting statistical analyses. Moreover, the PNs in the present trial were able to counsel many smokers during the intervention period, resulting in a large amount of data collected on individual smoking cessation trajectories (N=1175). As each individual PN used the counseling checklist to self-report their application of guideline steps during each consultation with these smokers, we obtained data on PNs' guideline adherence from at least one consultation per smoker and from on average six different smokers' counseling trajectories per PN. Our primary outcome measure therefore reflects PNs' guideline adherence during the entire 12-month trial period instead of at a single time point only (eg, at the end of the intervention period). Outcome measures composed of multiple measurement points are expected to provide a more reliable insight in the target behavior compared with a single measurement point [42]. A final strength is the fact that PNs' guideline adherence was assessed and analyzed from a step-based perspective, yielding more detailed insights than only taking overall guideline adherence into account.

Nevertheless, we also experienced some challenges during the trial. Our intention was to triangulate data on PNs' guideline adherence from both smokers' and PNs' perspectives to compute the primary outcome measure [21]. Unfortunately, only 33.3% of smokers counseled by PNs (391/1175) also participated in the trial themselves, ie, completed the baseline smoker questionnaire. Qualitative posttrial interviews with PNs (N=17) revealed issues such as time constraints of both smokers and PNs (ie, trial participation was not addressed during the consultation) and smokers' privacy concerns as reasons for the low participation rate among smokers (data not shown). As a result, conducting effect analyses with these data would be unreliable because of substantial loss of power and selective inclusion of smokers. Moreover, a comparison of smoker-reported data from baseline questionnaires with PN-reported data from counseling checklists revealed significantly higher adherence scores reported by smokers (t_{390}=−6.73, P<.001). As we observed possible ceiling effects in these adherence scores, smoker-reported data were deemed to be unreliable for being used as outcome measure. An explanation could be that smokers overestimated their PNs' guideline adherence, as they were recruited for participation in the trial by their PN, who also supported them with smoking cessation. Perhaps they were afraid that providing critical answers to questions about their PNs' performance could influence the relationship with their PN. Although smokers were informed that data were treated anonymously and were not reported back to PNs, the phenomenon of social desirability is often observed when collecting data from patients in a health care setting [43]. It, hence, seems that collecting patient data might not always be reliable and that more objective data collection methods are required. A second limitation was that 58 PNs (21.6%, 58/269) did not manage to recruit and counsel smokers during the intervention period, resulting in lacking data on these PNs' guideline adherence during consultations. As a result, these PNs could not be included in the effect analyses. Nevertheless, inclusion of data from the remaining PNs (N=211) still ensured adequate power (alpha=5%; beta=10%) to conduct

mixed regression analyses concerning their guideline adherence [21]. To examine the sensitivity of these results, the final regression models were repeated following an optimistic and pessimistic imputation scenario. This sensitivity analysis showed that results could have been sensitive to informative dropout (ie, missingness related to unobserved variables), as some main and interaction effects disappeared in both scenarios [44]. Yet, most results were similar to the results of the complete-case analyses, indicating their robustness.

Implications

In light of the issues described, it would be worthwhile to investigate additional methods to gather data on PNs' guideline adherence. One such method could be qualitative data collection, by conducting content analyses of audio or video recordings of smoker-PN consultations [45-47]. In the present trial, however, PNs were very reluctant in agreeing to make an audio recording of a smoking cessation consultation, resulting in one or more successful recordings for only 11.5% of PNs (31/269) and 42 recordings in total. During posttrial interviews, PNs reported barriers such as privacy concerns (of both PN and smoker) and the perception that making a recording would influence the interaction with the smoker during a consultation. As a consequence, PNs believed that the recordings would not provide reliable insight into their counseling approach; a phenomenon also described in other recent studies [48,49]. Another potential method to collect data on PNs' guideline adherence could be to ask PNs to respond to simulated practice situations or clinical vignettes (ie, case studies of smokers visiting their practice for cessation support) to assess their application of evidence-based guideline steps. Earlier studies applied this method, for instance, in written format concerning physiotherapists' guideline adherence [50] or video-based format concerning suicide guideline implementation [51]. Similarly, PNs could be provided with clinical vignettes that describe different types of smokers in terms of (1) Motivation status (eg, unmotivated, contemplating, and motivated to quit); (2) Perceived barriers toward quitting (eg, smoking partner, weight gain, and stress); and (3) Requests for pharmacotherapy (eg,

desire for particular medicine, refusing any pharmacotherapy, and favoring alternative medicine), to assess their guideline adherence in various situations that they could encounter in practice. Research with such clinical vignettes among PNs is needed to determine the potential value of using vignettes to reliably measure PNs' smoking cessation guideline adherence.

Furthermore, the results of the present trial illustrated that the CT e-learning program successfully improved smoking cessation guideline adherence of experienced PNs. Unfortunately, it was also established that usage of the program by PNs in the trial was quite limited (ie, three module visits on average). This means that the program's effectiveness could potentially be increased when implementation of the program by PNs (ie, number of modules visits) improves. During posttrial interviews, PNs mentioned that receiving a reminder to visit the program more often could be a potential strategy to stimulate program usage. Furthermore, a process evaluation could inform improving alterations to the program's tailored content to extend the program's effectiveness beyond experienced PNs. When such strategies would be combined with program implementation among a larger population of PNs, this could further substantiate the impact of the CT e-learning program on the quality of smoking cessation care in the Netherlands.

Conclusions

Providing PNs access to a novel CT e-learning program resulted in significantly better adherence to evidence-based smoking cessation guidelines among Dutch PNs experienced in smoking cessation counseling, compared with similar PNs without access to this program. More favorable improvements on behavioral predictors of guideline adherence among intervention group PNs may explain better adherence scores. To further substantiate the effectiveness of e-learning programs on guideline adherence by HCPs, alternative methods of collecting data on guideline adherence should be explored, and strategies are needed to promote program usage and to also support less experienced PNs to adhere to evidence-based smoking cessation guidelines. This could subsequently inform widespread implementation of the e-learning program among PNs.

Authors' Contributions

All authors significantly contributed to drafting this manuscript. ESS, HdV, and CH developed the proposal for the present research. DdR was responsible for conducting and reporting on the research, and MC assisted in data analysis. MC, ESS, HdV, and CH supervised the process and provided feedback. DdR, ESS, HdV, and CH were involved in developing the content of the e-learning program.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Results of backward logistic mixed regression analyses on practice nurses' (PNs') step-based guideline adherence.

[PDF File (Adobe PDF File), 510KB - jmir_v20i5e193_app1.pdf]

Multimedia Appendix 2

Odds ratios (95% CI) reflecting intervention effects of subgroup analyses of significant interaction effects per guideline step with Holm-Bonferroni corrected P values.

[[PDF File (Adobe PDF File), 35KB](#) - [jmir_v20i5e193_app2.pdf](#)]

Multimedia Appendix 3

Correlation matrix of practice nurse (PN) characteristics.

[[PDF File (Adobe PDF File), 28KB](#) - [jmir_v20i5e193_app3.pdf](#)]

Multimedia Appendix 4

CONSORT - EHEALTH checklist (V 1.6.1).

[[PDF File (Adobe PDF File), 490KB](#) - [jmir_v20i5e193_app4.pdf](#)]

References

1. WHO. WHO report on the global tobacco epidemic 2015: Raising taxes on tobacco. Geneva: World Health Organization; 2015.
2. USDHHS. The Health Consequences of Smoking - 50 Years of Progress. A Report of the Surgeon General. Atlanta, GA: 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; 2014.
3. Springvloet L, van Laar M. Roken onder volwassenen: kerncijfers 2016. Utrecht: Nationaal Expertisecentrum Tabaksontmoediging; 2017.
4. Kleinjan M, Bommelé J, Verdurmen J, van Laar M. Het bespreken van (stoppen met) roken door de huisarts en andere zorgverleners (tandartsen, medisch specialisten en verloskundigen). Utrecht: Trimbos-instituut, National Expertisecentrum Tabaksontmoediging; 2016.
5. Trimbos-instituut, Nederlands Huisartsen Genootschap. Richtlijn Behandeling van tabaksverslaving en stoppen met roken ondersteuning Herziening 2016. Utrecht: Trimbos-instituut; 2016.
6. Stead LF, Buitrago D, Preciado N, Sanchez G, Hartmann-Boyce J, Lancaster T. Physician advice for smoking cessation. Cochrane Database Syst Rev 2013 May 31(5):CD000165. [doi: [10.1002/14651858.CD000165.pub4](#)] [Medline: [23728631](#)]
7. Freund T, Everett C, Griffiths P, Hudon C, Naccarella L, Laurant M. Skill mix, roles and remuneration in the primary care workforce: who are the healthcare professionals in the primary care teams across the world? Int J Nurs Stud 2015 Mar;52(3):727-743 [[FREE Full text](#)] [doi: [10.1016/j.ijnurstu.2014.11.014](#)] [Medline: [25577306](#)]
8. de Ruijter D, Smit ES, de Vries H, Hoving C. Dutch practice nurses' adherence to evidence-based smoking cessation treatment guidelines. Fam Pract 2017 May 09;34(6):685-691. [doi: [10.1093/fampra/cmx039](#)] [Medline: [28486612](#)]
9. Miller M, Kearney N. Guidelines for clinical practice: development, dissemination and implementation. Int J Nurs Stud 2004 Sep;41(7):813-821. [doi: [10.1016/j.ijnurstu.2003.09.005](#)] [Medline: [15288803](#)]
10. Rice VH, Hartmann-Boyce J, Stead LF. Nursing interventions for smoking cessation. Cochrane Database Syst Rev 2013 Aug 12(8):CD001188. [doi: [10.1002/14651858.CD001188.pub4](#)] [Medline: [23939719](#)]
11. de Ruijter D, Smit ES, de Vries H, Goossens L, Hoving C. Understanding Dutch practice nurses' adherence to evidence-based smoking cessation guidelines and their needs for web-based adherence support: results from semistructured interviews. BMJ Open 2017 Mar 22;7(3):e014154 [[FREE Full text](#)] [doi: [10.1136/bmjopen-2016-014154](#)] [Medline: [28336746](#)]
12. Studts JL, Flynn SM, Dill TC, Ridner SL, Worth CT, Walsh SE, et al. Nurse practitioners' knowledge, attitudes, and clinical practices regarding treatment of tobacco use and dependence. J Nurse Pract 2010 Mar 09;6(3):212-219 [[FREE Full text](#)] [doi: [10.1016/j.nurpra.2009.06.003](#)]
13. Segaar D, Willemsen MC, Bolman C, De Vries H. Nurse adherence to a minimal-contact smoking cessation intervention on cardiac wards. Res Nurs Health 2007 Aug;30(4):429-444. [doi: [10.1002/nur.20204](#)] [Medline: [17654478](#)]
14. Berndt NC, Bolman C, de Vries H, Segaar D, van Boven I, Lechner L. Smoking cessation treatment practices: recommendations for improved adoption on cardiology wards. J Cardiovasc Nurs 2013;28(1):35-47. [doi: [10.1097/JCN.0b013e318231f424](#)] [Medline: [22048618](#)]
15. Kreuter MW, Wray RJ. Tailored and targeted health communication: strategies for enhancing information relevance. Am J Health Behav 2003 Dec;27(Suppl 3):S227-S232. [Medline: [14672383](#)]
16. de Vries H, Brug J. Computer-tailored interventions motivating people to adopt health promoting behaviours: introduction to a new approach. Patient Educ Couns 1999 Feb;36(2):99-105. [Medline: [10223015](#)]
17. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006;8(2):e10 [[FREE Full text](#)] [doi: [10.2196/jmir.8.2.e10](#)] [Medline: [16867965](#)]

XSL•FO
RenderX

18. Tate DF, Finkelstein EA, Khavjou O, Gustafson A. Cost effectiveness of internet interventions: review and recommendations. Ann Behav Med 2009 Aug;38(1):40-45 [FREE Full text] [doi: 10.1007/s12160-009-9131-6] [Medline: 19834778]

19. Smit ES, Candel MJ, Hoving C, de Vries H. Results of the PAS study: a randomized controlled trial evaluating the effectiveness of a web-based multiple tailored smoking cessation program combined with tailored counseling by practice nurses. Health Commun 2016 Sep;31(9):1165-1173. [doi: 10.1080/10410236.2015.1049727] [Medline: 26934538]

20. Taylor GM, Dalili MN, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2017 Sep 04;9:CD007078. [doi: 10.1002/14651858.CD007078.pub5] [Medline: 28869775]

21. de Ruijter D, Smit ES, de Vries H, Hoving C. Web-based computer-tailoring for practice nurses aimed to improve smoking cessation guideline adherence: a study protocol for a randomized controlled effectiveness trial. Contemp Clin Trials 2016 May;48:125-132. [doi: 10.1016/j.cct.2016.04.007] [Medline: 27103232]

22. Stanczyk NE, Bolman C, Muris JW, de Vries H. Study protocol of a Dutch smoking cessation e-health program. BMC Public Health 2011;11:847 [FREE Full text] [doi: 10.1186/1471-2458-11-847] [Medline: 22059446]

23. Smit ES, de Vries H, Hoving C. The PAS study: a randomized controlled trial evaluating the effectiveness of a web-based multiple tailored smoking cessation programme and tailored counselling by practice nurses. Contemp Clin Trials 2010 May;31(3):251-258. [doi: 10.1016/j.cct.2010.03.001] [Medline: 20215047]

24. de Vries H. An integrated approach for understanding health behavior; The I-Change model as an example. PBSIJ 2017 Mar 9;2(2):1-4. [doi: 10.19080/PBSIJ.2017.02.555585]

25. Elfeddali I, Bolman C, Candel MJ, Wiers RW, de Vries H. Preventing smoking relapse via web-based computer-tailored feedback: a randomized controlled trial. J Med Internet Res 2012;14(4):e109 [FREE Full text] [doi: 10.2196/jmir.2057] [Medline: 22903145]

26. Stanczyk N, Bolman C, van Adrichem M, Candel M, Muris J, de Vries H. Comparison of text and video computer-tailored interventions for smoking cessation: randomized controlled trial. J Med Internet Res 2014 Mar 03;16(3):e69 [FREE Full text] [doi: 10.2196/jmir.3016] [Medline: 24589938]

27. Smit ES, de Vries H, Hoving C. Effectiveness of a web-based multiple tailored smoking cessation program: a randomized controlled trial among Dutch adult smokers. J Med Internet Res 2012;14(3):e82 [FREE Full text] [doi: 10.2196/jmir.1812] [Medline: 22687887]

28. Bolman C, Eggers SM, van Osch L, Te Poel F, Candel M, de Vries H. Is action planning helpful for smoking cessation? Assessing the effects of action planning in a web-based computer-tailored intervention. Subst Use Misuse 2015;50(10):1249-1260. [doi: 10.3109/10826084.2014.977397] [Medline: 26440754]

29. de Ruijter D, Smit E, De Vries H, Hoving C. STER Studie. URL: http://www.sterstudie.nl [WebCite Cache ID 6jHiDBycb]

30. Bolman C, de Vries H, Mesters I. Factors determining cardiac nurses' intentions to continue using a smoking cessation protocol. Heart Lung 2002;31(1):15-24. [Medline: 11805745]

31. Fleuren M, Wiefferink K, Paulussen T. Determinants of innovation within health care organizations: literature review and Delphi study. Int J Qual Health Care 2004 Apr;16(2):107-123 [FREE Full text] [doi: 10.1093/intqhc/mzh030] [Medline: 15051705]

32. Sangster-Gormley E, Martin-Misener R, Downe-Wamboldt B, Dicenso A. Factors affecting nurse practitioner role implementation in Canadian practice settings: an integrative review. J Adv Nurs 2011 Jun;67(6):1178-1190. [doi: 10.1111/j.1365-2648.2010.05571.x] [Medline: 21261696]

33. Leitlein L, Smit ES, de Vries H, Hoving C. Factors influencing Dutch practice nurses' intention to adopt a new smoking cessation intervention. J Adv Nurs 2012 Oct;68(10):2185-2194. [doi: 10.1111/j.1365-2648.2011.05903.x] [Medline: 22141413]

34. Smit ES, de Vries H, Hoving C. Determinants of practice nurses' intention to implement a new smoking cessation intervention: the importance of attitude and innovation characteristics. J Adv Nurs 2013 Dec;69(12):2665-2674. [doi: 10.1111/jan.12153] [Medline: 23600904]

35. Nicastro E, Lo Vecchio A, Liguoro I, Chmielewska A, De Bruyn C, Dolinsek J, et al. The impact of e-learning on adherence to guidelines for acute gastroenteritis: a single-arm intervention study. PLoS One 2015;10(7):e0132213 [FREE Full text] [doi: 10.1371/journal.pone.0132213] [Medline: 26148301]

36. Bond SE, Crowther SP, Adhikari S, Chubaty AJ, Yu P, Borchard JP, et al. Design and implementation of a novel web-based e-learning tool for education of health professionals on the antibiotic Vancomycin. J Med Internet Res 2017 Mar 30;19(3):e93 [FREE Full text] [doi: 10.2196/jmir.6971] [Medline: 28360025]

37. van de Steeg L, IJkema R, Wagner C, Langelaan M. The effect of an e-learning course on nursing staff's knowledge of delirium: a before-and-after study. BMC Med Educ 2015;15:12 [FREE Full text] [doi: 10.1186/s12909-015-0289-2] [Medline: 25653115]

38. Waldorff FB, Steenstrup AP, Nielsen B, Rubak J, Bro F. Diffusion of an e-learning programme among Danish general practitioners: a nation-wide prospective survey. BMC Fam Pract 2008 Apr 25;9:24 [FREE Full text] [doi: 10.1186/1471-2296-9-24] [Medline: 18439279]

39. Jäger C, Steinhäuser J, Freund T, Baker R, Agarwal S, Godycki-Cwirko M, et al. Process evaluation of five tailored programs to improve the implementation of evidence-based recommendations for chronic conditions in primary care. Implement Sci 2016 Sep 13;11(1):123 [FREE Full text] [doi: 10.1186/s13012-016-0473-8] [Medline: 27624776]

XSL•FO

RenderX

40. Sarna L, Bialous SA, Wells M, Brook J. Impact of a webcast on nurses' delivery of tobacco dependence treatment. J Clin Nurs 2017 May 11;27(1-2):e91-e99. [doi: 10.1111/jocn.13875] [Medline: 28493616]

41. Ketelaar SM, Nieuwenhuijsen K, Gärtner FR, Bolier L, Smeets O, Sluiter JK. Mental Vitality @ Work: The effectiveness of a mental module for workers' health surveillance for nurses and allied health professionals, comparing two approaches in a cluster-randomised controlled trial. Int Arch Occup Environ Health 2014 Jul;87(5):527-538. [doi: 10.1007/s00420-013-0893-6] [Medline: 23893252]

42. Murphy KR, Davidshofer CO. Psychological Testing: Principles and Applications 6th ed. Upper Saddle River, NJ: Pearson/Prentice Hall; 2005.

43. Edwards C, Staniszweska S, Crichton N. Investigation of the ways in which patients' reports of their satisfaction with healthcare are constructed. Sociol Health Illn 2004 Mar;26(2):159-183 [FREE Full text] [doi: 10.1111/j.1467-9566.2004.00385.x] [Medline: 15027983]

44. Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Methods 2002 Jun;7(2):147-177. [Medline: 12090408]

45. Brandes K, van Weert JC. Implementing consultation audio-recordings and question prompt lists into routine cancer care: how can we address healthcare providers' barriers? Patient Educ Couns 2017 Jun;100(6):1029-1030. [doi: 10.1016/j.pec.2017.04.002] [Medline: 28455108]

46. Henry SG, Penner LA, Eggly S. Associations between thin slice ratings of affect and rapport and perceived patient-centeredness in primary care: comparison of audio and video recordings. Patient Educ Couns 2017 Jun;100(6):1128-1135. [doi: 10.1016/j.pec.2016.12.020] [Medline: 28087212]

47. Lorencatto F, West R, Christopherson C, Michie S. Assessing fidelity of delivery of smoking cessation behavioural support in practice. Implement Sci 2013 Apr 04;8:40 [FREE Full text] [doi: 10.1186/1748-5908-8-40] [Medline: 23557119]

48. van Bruinessen IR, Leegwater B, van Dulmen S. When patients take the initiative to audio-record a clinical consultation. Patient Educ Couns 2017 Aug;100(8):1552-1557. [doi: 10.1016/j.pec.2017.03.001] [Medline: 28302340]

49. Moloczij N, Krishnasamy M, Butow P, Hack TF, Stafford L, Jefford M, et al. Barriers and facilitators to the implementation of audio-recordings and question prompt lists in cancer care consultations: a qualitative study. Patient Educ Couns 2017 Jun;100(6):1083-1091 [FREE Full text] [doi: 10.1016/j.pec.2017.01.005] [Medline: 28117193]

50. Rutten GM, Degen S, Hendriks EJ, Braspenning JC, Harting J, Oostendorp RA. Adherence to clinical practice guidelines for low back pain in physical therapy: do patients benefit? Phys Ther 2010 Aug;90(8):1111-1122. [doi: 10.2522/ptj.20090173] [Medline: 20488978]

51. de Beurs DP, de Groot MH, de Keijser J, Mokkenstorm J, van Duijn E, de Winter RF, et al. The effect of an e-learning supported Train-the-Trainer programme on implementation of suicide guidelines in mental health care. J Affect Disord 2015 Apr 01;175:446-453 [FREE Full text] [doi: 10.1016/j.jad.2015.01.046] [Medline: 25679199]

Abbreviations

CT: computer-tailored
e-learning: electronic learning
GP: general practitioner
HCP: health care professional
ICM: I-Change Model
OR: odds ratio
PN: practice nurse
RCT: randomized controlled trial

XSL•FO
RenderX

XSL•FO

RenderX

<u>Original Paper</u>

Evaluation of App-Based Serious Gaming as a Training Method in Teaching Chest Tube Insertion to Medical Students: Randomized Controlled Trial

Patrick Haubruck[1*], MD; Felix Nickel[2*], MD; Julian Ober[1], BS; Tilman Walker[1], MD; Christian Bergdolt[1], MD; Mirco Friedrich[2], BS; Beat Peter Müller-Stich[2], MD; Franziska Forchheim[1], MD; Christian Fischer[1], MD; Gerhard Schmidmaier[1], MD; Michael C Tanner[1], MD

[1]Heidelberg Trauma Research Group, Center for Orthopedics, Trauma Surgery and Spinal Cord Injury, Trauma and Reconstructive Surgery, Heidelberg University Hospital, Heidelberg, Germany

[2]Department of General, Visceral and Transplantation Surgery, Heidelberg University Hospital, Heidelberg, Germany

[*]these authors contributed equally

Corresponding Author:
Patrick Haubruck, MD
Heidelberg Trauma Research Group
Center for Orthopedics, Trauma Surgery and Spinal Cord Injury, Trauma and Reconstructive Surgery
Heidelberg University Hospital
Schlierbacher Landstrasse 200 a
Heidelberg, 69121
Germany
Phone: 49 06221 ext 5635987
Fax: 49 62215626300
Email: patrick.haubruck@med.uni-heidelberg.de

Abstract

Background: The insertion of a chest tube should be as quick and accurate as possible to maximize the benefit and minimize possible complications for the patient. Therefore, comprehensive training and assessment before an emergency situation are essential for proficiency in chest tube insertion. Serious games have become more prevalent in surgical training because they enable students to study and train a procedure independently, and errors made have no effect on patients. However, up-to-date evidence regarding the effect of serious games on performance in procedures in emergency medicine remains scarce.

Objective: The aim of this study was to investigate the serious gaming approach in teaching medical students an emergency procedure (chest tube insertion) using the app Touch Surgery and a modified objective structural assessment of technical skills (OSATS).

Methods: In a prospective, rater-blinded, randomized controlled trial, medical students were randomized into two groups: intervention group or control group. Touch Surgery has been established as an innovative and cost-free app for mobile devices. The fully automatic software enables users to train medical procedures and afterwards self-assess their training effort. The module chest tube insertion teaches each key step in the insertion of a chest tube and enables users the meticulous application of a chest tube. In contrast, the module "Thoracocentesis" discusses a basic thoracocentesis. All students attended a lecture regarding chest tube insertion (regular curriculum) and afterwards received a Touch Surgery training lesson: intervention group used the module *chest tube insertion* and the control group used *Thoracocentesis* as control training. Participants' performance in chest tube insertion on a porcine model was rated on-site via blinded face-to-face rating and via video recordings using a modified OSATS tool. Afterwards, every participant received an individual questionnaire for self-evaluation. Here, trainees gave information about their individual training level, as well as previous experiences, gender, and hobbies. Primary end point was operative performance during chest tube insertion by direct observance.

Results: A total of 183 students enrolled, 116 students participated (63.4%), and 21 were excluded because of previous experiences in chest tube insertion. Students were randomized to the intervention group (49/95, 52%) and control group (46/95, 48%). The intervention group performed significantly better than the control group (Intervention group: 38.0 [I_{50}=7.0] points; control group:

XSL•FO
RenderX

30.5 [I_{50}=8.0] points; *P*<.001). The intervention group showed significantly improved economy of time and motion (*P*=.004), needed significantly less help (*P*<.001), and was more confident in handling of instruments (*P*<.001) than the control group.

Conclusions: The results from this study show that serious games are a valid and effective tool in education of operative performance in chest tube insertion. We believe that serious games should be implemented in the surgical curriculum, as well as residency programs, in addition to traditional learning methods.

Trial   Registration: German   Clinical   Trials   Register   (DRKS)   DRKS00009994; https://www.drks.de/drks_web/navigate.do?navigationId=trial.HTML&TRIAL_ID=DRKS00009994 (Archived by Webcite at http://www.webcitation.org/6ytWF1CWg)

Introduction

Clinical Background

No matter what the underlying cause, to increase the benefit for the patient, the insertion of a chest tube should be as quick and accurate as possible [1]. Therefore, profound anatomical knowledge, meticulous positioning, and standardized execution are crucial for the success of this therapy [1]. Due to the proximity of thoracic and abdominal organs, incorrect insertion of a chest tube might lead to potentially lethal complications [2]. Therefore, comprehensive training and assessment before an emergency situation are essential in acquiring proficiency in chest tube insertion (CTI) [1]. Over the last years, the education of junior doctors and medical students has become more diverse, voluminous, and challenging. Therefore, there is an increasing need for comprehensive, objective, and resource-sparing educational concepts that ensure a high-quality education of standardized procedures without impairing patient safety [3]. In particular, early and decisive training using an educational tool might standardize execution of CTI and ultimately improve emergency care of traumatized patients and patients' outcome [1].

Current Training Methods

Training methods for CTI cover theoretical instructions, the use of animal models, and teaching with real patients supervised by an experienced surgeon. However, these established training methods have significant limitations. First, this type of education is limited by an increasing shortage of manpower in hospitals [4,5], and second, teaching students and inexperienced doctors with the help of real patients is not always possible with regards to patient safety. Third, not all facilities have the capability of employing animal models. Additionally, these methods show limited objectiveness in the assessment of learning success, resulting in an inaccurate educational quality [6,7]. Therefore, novel training methods are needed that enable students and inexperienced doctors to study a procedure realistically and furthermore, self-assess their performance without the need of a supervising surgeon. Additionally, future training methods should aim at providing highest levels of education without endangering patient safety.

Serious Games for Teaching Medical Students

Serious games have become more prevalent in surgical training of physicians and medical students because of their ability to increase intrinsic motivation [8,9]. They enable students to study and train a procedure without need for a supervising surgeon. In addition, errors that are made during training with serious games have no effect on patients, and patient safety might be improved. Thus, because of the omnipresence of computer games in student's lives, the focus in educational research has shifted towards serious games in context with medical training [9-12].

In a systematic review by Graafland et al published in 2012 [13], the authors stated that simulation and serious gaming represent ideal teaching methods for optimization of residents' knowledge and skill before they are entrusted with procedures performed on real patients [13]. They also concluded that serious games may be used to train everyday clinical abilities such as surgical procedures [13]. In a consecutive study by Graafland et al [14], the authors examined the rate of acceptance among surgical educators and trainees regarding serious games as a training method. They found that serious games were viewed as positive by 78% of the participants, and 66% of the participants would play the game in their leisure time [14].

Defining serious games remains challenging as several classifications exist [15]. The first formal definition of the concept was provided by Abt in 1970, where he presented simulations and games to improve education. Since then, multiple classifications, both market- and purpose-based, have been postulated, and in 2002, Sawyer redefined the definition of serious games based on his idea of connecting a serious purpose to knowledge and technology from the video game industry [15]. In 2011, Djaouti et al [15] established a novel and comprehensive classification that combined the analysis of both "serious" and "game" dimensions: the gameplay/purpose/scope model [15]. Applying the classification onto the training app Touch Surgery (TS; G: type: game-based; Goals: match; Means: move, select / P: educative message broadcasting, mental training / S: market: health care; Target audience: medical students, physicians) identifies it as a serious game [15].

Validation of TS has been performed before for various specialist fields. Sugand et al (2015) published results regarding

the use of TS in intramedullary femoral nailing [16]. They demonstrated construct, content, and face validity of the intramedullary nailing module. Additionally, the authors stated that TS could be used to allow orthopedic trainees to learn operative steps and be subsequently tested before and after surgery [16]. TS was also evaluated positively by a study of Paro et al [17], who used TS in the context of open carpal tunnel release. In their study, they were also able to show positive results regarding construct, face, content, and acceptability validity of TS [17]. Moreover, TS has been validated for cognitive training and assessment of laparoscopic cholecystectomy [8,9], as well as for orbital floor reconstruction [18].

Objective

Up-to-date evidence regarding the effect of serious games on performance in procedures in trauma surgery and emergency medicine remains scarce. In this study, we sought to determine the influence of serious gaming on surgical training in CTI.

Methods

Study Design

The study was conducted from April 2017 to July 2017 at the Center for Orthopedics, Trauma Surgery and Spinal Cord Injury, Trauma and Reconstructive Surgery at the University of Heidelberg. We implemented the validated surgical training app TS (Kinosis Ltd, London, United Kingdom) [14] in a standardized and structured setting into medical student education to complement existing training methods and reduce necessary training staff and resources. In addition, we used an objective structural assessment of technical skills (OSATS) to advance and standardize training. The current OSATS tool for CTI was developed based on key steps of correct CTI that were modified and revised by a team of trauma and general surgeons [1]. This study was designed as a prospective, single-center, rater-blinded, two-arm, parallel group randomized controlled trial (RCT), and the study protocol was published in 2017 [1]. The trial is reported in accordance with Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth. The study was executed in concordance with the Declaration of Helsinki. Before the enrollment of participating students, approval was received from the local ethics committee of the University of Heidelberg (S-174/2016). In addition, the study was registered at the German Clinical Trials Register (DRKS) before commencement of the study (DRKS0009994). No changes to the trial design were performed after the commencement of the trial.

Randomization

To minimize selection bias, a blocked randomization, stratified by gender, was utilized to randomly assign (1:1 ratio) participants into either an intervention group or a control group. Randomization was performed at the beginning of day 1 before the lecture and theoretical introduction. An independent employee using sealed opaque envelopes performed randomization [1]. The same employee assessed whether students complied with the study instructions before

assessment—otherwise this employee was not involved in any other aspects of the study.

Primary End Points

The primary end point of this study was the operative performance during CTI based on the standardized and evaluated OSATS scoring tool as measured by direct observation of a blinded rater during the course [1].

Secondary End Points

Secondary end points of this study were the operative performance during CTI based on the standardized and evaluated OSATS scoring tool as measured by indirect video observation of two independent and blinded raters, as well as combined ratings of indirect and direct raters. Additionally, differences between on-site and video ratings were evaluated. Furthermore, analysis of the correlation of extracurricular activities and hobbies with operative performance was performed, as well as a time-dependent performance analysis. Another secondary end point was the subgroup analysis regarding gender-dependent differences in the operative performance [1]. There were no changes to the primary and secondary trial outcomes after commencement of the study.

Sample Size Determination

To detect differences with a significance level Cronbach alpha=.05 and a power of 1-β=0.8, a group size of N=45 was determined in a sample size determination performed before this study [1].

Statistical Analysis

Before the statistical analysis, all data were completely anonymized. Statistical analyses were carried out by SPSS statistics version 24.0 (IBM Corp). For analysis of nonparametric, nonrelated data, the Mann-Whitney U test was carried out. Correlation analysis for gender influences, as well as influence of leisure activities on operative performance, was calculated via Spearman correlation coefficient. Intraclass correlation coefficient for interrater reliability (IRR) between the "indirect" raters was measured via two-way random absolute agreement intraclass correlation analysis. For all tests, a P value less than .05 was considered statistically significant. Data is expressed as median values (xMed) and interquartile ranges (I_{50}). The presentation of the results is done via box and whisker plots.

Participants

According to our inclusion criteria, only medical students enrolled at the medical faculty of the University of Heidelberg during their clinical years (3rd-6th year) who reached the age of 18 years were included in the study. Participants having previous clinical experience and practice regarding CTI were excluded from this study. Participation in the study was offered as a voluntary training opportunity for CTI in context with the regular surgical curriculum. All participants had access to the app during the study because of provision of iPads with preloaded app and module. Each participating student received information about the study before participation. Furthermore, informed consent for anonymous data collection, as well as

anonymous recording of videos during the training sessions, was obtained for each participant.

Materials

The Utilized Serious Games

TS [19] has been established as an innovative and cost-free app for mobile devices [8,16] and can be downloaded from Google Play and iPhone operating system (iOS, Apple Inc) stores. The software enables users to train medical procedures in a rendered three-dimensional environment and then guides users through every stage of each procedure using touchscreen motion gestures. Hereafter, users can self-assess their training effort via active rehearsal of the steps of the procedure [20]. All participants used version 4.14.6, and there were no changes in content of the modules used during the course of the study. Currently, procedures in TS are divided into modules (over 100 procedures can be trained), and users learn procedure-specific steps such as patient positioning and access to the operating field [8]. Hereafter, in training mode, the app leads the user through each relevant step of the operation, instrument selection, and application in the specific operative procedure. Afterwards, users can switch to the self-assessment part of the software, and each step is assessed by multiple-choice questions training "cognitive decision making" [8]. The module "Chest tube insertion" was developed by Rafael J. Grossmann, an attending surgeon at Eastern Maine Medical Center, and consists of visualization of the safe zone for insertion, correct handling, and right sequence of instruments used, as well as positioning of the chest tube (Figure 1).

Therefore, the module teaches each key step in the insertion of a chest tube and enables users the meticulous application of a chest tube. In contrast, the module "Thoracocentesis" was developed by Shannon Toohey, a clinical instructor in the Emergency Department at UC Irvine, and discusses a basic thoracocentesis. The key steps: confirmation of pleural effusion using chest X-ray, anesthesia with local anesthetic, and pleural effusion aspiration are taught, and the handling of the used instruments is visualized (Figure 2).

This module was chosen because of its similarities in presentation, while allowing enough differentiation of the key aspects of OSATS score without risk of obfuscation. In addition, because of both a different approach and a different set of instruments, as well as a profoundly different procedure, both modules differ considerably, and thoracocentesis can be used as a control procedure without confounding the results of the study.

The Training

Training was conducted in context with the regular surgical curriculum. Participation in the study was voluntary. Subsequent to randomization, all participants received structured instructions regarding their respective training curriculum. In particular, participants of the control group were encouraged to further self-study using literature and available books, whereas participants of the intervention group were advised not to do so. Hereafter, all participants received the standardized theoretical training regarding CTI as part of a theoretical lecture by an experienced surgeon [1]. Participants were given access to the script of the lecture and information on where to find further material for self-studying CTI (eg, books and e-learning material). At the end of the lecture, the teaching surgeon instructed participants to further self-study the topic of the lecture as per initially delivered instructions. Thereby, participants of the control group were instructed to self-study CTI (topic of lecture), whereas participants of the intervention group received initial information not to do so regardless of instructions given at the lecture. Therefore, training in the control group was equivalent to the standard surgical education regarding CTI in our institution. Hereafter, students were introduced into the app-based serious games and the handling of TS [1]. On the afternoon of day 1, training with the app-based serious games was conducted in a training lesson lasting 120 min, supervised by experienced surgeons not involved in the randomization or analysis of study results. While participants of the intervention group used the module "Chest Drain Insertion" (Figure 1) for training of CTI, the participants of the control group used the module "Thoracocentesis" (Figure 2). Training, regardless of group, was conducted until participants reached an overall app-based score of 95% on performance of the assigned module. This was intended to assure trainees had performed the module while leaving enough margin for differentiation of scores during OSATS rating. Otherwise, participants in the control group attended every introduction and assessment included in the training sequence. A supervising experienced surgeon monitored performance of the assigned training modules to ensure adherence to each group-specific protocol [1]. Therefore, the intervention period was short, reducing the attrition bias [21]. Hereafter, participants had the rest of the day free for self-study, if applicable, as per initial instructions. Thereby, enough time was given to the participants to work through the provided material and further self-study CTI (Figure 3).

The Assessment

The OSATS for CTI was developed based on the CTI Scoring System by Hutton et al [22]. In particular, the OSATS was based on key steps of correct CTI that were modified and revised by a team of trauma and general surgeons [1,22]. As shown in Figure 4, the score consists of 10 key steps. Each key step is scored from 1 (worst) to 5 (best), based on a 5-point Likert scale [1]. The maximum possible score was 50 points in total, the minimal score was 10 points (Figure 4).

Figure 1. Screenshots of the Touch Surgery "Chest Tube Insertion" module. Panels A to D visualize the different key steps of the module. A: anatomical location of the safe surgical approach; B: illustration of the correct subcutaneous preparation; C: handling of instruments and tube; D: radiological control using a chest X-ray.

Figure 2. Screenshots of the Touch Surgery "Thoracocentesis" module. Panels A to D visualize the different key steps of the module. A: anatomical location of approach; B: correct administration of local anesthetics; C: aspiration of fluid; D: radiological control using a chest X-ray.

Figure 3. Visualization of the time schedule of the training and assessment part of this study. Randomization was performed before the first study-related interventions, and group-related instructions were given to individual participants. Introduction of both groups via a lecture occurred on the first day between 09:00 and 11:00 AM. In the afternoon (2:00-4:00 PM) of the same day, training using the app Touch Surgery was conducted under supervision of an experienced trauma surgeon. Afterwards, the rest of the day was free to provide enough time to self-study as instructed. At the beginning of day 2, the unblinded employee in charge of randomization assessed whether participants complied with the given instructions. Afterwards, operative performance was assessed utilizing the porcine model via a blinded on-site rater. CTI: chest tube insertion.
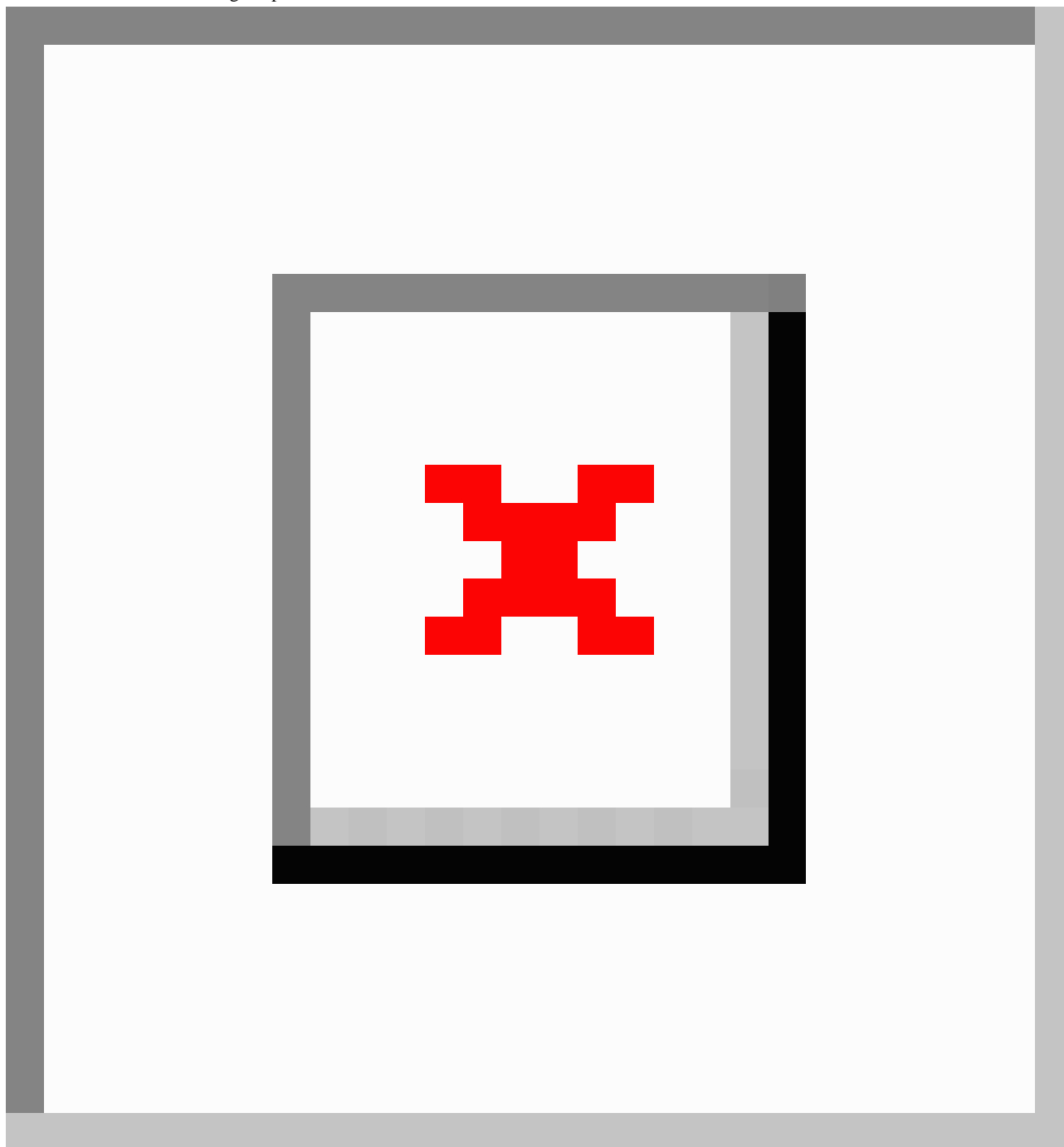
Before assessment, the intervention of interest was purposely not clarified, and participants were solemnly instructed to comply with their initially delivered instructions. At the beginning of the second day, participants were asked if they complied with their respective instructions (Figure 3). In particular, participants of the control group were asked if they had adequately studied CTI, and participants of the intervention group were asked whether they had not conducted further self-study. Participants that failed to comply with the instructions would have been excluded from the study (all participants complied with the instructions in our study). During assessment session, participants performed a CTI on a previously prepared porcine model [1]. A blinded on-site rater evaluated the performance of participants face-to-face using the modified OSATS tool (Figure 4) for CTI. Performance was videorecorded, showing only the porcine model and the hands of the participants. Furthermore, the on-site rater supervised the use of instruments and use of personal safety equipment, thereby guaranteeing participants' safety during the course of the study. Afterwards, two independent blinded video raters performed a

blinded, video-based evaluation using the same scoring tool [1]. All raters were experts from the Center for Orthopedics, Trauma and Reconstructive Surgery and Spinal Cord Injury, Heidelberg University. Afterwards, every participant received a personal questionnaire for self-evaluation. Herein, trainees gave information about their individual training level, as well as previous experiences, gender, and hobbies (see Multimedia Appendices 1 and 2) [1].

Figure 4. Visualization of the modified objective structural assessment of technical skills (OSATS) score for chest tube insertion (CTI).

	1 poor	2	3 sufficient	4	5 excellent
Correct identification of incision location	the chosen dissection plane deviates tremendously from the suggested site		the chosen dissection plane deviates slightly from the suggested site		4th or 5th intercostal space; mid axillary line
Correct plane of dissection subcutaneously	both distance or execution of tunneling lack accuracy		either distance or execution of tunneling lack accuracy		both distance and execution of tunneling are accurate
Blunt dissection on top side of rib	flawed dissection; not carried out on top side of rib		solid dissection carried out with minor errors		confident cut through the subcutaneous layers and intercostal muscles
Scissors or Clamp guarded with other hand during dissection and pulled out without closing the instrument	hazardous handling that might affect the patient		improvable handling		confident handling of the used instruments
Digital exploration of pleural cavity on chest wall to rule out adhesions	no digital exploration		finger inserted in pleural cavity		digital exploration in 360° with turning of the wrist rules out adhesions
Drain guarded with hand while being inserted	hazardous handling that might affect the patient		improvable handling		confident handling of the used instruments
Drain inserted into pleural cavity	tube advancement is carried out poorly.		tube advancement is carried out with minor errors		forceps unclamped in time and tube manually advanced.
Estimate made of drain length	estimate deviates tremendously from rater's opinion		estimate deviates slightly from rater's opinion		optimal estimate stated
Economy of time and motion	Many unnecessary or disorganized movements		Organized time and motion, some unnecessary movement		Maximum economy of movement and efficiency
Amount of help and assistance needed from tutor	Task couldn't be carried out without extensive assistance		Trainee only raises important questions in order to maximize performance		almost no assistance needed; task is carried out confidently

Results

Participants

From April 2017 to July 2017, 183 students enrolled in the emergency medicine course of our hospital, of which 116 students participated (63.4%) in this study; 21 participants had to be excluded from the analysis because of previous experiences in CTI (Figure 5).

Students were randomly assigned to the intervention group (49/95, 52%) and control group (46/95, 48%). All students participated in the clinical part of the medical curriculum (Table 1), and most students (84%, 80/95) were in their 6th semester. For the intervention group, the median age of participants was 22.0 years ($I_{50}=1.0$), and 33 participants were female (67%, 33/49). In the control group, the median age of participants was also 22.0 years ($I_{50}=3.0$), and 27 participants were female (59%, 27/46). Further data regarding the demographics of participants are depicted in Table 1.

Figure 5. Study flowchart based on Consolidated Standards of Reporting Trials (CONSORT) guidelines.
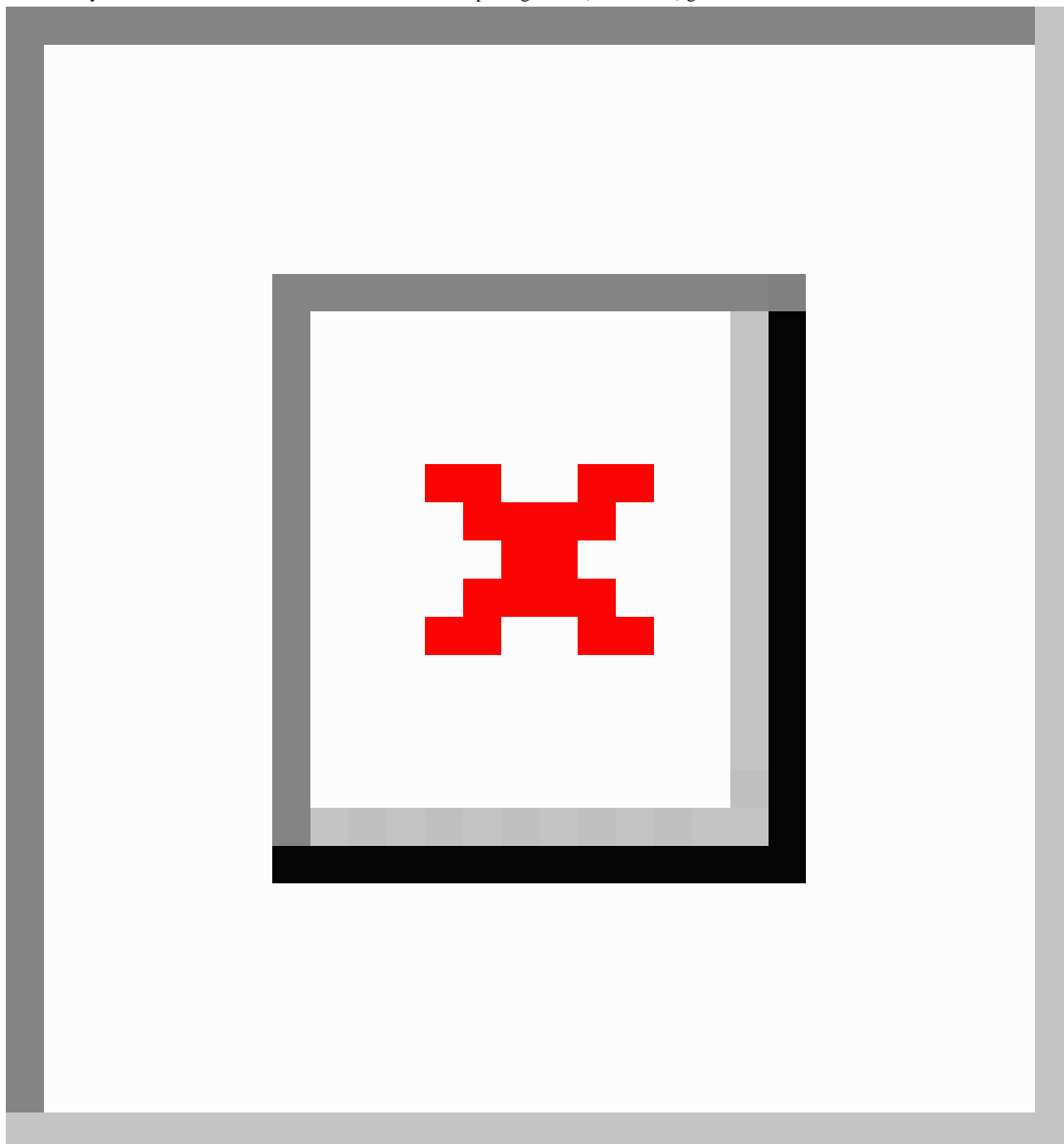
Table 1. Participants' demographics.

Characteristics	Intervention group (N=49)	Control group (N=46)
Sex, n (%)		
Male	16 (33)	19 (41)
Female	33 (64)	27 (59)
Age (years), median (I_{50})	22.0 (1.0)	22.0 (3.0)
Male	22.0 (2.0)	22.0 (3.0)
Female	22.0 (1.0)	22.0 (2.0)
Leisure activity, n (%)		
Playing video games (total)	38 (78)	38 (83)
Playing an instrument (total)	31 (63)	34 (74)
Regular sportive activity (total)	43 (88)	39 (85)
Previous experience in handicraft work (total)	24 (49)	31 (67)
Level of education (semester), n (%)		
5	1 (2)	0 (0)
6	41 (84)	39 (85)
7	0 (0)	1 (2)
8	7 (14)	5 (11)
10	0 (0)	1 (2)

Primary End Point

Operative Performance in Direct Observation

The primary end point of this study was the influence of serious games (TS) on the operative performance in CTI evaluated by the OSATS scoring tool measured via direct observation of a blinded rater during the course. Analysis of the data revealed that participants of the intervention group performed significantly better than participants of the control group in general (intervention group: 38.0 [I_{50}=7.0] points; control group: 30.5 [I_{50}=8.0] points; *P*<.001; Figure 6), as well as in each key step besides the "Correct plane of dissection subcutaneously."

Participants of the intervention group showed a significantly improved economy of time and motion (intervention group: 4.0 [I_{50}=1.0] vs control group: 3.0 [I_{50}=1.0]; *P*=.004) and needed significantly less help from the supervising surgeon (intervention group: 4.0 [I_{50}=1.0] vs control group: 2.0 [I_{50}=1.0]; *P*<.001). Furthermore, participants of the intervention group were more confident in the handling of the required instruments (intervention group: 3.0 [I_{50}=2.0] vs control group: 3.0 [I_{50}=2.0]; *P*<.001) than participants of the control group, and the digital exploration of the pleural cavity was performed significantly better (intervention group: 4.0 [I_{50}=2.0] vs control group: 2.0 [I_{50}=2.0]; *P*<.001; Table 2). Median time of performing a CTI was 4:15 min in the intervention group. Participants of the control group needed a median of 4:17 min.

Figure 6. Outcome of operative performance. A: box plot showing points reached in direct objective structural assessment of technical skills (OSATS) by intervention and control group. B: box plot showing points reached in "indirect" rating by intervention and control group. C: box plot showing points reached in average of all three ratings by intervention and control group. ConGr=control group and IntGr=intervention group.
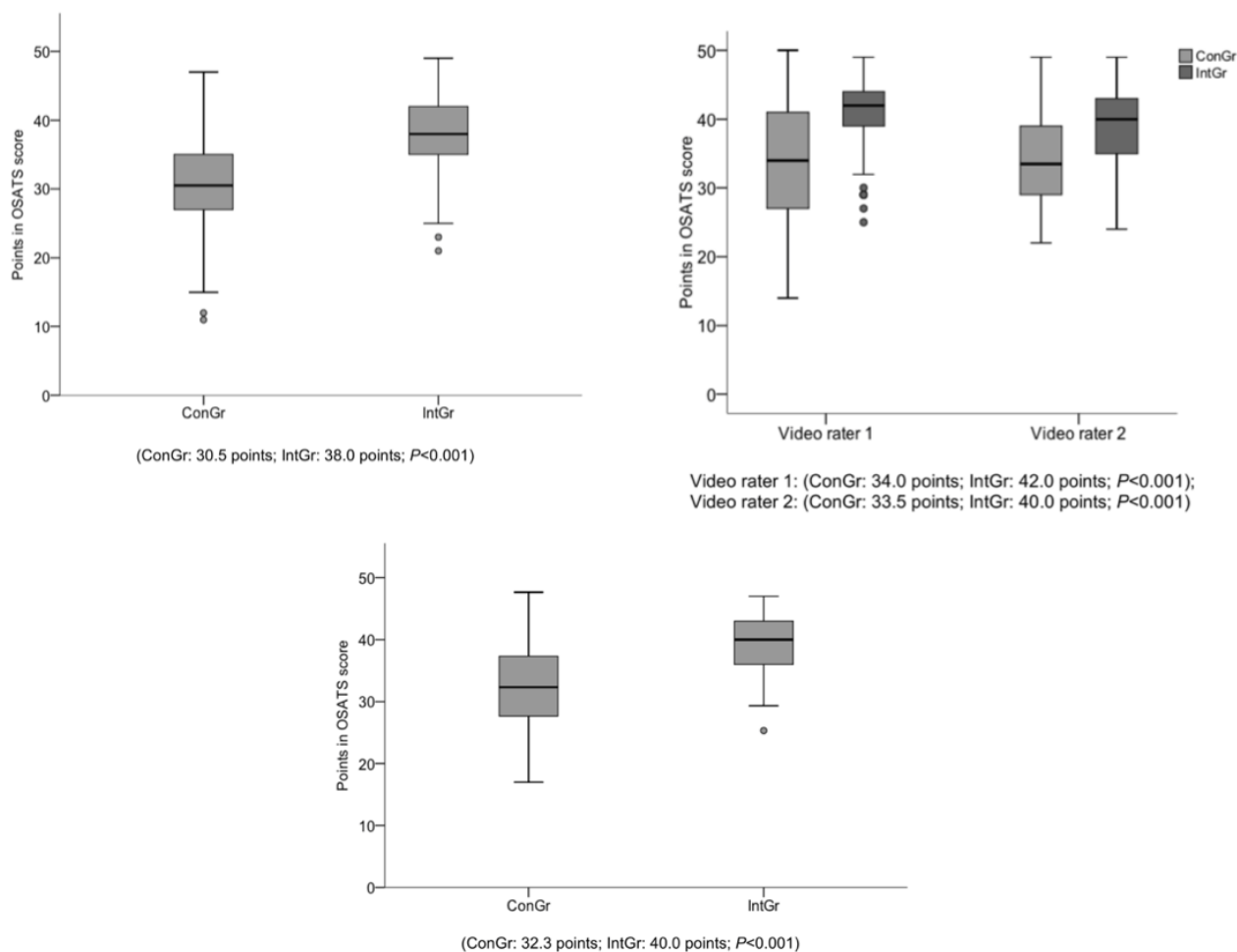
Table 2. Results of an objective structural assessment of technical skills (OSATS) subgroup analyses for direct rating.

Step of chest tube insertion	Intervention group, xMed (I_{50})	Control group, xMed (I_{50})	P value
Correct identification of incision location	5.0 (0.0)	5. (1.0)	.01[a]
Correct plane of dissection subcutaneously	4.0 (1.0)	3.0 (1.0)	.06
Blunt dissection on top side of rib	5.0 (1.0)	4.0 (2.0)	.02[a]
Scissors or clamp guarded with other hand during dissection and pulled out without closing the instrument	3.0 (2.0)	2.0 (1.0)	.03[a]
Digital exploration of pleural cavity on chest wall to rule out adhesions	4.0 (2.0)	2.0 (2.0)	<.001[a]
Drain guarded with hand while being inserted	4.0 (2.0)	3.0 (2.0)	<.001[a]
Drain inserted into pleural cavity	4.0 (1.0)	3.0 (1.0)	<.001[a]
Estimate made of drain length	5.0 (2.0)	3.0 (4.0)	<.001[a]
Economy of time and motion	4.0 (1.0)	3.0 (1.0)	.004[a]
Amount of help or assistance needed from tutor	4.0 (1.0)	3.0 (1.0)	<.001[a]
Total	38.0 (7.0)	30.5 (8.0)	<.001[a]

[a]A P value less than .05 is considered as statistically significant.

Secondary End Points

Operative Performance in Indirect Observation

Operative performance based on indirect video observation revealed significant differences between groups. In particular, participants of the intervention group performed significantly better than participants of the control group regardless of the video rater (rater 1: intervention group: 42.0 [I_{50}=7.0] vs control group: 34.0 [I_{50}=14.0]; P<.001; rater 2: intervention group: 40.0 [I_{50}=9.0] vs control group: 33.5 [I_{50}=10.0]; P<.001; Figure 6). Interestingly, participants of the intervention group were significantly better in the handling of the scissors and clamps during direct observation compared with the control group (intervention group: 3.0 [I_{50}=2.0] vs control group: 2.0 [I_{50}=1.0]; P=.03), whereas indirect observation revealed better tendencies without statistical significance (intervention group: rater 1: 4.0 [I_{50}=1.0], rater 2: 4.0 [I_{50}=1.0] vs control group: rater 1: 3.0 [I_{50}=2.0], rater 2: 3.0 [I_{50}=1.0]; rater 1: P=.098, rater 2: P=.38). Overall, indirect ratings revealed smaller differences between groups than the direct rating. The analysis of the IRR revealed an excellent correlation between the results from both indirect raters (correlation index: 0.929 [95% CI 0.894-0.953]) [23]. Furthermore, analysis between direct and indirect rating revealed a good correlation as well (correlation index: 0.723 [95% CI 0.423-0.858]). In addition, combined ratings of direct and indirect observation confirmed statistically significant differences (intervention group: 40.0 [I_{50}=7.2] vs control group: 32.3 [I_{50}=10.1]; P<.001; Figure 6).

Influence of Gender and Hobbies on Operative Performance

Gender of participants did not correlate with operative performance during CTI in the total study collective. Regular sportive activity had a significant positive correlation with the operative performance during CTI regardless of prior training (Spearman index: .214; P=.04). However, positive correlation of regular sportive activity was higher in the control group compared with the intervention group (Table 3). Furthermore, previous experience in handicraft work had no correlation with operative performance in the intervention group (Pearson index: −.071; P=.63), whereas within the control group, participants that had previous experience in handicraft work performed significantly better than participants that did not (Pearson index: .353; P=.02; Table 3). Somewhat surprisingly, the data from this study showed that playing video games had no correlation with operative performance during CTI in the total study collective (Spearman index: −.007; P>.95; Table 3).

Evaluation of Touch Surgery as a Serious Gaming Device by Participants

Participants were asked to evaluate TS as a serious gaming device after performing CTI. Participants rated TS as an efficient training device, and when asked if they would continue to train with TS, the majority responded positively. In addition, participants were asked to evaluate both serious gaming and traditional learning methods (lecture) regarding the training benefit, level of simulation regarding the reality, and benefit regarding the handling of operative situations. Analysis of our data revealed that participants scored serious games better than traditional learning methods in all aspects regardless of the module that was trained (Table 4).

Table 3. Correlation regarding the influence of leisure activities on the operative performance.

Type of leisure activity	Total study collective		Intervention group		Control group	
	Spearman index	P value	Spearman index	P value	Spearman index	P value
Regular sportive activity	.214	.038[a]	.129	.379	.287	.05
Previous experience in handicraft work	.055	.599	−.071	.631	.353	.02[a]
Playing video games	−.007	.949	−.073	.621	−.045	.78

[a]A P value less than .05 is considered as statistically significant.

Table 4. Participants' evaluation of training.

Factors of evaluation	Lecture, xMed (I_{50})	Training with Touch Surgery, xMed (I_{50})	P value
Training benefit	3.0 (2.0)	2.0 (1.0)	<.001[a]
Level of simulation regarding the reality	4.0 (2.0)	2.0 (1.0)	<.001[a]
Benefit regarding the handling of operative situations	4.0 (3.0)	3.0 (2.0)	<.001[a]

[a]A P value less than.05 is considered as statistically significant.

Discussion

Principal Findings

In this study, we sought to determine the influence of serious games on surgical training of CTI. We utilized the validated surgical training app TS in addition to a modified OSATS in an RCT. The results from this study indicated that serious gaming using a relevant content might be superior to traditional teaching methods including training with serious games with an irrelevant content regarding the operative performance in CTI. Participants from this study evaluated TS as an efficient

and motivating training device. Furthermore, most participants wanted to continue the training with TS.

The results from this study support previous findings [24-26]. Participants that trained with TS showed an improved operative performance when placing a chest tube both in direct and indirect evaluation by blinded raters using the OSATS scoring tool. Participants of the intervention group performed significantly better in key steps regarding patient safety. In particular, participants of the intervention group were more accurate in the digital exploration of the pleural cavity and needed significantly less help from the supervising surgeon. Therefore, training with TS might help in enabling students to perform surgical procedures with a higher degree of independency.

Participants of the intervention group performed significantly better in the handling of the scissors and clamps. This is noteworthy because of the fact that TS offers merely gesture-based controls of the instruments used in the respective module of the app and only visualization of their correct use. Differences in the handling of instruments might be caused by differences in mental practice. Mental practice utilizes the systemic use of mental imagery to rehearse an action before the actual performance without the need for physical movement [27,28]. Mental practice has been validated as an efficient training method in teaching laparoscopic surgery [27], as well as in learning basic surgical skills [29]. In training mode, TS instructs and visualizes the correct use of instruments regarding the necessary action, whereas in test mode, the user needs to utilize gesture-based controls to initiate the use of instruments. Therefore, the user needs to visualize the necessary action using the required instrument before each respective task of the procedure. This visualization process might induce mental practice [8]. Participants of the control group trained with a different set of instruments, and therefore, visualization of the correct use of the required instruments for CTI was missing. Thus, better performance in the use of the actual instruments might be because of stimulation of mental practice by use of TS.

In addition, mental practice has been successfully implemented in teaching and rehearsing complex psychomotor tasks in several domains such as sports or music [28]. Our data indicated that sports activity positively correlates with operative performance. However, this correlation was more distinct in the control group. Participants that exercise often might show better performance in mental practice and therefore benefit more from a theoretical lecture, whereas use of TS with relevant content might induce mental practice regardless of previous experiences in sports activities. Therefore, positive effects of sports activity regarding mental practice might be reduced in the intervention group because of induction of mental practice through the use of TS with relevant content and might explain the higher correlation of sports activity and operative performance in the control group.

The influence of gender on surgical skill acquisition is not clear yet [30]. Although studies of Schueneman et al [31] and Madan et al [32] found some differences between males and females in surgical skill acquisition [33], the findings of Kolozsvari et al [34], Grantcharov et al [35], Nickel et al [36], and Kowaleski

et al [37] support the results of our study that no correlation between operative performance and gender of participants could be found.

In our literature review, we found no evidence for the influence of handicraft on surgical performance. According to our data, there was no correlation between trainees' performance and experience in handicraft work for the intervention group, but we interestingly found a significant correlation for participants of the control group. On the basis of these results, the influence of experience in handicraft work on trainees' performance is not clear yet. Further investigation of this question to prove or disprove the hypothesis that experience in handicraft work could influence surgical performance is therefore needed.

Finally, we examined the correlation between the operative performance of the trainees and playing video games. Influence of video gaming on surgical skills has been described in various studies [38-40]. Rosser et al [41] for instance reported better performance of participants with experience in playing video games when training laparoscopic interventions. It should be noticed that in most cases, the influence of video gaming was proven for training laparoscopic interventions. In contrast to these findings, the study of Khatri et al [42] found no correlation between video gaming and surgical performance for orthopedic skill acquisition when examining dynamic hip screw simulation. Those findings support our study results. We found no correlation for both groups between playing video games and surgical performance in CTI. Regarding these results, we suppose that the influence of video gaming on the surgical performance depends on the type of surgical intervention. It might be assumable that, in contrast to laparoscopic interventions, video gaming has much lower influence on surgical performance on interventions in trauma surgery.

Participants from this study evaluated TS as an efficient and motivating training device. Furthermore, most participants wanted to continue the training with TS. According to Hutchinson et al, motivation to learn can be intrinsic (from the trainee) and extrinsic (from external influences). Intrinsic factors can be improvement of personal achievement, preparation for new situations, fun, and competition [43]. Serious gaming with TS might increase the intrinsic motivation of students by offering educational modules while providing a fun experience and because of the self-assessment, instant feedback regarding improvement. Therefore, implementation of TS in surgical education might increase students' motivation to self-educate.

Limitations

Despite relevant findings of this study regarding the training benefit of serious games with a relevant content in the context of CTI, our study has limitations. Participants of the control group received the initial lecture, as well as a related but irrelevant game-based study exercise. This might have led to confusion among participants of the control group leading to a poorer performance in the exercise and therefore limiting the results of the study. However, participants were instructed regarding their specific training program before the study. In addition, all students participated in the introductory lecture normally held on CTI, and students were given access to the script of the lecture and information on where to find further

material for self-study (eg, books and e-learning material). Furthermore, participants of the control group were specifically interviewed before the assessment if they complied with the instructions and had adequately studied CTI. Therefore, because of the design of the study, we believe the influence of confusion derived by the control procedure and the influence of the control procedure as a distractor to be minimal. Nonetheless, the results of this study might be limited by this possibility. Before assessment, the character of the tests and the intervention of interest were purposively not clarified. However, participants of the control group were given the instruction to further study CTI, and participants of intervention group were given a single intervention. Thereby, it can be assumed that students suspected CTI to be the intervention of interest. This might limit the results of our study as participants may have performed disproportionately well by preparing especially for CTI. However, the specific type of assessment and the parameters of interest remained unknown. In addition, participants were only assessed if they complied with the instructions given at the beginning of the study. Compliance was analyzed based on subjective statements of participants before the assessment. The subjective nature of the analysis might bias the results of this study by leaving the risk of an over- or underreporting of compliance. Evaluation of motivation and satisfaction regarding serious gaming as an educational method was based on the subjective self-evaluation of the participants. It is possible that there were inaccuracies between the different groups because of incorrect answers in the participants' self-assessment. Another

limitation lies in the fact that TS is only offered in English. Some participants struggled with the language barrier and needed help in translating the different instruments. However, supervising surgeons helped students regarding difficulties in translating specific terms. Therefore, we believe that the language barrier does not influence the results of the study.

Conclusions

The results from this study indicate that utilizing serious games with relevant content might provide a higher level of education in preparing medical students to perform CTI than traditional learning methods, including serious games with irrelevant content. Furthermore, training with serious games using a relevant content seems to improve the independency of students in performing a CTI. In addition, serious gaming might increase the intrinsic motivation of students by offering educational modules while providing a fun experience and because of the self-assessment, instant feedback regarding improvement. Thereby, implementation of TS in surgical education might increase student's motivation to self-educate. In conclusion, the results from this study led us to believe that serious games are a valid and effective tool in the education of medical students regarding the operative performance in CTI. We believe that serious games should be implemented in surgical training curricula of medical students in addition to traditional learning methods and might add benefit to the training curriculum of residency programs. However, further studies are needed to clarify the training benefit of TS during the education of residents to confirm this assumption.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questionnaires for self-assessment of students. Assessment for control group.

[PDF File (Adobe PDF File), 70KB - jmir_v20i5e195_app1.pdf]

Multimedia Appendix 2

Questionnaires for self-assessment of students. Assessment for study group.

[PDF File (Adobe PDF File), 69KB - jmir_v20i5e195_app2.pdf]

Multimedia Appendix 3

CONSORT‐EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 649KB - jmir_v20i5e195_app3.pdf]

References

1. Friedrich M, Bergdolt C, Haubruck P, Bruckner T, Kowalewski K, Müller-Stich BP, et al. App-based serious gaming for training of chest tube insertion: study protocol for a randomized controlled trial. Trials 2017 Feb 06;18(1):56 [FREE Full text] [doi: 10.1186/s13063-017-1799-5] [Medline: 28166840]
2. Menger R, Telford G, Kim P, Bergey MR, Foreman J, Sarani B, et al. Complications following thoracic trauma managed with tube thoracostomy. Injury 2012 Jan;43(1):46-50. [doi: 10.1016/j.injury.2011.06.420] [Medline: 21839442]
3. Faulkner H, Regehr G, Martin J, Reznick R. Validation of an objective structured assessment of technical skill for surgical residents. Acad Med 1996 Dec;71(12):1363-1365. [Medline: 9114900]

XSL•FO
RenderX

4. Adler G, v d Knesebeck JH. [Shortage and need of physicians in Germany? Questions addressed to health services research]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2011 Feb;54(2):228-237. [doi: 10.1007/s00103-010-1208-7] [Medline: 21290278]

5. Kasch R, Engelhardt M, Förch M, Merk H, Walcher F, Fröhlich S. [Physician shortage: how to prevent generation Y from staying away - results of a nationwide survey]. Zentralbl Chir 2016 Apr;141(2):190-196. [doi: 10.1055/s-0035-1557857] [Medline: 26394048]

6. van Hove PD, Tuijthof GJ, Verdaasdonk EG, Stassen LP, Dankelman J. Objective assessment of technical surgical skills. Br J Surg 2010 Jul;97(7):972-987. [doi: 10.1002/bjs.7115] [Medline: 20632260]

7. Winckel CP, Reznick RK, Cohen R, Taylor B. Reliability and construct validity of a structured technical skills assessment form. Am J Surg 1994 Apr;167(4):423-427. [Medline: 8179088]

8. Kowalewski K, Hendrie JD, Schmidt MW, Proctor T, Paul S, Garrow CR, et al. Validation of the mobile serious game application Touch Surgery™ for cognitive training and assessment of laparoscopic cholecystectomy. Surg Endosc 2017 Oct;31(10):4058-4066. [doi: 10.1007/s00464-017-5452-x] [Medline: 28281111]

9. Verdaasdonk EG, Dankelman J, Schijven MP, Lange JF, Wentink M, Stassen LP. Serious gaming and voluntary laparoscopic skills training: a multicenter study. Minim Invasive Ther Allied Technol 2009 Aug;18(4):232-238. [doi: 10.1080/13645700903054046] [Medline: 19575329]

10. Cowan B, Sabri H, Kapralos B, Moussa F, Cristancho S, Dubrowski A. A serious game for off-pump coronary artery bypass surgery procedure training. Stud Health Technol Inform 2011;163:147-149. [Medline: 21335779]

11. Nosek TM, Cohen M, Matthews A, Papp K, Wolf N, Wrenn G, et al. A serious gaming/immersion environment to teach clinical cancer genetics. Stud Health Technol Inform 2007;125:355-360. [Medline: 17377303]

12. Shewaga R, Knox A, Ng G, Kapralos B, Dubrowski A. Z-DOC: a serious game for Z-plasty procedure training. Stud Health Technol Inform 2013;184:404-406. [Medline: 23400192]

13. Graafland M, Schraagen JM, Schijven MP. Systematic review of serious games for medical education and surgical skills training. Br J Surg 2012 Oct;99(10):1322-1330. [doi: 10.1002/bjs.8819] [Medline: 22961509]

14. Graafland M, Bemelman WA, Schijven MP. Appraisal of face and content validity of a serious game improving situational awareness in surgical training. J Laparoendosc Adv Surg Tech A 2015 Jan;25(1):43-49. [doi: 10.1089/lap.2014.0043] [Medline: 25607899]

15. Djaouti D, Alvarez J, Jessel J. Classifying Serious Games: The G/P/S Model. In: Handbook of Research on Improving Learning and Motivation through Educational Games: Multidisciplinary Approaches. United States: IGI Global; Nov 2011:118-136.

16. Sugand K, Mawkin M, Gupte C. Validating Touch Surgery™: a cognitive task simulation and rehearsal app for intramedullary femoral nailing. Injury 2015 Nov;46(11):2212-2216. [doi: 10.1016/j.injury.2015.05.013] [Medline: 26094504]

17. Paro JA, Luan A, Lee GK. Validation of a cognitive task simulation and rehearsal tool for open carpal tunnel release. Arch Plast Surg 2017 May;44(3):223-227 [FREE Full text] [doi: 10.5999/aps.2017.44.3.223] [Medline: 28573097]

18. Khelemsky R, Hill B, Buchbinder D. Validation of a novel cognitive simulator for orbital floor reconstruction. J Oral Maxillofac Surg 2017 Apr;75(4):775-785. [doi: 10.1016/j.joms.2016.11.027] [Medline: 28012843]

19. Touchsurgery. About Touch Surgery URL: https://www.touchsurgery.com/about.html [accessed 2018-01-25] [WebCite Cache ID 6wjIyC37a]

20. Bartlett RD, Radenkovic D, Mitrasinovic S, Cole A, Pavkovic I, Denn PC, et al. A pilot study to assess the utility of a freely downloadable mobile application simulator for undergraduate clinical skills training: a single-blinded, randomised controlled trial. BMC Med Educ 2017 Dec 11;17(1):247 [FREE Full text] [doi: 10.1186/s12909-017-1085-y] [Medline: 29228934]

21. Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, et al. The prevention and treatment of missing data in clinical trials. United States: New England Journal of Medicine; Oct 04, 2012:1355-1360.

22. Hutton IA, Kenealy H, Wong C. Using simulation models to teach junior doctors how to insert chest tubes: a brief and effective teaching module. Intern Med J 2008 Dec;38(12):887-891. [doi: 10.1111/j.1445-5994.2007.01586.x] [Medline: 18284460]

23. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]

24. Amer KM, Mur T, Amer K, Ilyas AM. A mobile-based surgical simulation application: a comparative analysis of efficacy using a carpal tunnel release module. J Hand Surg Am 2017 May;42(5):389.e1-389.e9. [doi: 10.1016/j.jhsa.2017.02.008] [Medline: 28318742]

25. Bunogerane GJ, Taylor K, Lin Y, Costas-Chavarri A. Using touch surgery to improve surgical education in low- and middle-income settings: a randomized control trial. J Surg Educ 2017 Jul 13;75(1):231-237. [doi: 10.1016/j.jsurg.2017.06.016] [Medline: 28712686]

26. Sugand K, Mawkin M, Gupte C. Training effect of using Touch Surgery for intramedullary femoral nailing. Injury 2016 Feb;47(2):448-452. [doi: 10.1016/j.injury.2015.09.036] [Medline: 26596416]

27. Arora S, Aggarwal R, Sevdalis N, Moran A, Sirimanna P, Kneebone R. Development and validation of mental practice as a training strategy for laparoscopic surgery. Surg Endosc Jan 2010;24(1):179-187. [doi: 10.1007/s00464-009-0624-y] [Medline: 19633892]

XSL·FO

RenderX

28. Conlin A, Lea J, Bance M, Chadha N, Kilty S, Kozak F, et al. Mental practice in postgraduate training: a randomized controlled trial in mastoidectomy skills. J Otolaryngol Head Neck Surg 2016 Sep 15;45(1):46 [FREE Full text] [doi: 10.1186/s40463-016-0162-2] [Medline: 27634316]

29. Sanders CW, Sadoski M, van Walsum K, Bramson R, Wiprud R, Fossum TW. Learning basic surgical skills with mental imagery: using the simulation centre in the mind. Med Educ 2008 Jun;42(6):607-612. [doi: 10.1111/j.1365-2923.2007.02964.x] [Medline: 18435713]

30. Burgos CM, Josephson A. Gender differences in the learning and teaching of surgery: a literature review. Int J Med Educ 2014 Jun 15;5:110-124 [FREE Full text] [doi: 10.5116/ijme.5380.ca6b] [Medline: 25341220]

31. Schueneman AL, Pickleman J, Freeark RJ. Age, gender, lateral dominance, and prediction of operative skill among general surgery residents. Surgery 1985 Sep;98(3):506-515. [Medline: 4035571]

32. Madan AK, Harper JL, Frantzides CT, Tichansky DS. Nonsurgical skills do not predict baseline scores in inanimate box or virtual-reality trainers. Surg Endosc 2008 Jul;22(7):1686-1689. [doi: 10.1007/s00464-007-9691-0] [Medline: 18071808]

33. Roch PJ, Rangnick HM, Brzoska JA, Benner L, Kowalewski K, Müller PC, et al. Impact of visual-spatial ability on laparoscopic camera navigation training. Surg Endosc 2018 Mar;32(3):1174-1183. [doi: 10.1007/s00464-017-5789-1] [Medline: 28840317]

34. Kolozsvari NO, Andalib A, Kaneva P, Cao J, Vassiliou MC, Fried GM, et al. Sex is not everything: the role of gender in early performance of a fundamental laparoscopic skill. Surg Endosc 2011 Apr;25(4):1037-1042. [doi: 10.1007/s00464-010-1311-8] [Medline: 20734067]

35. Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Impact of hand dominance, gender, and experience with computer games on performance in virtual reality laparoscopy. Surg Endosc 2003 Jul;17(7):1082-1085. [doi: 10.1007/s00464-002-9176-0] [Medline: 12728373]

36. Nickel F, Hendrie JD, Kowalewski K, Bruckner T, Garrow CR, Mantel M, et al. Sequential learning of psychomotor and visuospatial skills for laparoscopic suturing and knot tying-a randomized controlled trial "The Shoebox Study" DRKS00008668. Langenbecks Arch Surg 2016 Sep;401(6):893-901. [doi: 10.1007/s00423-016-1421-4] [Medline: 27055853]

37. Kowalewski KF, Garrow CR, Proctor T, Preukschas AA, Friedrich M, Müller PC, et al. LapTrain: multi-modality training curriculum for laparoscopic cholecystectomy-results of a randomized controlled trial. Surg Endosc 2018 Feb 12:1-9 Epub ahead of print(forthcoming). [doi: 10.1007/s00464-018-6110-7] [Medline: 29435758]

38. Jalink MB, Goris J, Heineman E, Pierie JE, ten Cate Hoedemaker HO. The effects of video games on laparoscopic simulator skills. Am J Surg 2014 Jul;208(1):151-156. [doi: 10.1016/j.amjsurg.2013.11.006] [Medline: 24814309]

39. Jalink MB, Heineman E, Pierie JP, ten Cate Hoedemaker HO. The effect of a preoperative warm-up with a custom-made Nintendo video game on the performance of laparoscopic surgeons. Surg Endosc 2015 Aug;29(8):2284-2290. [doi: 10.1007/s00464-014-3943-6] [Medline: 25361658]

40. Nomura T, Matsutani T, Hagiwara N, Fujita I, Nakamura Y, Kanazawa Y, et al. Characteristics predicting laparoscopic skill in medical students: nine years' experience in a single center. Surg Endosc 2018 Jan;32(1):96-104. [doi: 10.1007/s00464-017-5643-5] [Medline: 28639038]

41. Rosser JC, Lynch PJ, Cuddihy L, Gentile DA, Klonsky J, Merrell R. The impact of video games on training surgeons in the 21st century. Arch Surg 2007 Feb;142(2):181-6; discusssion 186. [doi: 10.1001/archsurg.142.2.181] [Medline: 17309970]

42. Khatri C, Sugand K, Anjum S, Vivekanantham S, Akhtar K, Gupte C. Does video gaming affect orthopaedic skills acquisition? A prospective cohort-study. PLoS One 2014;9(10):e110212 [FREE Full text] [doi: 10.1371/journal.pone.0110212] [Medline: 25333959]

43. Hutchinson L. Educational environment. Br Med J 2003 Apr 12;326(7393):810-812 [FREE Full text] [doi: 10.1136/bmj.326.7393.810] [Medline: 12689981]

Abbreviations

CTI: chest tube insertion
IRR: interrater reliability
TS: Touch Surgery
RCT: randomized controlled trial
OSATS: objective structural assessment of technical skills

Original Paper

Parents' Experiences of Caring for Their Child at the Time of Discharge After Cardiac Surgery and During the Postdischarge Period: Qualitative Study Using an Online Forum

Jo Wray[1*], PhD; Katherine Brown[1*], MD; Jenifer Tregay[1*], PhD; Sonya Crowe[2*], PhD; Rachel Knowles[3*], PhD; Kate Bull[1*], MD; Faith Gibson[4,5*], PhD

[1]Charles West Division, Great Ormond Street Hospital for Children NHS Foundation Trust, London, United Kingdom

[2]Clinical Operational Research Unit, University College London, London, United Kingdom

[3]MRC Centre of Epidemiology for Child Health, UCL Institute of Child Health, London, United Kingdom

[4]Centre for Outcomes and Experience Research in Children's Health, Illness and Disability, Great Ormond Street Hospital for Children NHS Foundation Trust, London, United Kingdom

[5]School of Health Sciences, Faculty of Health and Medical Sciences, University of Surrey, Guildford, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Jo Wray, PhD
Charles West Division
Great Ormond Street Hospital for Children NHS Foundation Trust
Great Ormond Street
London, WC1N 3JH
United Kingdom
Phone: 44 207405 ext 9200
Email: jo.wray@gosh.nhs.uk

Abstract

Background: Congenital heart disease (CHD) is the most common class of birth defects, which encompasses a broad spectrum of severity ranging from relatively minor to extremely complex. Improvements in surgery and intensive care have resulted in an increasing number of infants with the most complex lesions surviving after surgery until the time of discharge from the hospital, but there remain concerns about out-of-hospital mortality, variability in how services are provided at the time of discharge and beyond, and difficulties experienced by some families in accessing care.

Objective: As part of a mixed-methods program of research, this study aimed to elicit parental experiences of caring for a child with CHD after hospital discharge following a cardiac surgery and collect information to inform interviews for a subsequent stage of the project.

Methods: A closed online discussion group was set up via the main Facebook page of the Children's Heart Federation (CHF), a national charity offering support to children with heart disease and their families. The discussion group was advertised through the charity's webpage, and interested participants were directed to the charity's Facebook page from where they could access the closed Facebook group and respond to questions posted. The CHF moderated the forum, and the research team provided questions to be posted on the forum. Responses were collated into a single transcript and subjected to thematic analysis.

Results: The forum was open for 4 months, and 91 participants (mean age 35 years, range 23-58 years, 89 females, 89 parents, and 2 grandparents) submitted demographic information and were given access to the closed forum group. A common experience of isolation emerged from the data, with descriptions of how that isolation was experienced (physical, social, knowledge) and its psychological impact, together with the factors that made it worse or better. Woven through this theme was the notion that parents developed expertise over time.

Conclusions: The use of an online forum provided a means for eliciting data from a large number of parents regarding their experiences of caring for their child after hospital discharge following cardiac surgery. Parents engaged with the forum and were able to articulate what went well and what went less well, together with sharing their stories and supporting each other through doing so. Some parents clearly found participating in the forum a positive experience in itself, demonstrating the potential of social media as a mechanism for providing support and reducing isolation. Information gained from the forum was used to shape questions for interviews with parents in a subsequent phase of the study. Furthermore, the themes identified in the online forum

XSL•FO
RenderX

have contributed to identifying ways of improving the provision of care and support for parents of high-risk babies following discharge after cardiac surgery.

Introduction

Congenital heart disease (CHD) affects approximately 8 in every 1000 live births, and every year in the United Kingdom, approximately 2800 infants undergo heart surgery, with those at the more severe end of the disease spectrum often leaving hospital requiring ongoing care, such as tube feeding, multiple medications, and daily monitoring of oxygen saturations [1,2]. Improvements in surgery and intensive care have resulted in increasing numbers of infants with the most complex lesions surviving after surgery until the time of discharge from hospital but this has been accompanied by concerns about out-of-hospital mortality, variability in how services are provided at the time of discharge and beyond, and difficulties experienced by some families in accessing care. To understand and address these concerns, a multidisciplinary mixed-methods program of research was undertaken (Figure 1), including a systematic review [3], quantitative analyses of national audit data [4,5], and interviews with key stakeholders (parents, health professionals, and charity helpline staff) [6-8]. The overall aim was to synthesize the information gathered from each of the data sources to identify ways in which care could be improved for high-risk infants with complex cardiac conditions at the time of discharge and during the postdischarge period [9].

The focus of this paper is to present a discrete element of the study that centered on an online forum method used to address the following 2 aims:

1. Elicit parental experiences of caring for a child with CHD after discharge following cardiac surgery
2. Collect information from parents of children with CHD to inform the in-depth interviews (Figure 1)

A further aim emerged, which was to comment on the use of an online forum in the context of this study.

In this online forum part of our study, we particularly sought to explore the parents' experiences of taking a baby home after congenital heart surgery. Elevated levels of distress, anxiety, and depression have been reported in parents of children who have undergone cardiac surgery, particularly those for whom their baby's surgery was recent and those with more complex heart conditions [10,11]. Parents are often discharged with suboptimal levels of information about their child's condition [12] and their children, particularly the most fragile infants, need skilled care once home after discharge [13], all of which can result in significant parental burden. A process of "safeguarding survival" has been described in parents of infants at highest risk of dying after discharge, whereby parents safeguard their child's survival through taking charge of their infant's care at home, protecting them from infection, and drawing on the support of others (extended family) to help with the care and monitoring of their child [14]. This has been developed further into a theory of "parenting under pressure," which highlights the specific challenges parents face but also focuses on parental ability, resourcefulness, and resilience [15]. However, although such research provides important and valuable insights into parenting fragile, high-risk infants, the reliance on face-to-face interviews as the mechanism for data collection has excluded many parents from participating and also prevented us from accessing the insights from parents sharing experiences and opening up to each other.

Parents of children with significant health conditions can find it challenging to participate in face-to-face research for a variety of practical reasons; however, those who are difficult to reach via traditional data collection methods, such as fathers, those who are more geographically remote, and those from ethnically and culturally diverse cultural backgrounds often have important and salient contributions to make. Recent years have seen a significant increase in the use of online social networks, with evidence suggesting that approximately two-thirds of adults in the United States and Europe now use social media [16,17]. Social networks provide a quick and easy means of sharing ideas, information, and opinions; have a broad population reach; and are widely used in every sphere of life [16,18]. Facebook is the dominant social network worldwide and many health-related groups have arisen on Facebook that are predominantly used for raising awareness, social support, and fundraising. Developments in technology have also resulted in an increasing use of electronic methods to collect data for research, and there is evidence to support their feasibility, the prompt responses of participants, the richness of data collected by these means, and fewer human errors [19-21]. One method of electronic data collection is the online forum, which allows asynchronous interactions, whereby participants are able to join discussions at their own convenience, in contrast to methods that require synchronous interactions, such as chat groups. They have been reported to be relatively easy to use, safe, accessible, and observable, and it has also been suggested that they offer a more comfortable mechanism for the discussion of sensitive or personal health issues and are a feasible alternative to more traditional research tools, such as face-to-face focus groups [22-24]. They also provide participants with a "free rein" to express their views, such that responses may include valuable detail beyond that which was originally asked. Furthermore, they offer flexibility to researchers and participants alike, thus reducing participant burden and pressures of time.

Figure 1. Elements of the research program. CHD: congenital heart disease.
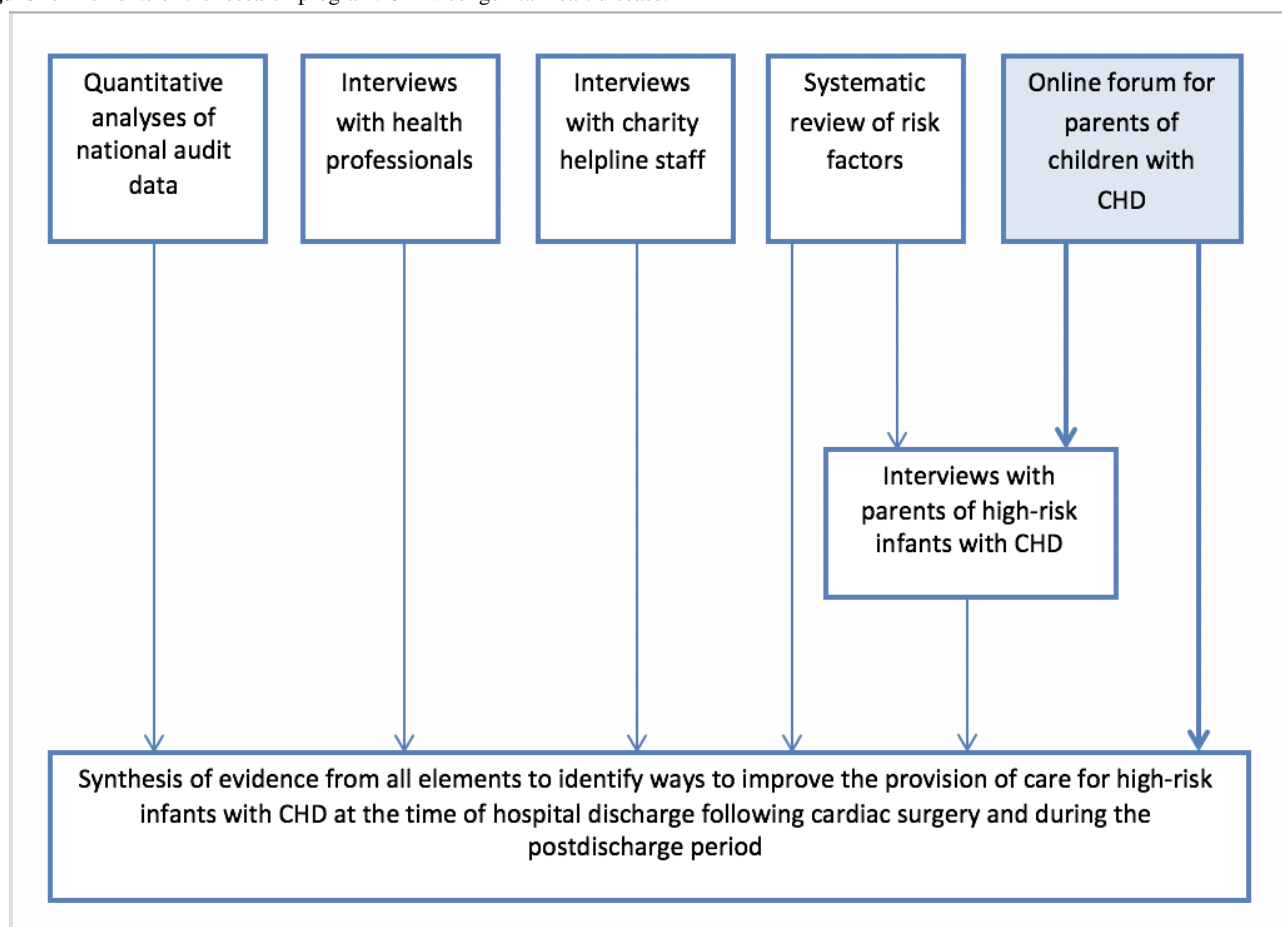
Methods

We used an Internet forum to elicit views on the information and support that parents were given at the time of discharge after their baby's surgery and during the postdischarge period and their experiences of caring for their baby and interacting with health professionals in primary, secondary, and tertiary services.

Data Collection

The online forum was facilitated via the Children's Heart Federation (CHF) through its Facebook page. CHF is a national charity and is an umbrella organization representing approximately 12,000 children with heart disease and their families. They have an active presence and following on Facebook and Twitter, and at the time of setting up the online forum, the CHF Facebook page had around 3120 members. CHF has experience of running online forums, and as they are a national charity, their forums are accessed by parents of children treated at each of the specialist pediatric cardiac centers in the UK, thus providing us with the potential to include a more representative population of parents than we could achieve otherwise.

CHF entered into a contractual arrangement with us, and a fixed fee was agreed upon for their role in facilitating and managing the forum. A closed online discussion group was set up via their main Facebook page. The discussion group was advertised through the charity's webpage, and interested participants were then directed to the charity Facebook page, where they could access more detailed information about the larger study and the online forum in particular. Information was provided about the rationale for the study, the role of the forum, how information from the forum would be used, and issues relating to confidentiality and anonymization of forum posts. The potential for touching on sensitive topics and that some bereaved families may contribute to the forum were acknowledged, and contact details for the CHF helpline were provided should participants want to talk to someone further about any issues that were raised on the forum. If people wanted to participate, they were required to provide some basic demographic information (their age, gender, ethnicity, and geographical region). Once this information was received by CHF, they were given access to a private or "closed" Facebook group and were able to begin responding to questions posted there. CHF was responsible for all day-to-day running and moderation of the forum in line with a standard operating procedure developed in collaboration with the research team, which included processes for managing inappropriate or offensive messaging and distressed users as well as procedures for running the forum. The research team provided questions to be posted on the forum at the start of the process, and CHF decided when new questions should be posted based on participant responses and the rate of responding. CHF was specifically asked to probe further if they noticed any of the following issues in participant responses:

1. Social and practical issues, for example, financial, educational, and transport issues

2. Issues to do with language or cultural differences
3. Difficulties accessing support in the community
4. Understanding information from health care providers

Forum responses were pseudonymized by CHF before being sent to the research team in a weekly update. CHF assigned participant numbers to individual respondents, but did not undertake any editing of responses before sending them to the research team.

Ethical approval was granted by the local National Research Ethics Service Committee London-Central (Reference: 12/LO/1398).

Data Analysis

Responses were collated into a single transcript (78 pages), and thematic analysis was used to analyze forum responses according to recommendations made by Coffey and Atkinson [25]. A conceptual model was built and discussed with all members of the research team, showing clearly the relationships between the themes. The analysis proceeded as follows:

1. The transcript was analyzed as a whole. It was read, and notes were made in the margins on interesting or significant points said (JT, FG, and JW).
2. Codes were attached to segments of data. These segments could be one word, a phrase, or a sentence. The codes were a summary of what a parent seemed to be referring to or describing.
3. After the transcript had been coded, segments of data with similar codes were brought together to create categories containing data that shared a common element.
4. The categories were discussed, refined, and used to generate themes (JT, FG, and JW). As meanings can change when phrases are isolated, the original contexts of the phrases within the themes were checked.
5. Each theme was given a name that aimed to capture all the elements within that theme.
6. Agreement and understanding were gauged during discussions until consensus was reached; there were in fact few discrepancies. Working in this manner, we safeguarded against an interpretation representing the subjectivity of the observer more than the object of study.

Results

Study Participants and Overview of Responses

The forum ran for 4 months, during which time a total of 91 participants (mean age 35 years, range 23-58 years, 89 females, 89 parents, and 2 grandparents) submitted demographic information and were given access to the closed forum group. Participants came from all over the United Kingdom and were predominantly of white British ethnicity (85 of 91 participants). Of these, 73 parents participated in the forum discussion and most responded to between 1 and 5 questions. Neither of the 2 grandparents contributed to the forum discussion. Although data were not collected about the children themselves, it was evident from the responses that they had a wide range of cardiac diagnoses, including as part of genetic syndromes, and that they spanned a wide age range (infant to teenager), although most

were of preschool age. A total of 19 questions were posted, although there was an overlap between some of the questions to probe for further detail and encourage other participants to respond. The first question had 43 responses and the final question had 39 responses, with the number of responses to individual questions ranging from 14 to 43. The questions with the most responses were those about information provided about caring for their baby at home (symptoms and how prepared they felt), community support (particularly experiences with general practitioners and health visitors [HVs]), and support from cardiac liaison nurses. Parents also responded to others' posts, offering support and sharing experiences, and toward the end of the forum, parents became more open in identifying with other parents' experiences.

Emerging from the data was an experience of isolation, as one parent said:

> *It's a pretty lonely place.*

This was described in terms of the way in which that isolation was experienced (physical, social, knowledge) and the resulting psychological impact, together with the factors that made that worse or better (challenging and mitigating factors). Threaded through this theme was the notion of time, revealing parents developing expertise, where parents moved from feeling overwhelmed and lacking in knowledge and skills to becoming "expert parents" with a corresponding increase in their knowledge and skills (Figure 2). We present here these themes drawing on salient quotes that best illustrate our findings.

Physical Isolation

A number of parents described being physically isolated as a result of having a baby with CHD:

> *[I] didn't see a soul…from leaving hospital to returning for review.*

The physical isolation was sometimes due to parents' own anxieties and concerns, particularly about the risk of infection. Some parents felt that they could not go out to places where they would be in contact with other parents and babies:

> *I felt it too risky to go to clinic, with infections.*

There were other parents who not only feared taking their baby out but also worried about visitors bringing infection into their home:

> *It was all such a huge shock, on discharge we were told she was like any other baby, but obviously there were differences. I was very scared about taking her out and catching a cold or something, or visitors bringing in germs.*

In other situations, parents identified that professionals—particularly HVs—were anxious about mothers taking their babies to a busy clinic, preferring instead to make home visits, but this in turn could compound the challenges of isolation that the families were facing:

> *[My health visitor] was nervous about me going to baby clinics etc for weekly weights but I needed to get out.*

Figure 2. Parental experiences of caring for their child with congenital heart disease.
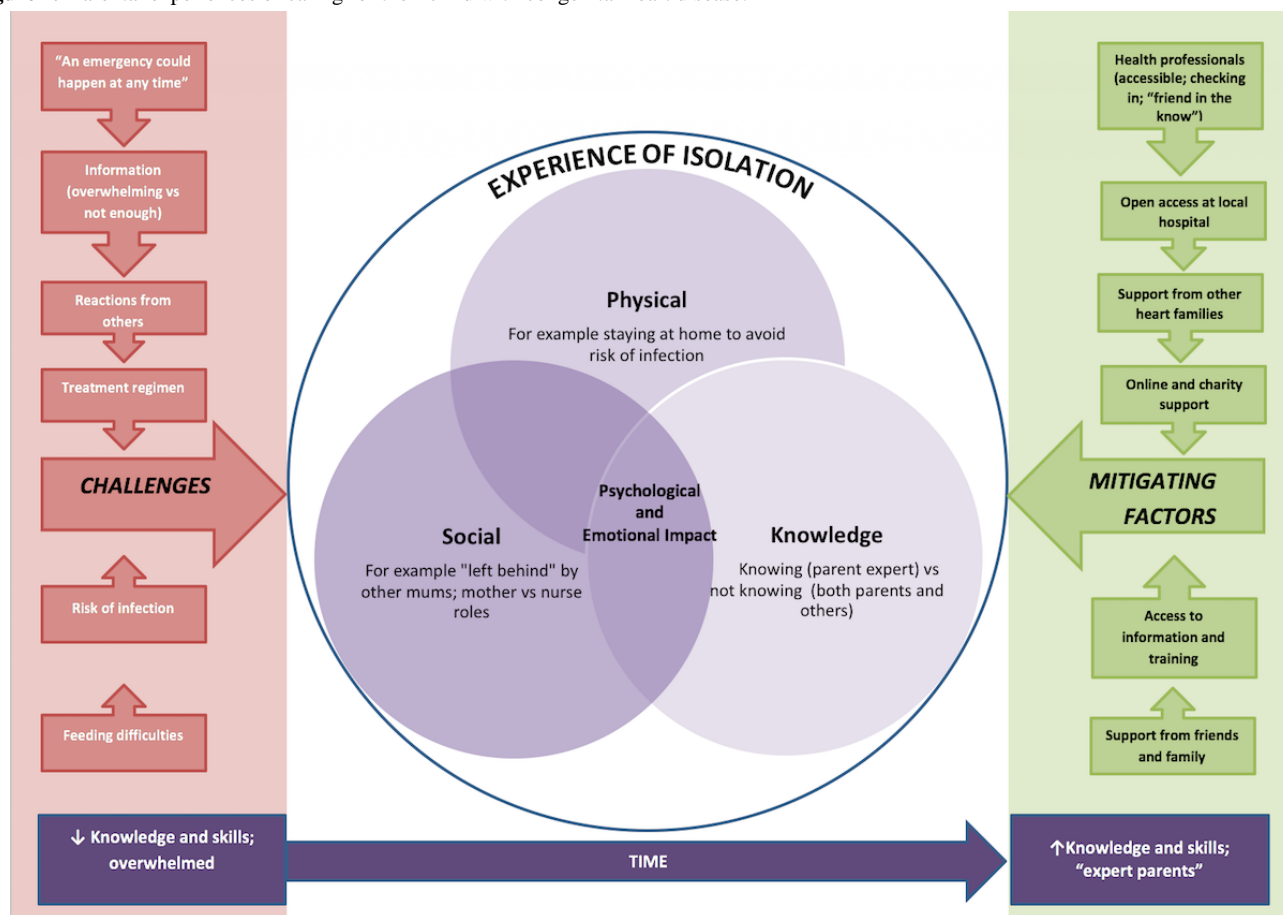
Even when the visits themselves were perceived positively, there was still an isolating element to them:

> ...home visits were lovely but made me feel more isolated.

Social Isolation

Parents described their feelings of social isolation in terms of interactions with other parents and about the support parents of a child with CHD received from support groups. Being unable to participate in "normal" mother and baby activities resulted in mothers feeling isolated from other new mothers, with the impact sometimes having consequences beyond the first weeks and months:

> Loads of mum formed their friendships at that time. I was left behind and still am because of his additional needs.

Mothers described how they wanted to be like others, but in several instances, the particular needs of their baby singled them out from other parents, with a resulting isolating impact:

> Went to mother and baby group once to get him weighed and never went again as was so upset by HV and other mums looking at your child's scar and not talking.

A number of parents were quite specific in their descriptions of what they considered to be failings in the support they were given and the isolating effect this had:

> Absolutely no support groups...or help...feel quite isolated sometimes.

In contrast, other parents reported feeling well supported by their local community team:

> Health visitor was good and did the best she could for us despite not having all the answers about concerns regarding baby's heart defect. She came almost every week to weigh him ...she contacted other professionals to seek advice or support for us. She was a welcome face at an isolated time and I looked forward to our chats.

Knowledge

Some parents described their own and others' knowledge, or lack of knowledge, as both challenging and isolating. Many commented that professionals in both secondary and primary care did not have sufficient information about CHD or their child's specific heart condition:

> There has been no one in the community or local hospital that had the answers.

And that they often knew more about their child than the local health professionals:

> I found that we know lots more than they [community health professionals] did about her condition, which was both understandable...and terrifying in equal measures.

Parents also described their own stress associated with knowledge, in terms of either feeling that they did not know enough about their child's condition or, conversely, the responsibility and burden of needing to communicate knowledge to local health professionals:

> [I am]...just sick of explaining to everyone. I wish someone could tell me about him not the other way around.

In contrast, others talked about the efforts made by health professionals in the community to become more knowledgeable about CHD or to get advice from other professionals. Some parents described the information they were given by the specialist center, highlighting the degree to which tertiary centers varied in the information they gave families about local services and support networks, both in terms of what they provided and how they provided it.

Psychological and Emotional Impact

Caring for a child after cardiac surgery can have a significant psychological and emotional impact, and parents described their anxiety, which, for some, developed into symptoms of post-traumatic stress or served to heighten their feelings of isolation. One parent talked about feeling unprepared for how life had changed and the tension between feeling grateful that her child had come through the surgery and was back home and not feeling able to tell the team how anxious she felt:

> Your whole life changes and no one tells you that...I think I had a bit of posttraumatic stress but I was so grateful my baby was home I didn't want to say how terrified I was constantly and how much I relived every moment.

It was evident that some parents felt disorientated and unsettled on returning home, with a number of parents describing that they felt in a "fog" when they got home after their baby's surgery:

> I was having panic attacks and quite "fog-like" for months, felt quite isolated but there seemed to be nobody really asking about the parents.

Many parents shared how lonely and scared they felt during those early weeks:

> Felt out of my depth and very scared.

Some had clearly been traumatized by their experiences, and described feeling unsupported:

> The child is discharged and the parents are left walking around in an often traumatised state with no suitable support.

And unable to share their experiences:

> It has been the most traumatic year of my life and yet I feel I can't really talk to anyone else about it.

Some parents also talked about their difficulty in perceiving themselves as a parent rather than just as someone providing medical care for their child because of the demands of the treatment regimen:

> For a long time I struggled to feel like his mummy and not a nurse because that was all I seemed to be doing...medications, meds and more meds.

Challenges

Parents provided insight into a number of challenges that they faced once home after their child's surgery, which contributed to and compounded the overall feeling of isolation. Some of these challenges were related to practical issues, such as information, with parents describing a "roller-coaster of information and procedures" and how they received a lot of information in the hospital but nothing about being at home after discharge:

> We got bombarded with info in hospital but once home there's no information as to what to look out for.

Other parents described a number of practical difficulties associated with aspects of the treatment regimen, such as getting prescriptions:

> At one point we were left without heart meds for one week. Doctor refused to prescribe...They told me they would phone the cardiologist to check. They didn't. Needless to say my baby was getting sicker and sicker. I ended up crying down the phone pleading for a prescription.

Parents conveyed a sense of these difficulties undermining local health services and reducing parents' trust and confidence in them.

Feeding was identified as a significant challenge by a number of the parents, which was related to the difficulties the baby had in feeding:

> I had...this baby that struggled to feed, cried if I touched her and lost weight constantly. All I got was "just persevere and top her up with her NG."

Challenges were also related to the time feeding took place and some of the practicalities for those children who were tube fed:

> The first few days were a nightmare...I ended up syringing milk into her mouth as she wouldn't take a bottle at all.

Some mothers also talked about breastfeeding, for example, one mother described feeling "lost and helpless" as she tried to feed her baby "without hurting her," whereas others discussed the lack of support they experienced around feeding issues:

> We had a horrendous 6 months and it was me that sought out support [for feeding issues] it wasn't just there for us.

Taking home a vulnerable baby after surgery was a further source of stress for a number of parents, related to the risk of complications, infections, or the fear of something going wrong, and parents conveyed a sense of waiting and watching for something to happen "…we feared an emergency could happen at any time" using words such as "horrendous," "scared to death," and "lost." Several parents also identified that a lack of training had contributed to their anxiety:

We weren't offered anything. I researched myself and paid to go on a babies/children first aid course. Think it would've been massively useful (if only to instil some confidence in us as parents that we could cope if a situation had arisen) to have had some basic training or advice.

Finally, parents described the challenge of dealing with the reactions of others, particularly professionals, to them and their baby. For example, one parent talked about how she felt that her GP thought she was an "over-reactive mum," whereas another parent felt that it was her baby whom her GP had issues with:

I found my doctor very "stand-offish" as if scared of him [baby].

Mitigating Factors

Although some parents described the challenges of caring for a baby with a heart condition after cardiac surgery, many of them also talked about the things that helped and those factors which lessened the feelings of isolation. A number of the mitigating factors were related to the same topics which were challenges for other parents. For example, in a number of cases parents were provided with, or had access to, training and information before they left the specialist center:

We were asked in the hospital to give her meds under the nurse's supervision so she could check we were administering them correctly. We were given all doses and medicines written down and plenty of syringes to take home. We were given a lot of info on care of her wound, what to do if she went blue etc and numbers for the CLN [cardiac liaison nurse], the ward and were told any queries just to call the ward direct which we did on a couple of occasions and got great and prompt advice.

Several parents were also given specific information about signs and symptoms to look for in their baby:

...told to look for blue lips. Fingers. If she got breathless tired sweaty while feeding also if oxygen levels go low.

A number of parents had open-access arrangements with their local hospital, which were clearly highly valued and helped to reassure parents:

We have open access at local hospitals and have [been] made to feel very welcome and nothing ever too big or small to come and see them.

Others described the accessibility of advice and the importance of that to them:

She was always on hand via phone or in person to answer questions and help explain stuff to us in layman's terms. I honestly believe she made a world of difference to our ability to cope.

Although some parents had experienced negative reactions from health professionals, others had a very positive experience of care and support after hospital discharge from professionals in primary care, for example:

No problem was too small for the GP, she would phone the hospital if she needed to while I was waiting and send me straight away if she was worried about her.

A number of parents also described how their experience of secondary care had been excellent:

Our community nurse was brilliant, she even gave me her home number just in case I needed her. The health visitor was great, ENT [ear, nose and throat] feeding specialist was really helpful and still helps if I need her.

In some cases, babies were admitted to their local hospital in a step-down arrangement after treatment at the tertiary center, which parents valued:

...our local hospital would never let us go straight home after discharge from [specialist center], they would have [baby] in for 1-2 nights so THEY knew what to expect and how to treat, they were nervous of his heart but did a great job in ensuring they knew everything about him.

Parents also valued ongoing support from the cardiac liaison nurses in the specialist centers. As one mother said:

She [cardiac liaison nurse] was a familiar face in a whirlwind of unknowns...a friend in the know.

Support from people other than health professionals was also identified as an important factor for reducing isolation and facilitating coping. Parents described 3 main sources of nonmedical support as being other "heart families," "online and charity support," and the "support of family and friends." One mother talked about wishing that she had had support:

...contact with someone who had "been there done that."

Another mother saw parents in similar situations to hers as her first port of call for support:

I will ask the other heart mums and dads first as they usually know what's what.

Although it was also evident that parents had to be ready to speak to other parents and that this was not the right approach for everyone, as one mother explained:

I'm still not at a point where seeking out other parents yet as I don't think I could share our story.

Online and charity contact were described as important and helpful sources of support, primarily as a means of having contact with other parents:

My lifeline throughout the whole experience was the "Heartline" charity forum (online) I got a tremendous amount of support and got in contact with 2 mums (both with heart children) who have supported me through the whole process.

The support from family and friends was mentioned less frequently than other sources of support, but those parents who did describe it saw it as an important facilitator of coping with the experience: "…it is the emotional support of friends and family that pulls you through," which was also attributed to

helping parents find some normality outside of their child's heart condition and care:

> *We used to go to our local to see friends just to try and get some normal life and conversation.*

Discussion

Study Findings

As far as we are aware, using a charity online forum as a systematic means of eliciting views from parents about their experiences of having a child with a health condition has not been undertaken previously within a research project. Collection of data using this method enabled us to reach a large number of potential participants, including those difficult-to-reach families who may find it more difficult or not wish to participate in projects that use more common methods of data collection such as focus groups, interviews, or questionnaires. Specifically, those families who were more isolated were able to participate, and this method offered a means of contributing to the project for those who were unable to participate in other strands of it. Furthermore, the ability to see other participants' posts may have had a positive impact on the reluctant responder and encouraged and empowered them to engage.

The 2 aims of this element of the project were to elicit parents' experiences of caring for a child with CHD and to collect information to inform the interviews with parents of high-risk infants. Both of these aims were achieved and, importantly, provided us with different information than that collected from other elements of the study. The population who participated in the online forum was broader than just parents of high-risk infants—rather, participants' responses were reflective of a wider spectrum of both diagnoses and age of the child—increasing the likely generalizability of our results to the wider population of parents of children with CHD.

In terms of eliciting parents' experiences of caring for a child with CHD, the main theme emerging from the forum was one of isolation, which parents described in terms of social and physical isolation and isolation related to knowledge. This finding was not as clearly articulated through the other methods of data collection in the wider project and provided unique insights not captured elsewhere. Physical and social isolation are common themes expressed by parents of children with health needs, including parents of children with other chronic illnesses [26], autism [27], and mothers of extremely preterm babies [28]. Parents also talked about the stress associated with knowledge about their child's condition, in terms of feeling that they did not have enough knowledge, particularly in the early stages after hospital discharge, but also the burden of responsibility as the "keeper" of the knowledge about their child's condition and the need to inform less knowledgeable professionals in primary care in particular. Over time, parents became "expert parents" in relation to the specifics of their child's condition and treatment, a phenomenon seen in parents of children with other chronic conditions [29,30], but parents also expressed some ambivalence about this role associated with the need to assume responsibility for informing health professionals outside the tertiary center about their child's condition.

The psychological and emotional impact of CHD on parents is well documented [31-34], and parents' descriptions of their anxiety and symptoms of post-traumatic stress on the online forum corroborate findings in the literature [35,36]. What we were able to additionally elicit, however, was parents' views about the specific challenges they faced following discharge after their child's surgery and their perceptions about what helped mitigate their feelings of isolation and psychological distress. Key themes that emerged were related to information, training, practical issues, and support, and examples for each of these themes were provided of things which went well and things which went badly. Our recommendations from this and other elements of the study have been reported elsewhere [7-9].

Our third aim was to comment on the usefulness of the online forum. We were confident at the outset that there was a fit with the method and the research purpose. We chose the online forum because we needed a national approach: to explore with a range of parents from different parts of the country what it was like taking a baby home after congenital heart surgery. The lack of geographical boundaries was an advantage, as without this approach, the researchers would have needed to make long-distance trips to generate the qualitative data. Reduction in costs was a further advantage, in terms of travel costs and labor; generation of an automatic transcript was initially considered a further asset. On reflection, 3 challenges emerged. The first challenge was the natural selection of specific participants, resulting in our not reaching parents from ethnically diverse populations, discussed further as a limitation in our study. The second challenge was the theoretical saturation, limited when using asynchronous interaction, where exchanges take place over time, not in real time, and this limited the opportunities for the researcher to continue to sample and code until new instances of variation for existing themes have ceased to emerge. The third challenge was the creation of an automatic transcript, which meant that some important aspects of qualitative research methodology were missing from our study, such as the nonverbal cues reported in verbatim transcripts. Despite these challenges, creating commitment online was achieved. Participants remained engaged with the forum, and although some participants only responded to 1 or 2 questions, a group of approximately 30 parents responded to questions throughout (although not necessarily every question).

Limitations

There were a number of limitations with this element of the study, which related to the method of collecting the data. The sample was predominantly white British respondents, most of whom were mothers, corroborating research which has identified that there are more female than male users of social media—particularly Facebook—and participants in online interactions tend to be predominantly white, younger, and highly educated [17,37]. One of the drivers for including this method of data collection was to enable parents who might otherwise not participate in research to have an opportunity to share their experiences. In particular, we were hoping that fathers and parents from ethnically diverse populations and those living in geographically difficult-to-reach areas would participate. Although parents came from a wide geographical area (all over the United Kingdom), we did not achieve our aim of capturing

the views of fathers or parents whose ethnicity was not white British. To participate, parents required access to the CHF website and familiarity with Facebook, which meant that parents who did not speak English and were not able or willing to use social media could not participate. Furthermore, participants required a certain level of computer skills because of the requirement to register and login to the site. It is probable that parents from a more ethnically diverse population face challenges and have different experiences, which this method of data collection prevented us from capturing. In particular, participants who have received poorer education and do not have English as their native language are more likely to find it challenging to access hospital or community support, they receive less information, and find it more difficult to understand the information given, all of which may have a negative impact on their experience of caring for their child after hospital discharge. The selection bias in our study may therefore mean that our findings are an under-representation of the difficulties that these parents face when they take their baby home after heart surgery. Furthermore, families who are in contact with charities offering support may be more likely to engage with this type of research, thus limiting the representativeness of the sample. Finally, we did not have any "entry criteria" to the forum, other than the requirement to provide some demographic details, which did not include information about the child's diagnosis. Although questions posted on the forum were related to discharge after infant cardiac surgery, some of the parents provided information and views about other stages of their journey (eg, after surgery in later childhood). Although we did not include these posts in our analysis where this was made explicit, it is possible that some posts were included where the time that parents were referring to was not identified as being other than following cardiac surgery in infancy.

We also had some challenges during the facilitation of the online forum, which was related to CHF staff leaving the organization during the time that the forum was running, which in turn meant that fewer probing questions were asked than intended. It was important that a neutral organization facilitated the forum rather than anyone from one of the specialist pediatric centers, particularly as at the time there was a national review taking place about the provision of pediatric cardiac services in the United Kingdom and whether services should be rationalized. It was also necessary that the forum was facilitated by an organization that had country-wide coverage and access to our specific target group (parents of children with CHD). CHF was clearly well placed to meet the requirements, and their staff turnover at a critical time for the project was unavoidable.

Online forums also have some inherent limitations compared with more traditional face-to-face methods of data collection. For example, nonverbal and contextual cues cannot be picked up on and specific participant comments cannot be probed, resulting in the potential loss of some richness of the data. Moreover, responses were generally much shorter than would be elicited in an interview. The automatically generated transcript, although being a benefit of this research method, was also not perfect, as has been identified previously in online forum research [38]. We did not collect information about the time spent on the site or number of visits participants made while the forum was running, and we also did not know if any participants had technical problems accessing the site at any time or whether potential participants failed to join the forum at all. However, it is also important to consider the aims of this part of the study and the fact that we were not conducting a qualitative study requiring data saturation but wanted to elicit information about parental experiences to inform the development of a topic guide for the in-depth interview in the next phase of the study.

Despite the limitations identified above, the online forum was a valuable component in our mixed-methods project. We have subsequently run another online forum for a different project, following a similar approach, in which we were able to address some of the challenges identified in this study related to the facilitation of the forum. However, reaching fathers and ethnically diverse families requires a different approach. We are aware of 1 support group in the United States that has designated part of its website specifically for fathers of children with complex heart disease, and 1 option would be to run a forum with this group, but we are not aware of anything similar for ethnically diverse families. Further work needs to be undertaken to identify more appropriate ways of engaging ethnic minorities in research, with specific attention given to the issues of culture and language.

Conclusions

Use of an online forum provided a means of eliciting the experiences of a large number of parents in caring for their child after discharge from hospital following cardiac surgery. Parents engaged with the forum and were able to articulate what went well and what went less well, as well as sharing their stories and supporting each other through doing so. At an individual level, some parents clearly found participating in the forum a positive experience in itself, as one parent said:

> *Thank you for this forum, it has helped reading other people's experiences and knowing I'm not alone.*

Although this was not one of the specific aims of the project, it demonstrates the potential of social media as a mechanism for providing support and reducing isolation. Information gained from the forum was used to shape the questions for the parent interviews in a subsequent phase of the study, particularly in terms of the barriers and facilitators parents experienced for accessing care after discharge. Furthermore, the themes identified in the online forum have contributed—in conjunction with other findings from the study—to identifying ways of improving the provision of care and support for parents of high-risk babies following discharge after cardiac surgery. Both the methods and results of the study have, we believe, wider generalizability and offer the potential of shared learning for other populations of parents with children with lifelong health conditions, particularly those with complex conditions who are receiving care in multiple settings, as well as the health professionals delivering that care. CHD can be an isolating experience for parents, particularly when their children are infants or have been discharged after surgery, and this has implications for the care and support of these children and families in primary, secondary, and tertiary care.

Conflicts of Interest

None declared.

References

1. van der Linde D, Konings EE, Slager MA, Witsenburg M, Helbing WA, Takkenberg JJ, et al. Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis. J Am Coll Cardiol 2011 Nov 15;58(21):2241-2247 [FREE Full text] [doi: 10.1016/j.jacc.2011.08.025] [Medline: 22078432]

2. National Institute for Cardiovascular Outcomes Research. Surgical Procedure Data - Summary 2015-2016 URL: https://nicor4.nicor.org.uk/chd/an_paeds.nsf/WSummaryYears?openview&RestrictToCategory=2015&start=1&count=500 [WebCite Cache ID 6twCGUhiY]

3. Tregay J, Wray J, Bull C, Franklin RC, Daubeney P, Barron DJ, et al. Unexpected deaths and unplanned re-admissions in infants discharged home after cardiac surgery: a systematic review of potential risk factors. Cardiol Young 2015 Jun;25(5):839-852. [doi: 10.1017/S1047951114002492] [Medline: 25547262]

4. Crowe S, Ridout DA, Knowles R, Tregay J, Wray J, Barron DJ, et al. Death and emergency readmission of infants discharged after interventions for congenital heart disease: a national study of 7643 infants to inform service improvement. J Am Heart Assoc 2016 May 20;5(5):e003369 [FREE Full text] [doi: 10.1161/JAHA.116.003369] [Medline: 27207967]

5. Knowles RL, Ridout D, Crowe S, Bull C, Wray J, Tregay J, et al. Ethnic and socioeconomic variation in incidence of congenital heart defects. Arch Dis Child 2017 Jun;102(6):496-502 [FREE Full text] [doi: 10.1136/archdischild-2016-311143] [Medline: 27986699]

6. Tregay J, Brown K, Crowe S, Bull C, Knowles R, Wray J. "I was so worried about every drop of milk" - feeding problems at home are a significant concern for parents after major heart surgery in infancy. Matern Child Nutr 2017 Dec;13(2):e12302. [doi: 10.1111/mcn.12302] [Medline: 26891904]

7. Tregay J, Brown KL, Crowe S, Bull C, Knowles RL, Smith L, et al. Signs of deterioration in infants discharged home following congenital heart surgery in the first year of life: a qualitative study. Arch Dis Child 2016 Oct;101(10):902-908 [FREE Full text] [doi: 10.1136/archdischild-2014-308092] [Medline: 26823534]

8. Tregay J, Wray J, Crowe S, Knowles R, Daubeney P, Franklin R, et al. Going home after infant cardiac surgery: a UK qualitative study. Arch Dis Child 2016 Apr;101(4):320-325 [FREE Full text] [doi: 10.1136/archdischild-2015-308827] [Medline: 26826171]

9. Crowe S, Knowles R, Wray J, Tregay J, Ridout DA, Utley M, et al. Identifying improvements to complex pathways: evidence synthesis and stakeholder engagement in infant congenital heart disease. Br Med J Open 2016 Dec 06;6(6):e010363 [FREE Full text] [doi: 10.1136/bmjopen-2015-010363] [Medline: 27266768]

10. Jackson AC, Frydenberg E, Liang RP, Higgins RO, Murphy BM. Familial impact and coping with child heart disease: a systematic review. Pediatr Cardiol 2015 Apr;36(4):695-712. [doi: 10.1007/s00246-015-1121-9] [Medline: 25618163]

11. Woolf-King SE, Anger A, Arnold EA, Weiss SJ, Teitel D. Mental health among parents of children with critical congenital heart defects: a systematic review. J Am Heart Assoc 2017 Feb 01;6(2):e004862 [FREE Full text] [doi: 10.1161/JAHA.116.004862] [Medline: 28151402]

12. Williams IA, Shaw R, Kleinman CS, Gersony WM, Prakash A, Levasseur SM, et al. Parental understanding of neonatal congenital heart disease. Pediatr Cardiol 2008 Nov;29(6):1059-1065 [FREE Full text] [doi: 10.1007/s00246-008-9254-8] [Medline: 18592297]

13. Staveski SL, Zhelva B, Paul R, Conway R, Carlson A, Soma G, et al. Pediatric cardiac surgery Parent Education Discharge Instruction (PEDI) program: a pilot study. World J Pediatr Congenit Heart Surg 2015 Jan;6(1):18-25. [doi: 10.1177/2150135114554659] [Medline: 25548339]

14. Rempel GR, Harrison MJ. Safeguarding precarious survival: parenting children who have life-threatening heart disease. Qual Health Res 2007 Jul;17(6):824-837. [doi: 10.1177/1049732307303164] [Medline: 17582024]

15. Rempel GR, Ravindran V, Rogers LG, Magill-Evans J. Parenting under pressure: a grounded theory of parenting young children with life-threatening congenital heart disease. J Adv Nurs 2013 Mar;69(3):619-630. [doi: 10.1111/j.1365-2648.2012.06044.x] [Medline: 22616855]

16. Chou WY, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]

XSL•FO

RenderX

17. Perrin A. Social media usage.: Pew Research Center; 2015. URL: https://static1.squarespace.com/static/56698c26e0327c6cd4b4a9c0/t/56cb504e356fb0ad1dc200dc/1456164942915/PI_2015-10-08_Social-Networking-Usage-2005-2015_FINAL.pdf [WebCite Cache ID 6twCc1KHO]

18. Gibbons MC, Casale CR. Reducing disparities in health care quality: the role of health IT in underresourced settings. Med Care Res Rev 2010 Oct;67(5 Suppl):155S-162S. [doi: 10.1177/1077558710376202] [Medline: 20829238]

19. Curl M, Robinson D. Hand-held computers in clinical audit: a comparison with established paper and pencil methods. Int J Health Care Qual Assur 1994;7(3):16-20. [doi: 10.1108/09526869410059709] [Medline: 10136781]

20. Fawcett J, Buhle Jr EL. Using the Internet for data collection. An innovative electronic strategy. Comput Nurs 1995;13(6):273-279. [Medline: 8529140]

21. Lakeman R. Using the Internet for data collection in nursing research. Comput Nurs 1997;15(5):269-275. [Medline: 9329228]

22. Anderson T, Kanuka H. On-Line forums 1: new platforms for professional development and group collaboration. J Comput Mediat Commun 1997;3(3). [doi: 10.1111/j.1083-6101.1997.tb00078.x]

23. Kramish CM, Meier A, Carr C, Enga Z, James AS, Reedy J, et al. Health behavior changes after colon cancer: a comparison of findings from face-to-face and on-line focus groups. Fam Community Health 2001 Oct;24(3):88-103. [Medline: 11563947]

24. Hsiung RC. The Best of Both Worlds: An Online Self-Help Group Hosted by a Mental Health Professional. Cyberpsychol Behav 2000;3(6):935-950. [doi: 10.1089/109493100452200]

25. Coffey A, Atkinson P. Making Sense of Qualitative Data:Complementary Research Strategies. Thousand Oaks, CA: Sage Publications; 1996.

26. Kratz L, Uding N, Trahms CM, Villareale N, Kieckhefer GM. Managing childhood chronic illness: parent perspectives and implications for parent-provider relationships. Fam Syst Health 2009 Dec;27(4):303-313. [doi: 10.1037/a0018114] [Medline: 20047354]

27. Woodgate RL, Ateah C, Secco L. Living in a world of our own: the experience of parents who have a child with autism. Qual Health Res 2008 Aug;18(8):1075-1083. [doi: 10.1177/1049732308320112] [Medline: 18650563]

28. Howe TH, Sheu CF, Wang TN, Hsu YW. Parenting stress in families with very low birth weight preterm infants in early infancy. Res Dev Disabil 2014 Jul;35(7):1748-1756. [doi: 10.1016/j.ridd.2014.02.015] [Medline: 24656293]

29. Nightingale R, Sinha MD, Swallow V. Using focused ethnography in paediatric settings to explore professionals' and parents' attitudes towards expertise in managing chronic kidney disease stage 3-5. BMC Health Serv Res 2014 Sep 18;14:403 [FREE Full text] [doi: 10.1186/1472-6963-14-403] [Medline: 25234741]

30. Smith J, Cheater F, Bekker H. Parents' experiences of living with a child with hydrocephalus: a cross-sectional interview-based study. Health Expect 2015 Oct;18(5):1709-1720 [FREE Full text] [doi: 10.1111/hex.12164] [Medline: 24341375]

31. Hearps SJ, McCarthy MC, Muscara F, Hearps SJ, Burke K, Jones B, et al. Psychosocial risk in families of infants undergoing surgery for a serious congenital heart disease. Cardiol Young 2014 Aug;24(4):632-639. [doi: 10.1017/S1047951113000760] [Medline: 23803336]

32. Lawoko S, Soares JJ. Psychosocial morbidity among parents of children with congenital heart disease: a prospective longitudinal study. Heart Lung 2006;35(5):301-314. [doi: 10.1016/j.hrtlng.2006.01.004] [Medline: 16963362]

33. Torowicz D, Irving SY, Hanlon AL, Sumpter DF, Medoff-Cooper B. Infant temperament and parental stress in 3-month-old infants after surgery for complex congenital heart disease. J Dev Behav Pediatr 2010 Apr;31(3):202-208 [FREE Full text] [doi: 10.1097/DBP.0b013e3181d3deaa] [Medline: 20375737]

34. Vrijmoet-Wiersma CM, Ottenkamp J, van Roozendaal M, Grootenhuis MA, Koopman HM. A multicentric study of disease-related stress, and perceived vulnerability, in parents of children with congenital cardiac disease. Cardiol Young 2009 Dec;19(6):608-614. [doi: 10.1017/S1047951109991831] [Medline: 19825253]

35. Fischer AL, Butz C, Nicholson L, Blankenship A, Dyke P, Cua CL. Caregiver anxiety upon discharge for neonates with congenital heart disease. Congenit Heart Dis 2012;7(1):41-45. [doi: 10.1111/j.1747-0803.2011.00600.x] [Medline: 22176595]

36. Helfricht S, Latal B, Fischer JE, Tomaske M, Landolt MA. Surgery-related posttraumatic stress disorder in parents of children undergoing cardiopulmonary bypass surgery: a prospective cohort study. Pediatr Crit Care Med 2008 Mar;9(2):217-223. [doi: 10.1097/PCC.0b013e318166eec3] [Medline: 18477936]

37. Pew Research Centre. 2017. Social media factsheet URL: http://www.pewinternet.org/fact-sheet/social-media/ [WebCite Cache ID 6twCtsVEj]

38. Im EO, Chee W. An online forum as a qualitative research method: practical issues. Nurs Res 2006;55(4):267-273 [FREE Full text] [Medline: 16849979]

Abbreviations

CHD: congenital heart disease
CHF: Children's Heart Federation
HV: health visitor

<u>Original Paper</u>

Alcohol Consumption Reduction Among a Web-Based Supportive Community Using the Hello Sunday Morning Blog Platform: Observational Study

Jessica Jane Louise Kirkman[1], BPsych (Hons), MClinPsych, PhD; Briony Leo[1], BPsych, MPsych (Counselling); Jamie Christopher Moore[1], Bachelor of Psychology and Criminology (Hons)

Hello Sunday Morning, Surry Hills, Australia

Corresponding Author:
Jessica Jane Louise Kirkman, BPsych (Hons), MClinPsych, PhD
Hello Sunday Morning
3/487 Elizabeth St
Surry Hills, 2010
Australia
Phone: 61 1300 403 196
Email: jessica@hellosundaymorning.org

Related Article:

This is a corrected version. See correction statement: http://www.jmir.org/2018/9/e11288/

Abstract

Background: Alcohol misuse is a major social and public health issue in Australia, with an estimated cost to the community of Aus $30 billion per annum. Until recently, a major barrier in addressing this significant public health issue is the fact that the majority of individuals with alcohol use disorders and alcohol misuse are not receiving treatment.

Objective: This study aimed to assess whether alcohol consumption changes are associated with participation in Hello Sunday Morning's blog platform, an online forum discussing experiences in abstaining from alcohol.

Methods: The study reports on Hello Sunday Morning participants who signed up for a 3-month period of abstinence from November 2009 to November 2016. The sample comprised 1917 participants (female: 1227/1917, 64.01%; male: 690/1917, 35.99%). Main outcome measures were Alcohol Use Disorders Identification Test (AUDIT) scores, mood, program engagement metrics, and slip-ups.

Results: Individuals who reported hazardous (preprogram AUDIT mean 11.92, SD 2.25) and harmful consumption levels (preprogram AUDIT mean 17.52, SD 1.08) and who engaged in the Hello Sunday Morning program reported a significant decrease in alcohol consumption, moving to lower risk consumption levels (hazardous, mean 7.59, SD 5.70 and harmful, mean 10.38, SD 7.43), 4 months following program commencement ($P<.001$). Those who reported high-risk or dependent consumption levels experienced the biggest reduction (preprogram mean 25.38, SD 4.20), moving to risky consumption (mean 15.83, SD 11.11), 4 months following program commencement ($P<.001$). These reductions in risk were maintained by participants in each group, 7 months following program commencement. Furthermore, those who engaged in the program more (as defined by more sign-ins, blogs posted, check-ins completed, and engagement with the community through likes and following) had lower alcohol consumption. Finally, those who experienced more slip-ups had lower alcohol consumption.

Conclusions: Participation in an online forum can support long-term behavior change in individuals wishing to change their drinking behavior. Importantly, reductions in AUDIT scores appeared larger for those drinking at high-risk and hazardous levels before program commencement. This has promising implications for future models of alcohol reduction treatment, as online forums are an anonymous, accessible, and cost-effective alternative or adjunct to treatment-as-usual. Further research is needed into the specific mechanisms of change within a Web-based supportive community, as well as the role of specific mood states in predicting risky drinking behavior.

XSL•FO
RenderX

Introduction

Background

Alcohol misuse is a major social and public health issue in Australia, with an estimated cost to the community of Aus $30 billion per annum [1,2]. In Australia, alcohol consumption is estimated to cause 3.2% of the total burden of disease, contributing approximately 188,000 disability adjusted life years, causing 5550 deaths per annum [1,2]. In high-income and middle-income countries, the costs associated with alcohol misuse amount to more than 1% of the gross national product, represented largely in social harm and health costs [3]. Government policies targeted at reducing alcohol consumption have focused on education about the risks and harms of alcohol consumption and reducing the accessibility and affordability of alcohol.

Although alcohol misuse and alcohol use disorder (AUD) can be seen as distinct from each other, it can be argued that they are on the same trajectory. Most theories of alcohol dependency describe high-risk drinkers experiencing nervous system adaptations from repeated alcohol exposure—a process which underlies the development of an alcohol addiction [4,5].

Until recently, a major barrier in addressing this significant public health issue is the fact that the majority of individuals with both high-risk drinking and AUDs are not receiving treatment [6]. For this population, a significant treatment gap exists whereby a *hidden* population of individuals with problematic alcohol use is dissuaded from seeking help due to a lack of accessibility of services, stigmatization, low motivation, or cost of treatment [7]. Furthermore, substantial evidence exists to show that those with less severe alcohol abuse are not offered interventions, and those who are gainfully employed are even less likely to receive any type of treatment [8]. What is notable though is that most alcohol-related harm arises from those who consume in an "at-risk" manner rather than those with an AUD [9].

This *hidden* population of individuals with problematic alcohol use is a prime candidate for internet-based interventions, as a first step toward seeking help [10]. It is this population who are difficult to engage in traditional treatment services but from whom internet-based interventions are ideal due to their low cost and relative convenience. Internet-based interventions have promising applications in the treatments of addiction, with advantages including anonymity, convenience, accessibility, cost-effectiveness, and privacy [11,12]. Internet therapies can also target clients in different stages of change [13], allowing potential clients to investigate treatment options without the shame and guilt associated with face-to-face interactions. Additionally, these therapies have the potential to be more consistent than face-to-face therapy delivery, with structured programs that are delivered with clinical fidelity [14].

Internet-based interventions can also provide the context for screening and brief intervention (SBI) for those drinking to "at-risk" levels. SBI has a growing evidence base for its effectiveness [15-18]; however, there is an absence of evidence for those experiencing more severe alcohol use problems [15,17,19]. Reviews of previous studies of electronic SBIs have also found inconsistent evidence as to their effectiveness with high-risk drinkers, with an acknowledgment that, although Web-based interventions appear to be well received, further controlled trials examining engagement and efficacy are needed within this population [20].

It is important to acknowledge that, in this emerging field, there are many terminologies to describe interventions delivered over the internet. Common terms include e-mental health, Web-based-therapy, mHealth, cybertherapy, and telemental health. eHealth describes the range of modalities that deliver therapeutic services—such as telephone, internet, email, and videoconferencing [21]. This differs from telerehab, which is rehabilitation-focused rather than a psychological or psychosocial intervention. For ease of reference, in this paper, the term eHealth will be used to describe services delivered online.

In terms of treatment, the existing evidence for eHealth shows promising results across multiple populations and demographics [22-26]. Again, the populations targeted in most of the research have been for problematic drinking rather than those with diagnosed AUDs, making it difficult to ascertain whether these interventions are useful for those with a serious drinking problem. More recent research that has focused on individuals with a specifically diagnosed AUD has found evidence for the efficacy of a stand-alone mobile phone alcohol reduction app, as well as internet-based interventions with bibliotherapy [27]. For both of these conditions, greater utilization of the system provided greater results over the course of the study, and participants had control over how quickly they progressed through their treatment—and were able to access the material when needed, in contrast to more traditional treatment services. The research into eHealth explores benefits such as cost-effectiveness, flexibility of design, replicability, interactivity, and ease of use, as well as the question of whether these interventions are as effective as face-to-face programs, and as adjuncts to treatment in primary care settings [28].

In terms of less structured or formalized Web-based interventions, the internet also provides individuals the opportunity to connect with others and participate in activities such as blogging. Blogging has gained popularity over the past 10 years, with most blogs fitting the description of regular, date stamped articles that represent a timeline, and are used by the author for personally oriented communication [29]. They differ from traditional print and digital media in terms of their flexibility, interactivity, informal structure, and engagement with readers, with users having the opportunity to express their

opinions on posts and in a public forum. Research into internet users has identified convenience, a sense of community, and information seeking as the major reasons for visiting blogs [30]. The impacts of blogging can also be behaviorally empowering and tie in with intrinsic motivations [31], as well as providing a cathartic space to process difficult emotions [32].

Research has also established that a person's social network can either increase or decrease their alcohol use [33]. Today, that includes an individual's online social network. These social networks can allow a user to post a blog with text and images or engage in group discussions [34]. Although the impacts of social networks and blogging on behavior change are inconsistent in relation to diet and exercise [35], their impact on reducing alcohol consumption has not been determined [34,35].

Hello Sunday Morning

Hello Sunday Morning (HSM) is an Australian social media health promotion "movement" that asks participants to publicly set a personal goal to stop drinking or reduce their consumption, for a set period of time, and to record their reflections and progress on blogs and social networks (see Figures 1 and 2). The action of setting a goal both motivates members and holds them accountable, while also creating a message about healthy relationships with alcohol within their peer group. Created in 2010, HSM has developed a Web-based platform that combines blogging, social media, and gamification (structured games that facilitate participation and engagement with the Web-based community). HSM is unique in that the program participants produce all of the content based on their individual experiences, using the vernacular of their peer group.

HSM has effectively built a strong culture that governs the norms and values of the community. Carah et al [36] suggested that HSM—like fitness apps, quit smoking initiatives, and mindfulness programs—is worth studying to understand how social media health promotion and treatment can overlap.

Preliminary research has been conducted on the HSM platform, exploring blog content and qualitative reports of change. First, text analytics of the blog posts on the platform show that participants typically begin with descriptions of their drinking practices and change over time to reflect their efforts at change, and their aspirations [37]. Early evaluations have also revealed that 84% of Australian HSM users reported completing the program time without a "slip-up" (defined as a drink before finishing the program), reduced their alcohol use, reported improved mental health, and experienced a change in their perceptions of alcohol over time [38]. Specifically, participants reported increased understanding of the negative effects of alcohol, a decrease in the desire to drink alcohol for fun, and a decrease in the likelihood to drink alcohol to relieve tension. Further evaluations have demonstrated members shift from being self-focused to reflecting on the role of alcohol in society and developing a desire to support others [38]. A study of high-risk Victorian HSM participants also found that the majority of participants reported low-risk drinking at 1 month and reported improved physical health, feeling positive about themselves, greater productivity, engagement in new activities, improved mental health, new or improved relationships, and financial savings [39].

More recent research on the HSM platform and its participants has revealed that 64% of participants were under the age of 40 years, were more likely to be female (which is the opposite of the Australian alcohol treatment population), and are riskier drinkers than other treatment-seeking populations [36]. These participants most commonly selected "fitness" and "mind and body" as their goal, whereas "sobriety" was the third most significant goal. Those drinking at high-risk levels were twice as likely to elect "alcohol"-related goals, compared with those drinking at lower risk levels. Carah et al [36] concluded that through qualitative analysis, they believe heavier drinking participants may use HSM as a treatment program. As such, it was suggested that further research is required to examine whether participants self-reported alcohol consumption changes after participation [36].

Figure 1. Sample blog post (Hello Sunday Morning, 2018).
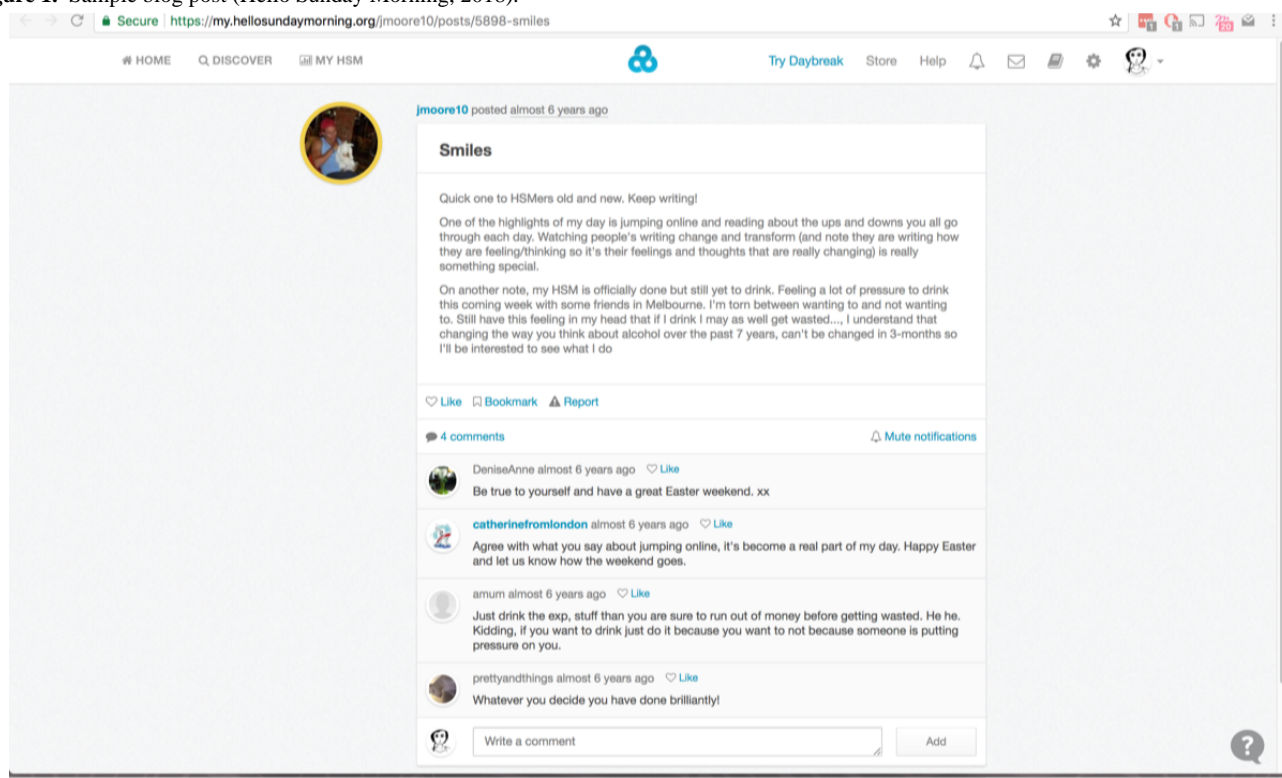
Figure 2. Sample profile of Hello Sunday Morning (HSM) member (Hello Sunday Morning, 2018).
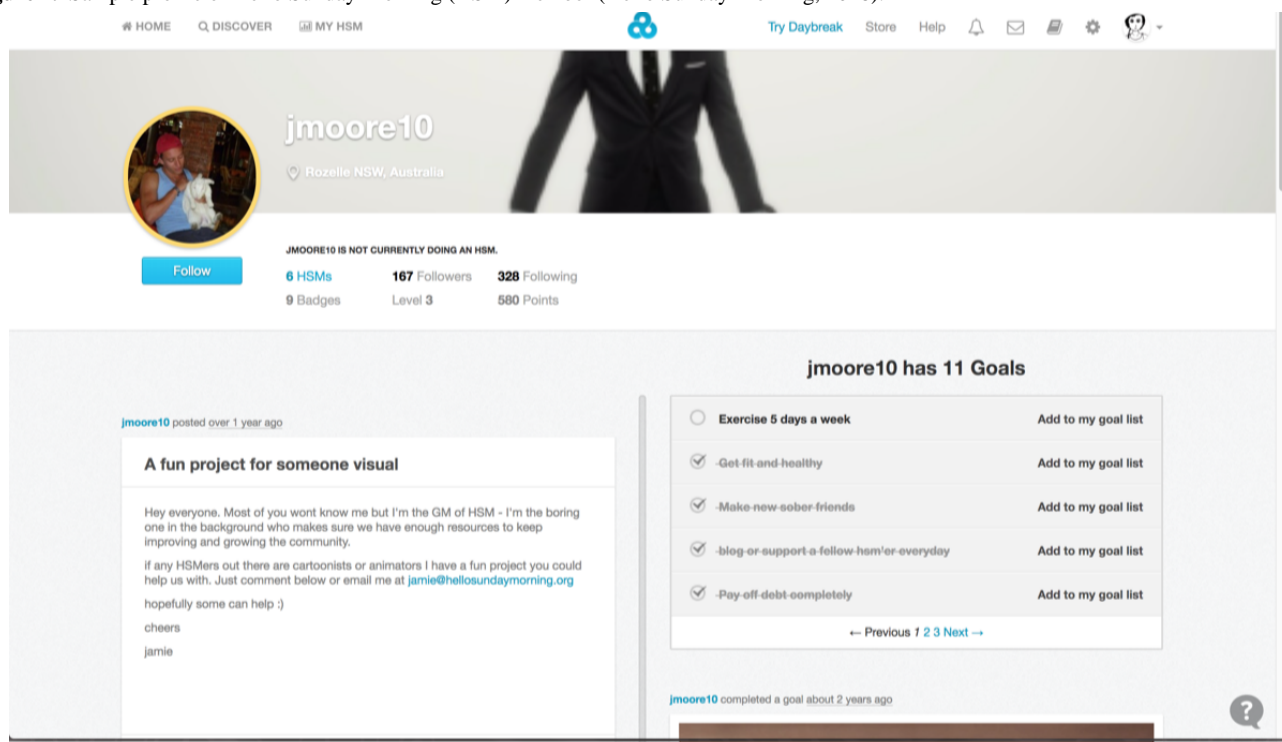
This Study

This study, therefore, aimed to address outstanding questions on whether alcohol consumption changes are associated with participation on the HSM platform and explore the reasons why this may be the case or what factors are driving the change. The primary outcome will be change in Alcohol Use Disorders Identification Test (AUDIT) scores. The AUDIT was developed by the World Health Organization and has been validated for use in Australia [40]. The AUDIT has been used as an outcome measure in previous evaluations of the HSM platform [38,39] as well as other clinical populations [41-44]. Although the gold standard for evaluating service efficacy is a randomized controlled trial design, such a design would be impractical when offering this service for the first time to the previously described *hidden* population who are not traditional treatment seekers. To

allow for a preliminary trial of the technology, the Web-based program was openly offered to all eligible participants. This study is the first in a series examining the service, its reach, and its therapeutic components, including a randomized controlled trial currently under way.

In this study, the following was hypothesized: (1) that participants' AUDIT score will significantly decrease from preprogram to 1-month follow-up (4 months after program commencement), and this will be maintained at 3-month follow-up (7 months after program commencement), and further, that this change will occur in participants across all 3 AUDIT risk zones; (2) that 1- and 3-month AUDIT change scores will be negatively correlated with program engagement metrics including the number of sign-ins, posts, comments received, comments given, followers, and people following; (3) that 1- and 3-month AUDIT change scores will be negatively correlated with mood and positively correlated with slip-ups.

Methods

Participants

This study reports data on HSM participants who have signed up for a 3-month period of abstinence from November 2009 to November 2016. Although some participants complete multiple HSM periods of abstinence, data reported in this study reflect participants' first engagement with the program. Participants classified as low-risk consumption (as measured by Zone 1 on the AUDIT) were excluded from the sample used in analyses, as they were not identified as having unhealthy drinking levels. The resulting sample consisted of program participants who signed up to complete one 3-month HSM experience, screened as Zones II, III, or IV on the AUDIT at program commencement, signed in at least once, and completed at least 1 follow-up survey (ie, 1- or 3-month).

Materials

Alcohol Use Disorders Identification Test

The AUDIT is a 10-item measure assessing alcohol consumption [40]. Participants are required to respond to items (eg, "How often do you have 6 or more drinks on one occasion?") on a 5-point Likert scale, with each question receiving a score with a range of 0-4 points, with a potential maximum total score of 40 points. In Australian usage of the AUDIT, a score between 0 and 7 indicates low-risk consumption, a score between 8 and 15 indicates risky or hazardous consumption, a score between 16 and 19 indicates harmful consumption, and a score of more than 20 indicates high-risk or dependent consumption [45].

Mood

Mood was measured by a single item ("How did you feel this week?") asked at each weekly check-in, with members responding on a 4-point visual scale of faces ranging from very happy to very sad.

Slip-Ups

Slip-ups were measured by a single item ("How did you go this week?") asked at each weekly check-in, with members

responding with either "I'm still going strong," indicating they did not drink, or "I slipped up," indicating they did drink.

Program Engagement

Program engagement is measured by a number of factors that were automatically logged within the website. This includes a count of the total number of sign-ins, posts made on the platform, comments received, comments given, followers, and people following the member. The ability to capture these metrics was intentionally created by the website developers, rather than relying on other metrics such as Google analytics.

Procedure

When participants register on the HSM Web platform, they accept the terms and conditions which state that information about their participation and alcohol use is used in research evaluations. Participants, therefore, provided informed consent for their deidentified data to be used for research purposes. Three months following the completion of their 3-month abstinence period, they are sent an email prompting them to check in to see how their relationship with alcohol has changed since signing up. They are offered a 25% discount on any HSM store purchase if they complete the follow-up survey within 48 hours.

Data Analysis

Results are presented for 2 different sets of data analyses. Data were cleaned and checked for appropriate assumptions of normality, sphericity, and outliers in line with the analyses conducted. First, a set of repeated measures analysis of variance (ANOVA) was conducted to explore reductions in AUDIT score following program completion. Then, separate bivariate Pearson correlations were conducted to determine associations with 1- and 3-month follow-up AUDIT change scores, with various program engagement metrics, and with mood and slip-ups. AUDIT change scores were computed by subtracting program commencement AUDIT score from 1- and 3-month follow-up AUDIT score, with the resulting variable indicating negative scores for AUDIT reduction and positive scores for AUDIT increase. All ANOVAs were conducted, with an overall alpha of .01 to provide a more conservative control of type-I error, with all correlational analyses conducted at an alpha of .05 as they were observational in nature.

Results

Demographic Information

Participants were 1917 adults with an average age of 46 years (SD 11.71). Overall, 64.00% (1227/1917) of participants were female and 35.02% (690/1917) were male, with the highest proportion reporting they reside in Australia (709/1917, 36.99%), participants followed by the United States (498/1917, 25.98%), the United Kingdom (249/1917, 12.99%), and Canada (153/1917, 7.98%). In total, 25.98% (498/1917) of participants screened as Zone II on the AUDIT, 20.41% (402/1917) screened as Zone III, and the remaining 54% of participants screened as Zone IV (1029/1917). Participants also spent, on average, 41 min and 45 seconds per visit to the HSM site and visited, on average, 6 pages.

Audit Score and Time in Program

To test hypothesis 1, a set of repeated measures ANOVAs was conducted to investigate any reductions in AUDIT score from program commencement to 1- and 3-month follow-up. Three separate ANOVAs were conducted to determine reduction in AUDIT score for participants who screened Zones II, III, or IV, respectively, at program commencement. Intention to treat data is reported, in that participants who failed to complete a follow-up survey at 1- or 3-month follow-up had their sign-up AUDIT score imputed at the follow-up points. Across all participants at program commencement, mean AUDIT score was 20.32 (SD 6.66), placing them on average in Zone IV of the AUDIT. At 1-month follow-up, the mean score was 12.60 (SD 9.95), and at 3-month follow-up, it was 8.90 (SD 8.86).

The assumption of sphericity was violated for all 3 ANOVAs; therefore, Greenhouse-Geisser F values and degrees of freedom are reported. For participants who screened Zone II at program commencement, there was a significant main effect of time, $F_{1.98,916.53}=401.28$, $P<.001$, $\eta_p^2=.45$, with pairwise comparisons indicating AUDIT scores reduced significantly from program commencement to 1-month follow-up and then reduced significantly again at 3-month follow-up (see Table 1 for descriptive statistics). The result showed that participants who screened Zone II at program commencement had reduced to Zone I (low-risk drinking) at 3-month follow-up.

For participants who screened Zone III at program commencement, there was a significant main effect of time, $F_{1.92,771.14}=437.16$, $P<.001$, $\eta_p^2=.52$, with pairwise comparisons indicating AUDIT scores reduced significantly from program commencement to 1-month follow-up and then reduced significantly again at 3-month follow-up (see Table 1 for descriptive statistics). The result showed that participants who screened Zone III at program commencement had reduced to borderline Zone I (low-risk drinking) at 3-month follow-up (M=7.67).

For participants who screened Zone IV at program commencement, there was a significant main effect of time, $F_{1.92,1975.28}=1067.23$, $P<.001$, $\eta_p^2=.51$, with pairwise comparisons indicating AUDIT scores reduced significantly from program commencement to 1-month follow-up and then reduced significantly again at 3-month follow-up (see Table 1 for descriptive statistics). The result showed that participants who screened Zone IV (high-risk or dependent drinking) at program commencement had reduced to Zone II (risky or hazardous drinking) at 3-month follow-up. Overall, 58.3% of the sample had a reduction at 1-month follow-up and 89.3% had a reduction at 3-month follow-up.

Table 1. Alcohol Use Disorders Identification Test (AUDIT) score change across time by pre-AUDIT screening level.

Consumption	Pre, mean (SD)	1 month, mean (SD)	3 months, mean (SD)
Zone II[a] (n=485)	11.92 (2.25)	7.59 (5.70)	5.39 (5.32)>
Zone III[b] (n=402)	17.52 (1.08)	10.38 (7.43)	7.67 (6.99)
Zone IV[c] (n=1029)	25.38 (4.20)	15.83 (11.11)	11.04 (10.14)

[a]Zone II: risky or hazardous consumption.

[b]Zone III: harmful consumption.

[c]Zone IV: high-risk or dependent consumption.

Table 2. Pearson correlations between Alcohol Use Disorders Identification Test (AUDIT) scores and engagement variables.

Variables	1-month AUDIT[a] score change	3-month AUDIT score change	Sign in count	# of posts	Likes received	Comments received	Comments given	Following	Followed	Check-ins
1-month AUDIT score change	—	−.66[d]	−.12[d]	−.10[c]	−.13[d]	−.05	−.03	−.14[d]	−.11[c]	−.22[d]
3-month AUDIT score change	—	—	−.09[d]	−.07[b]	−.12[d]	−.04	−.03	−.09[b]	−.09[b]	−.23[d]
Sign-in count	—	—	—	.60[d]	.57[d]	.68[d]	.67[d]	.43[d]	.66[d]	.29[d]
# of posts	—	—	—	—	.84[d]	.83[d]	.73[d]	.40[d]	.76[d]	.27[d]
Likes received	—	—	—	—	—	.84[d]	.72[d]	.54[d]	.81[d]	.23[d]
Comments received	—	—	—	—	—	—	.93[d]	.44[d]	.81[d]	.16[d]
Comments given	—	—	—	—	—	—	—	.37[d]	.78[d]	.12[c]
Following	—	—	—	—	—	—	—	—	.75[d]	.14[d]
Followed	—	—	—	—	—	—	—	—	—	.26[d]
Check-ins	—	—	—	—	—	—	—	—	—	—

[a]AUDIT: Alcohol Use Disorder Identification Test.

[b]$P<.05$.

[c]$P<.01$.

[d]$P<.001$.

Audit Score and Program Engagement

To test hypothesis 2, separate bivariate Pearson correlations were conducted to determine associations with 1- and 3-month follow-up AUDIT change scores, with various program engagement metrics. It was hypothesized that an increase in program engagement via following, sign-ins, posts, and comments would result in greater 1- and 3-month change scores.

Results indicate that, as expected, 1- and 3-month AUDIT change scores were negatively associated with more sign-ins, posts, received likes, following, followers, and check-ins (see Table 2), indicating the greater the reduction in AUDIT scores, the more of each of these engagement metrics participants had.

The number of received comments or comments given did not correlate with 1- or 3-month AUDIT change.

AUD Identification Test Score and Mood

To test hypothesis 3, that 1-month and 3-month AUDIT change would be negatively correlated with mood (ie, worse mood would result in higher AUDIT scores), and positively correlated with slip-ups (ie, more slip-ups would result in higher AUDIT scores), separate bivariate Pearson correlations were conducted. To run the analysis, the good and bad check rate was calculated, with good check indicating the proportion of check-ins where the participant indicated they had not had a drink that week, and bad check indicating the proportion of check-ins where the participant indicated they did have a drink that week.

Table 3. Pearson correlations between Alcohol Use Disorders Identification Test (AUDIT) scores, mood, and check-in ratings.

Variables	1-month AUDIT[a] score change	3-month AUDIT score change	Good check	Bad check	Positive check	Negative check
1-month AUDIT score change	—	.66[d]	.09[c]	−.25[d]	−.23[d]	−.02
3-month AUDIT score change	—	—	.07[b]	−.26[d]	−.24[d]	−.01
Good check	—	—	—	−.01	.08[b]	.68[d]
Bad check	—	—	—	—	.96[d]	.21[d]
Positive check	—	—	—	—	—	.17[d]
Negative check	—	—	—	—	—	—

[a]AUDIT: Alcohol Use Disorder Identification Test.

[b]$P<.05$.

[c]$P<.01$.

[d]$P<.001$.

Moreover, positive and negative check-in was calculated, with positive check-in representing the proportion of check-ins where the participant indicated they had a positive mood that week, and negative check-in representing the proportion of check-ins where the participant indicated they had a negative mood that week.

Results indicated that the good check rate was positively correlated with 1- and 3-month AUDIT change scores, and bad check was negatively correlated with 1- and 3-month AUDIT change scores (see Table 3). Contrary to expectations, this indicated that more slip-ups were correlated with lower 1-and 3-month AUDIT change scores, and less slip-ups were associated with higher 1- and 3-month AUDIT change scores. This indicated those who had reduced their AUDIT scores had more slip-ups during the program, than those whose AUDIT scores had increased. However, as expected, positive check-in was negatively correlated with AUDIT change scores, indicating that more check-ins where mood for the week was positive were associated with greater reduction in AUDIT scores. Interestingly, negative check-in (ie, the proportion of check-ins where mood that week was low) was not correlated with either 1- or 3-month AUDIT change scores.

Discussion

Principal Findings

Alcohol use is a major social and public health issue in Australia and around the world, with the major barriers for accessing treatment including accessibility of services, stigmatization, low motivation, and cost of treatment [7]. Alarmingly, there is a significant gap between those who exhibit problematic alcohol use and those that receive some form of treatment [6]. Furthermore, those with less severe alcohol abuse are not offered any form of intervention [8]. It is clear that emerging technologies may offer a low-cost, accessible alternative that may be able to support this *hidden* population; however, limited research exists in regard to the efficacy of blogging and social media engagement on alcohol consumption.

This study aimed to explore whether participation in the HSM platform is associated with changes in alcohol consumption. The HSM program asks participants to set a goal to stop or reduce their drinking for a set time, and to record their reflections and progress through blogs, while being connected with others within the community working toward similar goals.

As with previous research on the platform, participants were predominately female, and were beginning the program with predominately harmful or high-risk consumption levels [36]. Results revealed that individuals who reported hazardous and harmful consumption levels and who engaged in the HSM program reported a significant decrease in alcohol consumption, moving to low-risk consumption levels 4 months following program commencement. Those who reported high-risk or dependent consumption levels before program engagement experienced the biggest reduction, moving to risky consumption 4 months following program commencement. These reductions in risk were maintained by participants in each AUDIT screening zone, 7 months following program commencement. Given the

cost and wait times for traditional treatment programs, this is a significant finding, suggesting long-term behavior change can occur through the use of an easily accessible Web-based program. Furthermore, it is important to note that this program was more effective for those drinking to high-risk and hazardous levels.

As expected, those who engaged in the program more had significantly lower levels of alcohol consumption at 1- and 3-month follow up. In particular, significant improvements were seen for those who signed-in more, posted more blogs, received more "likes" from other community members, followed more community members, had more followers, and completed more check-ins.

It appears the engagement with the community and peer support was a key ingredient in the successful behavior change of HSM participants, with such peer-to-peer communities being described as one of the most transformational features of the internet [46]. Through this model, individuals with multiple barriers have the opportunity to connect and create supportive communities, with narrative expression having demonstrated psychological benefits for people experiencing chronic illness—allowing reflection, connection, and meaning-making [47-49].

The benefits of such expressive writing may include acknowledging and validating experiences, communicating the experience of recovery and relapse, and allowing for reflection of the current reality. The reflective nature of narrating personal experiences encourages self-disclosure and sharing of thoughts and emotions, as well as encouraging the participant to disclose facts about themselves and reduce inhibition [49], as well as providing an opportunity to externalize their inner experiences [50]. When this content is shared within a trusted and supportive community, it appears that participants feel validated, and in turn offer support and validation to other members of the community, thus creating a network of participants working toward the same goal of long-term behavior change [46,51,52].

The last hypothesis was that 1- and 3-month AUDIT scores would be negatively correlated with mood and positively correlated with slip-ups. However, results showed that negative mood was not related to alcohol consumption, but positive mood was significantly related to lower alcohol consumption. This is an interesting finding and warrants further investigation, particularly around what specific mood states predict drinking to excess, rather than simply positive versus negative emotions. Finally, results also show that more slip-ups were associated with lower alcohol consumption. This suggests that slip-ups provided more opportunities to learn about triggers, responses, and strategies to persist with rather than just focusing solely on staying sober. When supporting individuals experiencing a slip-up or lapse, it is therefore important to take advantage of the learning opportunity and facilitate processing, reframing, and learning rather than focusing on or trying to resolve feelings of weakness, guilt, and hopelessness.

Limitations

Despite the findings in this study, a number of limitations should be recognized. As noted, this study did not include a control group or benchmark group to compare to, and did not control

for variables such as additional treatment (psychological or pharmacological) and support. Although this limits the conclusions that can be drawn from the analyses, it has provided preliminary data to support the currently underway research projects utilizing best practice research designs. Additionally, although results showed participants were spending an average of 41 min per visit on the HSM site, the exact nature of how they spent their time on the site is unknown. They may have been predominately reading blog posts, writing their own blog posts, or commenting on other members' posts. Future research exploring the different functions of the program must address this limitation.

Finally, mood was measured by a single visual scale simply describing positive and negative mood through the use of a happy or sad face dichotomy. It is recommended that a more descriptive and validated measure is used in the future, to allow more reliable evaluation of mood and mood states. Although more comprehensive measures are recommended for assessment of mood, it can also be argued that traditional questionnaires assessing mood can present issues such as cognitive load, overall burden, and longer completion times [53,54]. Previous well-validated pictorial measures of mood include the Self-Assessment Manikin, a set of 3 pictorial assessment scales measuring pleasure, arousal, and dominance [55], as well as the Smileyometer [56], the AffectButton [57], and EmoCards [58]. Each of these measures aims for respondents to report their affective state quickly and accurately. An additional benefit to these scales is that they can be used reliably across cultures [59], as well as being time-efficient ways for respondents to convey their affective state.

Future measures may include both multidimensional tools to assess mood state, as well as brief, unidimensional pictorial measures. As mood disorders have a long-established link to alcohol use [60-62], it will be useful in future research to also utilize measures such as the Depression Anxiety and Stress Scale and the K10, which are more able to assess the multidimensional nature of mood disorders and co-occurring depression, anxiety, and stress.

Conclusions

The findings of this study highlight a number of future directions for program development and future investigation into the elements involved in the HSM program. These findings suggest that it is possible to bring about significant change in alcohol consumption through the use of an online platform and Web-based community. The large number of participants drinking to high-risk or dependency levels has, however, highlighted the need for further accessible targeted clinical support.

The results of this study indicate that being part of a supportive Web-based community, as well as having the opportunity to reflect on past experiences, may provide participants with the resources needed to create lasting behavior change. This is a crucial service for the *hidden* population of individuals with problematic alcohol use who do not seek out traditional services, or are not identified as being in-need of tailored support based on assessments or lack of, by their primary health providers. An added benefit to this is that the Web-based community is highly accessible, in that members are able to use it from their mobile phones or laptops, as well as utilizing the well-validated model of a supportive "sober" network. This can be compared with the traditional model of alcohol treatment, which is quite rigid in its service delivery and approaches—a shortcoming that is reflected in its poor engagement and outcomes [63]. It is hoped that with the introduction of new technologies, the social support model will continue to be utilized and the *hidden* population of high-risk drinkers will be able to access ongoing and effective support.

Conflicts of Interest

The authors of this paper are all employees of Hello Sunday Morning.

References

1. Health. The costs of tobacco, alcohol and illicit drug abuse to Australian society in 2004/2005 URL: https://www.health.gov.au/internet/drugstrategy/publishing.nsf/Content/34F55AF632F67B70CA2573F60005D42B/%24File/mono64.pdf [accessed 2018-05-07] [WebCite Cache ID 6zEhNlDed]

2. Laslett A, Catalano P, Chikritzhs T, Dale C, Doran C, Ferris J, et al. The range and magnitude of alcohol's harm to others. Fitzroy, Victoria: AER Centre for Alcohol Policy Research, Turning Point Alcohol and Drug Centre, Eastern Health; 2010.

3. Rehm J, Mathers C, Popova S, Thavorncharoensap M, Teerawattananon Y, Patra J. Global burden of disease and injury and economic cost attributable to alcohol use and alcohol-use disorders. Lancet 2009 Jun;373(9682):2223-2233. [doi: 10.1016/S0140-6736(09)60746-7]

XSL•FO

RenderX

4.   Corbin WR, Scott C, Leeman RF, Fucito LM, Toll BA, O'Malley SS. Early subjective response and acquired tolerance as predictors of alcohol use and related problems in a clinical sample. Alcohol Clin Exp Res 2013 Mar;37(3):490-497 [FREE Full text] [doi: 10.1111/j.1530-0277.2012.01956.x] [Medline: 23347236]

5.   Morean ME, Corbin WR. Subjective alcohol effects and drinking behavior: the relative influence of early response and acquired tolerance. Addict Behav 2008;33(10):1306-1313 [FREE Full text] [doi: 10.1016/j.addbeh.2008.06.007]

6.   Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bull World Health Organ 2004;82(11):858-866. [doi: 10.1590/S0042-96862004001100011]

7.   Room R, Babor T, Rehm J. Alcohol and public health. Lancet 2005;365(9458):519-530 [FREE Full text] [Medline: 15705462]

8.   de Graaf R, Ten Have M, van Dorsselaer S. The Netherlands Mental Health Survey and Incidence Study-2 (NEMESIS-2): design and methods. Int J Methods Psychiatr Res 2010 Jul;19(3):125-141. [doi: 10.1002/mpr.317] [Medline: 20641046]

9.   Health. 2009. Guidelines for the Treatment of Alcohol Problems URL: https://www.health.gov.au/internet/main/publishing.nsf/Content/0FD6C7C289CD31C9CA257BF0001F96BD/$File/AustAlctreatguidelines%202009.pdf [accessed 2018-05-06] [WebCite Cache ID 6zDCUm5pe]

10. Blankers M, Koeter M, Schippers GM. Evaluating real-time internet therapy and online self-help for problematic alcohol consumers: a three-arm RCT protocol. BMC Public Health 2009 Jan 14;9:16 [FREE Full text] [doi: 10.1186/1471-2458-9-16] [Medline: 19144162]

11. University of Sydney. 2009. Internet therapy for problem gambling: A literature review and environmental scan URL: http://www.greo.ca/Modules/EvidenceCentre/files/Monaghan%20et%20al%282009%29Internet_therapy_for_PG.pdf [accessed 2018-05-07] [WebCite Cache ID 6zEhaOzWy]

12. Monaghan S, Blaszczynski A. Internet-based interventions for the treatment of problem gambling. Toronto: Centre for Addiction and Mental Health; 2009.

13. Prochaska JO, DiClemente CC. Transtheoretical therapy: toward a more integrative model of change. Psychother Theory Res Pract 1982;19(3):276-288 [FREE Full text] [doi: 10.1037/h0088437]

14. Carroll KM, Rounsaville BJ. A vision of the next generation of behavioral therapies research in the addictions. Addiction 2007 Jun;102(6):850-862 [FREE Full text] [doi: 10.1111/j.1360-0443.2007.01798.x] [Medline: 17523974]

15. Bien TH, Miller WR, Tonigan JS. Brief interventions for alcohol problems: a review. Addiction 1993 Mar;88(3):315-336. [doi: 10.1111/j.1360-0443.1993.tb00820.x]

16. Landy MS, Davey CJ, Quintero D, Pecora A, McShane KE. A systematic review on the effectiveness of brief interventions for alcohol misuse among adults in emergency departments. J Subst Abuse Treat 2016 Feb;61:1-12. [doi: 10.1016/j.jsat.2015.08.004] [Medline: 26482134]

17. Moyer A, Finney JW, Swearingen CE, Vergun P. Brief interventions for alcohol problems: a meta-analytic review of controlled investigations in treatment-seeking and non-treatment-seeking populations. Addiction 2002;97(3):279-292. [doi: 10.1046/j.1360-0443.2002.00018.x]

18. Richardson LK, Frueh BC, Grubaugh AL, Egede L, Elhai JD. Current directions in video conferencing tele-mental health research. Clin Psychol (New York) 2009;16(3):323-338. [doi: 10.1111/j.1468-2850.2009.01170.x]

19. O'Donnell A, Anderson P, Newbury-Birch D, Schulte B, Schmidt C, Reimer J, et al. The impact of brief alcohol interventions in primary healthcare: a systematic review of reviews. Alcohol Alcohol 2014;49(1):66-78 [FREE Full text] [doi: 10.1093/alcalc/agt170] [Medline: 24232177]

20. Saitz R. Alcohol screening and brief intervention in primary care: absence of evidence for efficacy in people with dependence or very heavy drinking. Drug Alcohol Rev 2010 Nov;29(6):631-640 [FREE Full text] [doi: 10.1111/j.1465-3362.2010.00217.x] [Medline: 20973848]

21. Richardson LK, Frueh BC, Grubaugh AL, Egede L, Elhai JD. Current directions in videoconferencing tele-mental health research. Clinical Psychology 2009 Sep 1;16(3):323-338 Same as 18; delete [FREE Full text] [doi: 10.1111/j.1468-2850.2009.01170.x] [Medline: 20161010]

22. Bewick BM, Trusler K, Barkham M, Hill AJ, Cahill J, Mulhern B. The effectiveness of web-based interventions designed to decrease alcohol consumption--a systematic review. Prev Med 2008 Jul;47(1):17-26. [doi: 10.1016/j.ypmed.2008.01.005] [Medline: 18302970]

23. Rooke S, Thorsteinsson E, Karpin A, Copeland J, Allsop D. Computer-delivered interventions for alcohol and tobacco use: a meta-analysis. Addiction 2010 Aug;105(8):1381-1390. [doi: 10.1111/j.1360-0443.2010.02975.x] [Medline: 20528806]

24. Khadjesari Z, Murray E, Hewitt C, Hartley S, Godfrey C. Can stand-alone computer-based interventions reduce alcohol consumption? A systematic review. Addiction 2011;106(2):267-282. [doi: 10.1111/j.1360-0443.2010.03214.x] [Medline: 21083832]

25. Carey KB, Scott-Sheldon LA, Elliott JC, Garey L, Carey MP. Face-to-face versus computer-delivered alcohol interventions for college drinkers: a meta-analytic review, 1998 to 2010. Clin Psychol Rev 2012 Dec;32(8):690-703 [FREE Full text] [doi: 10.1016/j.cpr.2012.08.001] [Medline: 23022767]

26. Vernon ML. A review of computer-based alcohol problem services designed for the general public. J Subst Abuse Treat 2010 Apr;38(3):203-211 [FREE Full text] [doi: 10.1016/j.jsat.2009.11.001] [Medline: 20015607]

27. Gonzalez VM, Dulin PL. Comparison of a smartphone app for alcohol use disorders with an Internet-based intervention plus bibliotherapy: a pilot study. J Consult Clin Psychol 2015 Apr;83(2):335-345 [FREE Full text] [doi: 10.1037/a0038620] [Medline: 25622202]

28. Riper H, Spek V, Boon B, Conijn B, Kramer J, Martin-Abello K, et al. Effectiveness of E-self-help interventions for curbing adult problem drinking: a meta-analysis. J Med Internet Res 2011 Jun 30;13(2):e42 [FREE Full text] [doi: 10.2196/jmir.1691] [Medline: 21719411]

29. Kirchoff L, Bruns A, Nicolai T. Investigating the impact of the blogosphere: Using PageRank to determine the distribution of attention URL: http://snurb.info/files/talks2007/Blogosphere%20Impact.pdf [accessed 2017-11-28] [WebCite Cache ID 6vItlUnY5]

30. Johnson TJ, Kaye BK, Bichard SL, Wong WJ. Every blog has its day: politically-interested Internet users' perceptions of blog credibility. J Comput Mediat Commun 2007;13(1):100-122 [FREE Full text] [doi: 10.1111/j.1083-6101.2007.00388.x]

31. Sundar S, Edwards H, Hu H, Stavrositu C. Blogging for better health: Putting the "Public" back in public health. In: Tremayne M, editor. Blogging, citizenship, and the future of the media. New York: Routledge; 2007:83-97.

32. Baker JR, Moore SM. Blogging as a social tool: a psychosocial examination of the effects of blogging. Cyberpsychol Behav 2008 Dec;11(6):747-749. [doi: 10.1089/cpb.2008.0053] [Medline: 19072151]

33. Rosenquist JN, Murabito J, Fowler JH, Christakis NA. The spread of alcohol consumption behavior in a large social network. Ann Intern Med 2010 Apr 6;152(7):426-433 [FREE Full text] [doi: 10.7326/0003-4819-152-7-201004060-00007] [Medline: 20368648]

34. Mita G, Ni MC, Jull A. Effectiveness of social media in reducing risk factors for noncommunicable diseases: a systematic review and meta-analysis of randomized controlled trials. Nutr Rev 2016 Apr;74(4):237-247 [FREE Full text] [doi: 10.1093/nutrit/nuv106] [Medline: 26946250]

35. Maher CA, Lewis LK, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014 Feb;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]

36. Carah N, Meurk C, Hall W. Profiling Hello Sunday Morning: who are the participants? Int J Drug Policy 2015 Feb;26(2):214-216. [doi: 10.1016/j.drugpo.2014.07.019] [Medline: 25175599]

37. Hamley B, Carah N. Fare. 2012. One Sunday at a time: Evaluating Hello Sunday Morning URL: http://www.fare.org.au/wp-content/uploads/research/HSM-Evaluation-Final.pdf [accessed 2017-12-01] [WebCite Cache ID 6vNaxKfyw]

38. Hamley B, Carah N. Hello Sunday Morning. Research Report. Queensland: Hello Sunday Morning; 2011.

39. Pennay A, MacLean S, Rankin G. Hello Sunday Morning: alcohol, (non)consumption and selfhood. Int J Drug Policy 2016 Feb;28:67-75. [doi: 10.1016/j.drugpo.2015.10.008] [Medline: 26619752]

40. Babor T, Higgins-Biddle J, Saunders J, Monteiro MG. AUDIT: The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Health Care, 2nd Edition. Geneva: World Health Organization; 2001.

41. Bradley KA, Rubinsky AD, Lapham GT, Berger D, Bryson C, Achtmeyer C, et al. Predictive validity of clinical AUDIT-C alcohol screening scores and changes in scores for three objective alcohol-related outcomes in a Veterans Affairs population. Addiction 2016;111(11):1975-1984. [doi: 10.1111/add.13505]

42. Berger D, Williams EC, Bryson CL, Rubinsky AD, Bradley KA. Alcohol questionnaires and HDL: screening scores as scaled markers of alcohol consumption. Alcohol 2013 Sep;47(6):439-445 [FREE Full text] [doi: 10.1016/j.alcohol.2013.07.001] [Medline: 23886863]

43. Chavez LJ, Williams EC, Lapham G, Bradley KA. Association between alcohol screening scores and alcohol-related risks among female veterans affairs patients. J Stud Alcohol Drugs 2012 May;73(3):391-400 [FREE Full text] [doi: 10.15288/jsad.2012.73.391] [Medline: 22456244]

44. Rubinsky AD, Kivlahan DR, Volk RJ, Maynard C, Bradley KA. Estimating risk of alcohol dependence using alcohol screening scores. Drug Alcohol Depend 2010 Apr 1;108(1-2):29-36 [FREE Full text] [doi: 10.1016/j.drugalcdep.2009.11.009] [Medline: 20042299]

45. Alcohol Screen (AUDIT). Canberra, Australia: Australian Government, Department of Veterans Affairs; 2003. URL: http://at-ease.dva.gov.au/professionals/files/2012/12/Audit-AUDIT-Tool.pdf [accessed 2018-05-08] [WebCite Cache ID 6zGHeZ31f]

46. Ziebland S, Wyke S. Health and illness in a connected world: how might sharing experiences on the internet affect people's health? Milbank Q 2012 Jun;90(2):219-249 [FREE Full text] [doi: 10.1111/j.1468-0009.2012.00662.x] [Medline: 22709387]

47. Baikie K, Wilhelm K. Emotional and physical health benefits of expressive writing. Adv Psychiatr Treat 2005 Sep 1;11(5):338-346. [doi: 10.1192/apt.11.5.338]

48. Smyth JM, Lepore SJ, editors. The Writing Cure: How Expressive Writing Promotes Health and Emotional Well-Being. Washington, DC: American Psychological Association; 2002.

49. Pennebaker JW. Writing about emotional experiences as a therapeutic process. Psychol Sci 1997 May 1;8(3):162-166. [doi: 10.1111/j.1467-9280.1997.tb00403.x]

50. White M, Epston D. Narrative Means to Therapeutic Ends. New York: W.W Norton; 1990.

XSL•FO
RenderX

51.   Naslund JA, Grande SW, Aschbrenner KA, Elwyn G. Naturally occurring peer support through social media: the experiences of individuals with severe mental illness using YouTube. PLoS One 2014 Oct 15;9(10):e110171. [doi: 10.1371/journal.pone.0110171] [Medline: 25333470]

52.   Chung JE. Social networking in online support groups for health: how online social networking benefits patients. J Health Commun 2014 Apr;19(6):639-659. [doi: 10.1080/10810730.2012.757396] [Medline: 23557148]

53.   Curran SL, Andrykowski MA, Studts JL. Short Form of the Profile of Mood States (POMS-SF): psychometric information. Psychol Assess 1995;7(1):80-83. [doi: 10.1037/1040-3590.7.1.80]

54.   Poels K, Dewitte S. How to Capture the Heart? Reviewing 20 Years of Emotion Measurement in Advertising. Leven: Katolieke Universiteit; 2006.

55.   Sidowski JB, Johnson JH, Williams TA, editors. Behavioural treatment and bio-behavioural assessment: computer applications. In: Technology in Mental Health Care Delivery System. Norwood: Albex; 1980:129-139.

56.   Read JC. Validating the Fun Toolkit: an instrument for measuring children's opinions of technology. Cogn Tech Work 2007 May 22;10(2):119-128. [doi: 10.1007/s10111-007-0069-9]

57.   Broekens J, Brinkman W. AffectButton: a method for reliable and valid affective self-report. Int J Hum Comput Stud 2013 Jun;71(6):641-667. [doi: 10.1016/j.ijhcs.2013.02.003]

58.   Desmet P, Overbeeke K, Tax S. Designing products with added emotional value: development and appllcation of an approach for research through design. Des J 2001;4(1):32-47. [doi: 10.2752/146069201789378496]

59.   Blythe M, Monk A, Overbeeke K, Wright P, editors. Measuring emotion: development and application of an instrument to measure emotional responses to products. In: Funology: From Usability to Enjoyment. Dordrecht: Kluwer Academic Publishers; 2003:111-123.

60.   Grant BF. Comorbidity between DSM-IV drug use disorders and major depression: results of a national survey of adults. J Subst Abuse 1995 Jan;7(4):481-497. [doi: 10.1016/0899-3289(95)90017-9] [Medline: 8838629]

61.   Kessler RC, Nelson CB, McGonagle KA, Edlund MJ, Frank RG, Leaf PJ. The epidemiology of co-occurring addictive and mental disorders: implications for prevention and service utilization. Am J Orthopsychiatry 1996 Jan;66(1):17-31. [Medline: 8720638]

62.   Cooper ML, Frone MR, Russell M, Mudar P. Drinking to regulate positive and negative emotions: a motivational model of alcohol use. J Pers Soc Psychol 1995 Nov;69(5):990-1005. [Medline: 7473043]

63.   Kaskutas LA, Subbaraman MS, Witbrodt J, Zemore SE. Effectiveness of making alcoholics anonymous easier: a group format 12-step facilitation approach. J Subst Abuse Treat 2009 Oct;37(3):228-239 [FREE Full text] [doi: 10.1016/j.jsat.2009.01.004] [Medline: 19339148]

Abbreviations

ANOVA: analysis of variance
AUD: alcohol use disorder
AUDIT: Alcohol Use Disorders Identification Test
HSM: Hello Sunday Morning
SBI: screening and brief intervention

XSL•FO

RenderX

Original Paper

Phenotype Instance Verification and Evaluation Tool (PIVET): A Scaled Phenotype Evidence Generation Framework Using Web-Based Medical Literature

Jette Henderson[1], BA, MS; Junyuan Ke[2], BS; Joyce C Ho[2], BS, ME, PhD; Joydeep Ghosh[1], BTech, PhD; Byron C Wallace[3], BS, PhD

[1]The University of Texas at Austin, Austin, TX, United States
[2]Emory University, Atlanta, GA, United States
[3]Northeastern University, Boston, MA, United States

Corresponding Author:
Jette Henderson, BA, MS
The University of Texas at Austin
2501 Speedway, C0806
EER, 6.804
Austin, TX, 78712
United States
Phone: 1 5125855478
Email: jette@ices.utexas.edu

Abstract

Background: Researchers are developing methods to automatically extract clinically relevant and useful patient characteristics from raw healthcare datasets. These characteristics, often capturing essential properties of patients with common medical conditions, are called *computational phenotypes.* Being generated by automated or semiautomated, data-driven methods, such potential phenotypes need to be validated as clinically meaningful (or not) before they are acceptable for use in decision making.

Objective: The objective of this study was to present Phenotype Instance Verification and Evaluation Tool (PIVET), a framework that uses co-occurrence analysis on an online corpus of publically available medical journal articles to build clinical relevance evidence sets for user-supplied phenotypes. PIVET adopts a conceptual framework similar to the pioneering prototype tool PheKnow-Cloud that was developed for the phenotype validation task. PIVET completely refactors each part of the PheKnow-Cloud pipeline to deliver vast improvements in speed without sacrificing the quality of the insights PheKnow-Cloud achieved.

Methods: PIVET leverages indexing in NoSQL databases to efficiently generate evidence sets. Specifically, PIVET uses a succinct representation of the phenotypes that corresponds to the index on the corpus database and an optimized co-occurrence algorithm inspired by the Aho-Corasick algorithm. We compare PIVET's phenotype representation with PheKnow-Cloud's by using PheKnow-Cloud's experimental setup. In PIVET's framework, we also introduce a statistical model trained on domain expert–verified phenotypes to automatically classify phenotypes as clinically relevant or not. Additionally, we show how the classification model can be used to examine user-supplied phenotypes in an online, rather than batch, manner.

Results: PIVET maintains the discriminative power of PheKnow-Cloud in terms of identifying clinically relevant phenotypes for the same corpus with which PheKnow-Cloud was originally developed, but PIVET's analysis is an order of magnitude faster than that of PheKnow-Cloud. Not only is PIVET much faster, it can be scaled to a larger corpus and still retain speed. We evaluated multiple classification models on top of the PIVET framework and found ridge regression to perform best, realizing an average F1 score of 0.91 when predicting clinically relevant phenotypes.

Conclusions: Our study shows that PIVET improves on the most notable existing computational tool for phenotype validation in terms of speed and automation and is comparable in terms of accuracy.

Introduction

Computational Phenotyping

The rapidly expanding availability of electronic health records (EHRs) offers the promise to help clinicians better understand the populations they serve. The ability to efficiently characterize large volumes of healthcare data is essential to enabling clinicians to use this information effectively. Recently, machine learning and data mining researchers have attempted to address this need in several ways. One such line of work concerns developing methods to extract computational phenotypes from raw health data in an automated, high-throughput manner. Here we define a computational phenotype as a constellation of clinically interesting characteristics that delineates a cohesive group of patients. Such phenotypes can help clinicians reason about patient populations, identify patient cohorts, and identify and describe the progression of diseases within populations.

Although being able to extract phenotypes in a high-throughput manner constitutes a potentially important step in helping clinicians reason about their patient populations on a larger scale, this potential will be realized only if the identified phenotypes are clinically meaningful. Therefore, to increase the utility of data-driven phenotypes, some measure quantifying the inferred clinical meaningfulness should be reported alongside the phenotypes to help practitioners sort signal from noise. To address this need, we present Phenotype Instance Verification and Evaluation Tool (PIVET), a tool that uses analysis of open access (OA) PubMed (a corpus of online medical articles) to generate evidence sets and clinical relevance scores for candidate phenotypes. These evidence sets can be used by researchers when developing and tuning new computational phenotype methods; domain experts when they are validating candidate phenotypes; and eventually, clinicians examining the phenotypes associated with their patient populations.

PIVET is an improvement on a recently introduced prototype tool called PheKnow-Cloud [1]. PheKnow-Cloud, which earned the Distinguished Paper Award at the 2017 AMIA Joint Summits, demonstrated that the medical expertise contained in PubMed articles could be harnessed to build evidence sets for the clinical validity of candidate phenotypes. PIVET is built on the same conceptual framework as PheKnow-Cloud, but in PIVET, we have optimized each piece of PheKnow-Cloud's pipeline to deliver vast improvements in speed and interpretability without sacrificing the integrity of PheKnow-Cloud's phenotype evaluation.

The PheKnow-Cloud pipeline consists of three major steps: (1) representing each phenotype so occurrences of it and related terms in the corpus will be recognized (phenotypic representation), (2) analyzing the corpus using the phenotype representation (corpus analysis), and (3) calculating a clinical relevance score and designation (clinical validity determination). In the phenotype representation step, PIVET uses succinct and possibly more interpretable representations of terms contained within each phenotype. In the corpus analysis step, PIVET migrates from a brute force approach of analyzing the corpus to an approach that uses a NoSQL database to store and index the articles efficiently. PIVET then utilizes a variation of the Aho-Corasick algorithm to count appearances of the terms within each phenotype. Finally, in the clinical validity calculation step, PIVET streamlines the clinical relevance score analysis and uses a model, trained on domain expert–verified phenotypes, to classify the clinical relevance of supplied phenotypes. Through a combination of these improvements, PIVET runs an order of magnitude faster than PheKnow-Cloud without sacrificing the discriminative power of the original tool.

PheKnow-Cloud was developed to function in high-throughput phenotyping situations where a researcher has a large set of potential phenotypes to validate. Consequently, PheKnow-Cloud was built to run only in a batch setting. However, in clinical settings and some research settings, a user may only have a few new phenotypes to analyze, so we developed PIVET to run in either an online or batch environment. This improvement will allow clinicians to query PIVET even with single phenotypes, which could possibly help in decision-making processes. Additionally, it could help researchers tune their phenotype extraction algorithms. Thus, while the prototype tool demonstrated the analysis of medical articles could be used to evaluate candidate phenotypes, the improvements in speed and automation realized by PIVET make it useful in both research and clinical settings.

The paper is organized as follows. We first present research related to PIVET, including a description of the original prototype tool (PheKnow-Cloud). Next, we describe the PIVET framework, noting the important differences between PheKnow-Cloud and the new system. We then report the performance of PIVET on automatically generated phenotypes as well as domain expert–curated phenotypes and demonstrate how the framework can be used in an online setting. We conclude the paper with a discussion of the limitations of this work and thoughts on future directions.

Related Work

PubMed

PubMed Central (PMC) is an online collection comprising over 3 million biomedical and biological articles gathered from thousands of journals [2]. PMC is maintained and curated by the National Library of Medicine (NLM) at the US National Institute of Health [3].

In regard to phenotypes, researchers tend to use PubMed as an exploratory tool to *discover* new phenotypes rather than as a resource to *validate* candidate phenotypes. Boland et al orchestrated one of the few studies that used PubMed as a validation tool. They mined EHRs for patients with predefined disease codes and then compared the birth month and the disease of these patients with a group of control patients who did not have the disease codes present in their EHRs. They found a relationship between certain diseases and birth months in the case group [4]. They validated their results against papers retrieved from PubMed that mentioned disease and birth month. This study was novel in that it demonstrated PubMed could be utilized to provide feedback for and validation of results produced through automatic means.

More commonly, researchers use PubMed as tool to generate hypotheses and discover phenotypes and other biomedical issues

XSL•FO

RenderX

[5,6]. Multiple software packages such as LitInspector (Genomatix Software Suite) [7], PubMed.mineR (CSIR) [8], ALIBABA (Humboldt-Universität zu Berlin) [9], as well as python packages such as Pymedtermino (Paris 13 University) [10] and Biopython (Open Bioinformatics Foundation) [11] have been developed to help researchers extract and visualize PubMed. Other researchers have built tools to rank search results, discover topics and relationships within search results, visualize search results, and improve user interaction with PubMed [12].

Text Mining PubMed

Jensen et al give a thorough overview of how PubMed can be harnessed for information extraction and entity recognition [6]. Natural language processing (NLP) techniques form one approach to mining the literature. Some researchers have used NLP techniques on PubMed to discover disease-gene associations [13], and others have used PubMed in concert with additional data sources to generate phenotypes [14]. Collier et al used NLP techniques in conjunction with association rule mining to discover phenotypes using PubMed [15]. However, none of these approaches have sought to use PubMed as a validation tool for data-driven phenotypes.

Co-occurrence analysis, which is what PheKnow-Cloud and PIVET are built on, is more widely used because it is simple to implement and interpret. Researchers have applied co-occurrence strategies to generate phenotypes. Some have performed co-occurrence analysis on PubMed to study links between diseases [16], which can be viewed as a simple type of phenotype discovery. Others have explored relationships between phenotypes and genotypes [17,18]. In contrast to this work, our approach uses phenotypes as the starting point and performs co-occurrence analysis over the PMC corpus as a means of assessing their validity. We assume these phenotypes were induced over other sources (eg, EHRs) and not from PMC. Co-occurrence analysis has the drawback of not being able to explicitly model the type of relationship that exists between two or more terms (eg, negative or positive). However, we require the terms within a phenotype be positively related to one another, which aligns with the findings of publication bias research.

Publication bias is the tendency for the academic publishing ecosystem (eg, researchers, reviewers, and editors) to submit and publish articles that show positive relationships between the entities being studied. The nonrandom omission of results that is not based on the quality of the methodology but on the direction of the results is a well-studied area of research and has been shown to have a negative effect on research in many cases [19-24]. In general, publication bias introduces risks to researchers and to the general public to which research is applied

(via policies and treatment decisions). However, in PheKnow-Cloud and PIVET, this bias is a strength rather than a drawback because the current focus of PheKnow-Cloud and PIVET is on the presence of relationships within the user-supplied candidate phenotypes. Furthermore, as co-occurrence analysis does not attempt to infer information about the type of relationship or any causal information, the presence of publication bias allows for the assumption that when two phrases occur together, it may imply that a relationship exists [20,25,26].

PheKnow-Cloud Prototype

Phenotype evaluation via co-occurrence analysis of online articles was first introduced by Bridges et al [1]. Henderson and colleagues improved on the evaluation framework and developed a prototype tool implementing the approach called PheKnow-Cloud, which provided a Web interface for researchers and clinicians to interact with the technology [27]. We refer to the tool and framework introduced in these two works as PheKnow-Cloud. The input to the PheKnow-Cloud process is a set of potential phenotypes. Each phenotype consists of medical terms, which we refer to as phenotypic items, that are assumed to have been generated by an automatic high-throughput phenotyping process. PheKnow-Cloud builds evidence sets for batches of phenotypes based on co-occurrence analysis of the PubMed corpus (see [1,27] for details).

PheKnow-Cloud was developed as a proof-of-concept tool, and although it showed the PubMed corpus could be used to determine whether a phenotype was clinically valid, it had several drawbacks that PIVET addresses. One is the length of time the prototype method required to complete the analysis process; Table 1 compares the time that each method takes to perform each step. The computational bottlenecks for the prototype method are the co-occurrence generation and clinical relevance score analysis steps. The synonym generation step speed is determined by the number of requests that can be made to the NLM Medical Subject Headings (MeSH) database, which is an off-site system that places limits on the number of requests users can make in a given window of time. Overall, PIVET speeds up this process considerably. Another drawback of PheKnow-Cloud is that the clinical relevance scores for phenotypes are calculated only relative to all other phenotypes and must be used in a batch setting. In contrast, PIVET can analyze a single phenotype at a time, which makes it more flexible than PheKnow-Cloud. Finally, designating whether a candidate phenotype is clinically relevant or not is a manual process in PheKnow-Cloud. For PIVET, we built a classifier trained on a validated set of phenotypes. This classifier can be ported to other environments and can be used to automatically classify new, individual phenotypes.

Table 1. The time in seconds and (hours: minutes: seconds) each method used to complete task in phenotype generation process. All experiments were run on a machine with 3 AMD A6-5200 APU with Radeon (TM) HD Graphics processors, 8 GB of memory, 1 TB hard drive, running Ubuntu 14.04.5 LTS.

Task	PheKnow-Cloud	PIVET[a]
Synonym generation, seconds (hours:minutes:seconds)	7809 (02:10:09)	5948 (01:39:08)
Co-occurrence analysis, seconds (hours:minutes:seconds)	50,822 (14:07:02)	289 (00:04:59)
Lift analysis, seconds (hours:minutes:seconds)	2092 (00:34:52)	2 (00:00:02)
Total, seconds (hours:minutes:seconds)	60,723 (16:52:03)	6239 (01:43:59)

[a]PIVET: Phenotype Instance Verification and Evaluation Tool.

Methods

Methods Overview

In this section, we describe how PIVET performs co-occurrence analysis on an online corpus of publicly available journal articles to build evidence sets for phenotypes. This involves five components: (1) a database of phenotypes to analyze, (2) a database of the PubMed article corpus indexed by medical terms the articles contain, (3) an algorithm to generate and rank synonyms for the phenotypic items (phenotypic item representation), (4) a co-occurrence analysis module (corpus analysis), and (5) a clinical relevance scoring system (clinical validity determination). Figure 1 captures the PIVET workflow and the different components of the system. Both MongoDB (an open-source, document-based NoSQL database system) and MySQL (an open-source, relational database management system) are used to ensure consistency, durability, and efficiency.

Phenotype Extraction and Storage

PIVET can be used to analyze phenotypes generated from a variety of methods. Every phenotype analyzed by PIVET is stored in a MongoDB using a standardized representation to ensure consistency. We also created a simple parser to ingest new phenotypes that are stored in JavaScript Object Notation (JSON). The choice of JSON will also facilitate the eventual integration with a Web platform where users can provide new phenotypes. We populate the phenotype database with phenotypes from different sources (Figure 2).

For our purposes, we collected a total of 102 phenotypes from the following sources: (1) two high-throughput phenotyping algorithms, (2) a catalog of algorithms from a collaborative database, and (3) a peer-reviewed paper. Each phenotype we extracted was either derived by domain experts or validated as clinically relevant by domain experts.

The phenotype database includes 80 domain expert–verified phenotypes generated using two unsupervised, nonnegative tensor factorization models to perform automated phenotyping [28,29]. These were subsequently annotated by a panel of domain experts, and they were the phenotypes used to validate PheKnow-Cloud. The two automatic methods, Rubik [29] and Marble [28], extracted 30 and 50 candidate phenotypes, respectively, from the diagnoses and medications of 7744 deidentified patients from Vanderbilt University Medical Center recorded over a 5-year observation period.

Figure 1. Phenotype Instance Verification and Evaluation Tool (PIVET) analysis process. Phenotypes are collected in standardized format in a MongoDB (ie, "phenotype database"). For a single phenotype, synonyms for each phenotypic item in a phenotype are generated using the National Library of Medicine (NLM) Medical Subject Headings (MeSH) database and ranked based on their similarity to the phenotypic item (ie, "phenotypic item representation"). Co-occurrence analysis is performed on PubMed using the synonyms generated in the previous step (ie, "corpus analysis"). Lift analysis is performed, clinical relevance scores are calculated, and a classifier classifies the phenotype as clinically relevant or not (ie, "clinical validity determination"). The results of the analysis of the phenotype are presented to the viewer (ie, "phenotype evidence results").
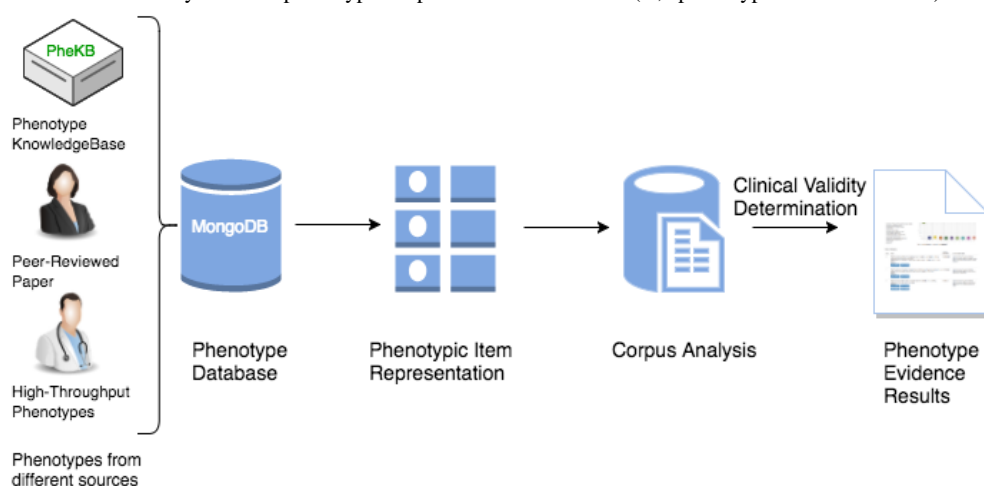
Figure 2. Database for storing phenotype information. The large cylinder at the top represents the phenotype database. The phenotype database consists of phenotypes (documents) extracted from three different sources (bottom). The first set of phenotypes, 80 in total, were generated by machine learning algorithms called Marble and Rubik and annotated for clinical relevance by 3 medical doctors. The second set of phenotypes, 13 in total, we refer to as gold standard phenotypes and come from Phenotype KnowledgeBase, an online repository of domain expert–developed phenotypes. The third set of phenotypes, 9 in total, we refer to as silver standard phenotypes and were derived by domain experts and extracted from a peer-reviewed journal article.
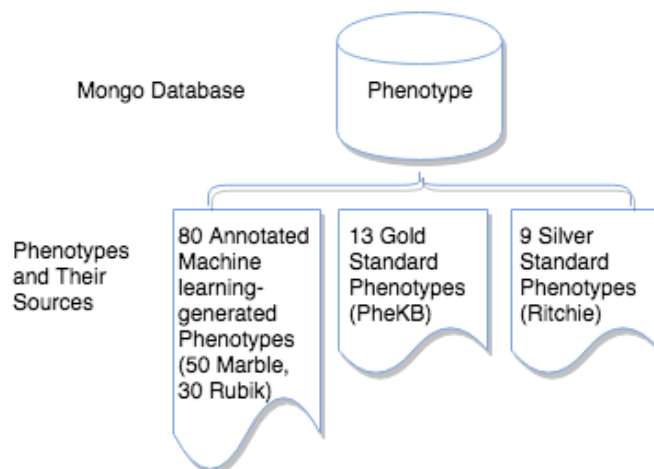
Each member of the panel assigned all phenotypes one of the following three labels: (1) *yes*, the candidate phenotype is clinically meaningful and therefore a phenotype; (2) *no*, the candidate phenotype is not clinically meaningful and therefore not a phenotype, or (3) *maybe*, the candidate phenotype is possibly clinically meaningful. Of the 80 combined Marble and Rubik phenotypes, the domain experts labeled 11 (14%, 11/80) as clinically meaningful, 62 (78%, 62/80) as possibly significant, and 7 (8%, 7/80) as not clinically meaningful. For the handful of phenotypes where the domain experts disagreed on the clinical relevance, the label that awarded the least amount of clinical significance was assigned. These annotated phenotypes were graciously shared by the authors of Rubik.

Additionally, the phenotype database includes two groups of domain expert–derived phenotypes. The first set, which we will refer to as the "gold standard" phenotypes, are from the Phenotype KnowledgeBase, an online phenotype knowledgebase that stores researchers' collaborations of electronic algorithms of phenotypes [30]. Gold standard phenotypes are developed by panels of domain experts across multiple sites. We manually extracted 13 phenotypes that have been reviewed and finalized by the Electronic Medical Records and Genomics phenotype working group. The second set of domain expert–derived phenotypes, which we will refer to as "silver standard" phenotypes, are the group of validated phenotype algorithms published by Ritchie et al [31]. Silver standard phenotypes are developed by a panel of domain experts at a single site. Nine phenotypes were manually extracted from the article. This peer-reviewed paper is not part of the article corpus. In summary, the full set of 102 phenotypes collected over the three different sources consists of 80 machine learning–extracted phenotypes validated by domain experts, 13 gold standard phenotypes, and 9 silver standard phenotypes.

PubMed Open Access Corpus

PIVET works by analyzing co-occurrences of phenotypic items within the PMC OA subset, an openly available online repository of medical articles, which constitutes roughly one-third of the total collection of articles in the PMC (over 1 million articles). The articles within the OA subset are copyright protected but have a flexible license concerning reuse. Trimmed down versions of the articles are stored in a MongoDB. We use the NoSQL database MongoDB because it is a document-based database without restrictive schema, ideal for storing articles that vary in content. Furthermore, MongoDB has been shown to outperform SQL-based databases in terms of read, write, and delete operations and scaling to larger datasets [32-34].

We limit the corpus in the database to those articles with attached MeSH terms; this amounts to 379,766 articles. MeSH is a hierarchical vocabulary curated by the NLM to index and catalog biomedical information [35]. There are 26,000 biomedical concepts or headings and over 200,000 supplementary concepts that form qualifiers for the headings. MeSH has two major benefits over the other existing ontologies. First, a large portion of the PubMed corpus has been manually annotated with MeSH labels. Expert indexers at the NLM assign MeSH terms to each article that best summarize the text. These terms are periodically reviewed and updated. We index the PMC database with the MeSH terms each article contains, and we represent each item in a phenotype with a set of MeSH terms, which is discussed in the next section. The index and phenotypic item representation combined with search optimization techniques described in the subsequent section speed up the co-occurrence analysis process considerably.

Phenotypic Item Representation: Constructing Medical Subject Headings Synonym Sets

Once the phenotypes are stored in the database, the next step is to build representations for the terms within each phenotype, which we refer to as "phenotypic items." Medical terms can have various synonyms (representations) across different articles. For example, the term "heart attack" can also be referred to as "cardiovascular stroke," "myocardial infraction," and "cardiogenic shock." Thus, it is important to generate a list of synonyms for each phenotypic item to achieve high recall within the PubMed corpus. PheKnow-Cloud built representations for each phenotypic item from related terms and concepts found in the following medical ontologies: MeSH, Systemized

Nomenclature of Medicine-Clinical Terms, and International Classification of Diseases-9 or -10. Further experiments indicated this approach can introduce noise into the representation. Instead, PIVET uses only MeSH terms to generate a phenotypic item representation for each phenotypic item with the following two-step process: (1) assign the most relevant MeSH term and (2) generate a ranked list of closely related MeSH terms.

To generate a candidate set of representations for a phenotypic item, PIVET first queries the NLM MeSH database using Biopython [11] with a cleaned version of the phenotypic item. The search returns a set of MeSH tree numbers. MeSH terms are formed into a hierarchical tree, where each MeSH term is assigned a node in the tree and labeled by a number. This number designates the MeSH term's place in the hierarchy. For example, the tree number of "hypertension" is C14.907.489, which indicates that it is a child of the node C14.907 ("vascular diseases"). Vascular diseases, in turn, is a child of node C14 ("cardiovascular diseases"). Gathering nodes with the prefix C14.907.489 gives a set of possible synonyms for the original phenotypic item "hypertension." Generally, this hierarchy gives a relatively straightforward method for finding synonyms and relevant concepts.

As the query does not rank the results (ie, it does not designate which tree number is most relevant to the search), it is necessary to identify the MeSH term that most closely matches the phenotypic item. For example, querying the phenotypic item "hypertension" returns the tree numbers that map to the natural language headings: "hypertension, malignant"; "hypertension, portal"; "hypertension, pulmonary"; "hypertension, renal"; "hypertension"; "masked hypertension"; "prehypertension"; etc (shown in Figure 3). PIVET designates the "most relevant synonym" for the original phenotypic item by finding the natural language heading associated with each of the tree numbers that most closely matches the original phenotypic item. Specifically, for each natural language heading or synonym, PIVET forms a set where each element is a word of the synonym and then finds size of the intersection between the set and the original

cleaned item, which has also been turned into a set. It also records the size difference between the two sets. For example, the phenotypic item "hypertension" and candidate synonym "hypertension, malignant" have an intersection of length one (ie, "hypertension") and a size difference of 1. However, PIVET would assign "hypertension" as the most relevant synonym because it has an intersection of size 1 and a set size difference of 0 with the original phenotypic item. In the event of a tie, the algorithm designates the tied candidate synonyms as the most relevant synonyms and builds the synonym sets for each.

The remaining synonyms are then ranked based on the percentage overlap between each candidate synonym and the most relevant synonym in our PubMed OA corpus. The percentage overlap, calculated as the number of times the candidate synonym appears with the most relevant synonym divided by the number of times the candidate synonym appears overall, serves as the relevance score to rank each synonym. The ranked list is then used to adjust the number of synonyms. An example of a ranked synonym set can be seen in Figure 3.

Corpus Analysis

The aim of the corpus analysis step is to gauge the strength of the relationship between items in a phenotype. However, it is unlikely all items in a phenotype will appear together, so instead, PIVET searches the corpus for occurrences of subsets of the phenotypic items (represented by their phenotypic item MeSH synonym sets as described in the last section). Through experimentation, we found only a small fraction of subsets of any phenotype occur in the article corpus. This means it is inefficient as well as computationally infeasible for even moderately sized phenotypes to look for all possible subsets (ie, the power set in this case has $2^{\wedge}|S| * n_1 * n_2 * \ldots * n_{|S|}$ elements, where $|S|$ is the cardinality of the phenotype and is the synonym set size for phenotypic item i).

Moreover, as the size of the subset increases, the likelihood of all the terms appearing in any given article diminishes. Therefore, it is not necessary to enumerate all the possible subsets.

Figure 3. Synonym generation process for the term "hypertension." First the National Library of Medicine (NLM) Medical Subject Headings (MeSH) database is queried with the term "hypertension," which returns a list of candidate MeSH terms. From this query result, the "most relevant synonym" is determined through a process of string matching between the original queried term and the candidate synonyms. In this case, the most relevant synonym is "hypertension." The candidate synonyms are then ranked based on the percentage overlap between PubMed articles that contain the MeSH term associated with the candidate synonym and the MeSH term of the most relevant synonym.

Candidate Synonyms

Query MeSH database →

Tree Number	Name
C08.381.423.847	Familial Primary Pulmonary Hypertension
C14.907.489.330	Hypertension, Malignant
C14.907.489.480	Hypertension, Pregnancy-Induced
C08.381.423	Hypertension, Pulmonary
C14.907.489.631	Hypertension, Renal
C14.907.489.631.485	Hypertension, Renovascular
C14.907.489	Hypertension
C14.907.489.861	Masked Hypertension
C13.703.395.249	Preeclampsia
C14.907.653	Prehypertension
C18.452.648.861.770	Pseudohypoaldosteronism
C14.907.489.907	White Coat Hypertension

Ranked Synonyms

Query PubMed corpus →

Name	Score
Hypertension	1.000
Hypertension, Malignant	0.600
White Coat Hypertension	0.400
Prehypertension	0.262
Hypertension, Renal	0.243
Pseudohypoaldosteronism	0.167
Masked Hypertension	0.125
Pre-eclampsia	0.069
Hypertension, Pregnancy-Induced	0.051
Hypertension, Renovascular	0.037
Familial Primary Pulmonary Hypertension	0.015
Hypertension, Pulmonary	0.008

XSL•FO
RenderX

Using this observation, we implement an algorithm inspired by the string-matching Aho-Corasick algorithm to search the space effectively [36], an approach also made popular by the a priori algorithm for finding association rules in data mining. We sketch the algorithm with a set comprised terms A, B, C, and D that we assume all occur individually in the corpus. We observe that if terms A and B, comprising a tuple (A,B), do not co-occur in any article together, then any larger subset also containing these two terms will necessarily have zero counts (eg, [A,B,C], [A,B,D], and [A,B,C,D]). As a result, only nonzero (feasible) co-occurrence subsets need to be expanded. A key insight for efficient expansion of an existing co-occurrence subset with nonzero counts is to join it with the associated tuple pairs with one overlapping term that have nonzero counts. For example, if the only nonzero tuple pairs are (A,C), (A,D), (B,C), (B,D), and (C,D), then the possible tuples with cardinality 3 are (A,C,D) and (B,C,D). As increasing the cardinality size of the tuple is equivalent to a join operation in a SQL database, PIVET uses MySQL to implement this portion of the analysis. After constructing the query tuples of MeSH terms in MySQL, PIVET then counts the number of articles where each tuple appears.

Additionally, we set a few more restrictions on the subset queries to make them even more efficient. For one, each subset is constructed using "different" phenotypic items to avoid arbitrary inflation of counts. If two or more phenotypic items contain identical MeSH synonym sets, a "super" phenotypic item is formed (eg, "tuberculosis of adrenal glands" and "tuberculosis of adrenal glands, bacteriological or histological examination not done" are merged together). In addition, terms for the same phenotypic item (eg, all MeSH terms associated with "myocardial infraction") are never paired with one other.

Given these tuple co-occurrence counts, the next step is to map the co-occurring subsets of phenotypic synonyms back to their phenotypic items. For example, if the synonym set for the phenotypic item "attention deficit disorder" contains two synonym terms "attention deficit and disruptive behavior disorders" and "attention deficit disorder with hyperactivity," then any tuple of cardinality 1 with either of these terms is collected, and the sum of the co-occurrences is then designated as the number of times the phenotypic item "attention deficient disorder" occurred. The aggregated co-occurrence counts for all the nonzero subsets of the phenotypic items are then used to calculate the clinical relevance scores for the phenotype.

Clinical Validity Determination

PIVET uses a two-step process to calculate the clinical relevance score: (1) obtain the *lift* (see below) for each co-occurring subset of phenotypic items and (2) classify the relevance of the phenotype based on features derived from the previous step. As in PheKnow-Cloud, PIVET uses lift to evaluate the strength of the relationship between the items in a phenotype. Given items I_1, I_2, …, I_N, lift is defined in equation 1 in Figure 4.

Lift is a widely used metric to measure the statistical independence of objects [37]. A lift of greater than 1 suggests a nonrandom relationship. Although there are many metrics (eg, support, gain, certainty, confidence, and coverage) that can help assess the plausibility of relationships between objects, lift has the benefit of being symmetric (ie, lift[A,B]=lift[B,A]), and

therefore, the order of the objects does not matter [38]. Another metric called leverage also has this symmetric property. However, unlike leverage, lift is not impaired by the "rare item problem," which refers to the property of a metric excluding objects that appear infrequently [39]. In the OA corpus, phenotypic items appear infrequently, so it is especially important to use a metric that does not suffer from the rare item problem. In PIVET, the lift calculation entails dividing the percentage of times items appear together in the corpus by the product of percentages of times each item appears individually in the corpus, which can be rewritten as equation 2 in Figure 4, where count(A) is the number of articles in the corpus that contain the set A, and D is the number of documents in the corpus.

It was observed in PheKnow-Cloud that the lift increases exponentially with the size of the co-occurrence set [1]. This is consistent with equation 2. For example, if a set of six items appears together then the fraction of counts will be multiplied by the size of the corpus to the fifth power. These lifts of larger co-occurring subsets drown out the lifts of smaller-sized subsets, which is not necessarily desirable. Thus, we must "normalize" the cardinality of co-occurrence sets. To this end, PheKnow-Cloud calculated the lift for any subset that occurred in the corpus without regard to whether the subset occurred in a phenotype, separated the lifts by the cardinality of the subsets, computed the SDs above the median within that cardinality, aggregated all the SDs above the median values back into the respective phenotypes, and averaged the SD values for each phenotype. This average served as the "clinical relevance score" for that phenotype. This implies that the relevance score will vary depending on the phenotype corpus, as phenotype scores are relative to other candidate phenotypes.

PIVET mitigates this issue inherent to PheKnow-Cloud normalization by including the number of tuples with zero co-occurrences. The number of subsets that had zero occurrences in the corpus is calculated using a simple combinatorial formula as shown in equation 3 in Figure 4, where S^j is the number of phenotypic items in phenotype j.

Including the zero occurrence counts for each cardinality pulls down the overall lift of the larger items (as it is improbable that large subsets of the phenotype will occur) and thus mitigates the impact of larger co-occurring subsets. Consequently, PIVET avoids the need to pool the phenotypic items across all the phenotypes and avoids unnecessary co-occurrence queries for tuples that do not occur in a phenotype. Perhaps more importantly, this implies that the relevance score is decoupled from the phenotype corpus and can be computed independently for a given phenotype.

The final step in the process is to classify the relevance of the phenotype. We compared four separate classification models: logistic regression, logistic regression with least absolute shrinkage and selection operator (lasso), ridge logistic regression, and k-nearest neighbors (k-NN) on the entire phenotype corpus to predict clinically significant vs not clinically significant. Gold and silver standard phenotypes are denoted as clinically significant because of their relatively small numbers.

Figure 4. Lift, number of zeros, and F1 equations.

$$\text{lift}(I_1, I_2, \ldots, I_N) = \frac{P(I_1 \cap I_2 \cap \cdots \cap I_{N-1} \cap I_N)}{P(I_1)P(I_2)\cdots P(I_{N-1})P(I_N)} \tag{1}$$

$$\text{lift}(I_1, I_2, \ldots, I_N) = \frac{\text{count}(I_1, I_2, \ldots, I_N)}{\text{count}(I_1)\text{count}(I_2)\cdots\text{count}(I_N)}D^{N-1} \tag{2}$$

$$\text{Size(zeros for phenotype } j) = \sum_{i=1}^{S^j} \binom{S^j}{i} - \text{size(co-occurrences of cardinality } i) \tag{3}$$

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{4}$$

The features we use are lift mean, lift median, and lift SD for each individual cardinality from 1, 2, 3, and 4 (12 features). We also include the overall lift mean, median, and SD (3 features) and the average cardinality of subsets of the phenotype with nonzero co-occurrences (16 features in total). Model-specific parameters (ie, K for k-NN and the regularization parameter for ridge and lasso) are chosen based on the best area under receiver operating characteristic via five-fold cross-validation.

In summary, the PIVET lift analysis differs from that performed by PheKnow-Cloud in two key ways. First, we eliminate the need to pool the lifts across the entire phenotype corpus, which means that phenotypes can be analyzed on an individual basis. Second, we introduce classification models to determine relevance based on lift-based features, removing the need to perform an exhaustive search to determine the clinical relevancy threshold.

Results

Results Overview

PIVET is evaluated using two different methods. The first compares the new framework with its predecessor, PheKnow-Cloud, on the set of phenotypes PheKnow-Cloud examined. Differences in computation time, synonym generation, and clinical relevance scores are quantitatively and qualitatively examined. This comparison shows that PIVET delivers clinical relevance determination performance comparable with that of PheKnow-Cloud in a fraction of the time. Furthermore, PIVET's performance justifies shifting from the old to the new framework.

In the second set of experiments, we demonstrate the full PIVET framework on the combined set of machine learning–generated phenotypes, gold standard phenotypes, and silver standard phenotypes. This experiment and discussion show how PIVET's classification method can be used to identify clinically relevant phenotypes from the pool of possibly clinically relevant phenotypes.

PheKnow-Cloud Versus Phenotype Instance Verification and Evaluation Tool Comparison

Phenotypic Item Representation

A subset comprising one-quarter of the PMC OA corpus is used to compare our framework's use of MeSH terms for the phenotypic item synonym sets with PheKnow-Cloud's phenotypic item synonym sets. This subset is identical to the one used in the original evaluation of Pheknow-Cloud (see [1] for more details regarding the construction of the dataset). We limit this subset to articles with MeSH terms, which results in a corpus of articles that comprises 7.85% of the PMC OA subset (94,673/1,206,506). We restrict the phenotypes in question to the 80 domain expert–verified, machine learning–generated phenotypes used in the original papers [1,27]. Table 2 shows the clinical validity annotations of these 80 phenotypes.

PIVET takes less than 2 hours to evaluate 80 phenotypes on the 8% PMC OA subset; PheKnow-Cloud required 17 hours for the same phenotypes. The breakdown of the computation time for the major components of the two frameworks is shown in Table 1. The phenotypic item representation process time is roughly the same for both PIVET and PheKnow-Cloud, and querying the NLM MeSH database remains the bottleneck. However, PIVET is 170 and 35 times faster for the corpus analysis and clinical relevance determination steps, respectively. Not only does PIVET provide an overall speedup of 10 times on the same article corpus, but the entire process does not need to be repeated to analyze new phenotypes.

Table 2. Counts of the 80 machine learning–generated phenotypes by clinical relevance annotation category.

Domain expert annotation category	Count, n (%)
Clinically significant	11 (14)
Possibly clinically significant	62 (78)
Not clinically significant	7 (8)

As discussed in an earlier section, the phenotypic item representation is different between the two frameworks. PIVET uses sets of MeSH terms to represent each phenotypic item, whereas PheKnow-Cloud's representative synonym sets are built from several ontologies that include the MeSH terms. Overall, PIVET finds more descriptive, discriminative, and possibly more interpretable representations of phenotypic items, whereas PheKnow-Cloud's synonym sets produced a sizeable number of less descriptive words in comparison. Figure 5 shows the top 50 PheKnow-Cloud-generated synonyms that were found in the corpus. Although PheKnow-Cloud excludes the first 30 most common terms from its co-occurrence analysis, the remaining 20 words are not discriminative. For example, the word "diseases" is associated with many of the phenotypic items but is too generic to be a meaningful representation of the items.

Further qualitative evidence of the nonspecific nature of the synonym sets produced by PheKnow-Cloud can be found by consideration of examples. Table 3 shows the synonyms for the phenotypic item "unspecified chest pain." Under the PheKnow-Cloud framework, although discriminative terms such as "unspecified chest pain" and "chest pain" are present in the synonym set, the terms "pain," "chest," and "unspecified" are words that will be present in many articles that do not actually refer to "unspecified chest pain." In contrast, under the PIVET framework, the MeSH term for "unspecified chest pain" is "chest pain," which while less specific than the original term, has the advantage that it will only be found in articles that mention chest pain.

Figure 5. Most common synonyms found in corpus using PheKnow-Cloud synonym generation process.
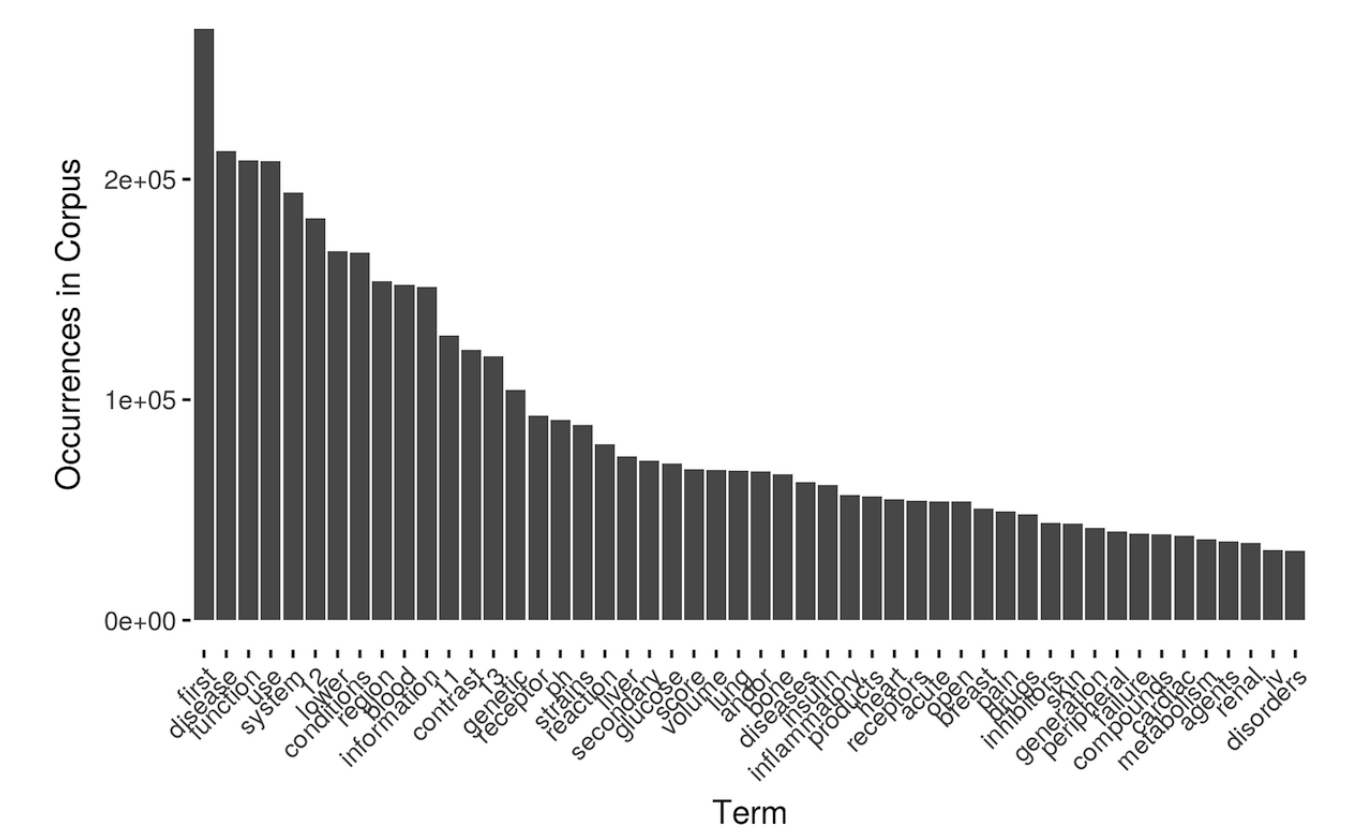
Table 3. Comparison of representation of the phenotypic item "unspecified chest pain" generated by PheKnow-Cloud (left column) and Phenotype Instance Verification and Evaluation Tool (PIVET; right column).

PheKnow-Cloud (synonyms)	PIVET[a] (MeSH[b] terms)
Unspecified chest pain	Chest pain
Chest pain	—
Unspecified chest	—
Pain	—
Chest	—
Unspecified	—

[a]PIVET: Phenotype Instance Verification and Evaluation Tool.

[b]MeSH: Medical Subject Headings.

XSL•FO

RenderX

In some cases, the synonym sets are reasonable representations of the item and similar for both frameworks. For example, PIVET and PheKnow-Cloud can capture the meaning of the phenotypic item "laxatives" (shown in Table 4). PheKnow-Cloud extracts synonyms that are close literal matches to the phenotypic item or specific kinds of laxatives. Similarly, PIVET finds a MeSH term that is an exact match to the phenotypic item and a specific example of the phenotypic item. When looking through the corpus for occurrences of the original term "laxatives," both frameworks should recover mentions of the original term.

Clinical Validity Determination

Next, we examine how PIVET's phenotype representation compares with that of PheKnow-Cloud in terms of identifying clinically relevant phenotypes. To do this, we instrumented PIVET to record co-occurrences in the same manner as PheKnow-Cloud. Table 5 summarizes the number of articles that are found under each framework. Although the PIVET MeSH representation identifies significantly fewer articles from the corpus, the articles have an 85% overlap with PheKnow-Cloud articles. In conjunction with Figure 5 and Table 2, the results suggest that not all of the PheKnow-Cloud articles are relevant or directly related to the phenotypic item. Thus, PIVET synonym sets may result in higher precision.

Finally, we compared the two frameworks' ability to discriminate between clinically significant and not significant

phenotypes using the process PheKnow-Cloud used. To do this, we first calculated the normalized lift for all the phenotypes using the synonyms sets generated by PheKnow-Cloud and PIVET. Figure 6 plots the pooled normalized lift values for the 80 phenotypes based on the annotated significance level. As we saw in the PheKnow-Cloud framework, under the PIVET representation, the distributions of normalized lift between significant and not significant phenotypes are not identical, which indicates that lift scores can be used to discriminate between significant and not significant phenotypes.

In the final step, we calculated clinical validity scores for each phenotype by taking the average of the normalized lift scores in each phenotype. An exhaustive search was performed on the clinical validity scores to determine the boundaries for PIVET and PheKnow-Cloud, which was the method used in PheKnow-Cloud that maximized the F1 score. F1 is computed as equation 4 in Figure 4.

We obtained an F1 score of 0.85 and 0.89 for PIVET and PheKnow-Cloud, respectively. Although the predictive performance of PIVET is slightly lower than that of PheKnow-Cloud, the performance loss is negligible when compared with the total run time of each framework (Table 1) on 8% of the PMC OA subset. Moreover, by mapping directly to MeSH terms, PIVET can leverage the "automatic" assignment of MeSH terms for all articles and can have a higher probability of capturing appearances of the original phenotypic item in the corpus.

Table 4. Comparison of representation of the phenotypic item "laxatives" generated by PheKnow-Cloud (left column) and Phenotype Instance Verification and Evaluation Tool (PIVET; right column).

PheKnow-Cloud (synonyms)	PIVET[a] (MeSH[b] terms)
Laxatives	Laxatives
Laxatives pharmacological action	Senna extract
Psyllium	—
Senna	—
Senna extract	—

[a]PIVET: Phenotype Instance Verification and Evaluation Tool.
[b]MeSH: Medical Subject Headings.

Table 5. Number of articles that each framework's synonym generation process found.

Synonym type	Number of articles
PIVET[a]	28,068
PheKnow-Cloud	79,786
PIVET and PheKnow-Cloud	23,901

[a]PIVET: Phenotype Instance Verification and Evaluation Tool.

Figure 6. Normalized lift comparison between Phenotype Instance Verification and Evaluation Tool (PIVET) and PheKnow-Cloud. Normalized lift is calculated as follows: the lift for any subset of phenotypic items that occurred in the corpus without regard to whether the subset occurred in a phenotype is calculated. Then the lifts are separated by the cardinality of the subsets, and the standard deviations above the median within that cardinality is computed (ie, this is the normalized lift). The boxplot depicts the normalized lift for the subsets that appeared in each type (ie, "maybe significant," "not significant," and "significant") of phenotype.
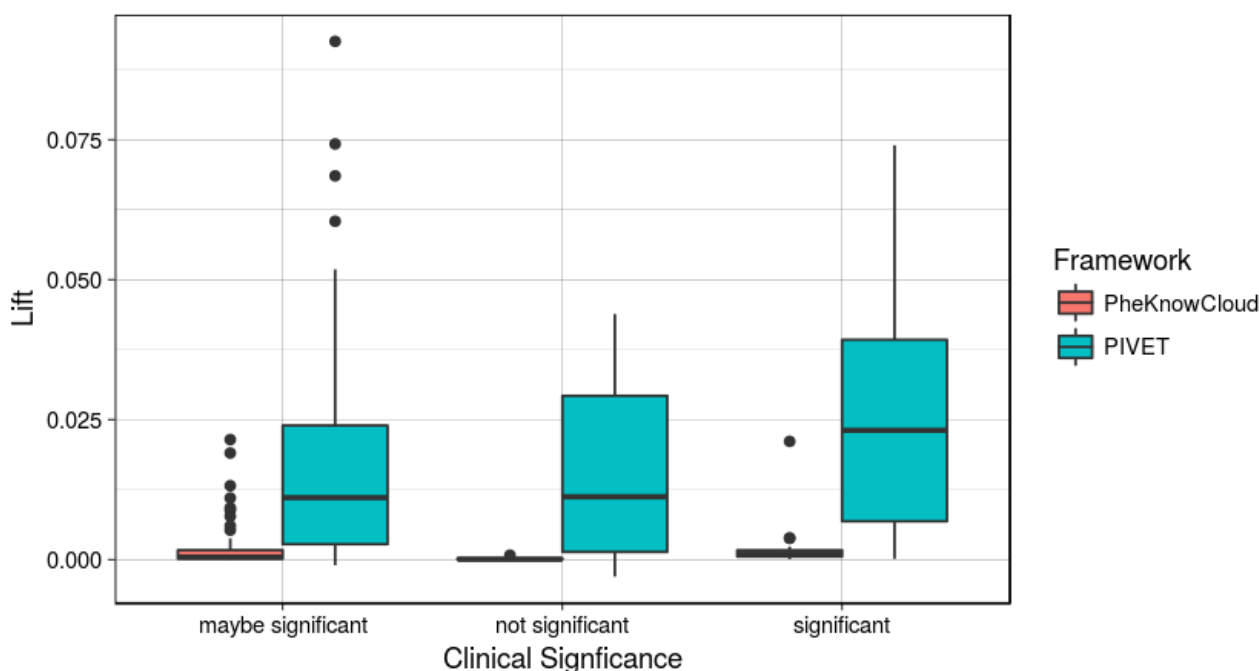
Phenotype Instance Verification and Evaluation Tool

In the first set of experiments, we demonstrated PIVET's synonym generation process results in discriminative performance comparable with that of PheKnow-Cloud in a fraction of the time. In the second set of experiments, we use PIVET's full framework (Figure 1) to predict which phenotypes are clinically valid and show how PIVET can be used to examine phenotypes that are possibly clinically valid.

Corpus Analysis: Classification Score Evaluation

We evaluated the ability of the PIVET classification system to identify clinically significant phenotypes. The entire phenotype corpus, including the gold and silver standard phenotypes, were analyzed using the entire PMC OA corpus. There is ambiguity regarding the "possibly significant" Marble and Rubik phenotypes, and they were therefore excluded from the training set. Thus, a total of 45 phenotypes were used to build the classifier, with 7 annotated as not significant.

The diversity of the phenotypes in our corpus yielded phenotypes that contained anywhere from 3 to 63 phenotypic items. The size of the phenotype sets impacted the cardinality of the nonzero co-occurrence tuples; thus, we limited the lift summary features to only include tuples up to 4 (the average across the phenotype corpus). Figure 7 illustrates the differences in the mean lift values between the various categories, with the gold and silver standard separated from the clinically significant group. The results show that the phenotypes that are clinically significant exhibited a higher (more positive) distribution in lift

mean compared with the nonsignificant phenotypes. Moreover, for co-occurrence cardinality less than 5, gold standard phenotypes generally had a higher lift. The figure suggests it is suitable to use the mean lift of tuples of cardinalities 2, 3, and 4 as individual features to distinguish the clinical significance of a phenotype.

Next, we used logistic regression to analyze the effect of the size of the synonym set. For each synonym set size ranging from 2 to 10, we used five-fold cross-validation to examine how the size of the synonym set generalizes to an unseen dataset for different metrics. Figure 8 plots the average precision, recall, and F1 score as a function of the synonym set size. The figure shows significant increases for all three metrics at synonym size 6, at which point an F1 score of 0.89, recall rate of 0.89, and a precision score of 0.88 are achieved. On the basis of these results, we used six synonyms for each phenotypic item for the remaining analysis.

We repeated the classification process using four models (logistic regression, k-NN, logistic regression with lasso, and ridge-regularized logistic regression) with six MeSH term synonyms for each phenotypic item. The results are shown in Table 6. Of the four classification models, ridge regression achieved the highest F1 score of 0.91 and an Area Under the Receiver Operating Curve score of 0.60. On the basis of these results, we use ridge regression as our classification model for the remaining results. Incorporating a classification model into the framework is an improvement over PheKnow-Cloud, which depended on an exhaustive search to obtain a boundary between clinically relevant and not clinically relevant phenotypes.

Figure 7. Log mean lift for co-occurrences of sizes 2, 3, 4, and 5 for each type of phenotype.
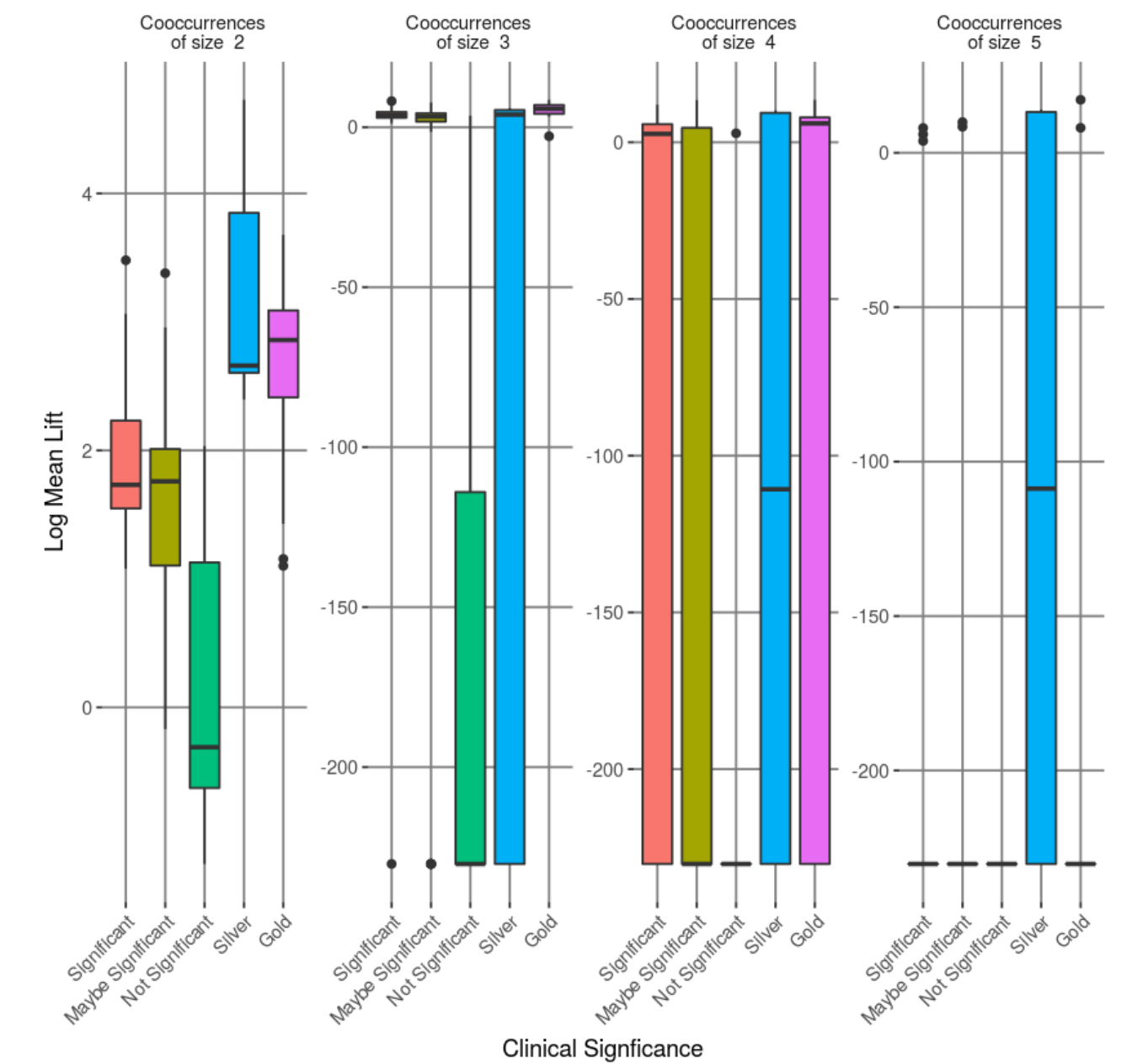
Figure 8. Classification scores for different sizes of synonyms using the Phenotype Instance Verification and Evaluation Tool (PIVET) framework.
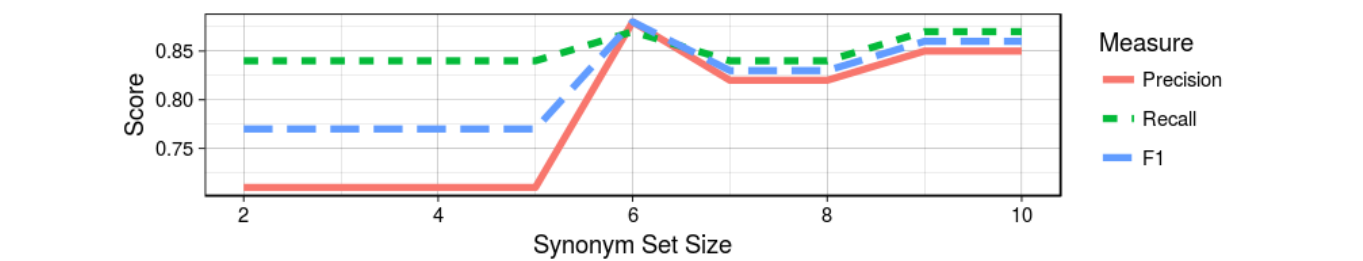
Table 6. Performance metrics for classification task to identify clinically relevant phenotypes using synonym sets of size 6.

Metric	Logistic regression	K-nearest neighbors	Lasso	Ridge regression
Area Under the Receiver Operating Curve	0.79	0.72	0.33	0.60
F-1	0.87	0.90	0.77	0.91

Table 7. Diagnoses and medications for candidate phenotypes along with domain expert annotations, classification score, and lift for two possibly significant phenotypes with high (top two rows) and low (bottom two rows) classification scores.

Diagnoses	Medications	Comment	Score	Lift
Hypotension, heart failure, cardiac dysrhythmias, unspecified chest pain, ischemic heart disease, hypertension, cardiomyopathy	Statins, proton pump inhibitors, gabapentin, noncardioselective beta blockers, sodium, group v antiarrhythmics, potassium-sparing diuretics	The arrhythmic heart patient	1	317.380
Disorders of fluid, electrolyte, and acid-base balance; other and unspecified anemias; hypertensive chronic kidney disease; hypertension; diabetes mellitus; type 2; other disorders of kidney and ureter; chronic kidney disease	Antiadrenergic agents, centrally acting, angiotensin receptor blockers, angiotensin converting enzyme inhibitors, selective immunosuppressants, loop diuretics, gabapentin	Heading toward dialysis	0.999	24683.383
Volume depletion; dehydration, nausea, or vomiting; hypopotassemia; abdominal pain	Heparins, antihistamines, 5HT3 receptor antagonists, minerals and electrolytes, narcotic analgesic combinations, proton pump inhibitors	Gastroenteritis	0.418	0.270
Disorders of fluid, electrolyte, and acid-base balance; other diseases of lung; hypotension; pleurisy, atelectasis, and pulmonary collapse; unspecified chest pain; other disorders of the kidney and ureter	Anticholinergic bronchodilators, loop diuretics	Lung diseases?	0.417	0.509

Clinical Validity Determination: Phenotype Instance Verification and Evaluation Tool Analysis of Possibly Clinically Significant Phenotypes

We demonstrate the potential of using PIVET to annotate phenotypes by examining the 62 "possibly clinically significant" phenotypes in our phenotype dataset. Using the PIVET classification ridge model, we predicted the clinical relevance scores of these ambiguous phenotypes. Table 7 shows the two extremes based on the averaged prediction score: phenotypes with the highest probability of being "clinically significant" (top two rows) and phenotypes with the lowest probability of being "clinically significant" (bottom two rows), as well as the annotator's comment on the phenotype and the average lift calculated by PIVET. The prediction scores seem to reflect the annotator's certainty, as the lowest prediction score is associated with a question mark, whereas the top two scoring phenotypes seem to capture a relevant concept. The results underscore the potential of PIVET system to help resolve uncertainties.

Discussion

Principal Findings

Several automated, high-throughput phenotype methods have been proposed to help clinicians quickly characterize and understand vast amounts of healthcare data. However, the potential for computational phenotyping to help physicians reason about patient populations will only be realized if the phenotypes generated are clinically meaningful. To increase the utility of such data-driven phenotype discovery, some measure of inferred clinical meaningfulness should be reported to help clinicians sort the signals from the noise. We developed PIVET to meet this need. PIVET generates evidence sets and clinical relevance scores for data-driven candidate phenotypes using the literature available in PubMed, a large online repository of biomedical articles.

We compared our framework with PheKnow-Cloud, its predecessor, and showed that PIVET improves the run time dramatically. In addition to scaling up to the entire PMC OA corpus, PIVET can analyze phenotypes individually and automatically assign clinical relevance scores that are independent of the other phenotypes in the corpus. Furthermore, there was anecdotal evidence that the PIVET synonym generation process was more discriminative and meaningful than its PheKnow-Cloud counterpart. In the future, one goal is to make PIVET available to researchers and clinicians. To this end, we plan to deploy a live version of the phenotype parser that users can interact with via a REST API and receive phenotype JSON files in return. We are currently investigating the best way to release PIVET for general use.

Possible Use Cases

For researchers developing models and algorithms to automatically extract phenotypes from EHRs without supervision, all phenotypes are possibly clinically significant before they have been validated. We envision PIVET being used by researchers to gain understanding into the phenotypes they have extracted. Outside a machine learning setting, there are several potential uses for PIVET. For example, a pharmaceutical company may uncover a potentially interesting pathway analysis or phenotype, and they can use PIVET to identify all the articles that have been previously published on the subject, as well as PIVET's clinical validity determination to decide if the pathway is worth pursuing and how much it can be trusted. Similarly, in a healthcare setting, a clinician could encounter an interesting group of patients and use PIVET to explore what pathways have been discovered with relation the set of patient characteristics. As in the pharmaceutical setting, PIVET's ability to deliver a clinical validity determination, as well as generate a body of evidence in the form relevant articles, can help clinicians reason about the patterns they encounter on a daily basis.

Limitations

One possible way to improve PIVET is to include more phenotypes when training the classifier. We continue to gather additional domain expert annotated phenotypes to include in

the framework. One limitation of this analysis was that all the gold and silver standard phenotypes were combined with the domain expert–labeled examples for classification purposes. As we continue to gather more gold and silver phenotypes, we plan to refine the classification process by incorporating this "annotation quality" information. We also plan to test new sets of features that incorporate interaction between the lift statistics and to examine different metrics for evaluating the clinical significance of candidate phenotypes.

Authors' Contributions

JH generated phenotype representations by mapping phenotypic items to MeSH terms and also optimized the corpus analysis and clinical validity determination modules. Additionally, she compared this work to a previous baseline. JK extracted phenotypes from raw documents and mapped them to a database format and built a classifier. JCH curated the article database and optimized the co-occurrence algorithm. BW and JG gave guidance on the co-occurrence algorithm and on system architecture.

Conflicts of Interest

None declared.

References

1. Bridges R, Henderson J, Ho JC, Wallace BC, Ghosh J. Automated verification of phenotypes using pubmed. 2016 Oct Presented at: ACM BCB; October 2-5, 2016; Seattle p. 595-602 URL: http://doi.acm.org/10.1145/2975167.2985844 [doi: 10.1145/2975167.2985844]

2. NCBI Resource Coordinators. Database resources of the national center for biotechnology information. Nucleic Acids Res 2018 Jan 04;46(D1):D8-D13 [FREE Full text] [doi: 10.1093/nar/gkx1095] [Medline: 29140470]

3. National Center for Biotechnology Information. NCBI. PMC FAQs URL: https://www.ncbi.nlm.nih.gov/pmc/about/faq/ [WebCite Cache ID 6tPHmk6mp]

4. Boland MR, Shahn Z, Madigan D, Hripcsak G, Tatonetti NP. Birth month affects lifetime disease risk: a phenome-wide method. J Am Med Inform Assoc 2015 Sep;22(5):1042-1053 [FREE Full text] [doi: 10.1093/jamia/ocv046] [Medline: 26041386]

5. Ananiadou S, Kell DB, Tsujii J. Text mining and its potential applications in systems biology. Trends Biotechnol 2006 Dec;24(12):571-579. [doi: 10.1016/j.tibtech.2006.10.002] [Medline: 17045684]

6. Jensen LJ, Saric J, Bork P. Literature mining for the biologist: from information retrieval to biological discovery. Nat Rev Genet 2006 Feb;7(2):119-129. [doi: 10.1038/nrg1768] [Medline: 16418747]

7. Frisch M, Klocke B, Haltmeier M, Frech K. LitInspector: literature and signal transduction pathway mining in PubMed abstracts. Nucleic Acids Res 2009 Jul;37(Web Server issue):W135-W140 [FREE Full text] [doi: 10.1093/nar/gkp303] [Medline: 19417065]

8. Rani J, Shah AB, Ramachandran S. pubmed.mineR: an R package with text-mining algorithms to analyse PubMed abstracts. J Biosci 2015 Oct;40(4):671-682 [FREE Full text] [Medline: 26564970]

9. Plake C, Schiemann T, Pankalla M, Hakenberg J, Leser U. AliBaba: pubmed as a graph. Bioinformatics 2006 Oct 01;22(19):2444-2445. [doi: 10.1093/bioinformatics/btl408] [Medline: 16870931]

10. Lamy J, Venot A, Duclos C. PyMedTermino: an open-source generic API for advanced terminology services. Stud Health Technol Inform 2015;210:924-928. [Medline: 25991291]

11. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available python tools for computational molecular biology and bioinformatics. Bioinformatics 2009 Jun 01;25(11):1422-1423 [FREE Full text] [doi: 10.1093/bioinformatics/btp163] [Medline: 19304878]

12. Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. Database (Oxford) 2011;2011:baq036 [FREE Full text] [doi: 10.1093/database/baq036] [Medline: 21245076]

13. Kim J, Kim J, Lee H. An analysis of disease-gene relationship from medline abstracts by digsee. Sci Rep 2017 Jan 05;7:40154 [FREE Full text] [doi: 10.1038/srep40154] [Medline: 28054646]

14. Alnazzawi N, Thompson P, Batista-Navarro R, Ananiadou S. Using text mining techniques to extract phenotypic information from the PhenoCHF corpus. BMC Med Inform Decis Mak 2015 Jun;15 Suppl 2:S3 [FREE Full text] [doi: 10.1186/1472-6947-15-S2-S3] [Medline: 26099853]

15. Collier N, Groza T, Smedley D, Robinson PN, Oellrich A, Rebholz-Schuhmann D. PhenoMiner: from text to a database of phenotypes associated with OMIM diseases. Database (Oxford) 2015;2015:- [FREE Full text] [doi: 10.1093/database/bav104] [Medline: 26507285]

XSL•FO
RenderX

16. Rajpal DK, Qu XA, Freudenberg JM, Kumar VD. Mining emerging biomedical literature for understanding disease associations in drug discovery. Methods Mol Biol 2014;1159:171-206. [doi: 10.1007/978-1-4939-0709-0_11] [Medline: 24788268]

17. Pletscher-Frankild S, Pallejà A, Tsafou K, Binder JX, Jensen LJ. DISEASES: text mining and data integration of disease-gene associations. Methods 2015 Mar;74:83-89 [FREE Full text] [doi: 10.1016/j.ymeth.2014.11.020] [Medline: 25484339]

18. Xu D, Zhang M, Xie Y, Wang F, Chen M, Zhu KQ, et al. DTMiner: identification of potential disease targets through biomedical literature mining. Bioinformatics 2016 Dec 01;32(23):3619-3626 [FREE Full text] [doi: 10.1093/bioinformatics/btw503] [Medline: 27506226]

19. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. Cochrane Database Syst Rev 2009(1):MR000006. [doi: 10.1002/14651858.MR000006.pub3] [Medline: 19160345]

20. Dickersin K. The existence of publication bias and risk factors for its occurrence. JAMA 1990 Mar 9;263(10):1385-1389. [Medline: 2406472]

21. Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, 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]

22. Song F, Hooper L, Loke Y. Publication bias: what is it? how do we measure it? how do we avoid it? OAJCT 2013 Jul;5:71-81 [FREE Full text] [doi: 10.2147/OAJCT.S34419]

23. Ekmekci PE. An increasing problem in publication ethics: publication bias and editors' role in avoiding it. Med Health Care Philos 2017 Jun;20(2):171-178. [doi: 10.1007/s11019-017-9767-0] [Medline: 28342053]

24. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS One 2008 Aug 28;3(8):e3081 [FREE Full text] [doi: 10.1371/journal.pone.0003081] [Medline: 18769481]

25. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. Lancet 1991 Apr 13;337(8746):867-872. [Medline: 1672966]

26. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. BMJ 1997 Sep 13;315(7109):640-645 [FREE Full text] [Medline: 9310565]

27. Henderson J, Bridges R, Ho JC, Wallace BC, Ghosh J. PheKnow-Cloud: a tool for evaluating high-throughput phenotype candidates using online medical literature. AMIA Jt Summits Transl Sci Proc 2017 Jan;2017:149-157 [FREE Full text] [Medline: 28815124]

28. Ho J, Ghosh J, Sun J. Marble: high-throughput phenotyping from electronic health records via sparse nonnegative tensor factorization. 2014 Presented at: KDD; August 24-27, 2014; New York p. 115-124. [doi: 10.1145/2623330.2623658]

29. Wang Y, Chen R, Ghosh J, Denny J, Kho A, Chen Y, et al. Rubik: knowledge guided tensor factorization and completion for health data analytics. 2015 Presented at: KDD; August 10-13, 2015; Sydney, NSW p. 1265-1274. [doi: 10.1145/2783258.2783395]

30. Kirby JC, Speltz P, Rasmussen LV, Basford M, Gottesman O, Peissig PL, et al. PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability. J Am Med Inform Assoc 2016 Nov;23(6):1046-1052 [FREE Full text] [doi: 10.1093/jamia/ocv202] [Medline: 27026615]

31. Ritchie MD, Denny JC, Crawford DC, Ramirez AH, Weiner JB, Pulley JM, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. Am J Hum Genet 2010 Apr 09;86(4):560-572 [FREE Full text] [doi: 10.1016/j.ajhg.2010.03.003] [Medline: 20362271]

32. Li Y, Manoharan S. A performance comparison of SQL and NoSQL databases. 2013 Aug Presented at: PACRIM; August 27-29, 2013; Victoria, BC, Canada p. 15-19. [doi: 10.1109/PACRIM.2013.6625441]

33. Boicea A, Radulescu F, Agapin L. MongoDB vs oracle-database comparison. 2012 Sep Presented at: EIDWT; September 19-21, 2012; Bucharest, Romania p. 330-335. [doi: 10.1109/EIDWT.2012.32]

34. Indrawan-Santiago M. Database research: are we at a crossroad? reflection on NoSQL. Database research; 2012 Presented at: NBiS; September 26-28, 2012; Melbourne, VIC, Australia p. 45-51. [doi: 10.1109/NBiS.2012.95]

35. Lipscomb CE. Medical subject headings (MeSH). Bull Med Libr Assoc 2000 Jul;88(3):265-266 [FREE Full text] [Medline: 10928714]

36. Aho AV, Corasick MJ. Efficient string matching: an aid to bibliographic search. Commun ACM 1975;18(6):333-340. [doi: 10.1145/360825.360855]

37. Brin S, Motwani R, Ullman J, Tsur S. Dynamic itemset counting and implication rules for market basket data. 1997 Presented at: SIGMOD; May 11-15, 1997; Tucson p. 255-264. [doi: 10.1145/253262.253325]

38. Ventura S, Luna J. Pattern mining with evolutionary algorithms. Heidelberg: Springer; 2016.

39. Sheikh L. Interesting measures for mining association rules. 2004 Presented at: INMIC; December 24-26, 2004; Lahore, Pakistan p. 641-644. [doi: 10.1109/INMIC.2004.1492964]

Abbreviations

EHR: electronic health record
JSON: JavaScript Object Notation
k-NN: k-nearest neighbors
lasso: least absolute shrinkage and selection operator
MeSH: Medical Subject Headings
NLM: National Library of Medicine
NLP: natural language processing
OA: open access
PIVET: Phenotype Instance Verification and Evaluation Tool
PMC: PubMed Central

Review

Artificial Intelligence for Diabetes Management and Decision Support: Literature Review

Ivan Contreras[1], PhD; Josep Vehi[1,2], PhD

[1]Modeling, Identification and Control Laboratory, Institut d'Informatica i Aplicacions, Universitat de Girona, Girona, Spain
[2]Centro de Investigación Biomédica en Red de Diabetes y Enfermadades Metabólicas Asociadas, Girona, Spain

Corresponding Author:
Josep Vehi, PhD
Modeling, Identification and Control Laboratory
Institut d'Informatica i Aplicacions
Universitat de Girona
Campus Montilivi. Edifici P4
Girona, 17003
Spain
Phone: 34 620131826
Email: josep.vehi@udg.edu

Abstract

Background: Artificial intelligence methods in combination with the latest technologies, including medical devices, mobile computing, and sensor technologies, have the potential to enable the creation and delivery of better management services to deal with chronic diseases. One of the most lethal and prevalent chronic diseases is diabetes mellitus, which is characterized by dysfunction of glucose homeostasis.

Objective: The objective of this paper is to review recent efforts to use artificial intelligence techniques to assist in the management of diabetes, along with the associated challenges.

Methods: A review of the literature was conducted using PubMed and related bibliographic resources. Analyses of the literature from 2010 to 2018 yielded 1849 pertinent articles, of which we selected 141 for detailed review.

Results: We propose a functional taxonomy for diabetes management and artificial intelligence. Additionally, a detailed analysis of each subject category was performed using related key outcomes. This approach revealed that the experiments and studies reviewed yielded encouraging results.

Conclusions: We obtained evidence of an acceleration of research activity aimed at developing artificial intelligence-powered tools for prediction and prevention of complications associated with diabetes. Our results indicate that artificial intelligence methods are being progressively established as suitable for use in clinical daily practice, as well as for the self-management of diabetes. Consequently, these methods provide powerful tools for improving patients' quality of life.

Introduction

Overview

Diabetes mellitus refers collectively to a group of diseases resulting from dysfunction of the glucoregulatory system [1]. Hyperglycemia, the hallmark of diabetes, is the primary consequence of this dysregulation. Chronic hyperglycemia in diabetes is associated with long-term complications involving tissue damage and organ failure, which can decrease life expectancy and even cause death. The International Diabetes Federation estimates that, by 2017, diabetes affected 425 million people worldwide, of whom, 4 million died in the same year. These figures are expected to increase dramatically in the coming decades, placing a rising burden on health care systems [2].

Most diabetes can be categorized into 3 subgroups: type 1 diabetes (T1D), type 2 diabetes (T2D), and gestational diabetes (GDM). Over the long term, T2D patients become resistant to the normal effects of insulin and gradually lose their capacity to produce enough of this hormone. A wide range of therapeutic

options are available for patients with T2D. At the early stages of disease, they commonly receive medications that improve insulin secretion or insulin absorption, but eventually they must receive external doses of insulin. On the other hand, T1D patients have severe impairments in insulin production, and must use external insulin exclusively to manage their blood glucose (BG). Treatment of T1D requires consistent doses of insulin through multiple daily injections (MDIs) or continuous subcutaneous insulin infusion (CSII) using a pump. GDM is treated similarly to T2D, but only occurs during pregnancy due to the interaction between insulin and hormones released by the placenta.

In each class of diabetes, timely diagnosis, education of patients in self-management, and continuous medical care are required to prevent acute complications (eg, ketoacidosis) and minimize the risk of long-term complications (eg, nephropathy, retinopathy, diabetic foot, cardiovascular disease, or stroke). In addition to medication, management of diabetes requires adherence to an array of self-care behaviors that are often very burdensome for patients: carefully scheduling meals, counting carbohydrates, exercising, monitoring BG levels, and adjusting endeavors on a daily basis. The effects of nonadherence to recommended treatment are not immediately evident and long-term complications may take years to develop. Accordingly, diabetes therapy is complex, and therapeutic decisions need to take into account diverse medical factors and lifestyle-related activities that must be optimized to improve diabetic patients' quality of life.

Artificial intelligence (AI) is a quickly growing field, and its applications to diabetes research are growing even more rapidly as shown in Figure 1, which is a gross estimate of the number of related articles in the Google Scholar database.

In the literature, intelligent algorithms are widely used in data driven methods to support advanced analysis and provide individualized medical aid. There is also evidence that an increasing number of health care companies are applying these techniques [3]. Short-term prospects indicate they are likely to have considerable success in clinical practice. The main reasons for this growth include the explosive increase in the amount of available information, along with the improved performance of intelligent methodologies capable of handling and processing this information, both of which have led to the development of tools and applications which can enhance the effective management of complicated diseases, including diabetes and cancer.

Over the last decade, the entire paradigm of diabetes management has been transformed due to the integration of new technologies such as continuous glucose monitoring (CGM) devices and the development of the artificial pancreas (AP), along with the exploitation of data acquired by applying these novel tools. AI is attracting increased attention in this field because the amount of data acquired electronically from patients suffering from diabetes has grown exponentially. By means of complex and refined methods, AI has been shown to provide useful management tools to deal with these incremental repositories of data. Thus, AI has played a key role in the recognition of these systems as routine therapeutic aids for patients with diabetes.

The literature offers ample evidence of the use of artificial intelligence methods in the field of diabetes, such as in general surveys [4,5] or in particular domains, for example early diagnosis [6]. In this manuscript, we describe the latest efforts and advances in the application of AI methodologies to diabetes management and decision support. Background information on AI methods is provided in the remainder of the Introduction section.

Figure 1. The number of published articles in Google Scholar that include the terms "diabetes" and "artificial intelligence.".
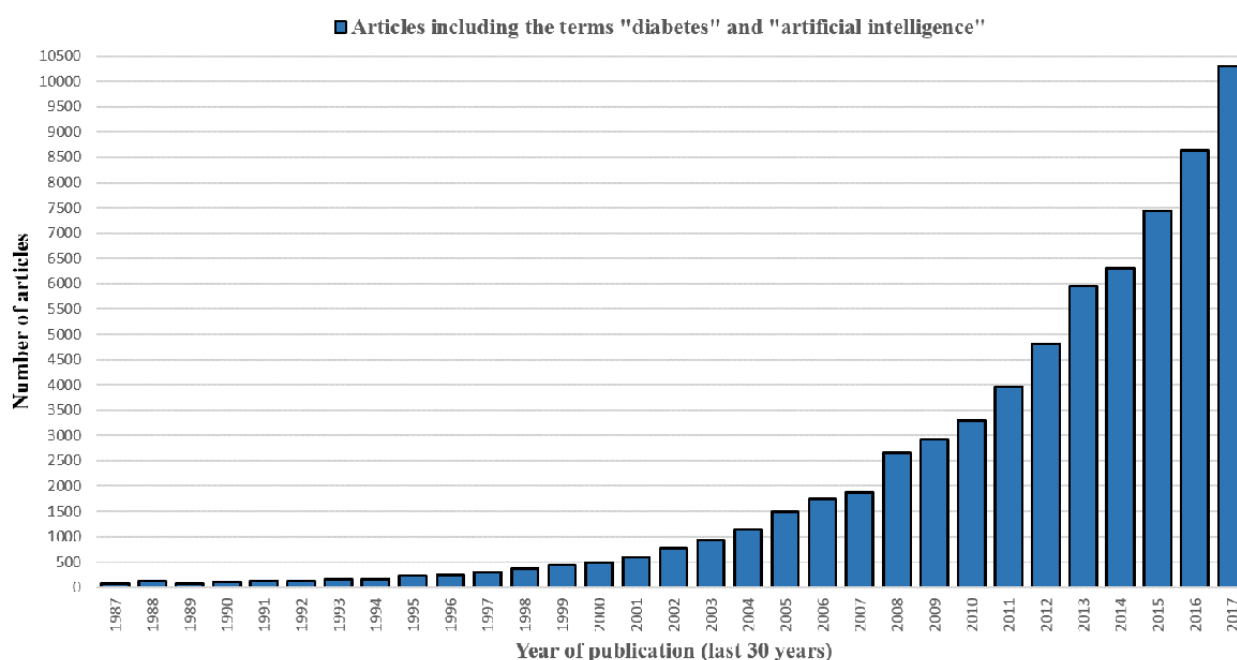
We then provide a detailed description of the methodological approach used in this review, report the results of a literature analysis, discuss studies on various sub-topics, and conclude with a brief summary and a discussion of future challenges.

Artificial Intelligence Techniques

Defining the concept of AI, computational intelligence, or machine intelligence is not a trivial undertaking. In this paper, we refer to AI as a branch of computer science that aims to create systems or methods that analyze information and allow the handling of complexity in a wide range of applications (in this case, diabetes management). Although the application of AI algorithms involves highly technical and specialized knowledge, this has not prevented AI from becoming an essential part of the technology industry and making contributions to major advances within the field. This section will provide a short overview of several well-known computational intelligence paradigms. For a more in-depth discussion of various intelligent algorithms, theoretical results, and applications, the reader is referred to the following book by Nilsson [7]. In this study, we categorized methodologies with respect to the objective sought: to explore and discover information, to learn using information, or to extract conclusions from information (see Figure 2).

Learning from Knowledge

Acquisition of knowledge is a key requirement of solutions intended to exhibit intelligent behavior. Because learning is an effective way to introduce such knowledge, most AI studies to date have employed learning techniques (see Figure 3). The primary aim of learning from knowledge is to allow computers to learn automatically without human intervention or assistance. This process could involve any method that includes some inductive component, ranging from a simple Kalman filter to a complex convolutional neural network. No method is inherently better than any other; each is more or less well-suited to different scenarios, for example, a softer learning curve, faster execution, or more flexible solutions. Furthermore, the performances of various methods are closely related to the quality and quantity of data: when more information is gathered, and less noise is present in the data, better solutions can be obtained. The most important families of techniques are artificial neural networks (ANNs), support vector machines (SVMs), random forest (RF), evolutionary algorithms (EAs), deep learning, Naïve Bayes, decision trees, and regression algorithms.

Exploration and Discovery of Knowledge

The discovery of knowledge revolves around the exploration and creation of algorithms for retrieving potential information from databases, commonly referred to as knowledge discovery in databases (KDD). The primary objective of KDD is identification of valid, potentially useful, and understandable information. KDD involves evaluation and interpretation of patterns and models for making decisions about what does and does not constitute knowledge, that is, distinguishing between data that are useful and those that are (in the context of interest) useless. Therefore, KDD requires broad and deep knowledge about the area of study.

The overall KDD process may be characterized into 6 steps in the cross industry standard process for data mining (CRISP-DM) model (Figure 4) [8]: business understanding, data understanding, data preparation, data modeling, evaluation of the model, and deployment. The application of data mining modelling is the most technical stage of the process. Techniques for data mining have taken much of their inspiration from learning algorithms and statistics, although the two types of approaches have different objectives. The most important data mining tasks involve the detection of anomalies, identification of dependencies between variables, regression, clustering, and classification. Some examples of representative techniques for this process are k-means, the k-nearest neighbor algorithm, and hierarchical clustering (HC).

Figure 2. A taxonomy of some of the best known artificial intelligence methods.
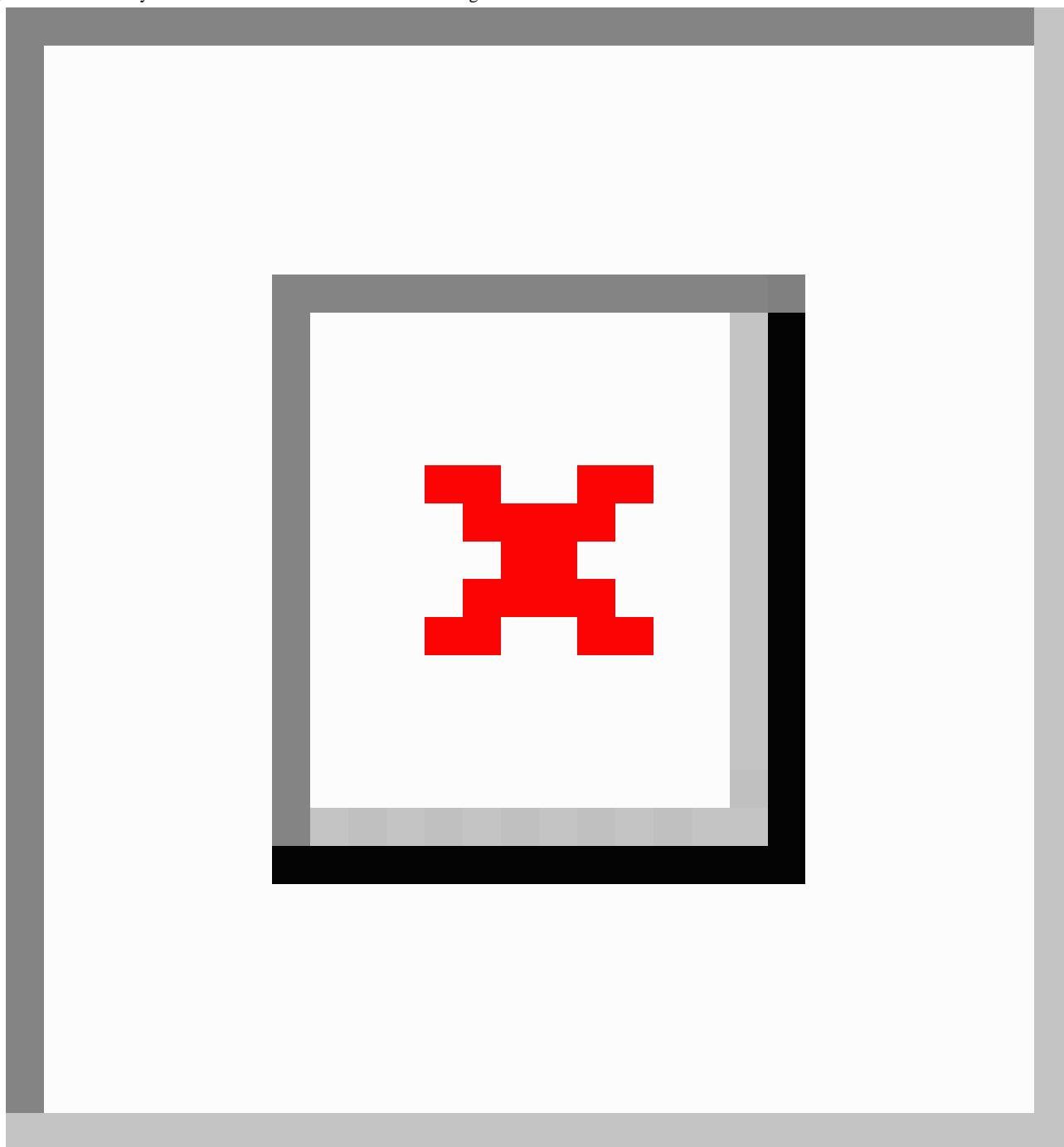
Figure 3. A general diagram of the learning algorithm process.
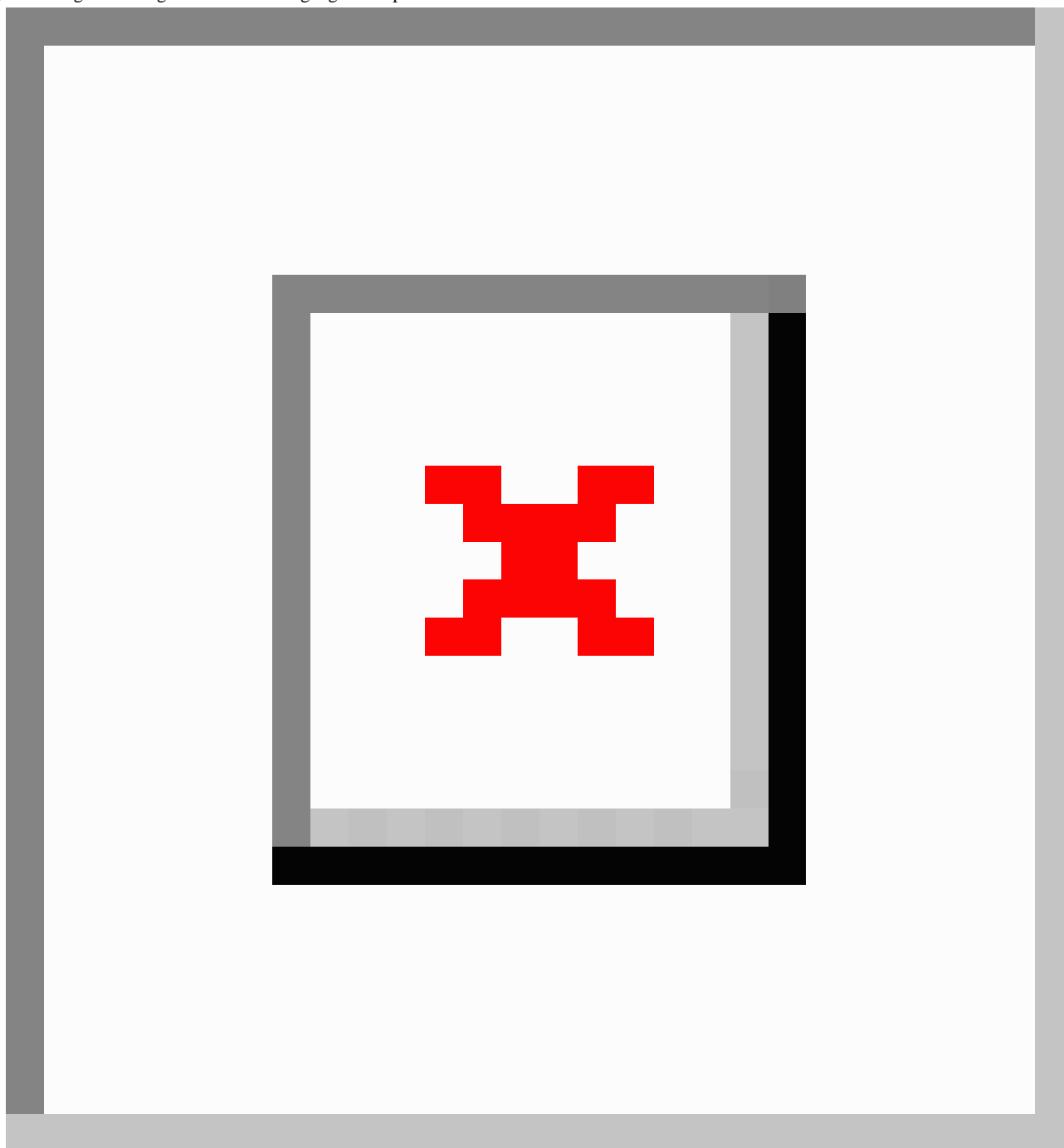
Figure 4. General CRISP-DM model for the knowledge discoveryin databases (KDD) process.

XSL•FO
RenderX
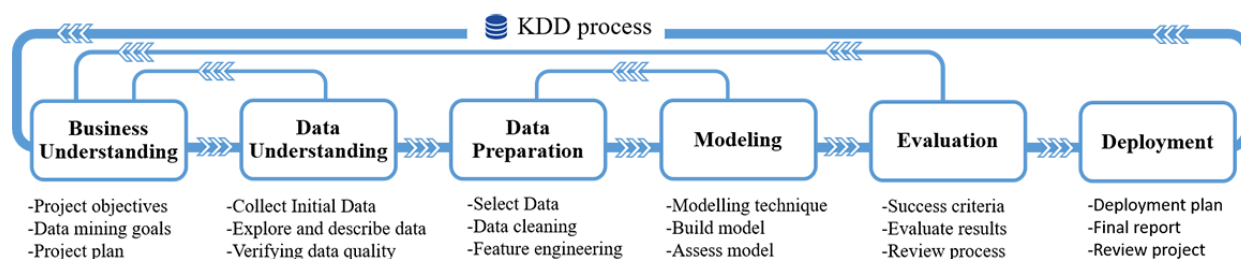
Reasoning from Knowledge

In this discourse, the idea of reasoning from knowledge denotes the creation of precise and effective ways to generate inferences in more precise and robust ways. Thus, reasoning from knowledge involves the use of logical techniques such as deduction and induction to generate conclusions from the available knowledge. The primary objective of systems that implement reasoning mechanisms is to perform tasks at a human-expert level in a narrow, specialized manner within the domain of interest. Such systems commonly apply heuristics to guide reasoning and reduce the search space of possible solutions.

These systems are based on 3 main components. First, a knowledge acquisition system is used to gather and collect inferences that can be used for further development. In this context, such a system is used to extract new rules and gather information. Second, a knowledge base, characterized by rules and information, is used for problem solving. Important aspects here include relations, conditions, recommendations, directives, and strategies. Finally, the inference engine links the knowledge base with the gathered information. Overall, this process facilitates reasoning, whereby the system becomes able to facilitate the realization of the anticipated solution. It is possible for a system structured on this basis to transfer expert knowledge directly to the knowledge base. This, in turn, helps to build new solutions based on previous cases, or to deal with ambiguous concepts and uncertainty. Representatives of these tasks include rule-based reasoning, case-based reasoning (illustrated in Figure 5), and fuzzy logic.

Figure 5. The case-based reasoning circle.
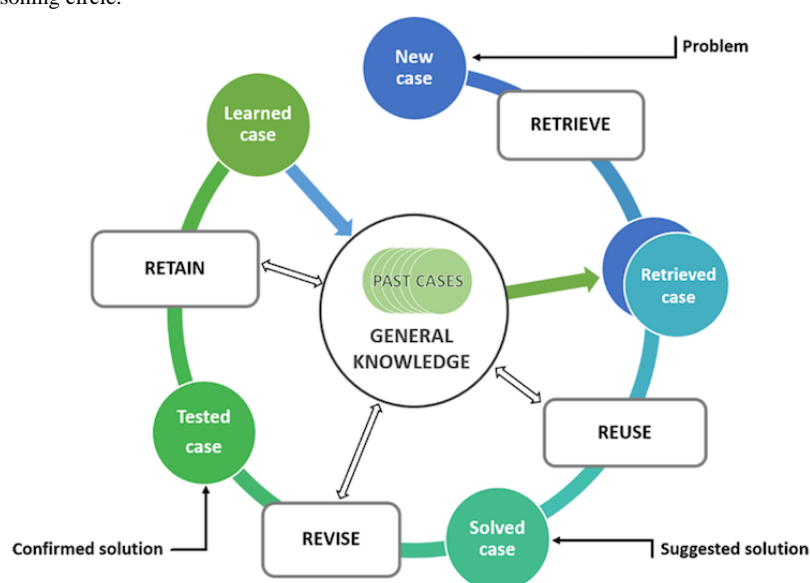
Methods

A review of the literature was conducted using the PubMed database. The selection of this bibliographic system as the primary data source was motivated by the sharp increase in the number of articles in the database and the strong link between these articles and the health care sector. PubMed has been validated as a reliable tool for retrieving information on medical research and clinical applications. Only English-language documents published between 2010 and 2018 were considered. The search terms listed in Textbox 1 were used to identify terms in the abstracts, titles, and keywords of the documents.

The search terms were explored and combined, yielding 1841 "hits." The terms "diabetes," "management," "artificial pancreas," and "blood glucose" were combined with the remaining terms using a conjunctive operator, and these terms were used as keywords to create individual datasets comprising all references to the following phrases: "artificial intelligence" (186), "computational intelligence" (179), "machine learning" (88), "data mining" (111), "deep learning" (3), "k-means" (9), "fuzzy logic" (24), "heuristic" (10), "clustering analysis" (281), "Bayes" (19), "decision tree" (67), "random forest" (21), "particle swarm optimization" (7), "pattern recognition" (31), "genetic algorithm" (43), "supervised algorithm" (14), "unsupervised algorithm" (9), "knowledge-based" (14), "case-based reasoning" (11), "decision support system" (71), "self-organizing map" (4), "evolutionary computation" (2), "neural network" (72), "natural language processing" (34), "reinforcement learning" (6), clustering (510), and "support vector machine" (23). Each of these datasets was then combined to build an objective dataset of articles. A comprehensive review was performed of all references cited in the datasets. Finally, the bibliography of the reviewed articles was thoroughly explored to find titles with relevance to the main focus of this study.

The complete method is summarized in Figure 6. The resultant final collection of articles was divided into various categories, designed to assist in the grouping of studies according to their shared and specific characteristics. Over the course of the systematic review, the subcategories were fused and fixed to accommodate the merging of information.

Textbox 1. The terms used in the search queries.

Artificial intelligence; Artificial neural network; Artificial pancreas; Blood glucose; Case-based reasoning; Cluster analysis; Clustering; Computational intelligence; Data mining; Decision support systems; Decision tree; Deep learning; Diabetes; Evolutionary computation; Fuzzy logic; Genetic algorithm; Heuristic; K-means; Knowledge-based; Machine learning; Management; Natural language processing; Naïve Bayes; Particle swarm; Pattern recognition; Random forest; Reinforcement learning; Self-organizing map; Supervised learning; Support vector machine; Unsupervised learning

Figure 6. Summary of the review process and classification of articles into a set of subdomains.
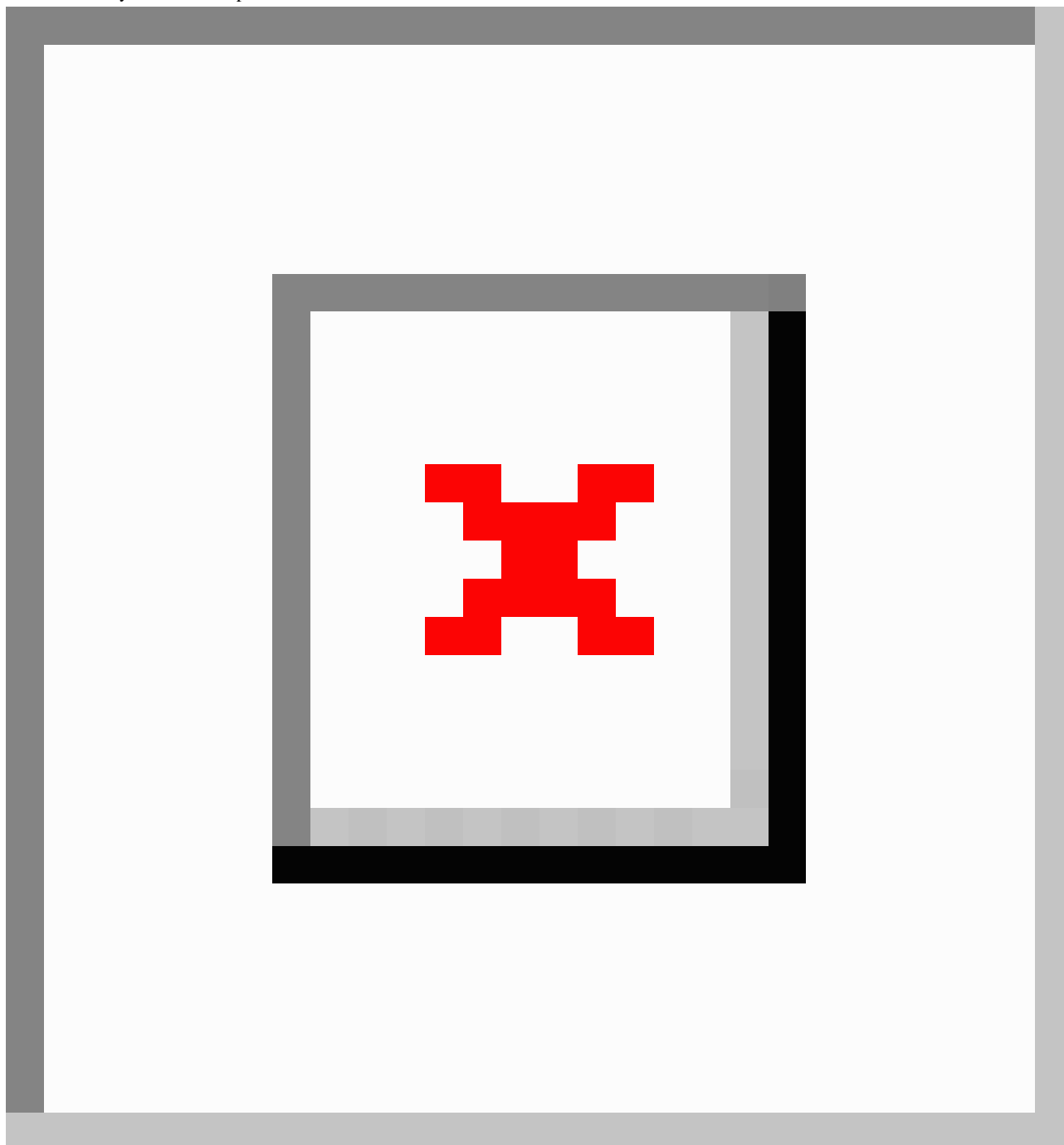
Results

Main Findings

Ultimately, 141 papers were included in the review. The potential of AI to enable diabetes solutions has been investigated in the context of multiple critical management issues. In this section, we use the following proposed diabetes management categories to summarize the latest contributions and the results described in the reviewed articles:

- Blood glucose control strategies
- Blood glucose prediction
- Detection of adverse glycemic events
- Insulin bolus calculators and advisory systems
- Risk and patient personalization

- Detection of meals, exercise and faults
- Lifestyle and daily-life support in diabetes management

As seen in Figure 7, the majority of the papers were published in the last 3 years, reflecting a clear acceleration in the application of AI techniques to diabetes management. In the following, contributions to each of the subdomains are detailed and discussed.

Blood Glucose Control Strategies

Development of the AP has been intensively pursued over the past decade. An AP consists of an automated system that mimics islet physiology, including a glucose sensor, a closed-loop control algorithm, and an insulin infusion device. The ultimate goal of an AP system is to improve overall diabetes management and reduce the frequency of life-threatening events associated with T1D. The algorithms used by the AP to calculate insulin dosage have been intensively investigated, either using data from diabetic patients or computer simulated patients, commonly named virtual patients (VP). The major candidate algorithms are derived from traditional control engineering theory; however, AI has become more established over the past few years and could ultimately provide better candidates to meet the challenges of an AP [9].

Although AI and control engineering have converged to some extent as the two fields incrementally exchange methods, here we will focus on studies dealing with closed-loop algorithms based on AI techniques. We direct interested readers to a recent comprehensive review on AP systems [10].

Three main AI methodologies have been established as control techniques in recent years: FL, ANNs, and reinforcement learning (RL). Most alternatives to control engineering algorithms are based on FL. Controllers apply FL theory to imitate the lines of reasoning of diabetes caregivers. Thus, the primary benefit of FL over classic control engineering is the ability to deal with nonlinearities and uncertainties. However fuzzy logic systems have not yet been proven to clearly outperform well-tuned classical approaches.

MD-Logic [11,12] was developed by authors who sought to individualize glycemic control using a fuzzy controller. Two feasibility studies were conducted in cohorts of 7 T1D patients to introduce the methodology and test the viability of the controller. Subsequently, a randomized crossover trial was conducted in 12 T1D patients [13]. The results suggested that the fuzzy method could improve nocturnal BG control without increasing the risk of hypoglycemia. Following the success of these feasibility studies, the authors performed a randomized crossover study of 56 young patients over 3 days [14]. The results confirmed a reduced rate of nocturnal hypoglycemia and superior glycemic control in comparison with insulin pump treatment. In a home-based randomized trial of 15 T1D patients [15], the authors compared the fuzzy AP and sensor-augmented pump over 4 nights; the results confirmed the feasibility, safety, and efficiency of their approach in a home setting. Later, an extended study of 24 T1D patients during 6 weeks of nocturnal control demonstrated the safety and effectiveness of long-term use of a FL-based controller. In a recent clinical trial evaluating remote patient monitoring of the FL controller, the AP was tested in 75 T1D patients for 4 consecutive nights. The results demonstrated safe and efficient glycemic control. Further studies will evaluate the MD-Logic controller implemented in MiniMEd 690G [16].

Figure 7. Number of articles reviewed according to subdomain and year of publication (BG: blood glucose).
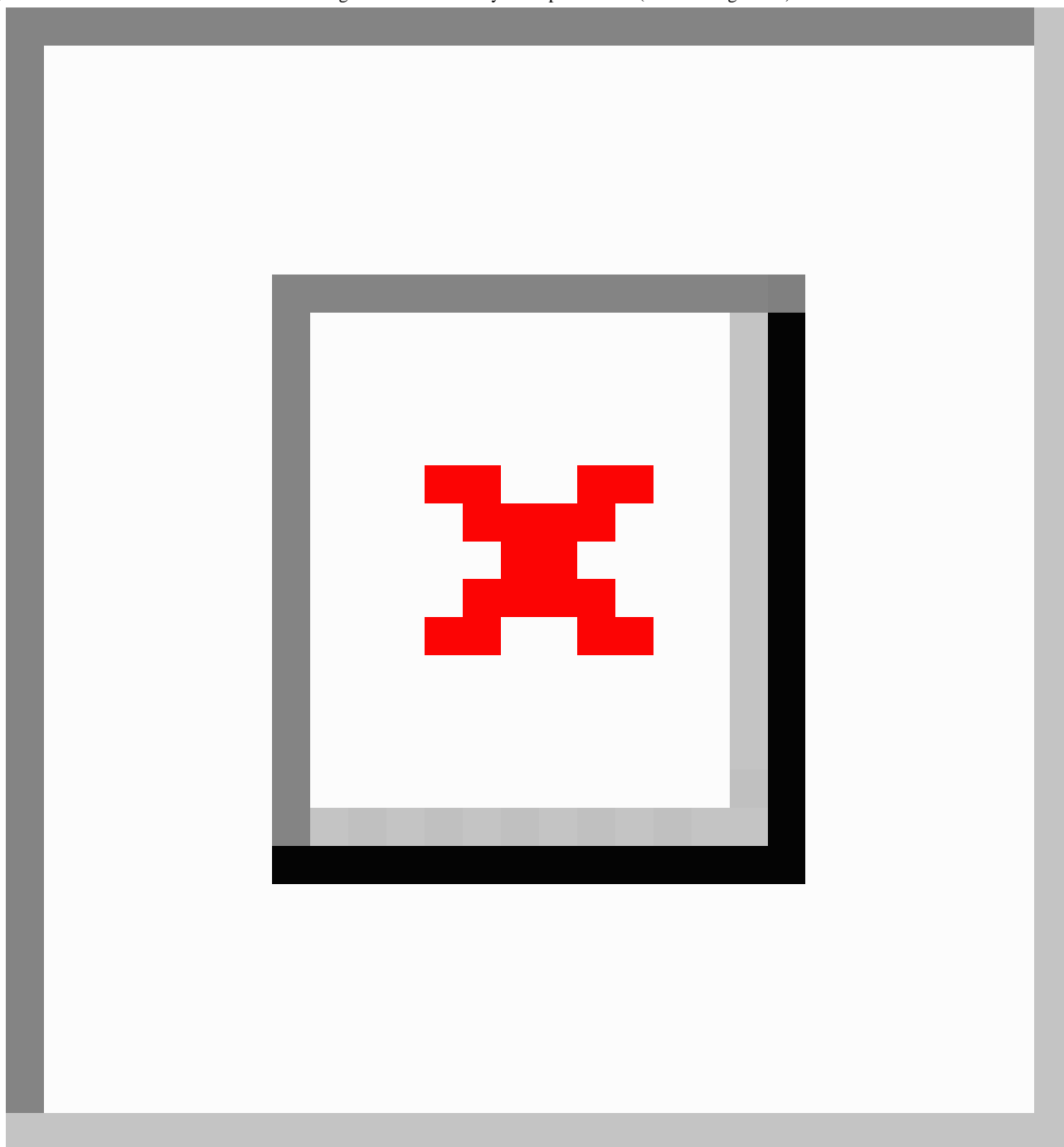
Other research groups have also investigated the application of FL to BG control. For example, Mauseth et al [17] reported a FL controller designed to personalize glycemic control. They tested it in 30 virtual patients on the UVA/Padova T1D simulator. Next, to demonstrate the feasibility of their approach, they conducted a pilot study in 12 T1D patients [18]. In a later study, they proposed stressing a fuzzy controller with high-fat meals and exercise and tested this approach in a trial with 10 T1D patients [19]. The results revealed deficits in their previous approach and ultimately led to improvements in the FL controller.

Other FL approaches have been tested using virtual patients or simulations [20-24]. For example, Miller et al reported a fuzzy controller combined with a learning algorithm that extracted initial patient profiles using open-loop data [23]. Another example was provided by Dinani et al [24], who suggested combining fuzzy and sliding-mode controllers, with the goal of using feedback to govern the insulin delivery rate more aggressively.

Another AI method that has been increasingly adopted in the area of control algorithms is the so-called reinforcement algorithm [25]. Daskalaki et al presented an adaptive, patient-specific BG control strategy based on the Actor-Critic learning approach and tested their approach on 28 virtual T1D patients (adults, adolescents, and children) [26].

Another AP study based on RL was proposed by Daskalaki et al, who proposed using an Actor-Critic algorithm to optimize

insulin infusion for personalized glucose regulation, and evaluated the system using virtual patients [27]. The results revealed that their novel tuning method decreased the risk of severe hypoglycemia, especially in patients with low insulin sensitivity. Other AI-related techniques used to support the development of the AP included modeling of glucose metabolism with an SVM [28].

Over this period, several groups proposed complementary AI algorithms to support AI controllers. Fereydouneyan et al proposed assisting the controller with a genetic algorithm (GA) that optimizes the values for two inputs and one output membership function [22], with the goal of preventing fluctuations caused by derivatives in fuzzy design. Another GA was used in the work of Catalogna et al to support an ANN controller [29]. In that case, the proposed GA optimizes network topology and learning features, instead of using the trial-and-error approach commonly adopted in ANN topology determination. In line with previous studies, Khooban et al proposed a controller assisted by particle swarm optimization that optimizes the parameters of the glucose-insulin model [21]. Following the trend of combining methods to control BG with metaheuristics, Yadav et al proposed the use of a Cuckoo search algorithm to optimize the gain of the controller [20]. Optimization of AI techniques, aside from their use in conjunction with AI controllers, have been used in other studies to determine controller parameters. For example, Tang et al used GA to mine information from patients' medical histories to generate multiple customized models [30], and Greenwood et al proposed the use of particle swarm optimization to adjust the function cost of an economic predictive control model [31].

Complementary to advances in control algorithms, efforts have been made to improve models that attempt to capture glucose-insulin dynamics. Since the publication of the seminal paper in this field [32], interest has increased in applying neural networks to identification and control of nonlinear systems. Zarkogianni et al [33,34] developed a recurrent neural network trained with a real-time recurrent learning algorithm that models the BG kinetics of T1D patients and predicts BG levels using information about meal intake, BG measurements, and infused insulin. González-Olvera et al investigated the use of a fuzzy neural network in an attempt to combine the best properties of ANNs and fuzzy systems [35]. Specifically, they implemented a learning system that combines input signals, infused insulin,

and BG measurements with a membership initialization based on the fuzzy c-means algorithm. The following year, Alanis et al performed a more rigorous study using a recurrent neural network trained to model BG in T1D patients [36]. The approach they implemented considered glucose absorption via carbohydrates consumption and insulin infusion as inputs. These 3 neural network studies proposed a model of insulin dynamics as the first step in the design of a control scheme, and then validated the model using T1D patient data. Their results demonstrated that they were indeed able to capture BG dynamics. More recently, 2 studies presented a complete control scheme using neural networks [29,37]. Both studies sought to determine whether this technique could predict and control BG excursions in T1D using patient-specific models. The first approach was tested using the hyperinsulinemic-euglycemic clamp technique in 34 rats, and the second was tested using *in silico* data including BG measurements, administered insulin, exercise information, and ingested food. The results of both approaches demonstrated that ANNs are accurate methods for regulation of glucose levels.

Blood Glucose Prediction

The ability to anticipate BG excursions could provide early warnings regarding ineffective or poor treatments. Thus, information collected from new technologies for diabetes management, such as the CGM devices, could lead to real-time predictions of future glucose levels. Prediction of BG levels is challenging due to the number of physiological factors involved, such as delays associated with absorption of food and insulin, and the lag associated to measurements in the interstitial tissue. Errors of the CGM also increase the difficulty of predicting BG values (approximately 9% of the mean absolute relative difference for the best sensors [38]).

The results of this section are presented in Table 1, which captures the critical information from all studies in which AI methods were used to predict BG values. The table was designed to provide quick access to information about current technologies being tested. We outline the features of each study using key information, including prediction horizon (PH) in minutes, objective population criteria, number of participants in the cohort, mean number of monitored days per patient, mean number of monitored hours per day, existence of monitoring during the overnight period, type of monitoring technology, and information about physical activity.

Table 1. Summary of reviewed studies addressing blood glucose prediction: prediction horizon in minutes, objective population criteria, number of participants in the cohort, mean number of monitored days per patient, mean number of monitored hours per day, type of monitoring technology, existence of monitoring during the overnight period (O) and inclusion of exercise or physical activity information (E).

Prediction horizon (min)	Population	Cohort	Days	Time	O	E	Method	Ref	Year
15, 30, 45	T1D[a]	15	28	10 h	✓		ANN[b]	[39]	2010
30	T1D	12	10	24 h	✓		ANN	[33]	2010
75	Critical Care	1	16	15 h	✓		ANN	[40]	2010
75	T1D	27	5	24 h	✓		ANN	[41]	2011
30	T1D	5VP[c], 1	7	24 h	✓	✓	ANN	[42]	2011
30, 45	T1D	30VP	8	24 h	✓		ANN	[43]	2012
15, 30, 60, 120	T1D	27	13	24 h	✓	✓	RF[d]	[44]	2012
30	T1D	20VP, 9	11, 7	24 h	✓		ANN	[45]	2012
30	T1D	10	3	24 h	✓		SVM[e], RA[f], ANN	[46]	2013
15, 30, 45	T1D	23	6.1	24 h	✓		RA, ANN	[47]	2013
15, 30, 60, 120	T1D	27	13	24 h	✓	✓	SVR	[48]	2013
30	T1D	20	3	24 h	✓		ANN	[49]	2015
15, 30, 45, 60	T1D	6	11	24 h	✓	✓	ANN	[50]	2015
30, 60, 120	T1D	10	6	24 h	✓	✓	ANN	[51]	2015
30	T1D	15	13	24 h	✓	✓	ANN	[52]	2015
30	T1D	5VP, 1	30	24 h	✓		SVR	[53]	2016
—	T2D[g]	346	1	—			ANN	[54]	2016
5, 15, 30, 45, 60	T1D	15	13	24 h	✓	✓	Kernel	[55]	2016
60	T1D	5	90	24 h			EA[h]	[56]	2016
1440	T1D, T2D	8	3	24 h		✓	DT[i]	[57]	2016
30	T1D	3	10	24 h	✓		EA	[58]	2016
30,60	T1D	17	6	24 h	✓		RA	[59]	2017
60, 120, 150, 180	T1D	20VP	14	24 h	✓		EA	[60]	2017
0	T2D	3	23	—			NB[j]	[61]	2017
30, 60, 90, 120	T1D	10	10	24 h	✓		KNN[k], RF, EA	[62]	2017
120	T1D	100VP	14	24 h	✓		EA	[63]	2017
30, 60, 90	T1D & T2D	106	<7	24 h	✓		RA and ANN	[64]	2017

[a]T1D: type 1 diabetes.

[b]ANN: artificial neural network.

[c]VP: virtual patient.

[d]RF: random forest.

[e]SVM: Support Vector Machine

[f]RA: regression algorithm.

[g]T2D: type 2 diabetes.

[h]EA: evolutionary algorithm.

[i]DT: decision tree.

[j]NB: Naïve Bayes.

[k]KNN: k-nearest neighbor.

Finally, we highlight the main AI methods applied in the studies, the bibliographic reference, and the year of publication. AI for BG prediction has been addressed in as many as 13 parallel lines of research. Most of these studies focused on T1D because of the inherent utility of AI in this condition and the availability of high-frequency data collected from patients using a CGM

XSL•FO

RenderX

device. The results of our review reveal the range of PHs explored, from 5 to 180 minutes. Short-term predictions were the most frequently explored: 38 out of 49 studies (76%) used PHs below 60 minutes. ANN approaches were the most widely applied methodology, but other machine learning methodologies such as RF, SVM, or RAs are being adopted with increasing frequency.

Detection of Adverse Glycemic Events

As with BG prediction, glycemic episode detection encompasses a set of tools that deal with the complexity of effective BG control. However, in this section we will not address glucose values, but instead focus on the appearance of hyperglycemic or hypoglycemic events. These tools enable us to detect the occurrence of glycemic episodes and give us the opportunity to respond promptly to their effects. In contrast to the previous section, most of the reviewed studies on this topic focus on detecting hyperglycemia or hypoglycemia in situations when it is not possible to effectively monitor BG. Therefore, most of these studies deal with real-time approaches rather than predictions of future events. We summarize the studies dealing with detection of BG excursions in Table 2. Each scenario is represented by the following features: prediction horizon in minutes, objective population criteria, number of participants in the cohort, mean number of monitored days per patient, mean number of monitored hours per day, type of monitoring technology, existence of monitoring during the overnight period, and inclusion of exercise or physical activity information. Finally, we highlight the main AI methods applied in each study, the bibliographic reference, and the year of publication. The results revealed that nine of the 14 studies (64%) reported real-time detection systems, and 10 of these studies (93%) were specifically focused on T1D. Table 2 shows that over six of the approaches that exclusively addressed T1D (60%) gathered data from CGM sensors, whereas the remainder used an electroencephalogram (EEG) or self-monitoring blood glucose (SMBG) measurements. Studies focusing on T1D were performed with fewer than 15 patients, whereas studies of T2D included larger cohorts. Sensitivity and specificity were the most common outcomes used to assess the quality of approaches to glycemic detection. Although this section contains fewer papers than the one on BG prediction, we identified more than 10 research groups contributing to this topic. In particular, researchers at the Centre for Health Technologies (Faculty of Engineering and Information Technology, Sydney, Australia) have published five studies on this topic over the last 7 years.

Insulin Bolus Calculators and Advisory Systems

The most common insulin therapies for diabetics, continuous subcutaneous insulin infusion (CSII) and multiple daily insulin injections (MDI), operate according to similar principles [79]. Both utilize basal insulin (injection of long-acting basal insulin and infusion at a constant basal rate, respectively) and bolus insulin (injection of quick-acting bolus insulin and meal boluses, respectively) to cover meals or snacks. The calculation of correct insulin doses and the estimation of the amount of carbohydrates is a regular task in the daily life of many insulin-dependent patients. Bolus advisors are based on previous insulin doses, BG measurements, planned carbohydrate estimates, and other

patient-specific parameters, including insulin-to-carbohydrate ratio and insulin sensitivity. Manually calculating bolus doses and counting carbohydrates can be complex and challenging because individuals must consider multiple parameters to achieve satisfactory glucose control, and miscalculation of these values may result in persistent glycemic episodes.

To support carbohydrate estimation and determination of insulin doses by patients, tools for providing bolus advice and carbohydrate estimates are increasingly being adopted. These tools seek to increase the accuracy of mealtime and correction boluses. AI has been used to provide sets of tools to improve the accuracy of carbohydrate estimates and to calculate the optimal insulin bolus for the ingested meal

We identified several studies that applied AI to systems aimed at supporting patient decisions by issuing advice regarding meals, exercise, or medication. Research groups at the Imperial College London performed an extensive study of an insulin bolus calculator based on case-based reasoning methodology [80-84]. Their approach, which manages various dynamically optimized diabetes scenarios, was proven in a clinical trial (NCT02053051) to be a safe decision support tool. Additionally, this approach was demonstrated to improve glycemic control in diabetes management when it was combined with an AP system [84]. A similar approach was presented recently by another group [85], which also proposed an insulin bolus calculator based on case-based reasoning but, in contrast to other bolus calculators, it used a novel temporal retrieval algorithm. The Center for Biomedical Engineering Research at the University of Bern performed several important and extensive studies [86-90] investigating the GoCARB system, which provides dietary advice to diabetic patients based on automatic carbohydrate counting. Their approach is based on the use of computer vision techniques, such as feature extraction and SVM, and pilot studies show it to be an excellent assistive tool. We have also found several studies that validated their approach using the UVA/Padova patient simulator. Srinivasan et al proposed the use of a set of insulin delivery profiles optimized by a PSO to find the optimal open- and closed-loop profiles for various meal compositions [91]. More recently, another study [92] presented an approach based on ANN to optimize bolus calculation by patients using CGM. The results revealed that it was better at reducing the blood glucose risk index value than other approaches. Finally, Lee et al proposed an advisory treatment system that provides insulin, meal, and exercise recommendations [93]. Their study, which compared rule-based reasoning and k-nearest neighbor algorithms, concluded that the k-nearest neighbor algorithm was best suited to this approach.

Risk and Patient Stratification

Most commercially available tools and protocols for managing diabetes are based on general models of the diabetic population or involve subsets of patients defined by simple clusterization features and easily identifiable characteristics. However, the daily lives of diabetic patients are determined by a wide range of management scenarios that are not represented in these general models. Insulin-dependent patients must manage a highly complex process to maintain suitable levels of BG.

Table 2. Summary of reviewed studies addressing detection of adverse glycemic events: prediction horizon (PH) in minutes, objective population criteria, number of participants in the cohort, mean number of monitored days per patient, mean number of monitored hours per day, type of monitoring technology, existence of monitoring during the overnight period (O), and inclusion of exercise or physical activity information (E),

PH (min)	Population	Cohort	Days	Time	Measurements	O	E	Method	Ref[a]	Year
0 min	T1D[b]	6	1 day	10 h	EEG[c]	✓		ANN[d]	[65]	2010
0 min	T1D	30	80-247 days	—	SMBG[e]		✓	RF[f], SVM[g]	[66]	2012
0 min	T1D	15	1 day	10 h	CGM[h]	✓		ANN, PSO[i]	[67]	2012
0 min	T1D	10	30 days	4 h	CGM			SVM	[68]	2013
30, 60 min	T1D	15	12.5 days	24 h	CGM	✓	✓	SVM	[69]	2013
0 min	T1D	10	4.5 days	6 h	CGM			SVM	[70]	2013
30 min	T1D	10	17.3 days	24 h	CGM	✓		DT[j]	[71]	2013
24 h	T2D[k]	163	—[l]	—[l]	SMBG			RF	[72]	2015
0 min	T1D	15	1 day	4 h	CGM			ANN	[73]	2014
0 min	T1D	10	—[m]	—[m]	SMBG, ECG	✓		ANN	[74]	2014
2, 7, 30, 61-90 days	T1D, T2D	201, 323	—[n]	—[n]	SMBG			Pattern recognition	[75]	2014
0 min	T1D	15	1 day	10 h	ECG	✓	✓	ANN	[76]	2016
Past events	T2D	119695	>12 days	—	EHR[o]			NLP[p]	[77]	2016
0 min	T1D, T2D	500	1 day	2 h	SMBG			DT, ANN	[78]	2017

[a]Ref: reference.

[b]T1D: type 1 diabetes.

[c]EEG: electroencephalogram.

[d]ANN: artificial neural network.

[e]SMBG: self-monitoring blood glucose.

[f]RF: random forest.

[g]SVM: support vector machine.

[h]CGM: continuous glucose monitoring.

[i]PSO: particle swarm optimization.

[j]DT: decision tree.

[k]T2D: type 2 diabetes.

[l]344 data points.

[m]18 data points.

[n]787 data points.

[o]EHR: electronic health record.

[p]NLP: natural language processing.

Treatment of diabetes is governed by diverse factors, implying high intra- and interpatient variability [94]. Exercise, nutrition disturbances, age, and cardiovascular complications are just some of the long list of factors that can dramatically impact quality of life and undermine medication adherence even when patients follow their treatment regimen strictly. Such patient variability severely limits the use of general models, which cannot capture the specific physiological behaviors of individuals. Thus, an important step toward better risk detection and intervention is personalization of the system. Over the past decade, major research efforts have been devoted to developing management tools capable of stratifying patients in different segments of the population. Risk assessment and patient stratification methods are important to improving the management of diabetes, and therefore the overall health outcomes of diabetic patients, and consequently have attracted a greater share of attention from the medical community.

This category gathers all reviewed papers that systematically identified individual patients and their risk factors to manage and coordinate their care based on specific conditions and on evidence-based guidelines. Table 3 outlines the type of stratification together with the specific challenge. Main characteristics, such as number of years, cohort, and objective population, are also included. Finally, the table defines the AI methodology applied, bibliographic reference, and year of publication.

XSL•FO

RenderX

Table 3. Summary of studies addressing risk and patient stratification.

Stratification	Challenge	Period	Cohort	Population	Methods	Year	Ref[a]
Complications	Group risks of retinopathy	5 years	55	T1D[b]	DT[c], ANN[d]	2010	[95]
Disease complexity	Group combinations of comorbid conditions	2 years	15480	Chronic diseases	Hierarchical clustering	2011	[96]
Disease complexity	Group management profiles	3 months	239	T1D	Hierarchical clustering	2011	[97]
Disease complexity	Group management profiles	3.5 months	70	T2D[e]	K-means	2012	[98]
Complications	Group biomechanical foot profiles	6 months	97	T1D, T2D	K-means	2013	[99]
Disease complexity	Group by drug purchases	7 years	953	T2D	Knowledge discovery	2015	[100]
Complications	Group risks of renal disease	3 years	109	T2D	K-means	2015	[101]
Complications	Group risks of complications	10 years	1441	T1D	Learning models	2015	[102]
Complications	Group risks of complications	5 years	84	T1D, T2D	RA[f] and ANN	2015	[103]
Complications	Group risks of retinopathy	<1 year	345	T2D	SVM[g], RF[h], DT, NB[i]	2015	[104]
Disease complexity	Groups of blood glucose profiles	4 months	10	T1D	Hierarchical clustering	2016	[105]
Complications	Group personal networks types	5 years	1862	T2D	RA, K-means;	2016	[106]
Disease progression	Group risks of T2D progression	5 years	24331	T2D	NB	2016	[107]
Complications	Group risks of retinopathy	2 years	323378	T2D	RF	2017	[108]
Disease complexity	Group blood glucose profiles	2 years	27005	T2D	RF	2017	[109]
Disease complexity	Groups of HbA1c profiles	5 years	684	T2D	RAs, KNN	2017	[110]
Weight intervention	Group of BMI profiles	31 years	2540	T2D	RA	2017	[111]
Complications	Group by retinopathy, neuropathy, or nephropathy	3,5,7 years	943	T2D	RF, RA	2018	[112]

[a]ref: reference.

[b]T1D: type 1 diabetes.

[c]DT: decision tree.

[d]ANN: artificial neural network.

[e]T2D: type 2 diabetes.

[f]RA: regression algorithm.

[g]SVM: support vector machine.

[h]RF: random forest.

[i]NB: Naïve Bayes.

Detection of Meals, Exercise, and Faults

Because people with both types of diabetes need support to successfully manage their disease, solutions with higher accuracy that require less user interaction are associated with higher-quality diabetes treatments. Tools or algorithms capable of early detection of critical events affecting glycemic control, such as exercise, a meal, or an infusion set failure, are critical for systematic automation of both closed-loop and open-loop systems. Insulin-dependent patients monitoring their glucose with CGM devices use BG measurements to calculate insulin infusion rates. Consequently, failure of these devices can lead to episodes of hyperglycemia or hypoglycemia. Leal et al proposed an approach using SVM to detect correct and incorrect measurements in real-time CGM [113]. They tested their system on 23 critically ill patients and obtained promising results in patients with sepsis or septic shock. The same objective was pursued in the work performed by Turksoy et al [114], who used a k-nearest neighbor algorithm for the diagnosis of faults

and the data from 51 patients to validate the performance of their approach. For the detection of inaccurate measurements by glucose meters, another study [115] developed an SVM algorithm to minimize the effect of hematocrit on glucose measurement and tested their method on 400 BG samples.

Physical activity offers multiple benefits for diabetic patients, but also complicates the management of diabetes, especially in T1D patients. Some of the factors affecting BG dynamics during exercise include the intensity, duration, and type of exercise, insulin on board, and the carbohydrate absorption rate. Tools and systems focused on automated detection of exercise could improve the accuracy of treatments. Turksoy et al also proposed the use of a k-nearest neighbor classification algorithm to automatically detect exercise type and intensity in an AP system [116]. They tested their approach in 5 T1D patients and reported a sensitivity of 98.37%. Similarly, Jacobs et al proposed a regression model to automatically detect physical exercise in patients carrying an accelerometer and a heart rate sensor [117].

XSL•FO
RenderX

The system was assessed in 13 T1D patients, yielding a sensitivity of 97.2% and a specificity of 99.5%.

Meal detection is important in AP systems that do not permit manual meal announcements and as a safety system for patients who may forget to enter meal information manually. Turksoy et al have also investigated the development of a meal detection system based on analysis of CGM signals using an unscented Kalman filter and a fuzzy system to estimate the carbohydrates content [118-120]. Their approach was validated *in silico* with 30 T1D patients using the UVA/Padova simulator, which revealed a sensitivity of 91.3% and an error of 23.1% in carbohydrate estimation; and *in vivo* using data from 11 T1D patients, which revealed a sensitivity of 93.5% for meals and 68.0% for snacks.

Insulin pump failure may result in prolonged hyperglycemia or diabetic ketoacidosis. Early detection of failures could minimize the associated risk. Cescon et al proposed the use of a time-varying autoregressive model to develop a patient alert system [121]. Validation with data from 9 T1D patients during 18 weeks of infusion set wear revealed that the system had 50% sensitivity and 66% specificity.

Lifestyle and Daily-Life Support in Diabetes Management

Lifestyle management is a fundamental aspect of diabetes care. Sedentary living, stress, nonadherence to medication, lack of regular medical examinations, and bad habits can lead to discontinuation of treatment for patients with diabetes. From the time of diagnosis, patients are required to optimize their lifestyles to manage complications and other comorbid conditions, with the overall goal of enhancing their own care. Current technologies and data warehouses enable solutions that model data and make quality decisions based upon them. Decision support systems (DSSs) consist of tools focused on helping patients or doctors to manage diabetes therapies. These systems usually have monitoring features that facilitate systematic recording of information about diet, physical activity, medication, glucose measurements, etc and combine it with tools to support both patients and clinicians, with the overall goal of enhancing therapeutic outcomes.

Multiple studies aimed at developing DSSs to manage diabetes have been proposed since 2010. One of the most productive approaches is the METABO project [122-127]. This project involves monitoring and advanced features including tools to prevent future excursions, dynamically optimize care pathways, extract patterns via knowledge discovery, and guide weight loss programs. The authors conducted several pilot trials, including usability tests in 36 T1D patients. The MOSAIC project [100,110,112,128], another important project, is focused on the development of a DSS for T2D management, with a special focus on the risk assessment of related complications using data mining methods. Another daily-life support system has advanced tools, such as a recommender system that employs case-based reasoning and an integrated BG prediction tool based on evolutionary computation [129].

Recently, Everett et al presented a DSS using machine learning to promote adherence to physical activity and weight reduction

[130]. Authors validated the system with 55 patients with prediabetes. Previously, Yom-Tov et al proposed a DSS based on a RL algorithm that automatically sends messages to patients who are following a personalized plan for physical exercise [131]. The approach was validated in 27 sedentary T2D patients. Daily-life support systems using AI tools for GDM were also investigated. A weight management proposal was presented in the MediClass system [132]. The system, which is based on the application of a natural language processing (NLP) algorithm, was validated during the postpartum visits of 600 GDM patients. Rigla et al also investigated tools for GDM patients [133]. They proposed a mobile app based on an AI-augmented telemedicine DSS as a tool for helping GDM patients. Later, they presented a platform to remotely evaluate patients using a classifier based on a clustering algorithm and a decision tree learning algorithm [134]. The system was evaluated in 90 GDM patients. The results showed a reduction in the time devoted by clinicians to patients and in face-to-face visits per patient.

Six other studies have proposed alternatives to the manual creation of patient care workflows. The studies offer support for the design and deployment of diabetes management protocols, as well as ways to continuously improve patient tracking throughout the entire process. Cleveringa et al presented a system aimed at decreasing cardiovascular risk of T2D patients by optimizing patient care workflows [135,136]. The authors validated their system by administering questionnaires to 3391 T2D patients. Miller et al used a machine learning approach to extract information from drug prescriptions from electronic health record (EHR) data and identify factors associated with patient care flow deviations [23]. Another DSS with care flow tools was presented in the work of Alotaibi et al [137]. This system focuses on the management of T2D patients using advanced features, such as computerized alerts and reminders. It was tested in 20 T2D patients for 6 months and resulted in reduced HbA_{1c} levels and improved diabetes awareness. Fernandez-Llatas et al proposed using data mining methods to enable the dynamic design of care protocols but highlighted the need for mechanisms to reduce the Spaghetti Effect and make DSSs usable by experts [138]. Contreras et al developed a diabetes management system to integrate a series of AI models and tools with an engine to manage diabetes patient care flows [139]. Finally, Suh et al proposed a dynamic care flow system that applied data clustering together with rule mining techniques to prioritize required user tasks [140].

Other tools have been proposed for improving daily-life support for diabetes therapies. Four different tools have been designed to analyze online discussion forums and social networks to extract relevant information. First, Grieves et al compared multiple machine learning techniques (decision trees, SVM, bagging, and Bayes) to analyze patients' online comments with the aim of predicting patients' assessment of hospital performance [141]. Second, Valdez et al proposed using a k-means clustering analysis to identify communication patterns both on and off Facebook [142]. They validated their tool in a cohort of 700 T2D patients. Third, Chen et al proposed clustering based on repeated bisecting k-means with the goal of obtaining patient experience information, including emotional and temporal aspects of diabetes management [143]. Finally,

XSL•FO

RenderX

Hamon et al proposed using NLP methods to extract information about patients' skill in managing diabetes [144].

Furthermore, tools have been developed to analyze clinical appointments, medication, and therapy adherence. For example, a machine learning approach to examine medication adherence thresholds and risk of hospitalization was implemented [145]. The system implemented in the study reported by Fioravanti et al promotes patient empowerment and adherence to therapy based on the automatic generation of feedback messages [146]. Greaves et al proposed a clinical DSS that issues medication interaction alerts based on clusters with similar management recommendations [147]. A fuzzy approach was also presented by Eghbali-Zarch et al to address the problem of medication selection in T2D patients [148]. Finally, Kurasawa et al proposed a machine learning algorithm to predict missed clinical appointments and help patients continue regular doctor visits [149].

Discussion

By systematically examining high-quality articles in the PubMed database, we identified a series of studies with the goal of evaluating the latest efforts of AI-enabled solutions for diabetes management. The topics we reviewed suggest that prediction and prevention are currently being revitalized and reinforced by AI applications, whereas "safety and failure detection" has been less extensively reviewed, constituting fewer than 6% of the studies we encountered. Similarly, few investigations have delved into the application of AI techniques to early detection of critical issues such as exercise, meals, infusion set failures, and so forth. Exploiting the latest AI techniques to improve the safety of both AP systems and open-loop tools has the potential to dramatically improve performance. By contrast, research on closed-loop systems, representing 31 out of 141 of the reviewed studies (22%), has been the most productive area for AI applications. Most of these efforts addressed fuzzy techniques, but the application of other methodologies has begun to attract increasing interest. In our opinion, researchers in this field should continue to take advantage of the latest improvements in AI and to combine them with development of the AP. A considerable number of the reviewed studies, 41 out of 141, investigated BG, either to develop models that enable accurate predictions of BG concentrations (27 studies) or to detect possible BG events (14 studies). Multiple studies reported accurate prediction and detection tools that promise to improve management resources for current and future therapies. These tools include bolus advisors, as well as both lifestyle and patient stratification.

Our findings show the increasing importance of AI methods for diabetes management. We think these methods will encourage further research into the use of AI methods to extract knowledge from diabetic data. In general, the most striking advances in the application of AI techniques come from data-driven methods that learn from large datasets. The ability to collect information from individual diabetic patients has led to a shift in diabetes management systems; accordingly, systems that lack access to valuable data will face substantial hurdles. Diabetes management is geared toward tailored management of therapies, at the level of smaller strata of patients or even individuals. Thus, management protocols provided to diabetic patients should be tailored to address their needs at various points during their illness. Furthermore, the availability of genetic data, such as that provided by metabolomics analysis, has also empowered the application of AI methods to personalization of diabetes management.

The increased availability of digitized health data from diabetic populations, along with the emerging applications of AI and research trends such as the AP and personalized medicine, suggests that we are moving toward a new paradigm for management of diabetes. This new outlook proposes to achieve custom delivery of diabetes care while tailoring professional practices, medical decisions, and treatments to individual patients. On the other hand, the inclusion of intelligent algorithms in decision making has ethical implications that should be addressed by physicians and scientists. The ethical risks associated with the release of personal data should also be investigated. For example, the increasingly frequent use of health apps and the potential use of tools based on AI by insurance companies could lead to discrimination or the exclusion (or both) of some citizens from health services.

A large number of studies have already been published on the application of AI to diabetes in a broad range of management domains. Our dive into PubMed demonstrates an acceleration in the pace of research on AI-powered tools designed to predict and prevent the complications associated with diabetes. Although the available technologies and methods for diabetes management are growing exponentially in terms of quantity and quality, the potential of AI to boost effective and accurate management of diabetes has already been demonstrated in both open- and closed-loop therapies. Research in this field should continue and should seek to discover the opportunities and advantages of applying AI methodologies in diabetes management that differentiate these strategies from other classical approaches.

Conflicts of Interest

None declared.

XSL•FO

RenderX

References

1. American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care 2010 Jan;33 Suppl 1:S62-S69 [FREE Full text] [doi: 10.2337/dc10-S062] [Medline: 20042775]
2. IDF Diabetes Atlas 8th edition Internet. 2017 URL: https://www.idf.org/e-library/epidemiology-research/diabetes-atlas/ 134-idf-diabetes-atlas-8th-edition.html [accessed 2018-05-23] [WebCite Cache ID 6zd9GKIXJ]
3. AI Heatmap: Healthcare Emerges As Hottest Area For Deals To Artificial Intelligence Startups Internet. 2016 URL: https:/ /www.cbinsights.com/research/artificial-intelligence-investment-heatmap/ [accessed 2018-05-23] [WebCite Cache ID 6zd9TF8UK]
4. Rigla M, García-Sáez G, Pons B, Hernando M. Artificial Intelligence Methodologies and Their Application to Diabetes. J Diabetes Sci Technol 2018 Mar;12(2):303-310. [doi: 10.1177/1932296817710475] [Medline: 28539087]
5. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine Learning and Data Mining Methods in Diabetes Research. Comput Struct Biotechnol J 2017;15:104-116 [FREE Full text] [doi: 10.1016/j.csbj.2016.12.005] [Medline: 28138367]
6. Shankaracharya, Odedra D, Samanta S, Vidyarthi A. Computational intelligence in early diabetes diagnosis: a review. Rev Diabet Stud 2010;7(4):252-262 [FREE Full text] [doi: 10.1900/RDS.2010.7.252] [Medline: 21713313]
7. Nilsson N. Artificial Intelligence: a new synthesis. Burlington: Morgan Kaufmann Publishers; 1998:a.
8. Chapman P, Clinton J, Kerber R, Khabaza T, Reinartz T, Shearer C, et al. CRISP-DM 1.0 Step-by-step data mining guide. 2000. URL: http://www.crisp-dm.org/CRISPWP-0800.pdf [accessed 2018-05-23] [WebCite Cache ID 6zd94Fo1n]
9. Bothe M, Dickens L, Reichel K, Tellmann A, Ellger B, Westphal M, et al. The use of reinforcement learning algorithms to meet the challenges of an artificial pancreas. Expert Rev Med Devices 2013 Sep;10(5):661-673. [doi: 10.1586/17434440.2013.827515] [Medline: 23972072]
10. Bertachi A, Ramkissoon C, Bondia J, Vehí J. Automated blood glucose control in type 1 diabetes: A review of progress and challenges. Endocrinología, Diabetes y Nutrición 2018 Mar;65(3):172-181. [doi: 10.1016/J.ENDINU.2017.10.011]
11. Nimri R, Atlas E, Ajzensztejn M, Miller S, Oron T, Phillip M. Feasibility study of automated overnight closed-loop glucose control under MD-logic artificial pancreas in patients with type 1 diabetes: the DREAM Project. Diabetes Technol Ther 2012 Aug;14(8):728-735. [doi: 10.1089/dia.2012.0004] [Medline: 22853723]
12. Atlas E, Nimri R, Miller S, Grunberg E, Phillip M. MD-logic artificial pancreas system: a pilot study in adults with type 1 diabetes. Diabetes Care 2010 May;33(5):1072-1076 [FREE Full text] [doi: 10.2337/dc09-1830] [Medline: 20150292]
13. Nimri R, Danne T, Kordonouri O, Atlas E, Bratina N, Biester T, et al. The "Glucositter" overnight automated closed loop system for type 1 diabetes: a randomized crossover trial. Pediatr Diabetes 2013 May;14(3):159-167. [doi: 10.1111/pedi.12025] [Medline: 23448393]
14. Phillip M, Battelino T, Atlas E, Kordonouri O, Bratina N, Miller S, et al. Nocturnal glucose control with an artificial pancreas at a diabetes camp. N Engl J Med 2013 Feb 28;368(9):824-833. [doi: 10.1056/NEJMoa1206881] [Medline: 23445093]
15. Nimri R, Muller I, Atlas E, Miller S, Kordonouri O, Bratina N, et al. Night glucose control with MD-Logic artificial pancreas in home setting: a single blind, randomized crossover trial-interim analysis. Pediatr Diabetes 2014 Mar;15(2):91-99. [doi: 10.1111/pedi.12071] [Medline: 23944875]
16. Clinical trial. 2018. A Crossover Study Comparing Two Automated Insulin Delivery System Algorithms in Adolescents and Young Adults With Type 1 Diabetes URL: https://clinicaltrials.gov/ct2/show/NCT03040414 [accessed 2018-05-23] [WebCite Cache ID 6zd69AU6T]
17. Mauseth R, Wang Y, Dassau E, Kircher R, Matheson D, Zisser H, et al. Proposed clinical application for tuning fuzzy logic controller of artificial pancreas utilizing a personalization factor. J Diabetes Sci Technol 2010 Jul 01;4(4):913-922 [FREE Full text] [doi: 10.1177/193229681000400422] [Medline: 20663457]
18. Mauseth R, Hirsch I, Bollyky J, Kircher R, Matheson D, Sanda S, et al. Use of a "fuzzy logic" controller in a closed-loop artificial pancreas. Diabetes Technol Ther 2013 Aug;15(8):628-633. [doi: 10.1089/dia.2013.0036] [Medline: 23829285]
19. Mauseth R, Lord S, Hirsch I, Kircher R, Matheson D, Greenbaum C. Stress Testing of an Artificial Pancreas System With Pizza and Exercise Leads to Improvements in the System's Fuzzy Logic Controller. J Diabetes Sci Technol 2015 Sep 14;9(6):1253-1259 [FREE Full text] [doi: 10.1177/1932296815602098] [Medline: 26370244]
20. Yadav J, Rani A, Singh V. Performance Analysis of Fuzzy-PID Controller for Blood Glucose Regulation in Type-1 Diabetic Patients. J Med Syst 2016 Dec;40(12):254. [doi: 10.1007/s10916-016-0602-6] [Medline: 27714563]
21. Khooban M, Nazari Maryam Abadi D, Alfi A, Siahi M. Swarm optimization tuned Mamdani fuzzy controller for diabetes delayed model. Turk J Elec Eng & Comp Sci 2013;21:2110-2126. [doi: 10.3906/elk-1202-21]
22. Fereydouneyan F, Zare A, Mehrshad N. Using a fuzzy controller optimized by a genetic algorithm to regulate blood glucose level in type 1 diabetes. J Med Eng Technol 2011 Jul;35(5):224-230. [doi: 10.3109/03091902.2011.569050] [Medline: 21557700]
23. Miller S, Nimri R, Atlas E, Grunberg EA, Phillip M. Automatic learning algorithm for the MD-logic artificial pancreas system. Diabetes Technol Ther 2011 Oct;13(10):983-990. [doi: 10.1089/dia.2010.0216] [Medline: 21774690]
24. Dinani S, Zekri M, Kamali M. Regulation of Blood Glucose Concentration in Type 1 Diabetics Using Single Order Sliding Mode Control Combined with Fuzzy On-line Tunable Gain, a Simulation Study. J Med Signals Sens 2015;5(3):131-140 [FREE Full text] [Medline: 26284169]

25. Bothe M, Dickens L, Reichel K, Tellmann A, Ellger B, Westphal M, et al. The use of reinforcement learning algorithms to meet the challenges of an artificial pancreas. Expert Rev Med Devices 2013 Sep;10(5):661-673. [doi: 10.1586/17434440.2013.827515] [Medline: 23972072]

26. Daskalaki E, Diem P, Mougiakakou S. An Actor-Critic based controller for glucose regulation in type 1 diabetes. Comput Methods Programs Biomed 2013 Feb;109(2):116-125. [doi: 10.1016/j.cmpb.2012.03.002] [Medline: 22502983]

27. Daskalaki E, Diem P, Mougiakakou S. Model-Free Machine Learning in Biomedicine: Feasibility Study in Type 1 Diabetes. PLoS One 2016;11(7):e0158722 [FREE Full text] [doi: 10.1371/journal.pone.0158722] [Medline: 27441367]

28. Ewings S, Sahu S, Valletta J, Byrne C, Chipperfield A. A Bayesian network for modelling blood glucose concentration and exercise in type 1 diabetes. Stat Methods Med Res 2015 Jun;24(3):342-372. [doi: 10.1177/0962280214520732] [Medline: 24492795]

29. Catalogna M, Cohen E, Fishman S, Halpern Z, Nevo U, Ben-Jacob E. Artificial neural networks based controller for glucose monitoring during clamp test. PLoS One 2012;7(8):e44587 [FREE Full text] [doi: 10.1371/journal.pone.0044587] [Medline: 22952998]

30. Tang F, Wang Y. Economic Model Predictive Control of Bihormonal Artificial Pancreas System Based on Switching Control and Dynamic R-parameter. J Diabetes Sci Technol 2017 Nov;11(6):1112-1123. [doi: 10.1177/1932296817721519] [Medline: 28728434]

31. Greenwood NJC, Gunton JE. A computational proof of concept of a machine-intelligent artificial pancreas using Lyapunov stability and differential game theory. J Diabetes Sci Technol 2014 Jul;8(4):791-806 [FREE Full text] [doi: 10.1177/1932296814536271] [Medline: 25562888]

32. Narendra KS, Parthasarathy K. Identification and control of dynamical systems using neural networks. IEEE Trans Neural Netw 1990;1(1):4-27. [doi: 10.1109/72.80202] [Medline: 18282820]

33. Mougiakakou S, Bartsocas C, Bozas E, Chaniotakis N, Iliopoulou D, Kouris I, et al. SMARTDIAB: a communication and information technology approach for the intelligent monitoring, management and follow-up of type 1 diabetes patients. IEEE Trans Inf Technol Biomed 2010 May;14(3):622-633. [doi: 10.1109/TITB.2009.2039711] [Medline: 20123578]

34. Zarkogianni K, Vazeou A, Mougiakakou S, Prountzou A, Nikita K. An insulin infusion advisory system based on autotuning nonlinear model-predictive control. IEEE Trans Biomed Eng 2011 Sep;58(9):2467-2477. [doi: 10.1109/TBME.2011.2157823] [Medline: 21622071]

35. González-Olvera MA, Gallardo-Hernández AG, Tang Y, Revilla-Monsalve M, Islas-Andrade S. A discrete-time recurrent neurofuzzy network for black-box modeling of insulin dynamics in diabetic type-1 patients. Int J Neural Syst 2010 Apr;20(2):149-158. [doi: 10.1142/S0129065710002322] [Medline: 20411597]

36. Alanis A, Leon B, Sanchez E, Ruiz-Velazquez E. Blood glucose level neural model for type 1 diabetes mellitus patients. Int J Neural Syst 2011 Dec;21(6):491-504. [doi: 10.1142/S0129065711003000] [Medline: 22131301]

37. Fernandez DCJ, Gonzalez-Perez S, Ramos-Diaz JC. Artificial neural networks for closed loop control of in silico and ad hoc type 1 diabetes. Comput Methods Programs Biomed 2012 Apr;106(1):55-66. [doi: 10.1016/j.cmpb.2011.11.006] [Medline: 22178070]

38. Bailey TS, Chang A, Christiansen M. Clinical accuracy of a continuous glucose monitoring system with an advanced algorithm. J Diabetes Sci Technol 2015 Mar;9(2):209-214 [FREE Full text] [doi: 10.1177/1932296814559746] [Medline: 25370149]

39. Pérez-Gandía C, Facchinetti A, Sparacino G, Cobelli C, Gómez EJ, Rigla M, et al. Artificial neural network algorithm for online glucose prediction from continuous glucose monitoring. Diabetes Technol Ther 2010 Jan;12(1):81-88. [doi: 10.1089/dia.2009.0076] [Medline: 20082589]

40. Pappada S, Borst M, Cameron B, Bourey R, Lather J, Shipp D, et al. Development of a neural network model for predicting glucose levels in a surgical critical care setting. Patient Saf Surg 2010 Sep 09;4(1):15 [FREE Full text] [doi: 10.1186/1754-9493-4-15] [Medline: 20828400]

41. Pappada S, Cameron B, Rosman P, Bourey R, Papadimos T, Olorunto W, et al. Neural network-based real-time prediction of glucose in patients with insulin-dependent diabetes. Diabetes Technol Ther 2011 Feb;13(2):135-141. [doi: 10.1089/dia.2010.0104] [Medline: 21284480]

42. Zecchin C, Facchinetti A, Sparacino G, De Nicolao G, Cobelli C. A new neural network approach for short-term glucose prediction using continuous glucose monitoring time-series and meal information. In: Conf Proc IEEE Eng Med Biol Soc. 2011 Presented at: Annu Int Conf IEEE Eng Med Biol Soc Internet IEEE; 2011; Boston p. 5653-5656. [doi: 10.1109/IEMBS.2011.6091368]

43. Daskalaki E, Prountzou A, Diem P, Mougiakakou S. Real-time adaptive models for the personalized prediction of glycemic profile in type 1 diabetes patients. Diabetes Technol Ther 2012 Feb;14(2):168-174. [doi: 10.1089/dia.2011.0093] [Medline: 21992270]

44. Georga E, Protopappas V, Polyzos D, Fotiadis D. A predictive model of subcutaneous glucose concentration in type 1 diabetes based on Random Forests. In: Conf Proc IEEE Eng Med Biol Soc. 2012 Presented at: Annu Int Conf IEEE Eng Med Biol Soc Internet IEEE; 2012; San Diego p. 2889-2892. [doi: 10.1109/EMBC.2012.6346567]

45. Zecchin C, Facchinetti A, Sparacino G, De NG, Cobelli C. Neural network incorporating meal information improves accuracy of short-time prediction of glucose concentration. IEEE Trans Biomed Eng 2012 Jun;59(6):1550-1560. [doi: 10.1109/TBME.2012.2188893] [Medline: 22374344]

46. Wang Y, Wu X, Mo X. A novel adaptive-weighted-average framework for blood glucose prediction. Diabetes Technol Ther 2013 Oct;15(10):792-801 [FREE Full text] [doi: 10.1089/dia.2013.0104] [Medline: 23883406]

47. Daskalaki E, Nørgaard K, Züger T, Prountzou A, Diem P, Mougiakakou S. An early warning system for hypoglycemic/hyperglycemic events based on fusion of adaptive prediction models. J Diabetes Sci Technol 2013 May 01;7(3):689-698 [FREE Full text] [doi: 10.1177/193229681300700314] [Medline: 23759402]

48. Georga EI, Protopappas VC, Ardigo D, Marina M, Zavaroni I, Polyzos D, et al. Multivariate prediction of subcutaneous glucose concentration in type 1 diabetes patients based on support vector regression. IEEE J Biomed Health Inform 2013 Jan;17(1):71-81. [doi: 10.1109/TITB.2012.2219876] [Medline: 23008265]

49. Zecchin C, Facchinetti A, Sparacino G, Cobelli C. Jump neural network for real-time prediction of glucose concentration. Methods Mol Biol 2015;1260:245-259. [doi: 10.1007/978-1-4939-2239-0_15] [Medline: 25502386]

50. Zarkogianni K, Mitsis K, Arredondo M, Fico G, Fioravanti A, Nikita K. Neuro-fuzzy based glucose prediction model for patients with Type 1 diabetes mellitus. 2014 Presented at: IEEE-EMBS Int Conf Biomed Heal Informatics; 2014; United States p. 252-255. [doi: 10.1109/BHI.2014.6864351]

51. Zarkogianni K, Mitsis K, Litsa E, Arredondo M, Fico G, Fioravanti A, et al. Comparative assessment of glucose prediction models for patients with type 1 diabetes mellitus applying sensors for glucose and physical activity monitoring. Med Biol Eng Comput 2015 Dec;53(12):1333-1343. [doi: 10.1007/s11517-015-1320-9] [Medline: 26049412]

52. Georga E, Protopappas V, Polyzos D, Fotiadis D. Online prediction of glucose concentration in type 1 diabetes using extreme learning machines. In: Conf Proc IEEE Eng Med Biol Soc. 2015 Presented at: IEEE Eng Med Biol Soc Annu Conf Internet; 2015; Milano p. 3262-3265. [doi: 10.1109/EMBC.2015.7319088]

53. Reymann M, Dorschky E, Groh B, Martindale C, Blank P, Eskofier B. Blood glucose level prediction based on support vector regression using mobile platforms. In: Conf Proc IEEE Eng Med Biol Soc. 2016 Dec Presented at: 38th Annu Int Conf IEEE Eng Med Biol Soc; 2016; Orlando p. 2990-2993. [doi: 10.1109/EMBC.2016.7591358]

54. Wang J, Wang F, Liu Y, Xu J, Lin H, Jia B, et al. Multiple Linear Regression and Artificial Neural Network to Predict Blood Glucose in Overweight Patients. Exp Clin Endocrinol Diabetes 2016 Jan;124(1):34-38. [doi: 10.1055/s-0035-1565175] [Medline: 26797861]

55. Georga E, Principe J, Polyzos D, Fotiadis D. Non-linear dynamic modeling of glucose in type 1 diabetes with kernel adaptive filters. In: Conf Proc IEEE Eng Med Biol Soc. 2016 Dec Presented at: 38th Annu Int Conf IEEE Eng Med Biol Soc; 2016; Orlando p. 5897-5900. [doi: 10.1109/EMBC.2016.7592070]

56. Contreras I, Vehi J. Mid-term prediction of blood glucose from continuous glucose sensors, meal information and administered insulin. : Springer; 2016 Presented at: XIV Mediterr Conf Med Biol Eng Comput - MEDICON 2016; 2016; Cyprus. [doi: 10.1007/978-3-319-32703-7_221]

57. Schindelboeck D, Praus F, Gall W. A Diabetes Self-Management Prototype in an AAL-Environment to Detect Remarkable Health States. Stud Health Technol Inform 2016;223:273-280. [Medline: 27139414]

58. Colmenar J, Winkler S, Kronberger G, Maqueda E, Botella M, Hidalgo J. Predicting Glycemia in Diabetic Patients By Evolutionary Computation and Continuous Glucose Monitoring. : ACM Press; 2016 Presented at: Proc Genet Evol Comput Conf Companion - GECCO 2016; 2016; New York p. 1393-1400. [doi: 10.1145/2908961.2931734]

59. Frandes M, Timar B, Timar R, Lungeanu D. Chaotic time series prediction for glucose dynamics in type 1 diabetes mellitus using regime-switching models. Sci Rep 2017 Jul 24;7(1):6232-6241 [FREE Full text] [doi: 10.1038/s41598-017-06478-4] [Medline: 28740090]

60. Contreras I, Vehi J, Visentin R, Vettoretti M. Towards Decision Support Systems Based upon Scenario Profile Analysis: A Hybrid Clustering Prediction for Type 1 Diabetes Aid. 2017 Presented at: 2017 IEEE/ACM Int Conf Connect Heal Appl Syst Eng Technol; 12/07/2017; Philadelphia. [doi: 10.1109/CHASE.2017.61]

61. Albers D, Levine M, Gluckman B, Ginsberg H, Hripcsak G, Mamykina L. Personalized glucose forecasting for type 2 diabetes using data assimilation. PLoS Comput Biol 2017 Apr;13(4):e1005232 [FREE Full text] [doi: 10.1371/journal.pcbi.1005232] [Medline: 28448498]

62. Hidalgo J, Colmenar J, Kronberger G, Winkler S, Garnica O, Lanchares J. Data Based Prediction of Blood Glucose Concentrations Using Evolutionary Methods. J Med Syst 2017 Aug 08;41(9):142. [doi: 10.1007/s10916-017-0788-2] [Medline: 28791547]

63. Contreras I, Oviedo S, Vettoretti M, Visentin R, Vehí J. Personalized blood glucose prediction: A hybrid approach using grammatical evolution and physiological models. PLoS One 2017;12(11):e0187754 [FREE Full text] [doi: 10.1371/journal.pone.0187754] [Medline: 29112978]

64. Fiorini S, Martini C, Malpassi D, Cordera R, Maggi D, Verri A, et al. Data-driven strategies for robust forecast of continuous glucose monitoring time-series. In: Conf Proc IEEE Eng Med Biol Soc. 2017 Dec Presented at: Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS; 11-15/07/2017; Jeju p. 1680-1683. [doi: 10.1109/EMBC.2017.8037164]

65. Nguyen H, Jones T. Detection of nocturnal hypoglycemic episodes using EEG signals. 2016 Nov Presented at: Annu Int Conf IEEE Eng Med Biol; 2010; Buenos Aires p. 4930-4933. [doi: 10.1109/IEMBS.2010.5627233]

66.   Skrøvseth SO, Årsand E, Godtliebsen F, Hartvigsen G. Mobile phone-based pattern recognition and data analysis for patients with type 1 diabetes. Diabetes Technol Ther 2012 Dec;14(12):1098-1104 [FREE Full text] [doi: 10.1089/dia.2012.0160] [Medline: 23035775]

67.   San PP, Ling SH, Nguyen H. Intelligent detection of hypoglycemic episodes in children with type 1 diabetes using adaptive neural-fuzzy inference system. Conf Proc IEEE Eng Med Biol Soc 2012;2012:6325-6328. [doi: 10.1109/EMBC.2012.6347440] [Medline: 23367375]

68.   Jensen M, Christensen T, Tarnow L, Mahmoudi Z, Johansen M, Hejlesen O. Professional continuous glucose monitoring in subjects with type 1 diabetes: retrospective hypoglycemia detection. J Diabetes Sci Technol 2013 Jan 01;7(1):135-143. [doi: 10.1177/193229681300700116] [Medline: 23439169]

69.   Georga E, Protopappas V, Ardigò D, Polyzos D, Fotiadis D. A glucose model based on support vector regression for the prediction of hypoglycemic events under free-living conditions. Diabetes Technol Ther 2013 Aug;15(8):634-643. [doi: 10.1089/dia.2012.0285] [Medline: 23848178]

70.   Jensen M, Christensen T, Tarnow L, Seto E, Dencker JM, Hejlesen O. Real-time hypoglycemia detection from continuous glucose monitoring data of subjects with type 1 diabetes. Diabetes Technol Ther 2013 Jul;15(7):538-543. [doi: 10.1089/dia.2013.0069] [Medline: 23631608]

71.   Eljil K, Qadah G, Pasquier M. Predicting hypoglycemia in diabetic patients using data mining techniques. 2013 Presented at: 9th Int Conf Innov Inf Technol Internet IEEE; 2013; Abu Dhabi p. 538-543. [doi: 10.1109/Innovations.2013.6544406]

72.   Sudharsan B, Peeples M, Shomali M. Hypoglycemia prediction using machine learning models for patients with type 2 diabetes. J Diabetes Sci Technol 2015 Jan;9(1):86-90 [FREE Full text] [doi: 10.1177/1932296814554260] [Medline: 25316712]

73.   San P, Ling S, Nuryani, Nguyen H. Evolvable rough-block-based neural network and its biomedical application to hypoglycemia detection system. IEEE Trans Cybern 2014 Aug;44(8):1338-1349. [doi: 10.1109/TCYB.2013.2283296] [Medline: 24122616]

74.   Otto E, Tannan V. Evaluation of the utility of a glycemic pattern identification system. J Diabetes Sci Technol 2014 Jul;8(4):830-838 [FREE Full text] [doi: 10.1177/1932296814532210] [Medline: 24876425]

75.   Nguyen LL, Su S, Nguyen H. Neural network approach for non-invasive detection of hyperglycemia using electrocardiographic signals. Conf Proc IEEE Eng Med Biol Soc 2014;2014:4475-4478. [doi: 10.1109/EMBC.2014.6944617] [Medline: 25570985]

76.   Phyo PS, Sai HL, Nguyen H. Deep learning framework for detection of hypoglycemic episodes in children with type 1 diabetes. Conf Proc IEEE Eng Med Biol Soc 2016 Dec;2016:3503-3506. [doi: 10.1109/EMBC.2016.7591483] [Medline: 28269053]

77.   Nunes A, Yang J, Radican L, Engel S, Kurtyka K, Tunceli K, et al. Assessing occurrence of hypoglycemia and its severity from electronic health records of patients with type 2 diabetes mellitus. Diabetes Res Clin Pract 2016 Nov;121:192-203. [doi: 10.1016/j.diabres.2016.09.012] [Medline: 27744128]

78.   Zhang Y, Zhu J, Liang Y, Chen H, Yin S, Chen Z. Non-invasive blood glucose detection system based on conservation of energy method. Physiol Meas 2017 Feb;38(2):325-342. [doi: 10.1088/1361-6579/aa50cf] [Medline: 28107204]

79.   Walsh J, Roberts R. Pumping insulin: verything you need to succeed on an insulin pump Internet. San Diego: Torrey Pines Pr; 2010:1-288.

80.   Pesl P, Herrero P, Reddy M, Xenou M, Oliver N, Johnston D, et al. An Advanced Bolus Calculator for Type 1 Diabetes: System Architecture and Usability Results. IEEE J Biomed Health Inform 2016 Jan;20(1):11-17. [doi: 10.1109/JBHI.2015.2464088] [Medline: 26259202]

81.   Herrero P, Pesl P, Reddy M, Oliver N, Georgiou P, Toumazou C. Advanced Insulin Bolus Advisor Based on Run-To-Run Control and Case-Based Reasoning. IEEE J Biomed Health Inform 2015 May;19(3):1087-1096. [doi: 10.1109/JBHI.2014.2331896] [Medline: 24956470]

82.   Reddy M, Pesl P, Xenou M, Toumazou C, Johnston D, Georgiou P, et al. Clinical Safety and Feasibility of the Advanced Bolus Calculator for Type 1 Diabetes Based on Case-Based Reasoning: A 6-Week Nonrandomized Single-Arm Pilot Study. Diabetes Technol Ther 2016 Dec;18(8):487-493. [doi: 10.1089/dia.2015.0413] [Medline: 27196358]

83.   Pesl P, Herrero P, Reddy M, Oliver N, Johnston D, Toumazou C, et al. Case-Based Reasoning for Insulin Bolus Advice. J Diabetes Sci Technol 2017 Dec;11(1):37-42 [FREE Full text] [doi: 10.1177/1932296816629986] [Medline: 26862136]

84.   Herrero P, Bondia J, Adewuyi O, Pesl P, El-Sharkawy M, Reddy M, et al. Enhancing automatic closed-loop glucose control in type 1 diabetes with an adaptive meal bolus calculator - in silico evaluation under intra-day variability. Comput Methods Programs Biomed 2017 Jul;146:125-131. [doi: 10.1016/j.cmpb.2017.05.010] [Medline: 28688482]

85.   Brown D, Aldea A, Harrison R, Martin C, Bayley I. Temporal case-based reasoning for type 1 diabetes mellitus bolus insulin decision support. Artif Intell Med 2018 Dec;85:28-42. [doi: 10.1016/j.artmed.2017.09.007] [Medline: 28986108]

86.   Rhyner D, Loher H, Dehais J, Anthimopoulos M, Shevchik S, Botwey R, et al. Carbohydrate Estimation by a Mobile Phone-Based System Versus Self-Estimations of Individuals With Type 1 Diabetes Mellitus: A Comparative Study. J Med Internet Res 2016 May 11;18(5):e101 [FREE Full text] [doi: 10.2196/jmir.5567] [Medline: 27170498]

87. Bally L, Dehais J, Nakas C, Anthimopoulos M, Laimer M, Rhyner D, et al. Carbohydrate Estimation Supported by the GoCARB System in Individuals With Type 1 Diabetes: A Randomized Prospective Pilot Study. Diabetes Care 2017 Dec;40(2):e6-e7. [doi: 10.2337/dc16-2173] [Medline: 27899490]

88. Agianniotis A, Anthimopoulos M, Daskalaki E, Drapela A, Stettler C, Diem P, et al. GoCARB in the Context of an Artificial Pancreas. J Diabetes Sci Technol 2015 May;9(3):549-555 [FREE Full text] [doi: 10.1177/1932296815583333] [Medline: 25904142]

89. Anthimopoulos M, Dehais J, Shevchik S, Ransford B, Duke D, Diem P, et al. Computer vision-based carbohydrate estimation for type 1 patients with diabetes using smartphones. J Diabetes Sci Technol 2015 May;9(3):507-515 [FREE Full text] [doi: 10.1177/1932296815580159] [Medline: 25883163]

90. Anthimopoulos MM, Gianola L, Scarnato L, Diem P, Mougiakakou SG. A food recognition system for diabetic patients based on an optimized bag-of-features model. IEEE J Biomed Health Inform 2014 Jul;18(4):1261-1271. [doi: 10.1109/JBHI.2014.2308928] [Medline: 25014934]

91. Srinivasan A, Lee JB, Dassau E, Doyle FJ. Novel insulin delivery profiles for mixed meals for sensor-augmented pump and closed-loop artificial pancreas therapy for type 1 diabetes mellitus. J Diabetes Sci Technol 2014 Sep;8(5):957-968 [FREE Full text] [doi: 10.1177/1932296814543660] [Medline: 25049364]

92. Cappon G, Vettoretti M, Marturano F, Facchinetti A, Sparacino G. A Neural-Network-Based Approach to Personalize Insulin Bolus Calculation Using Continuous Glucose Monitoring. J Diabetes Sci Technol 2018 Mar;12(2):265-272. [doi: 10.1177/1932296818759558] [Medline: 29493356]

93. Lee M, Gatton TM, Lee KK. A MonitoringAdvisory System for Diabetes Patient Management Using a Rule-Based Method and KNN. Sensors Internet 2010 Apr;10(4):3934-3953. [Medline: 22319334]

94. Bode BW. Glycemic Variability and the Role It Should Play in Diabetes Management and Blood Glucose Monitoring. US Endocrinol Internet 2008;4(2):67. [doi: 10.17925/USE.2008.04.2.67]

95. Skevofilakas M, Zarkogianni K, Karamanos B, Nikita K. A hybrid Decision Support System for the risk assessment of retinopathy development as a long term complication of Type 1 Diabetes Mellitus. Conf Proc IEEE Eng Med Biol Soc 2010;2010:6713-6716. [doi: 10.1109/IEMBS.2010.5626245] [Medline: 21096083]

96. Newcomer S, Steiner J, Bayliss E. Identifying subgroups of complex patients with cluster analysis. Am J Manag Care 2011 Aug 01;17(8):e324-e332 [FREE Full text] [Medline: 21851140]

97. Rohan J, Delamater A, Pendley J, Dolan L, Reeves G, Drotar D. Identification of self-management patterns in pediatric type 1 diabetes using cluster analysis. Pediatr Diabetes 2011 Nov;12(7):611-618. [doi: 10.1111/j.1399-5448.2010.00752.x] [Medline: 21446925]

98. Namayanja J, Janeja V. An assessment of patient behavior over time-periods: a case study of managing type 2 diabetes through blood glucose readings and insulin doses. J Med Syst 2012 Nov;36 Suppl 1:S65-S80. [doi: 10.1007/s10916-012-9894-3] [Medline: 23104240]

99. Deschamps K, Matricali G, Roosen P, Desloovere K, Bruyninckx H, Spaepen P, et al. Classification of forefoot plantar pressure distribution in persons with diabetes: a novel perspective for the mechanical management of diabetic foot? PLoS One 2013;8(11):e79924 [FREE Full text] [doi: 10.1371/journal.pone.0079924] [Medline: 24278219]

100. Sacchi L, Dagliati A, Segagni D, Leporati P, Chiovato L, Bellazzi R. Improving risk-stratification of Diabetes complications using temporal data mining. Conf Proc IEEE Eng Med Biol Soc 2015;2015:2131-2134. [doi: 10.1109/EMBC.2015.7318810] [Medline: 26736710]

101. Migliore C, Vorderstrasse A, Pan W, Melkus G. Renal Disease Risk Factors Among Risk Groups Comprised of African American Women With Type 2 Diabetes: A Secondary Analysis. Diabetes Educ 2015 Oct 22;41(5):569-581. [Medline: 26202051]

102. Lagani V, Chiarugi F, Manousos D, Verma V, Fursse J, Marias K, et al. Realization of a service for the long-term risk assessment of diabetes-related complications. J Diabetes Complications 2015 Jul;29(5):691-698. [doi: 10.1016/j.jdiacomp.2015.03.011] [Medline: 25953402]

103. Sangi M, Win K, Shirvani F, Namazi-Rad M, Shukla N. Applying a novel combination of techniques to develop a predictive model for diabetes complications. PLoS One 2015;10(4):e0121569 [FREE Full text] [doi: 10.1371/journal.pone.0121569] [Medline: 25902317]

104. Huang G, Huang K, Lee T, Weng J. An interpretable rule-based diagnostic classification of diabetic nephropathy among type 2 diabetes patients. BMC Bioinformatics 2015;16 Suppl 1:S5 [FREE Full text] [doi: 10.1186/1471-2105-16-S1-S5] [Medline: 25707942]

105. Contreras I, Quirós C, Giménez M, Conget I, Vehi J. Profiling intra-patient type I diabetes behaviors. Comput Methods Programs Biomed 2016 Nov;136:131-141. [doi: 10.1016/j.cmpb.2016.08.022] [Medline: 27686710]

106. Vassilev I, Rogers A, Kennedy A, Wensing M, Koetsenruijter J, Orlando R, et al. Social Network Type and Long-Term Condition Management Support: A Cross-Sectional Study in Six European Countries. PLoS One 2016;11(8):e0161027 [FREE Full text] [doi: 10.1371/journal.pone.0161027] [Medline: 27536988]

107. Anderson J, Parikh J, Shenfeld D, Ivanov V, Marks C, Church B, et al. Reverse Engineering and Evaluation of Prediction Models for Progression to Type 2 Diabetes: An Application of Machine Learning Using Electronic Health Records. J Diabetes Sci Technol 2015 Dec 20;10(1):6-18 [FREE Full text] [doi: 10.1177/1932296815620200] [Medline: 26685993]

108. DuBrava S, Mardekian J, Sadosky A, Bienen EJ, Parsons B, Hopps M, et al. Using Random Forest Models to Identify Correlates of a Diabetic Peripheral Neuropathy Diagnosis from Electronic Health Record Data. Pain Med 2017 Dec 01;18(1):107-115. [doi: 10.1093/pm/pnw096] [Medline: 27252307]

109. McCoy R, Ngufor C, Van HH, Caffo B, Shah N. Trajectories of Glycemic Change in a National Cohort of Adults With Previously Controlled Type 2 Diabetes. Med Care 2017 Nov;55(11):956-964. [doi: 10.1097/MLR.0000000000000807] [Medline: 28922296]

110. Dagliati A, Malovini A, Decata P, Cogni G, Teliti M, Sacchi L, et al. Hierarchical Bayesian Logistic Regression to forecast metabolic control in type 2 DM patients. AMIA Annu Symp Proc 2016;2016:470-479 [FREE Full text] [Medline: 28269842]

111. Contreras I, Vehi J. Mid-Term Prediction of Blood Glucose from Continuous Glucose Sensors, Meal Information and Administered Insulin. XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016 2016 Sep 17:1137-1143. [doi: 10.1007/978-3-319-32703-7_222]

112. Dagliati A, Marini S, Sacchi L, Cogni G, Teliti M, Tibollo V, et al. Machine Learning Methods to Predict Diabetes Complications. J Diabetes Sci Technol 2017 May 12;12(2):295-302. [doi: 10.1177/1932296817706375]

113. Leal Y, Gonzalez-Abril L, Lorencio C, Bondia J, Vehi J. Detection of correct and incorrect measurements in real-time continuous glucose monitoring systems by applying a postprocessing support vector machine. IEEE Trans Biomed Eng 2013 Jul;60(7):1891-1899. [Medline: 23380841]

114. Turksoy K, Roy A, Cinar A. Real-Time Model-Based Fault Detection of Continuous Glucose Sensor Measurements. IEEE Trans Biomed Eng 2017 Dec;64(7):1437-1445. [doi: 10.1109/TBME.2016.2535412] [Medline: 26930674]

115. Shin J, Park H, Cho S, Nam H, Lee KJ. A correction method using a support vector machine to minimize hematocrit interference in blood glucose measurements. Comput Biol Med 2014 Sep;52:111-118. [doi: 10.1016/j.compbiomed.2014.06.005] [Medline: 25033022]

116. Turksoy K, Paulino T, Zaharieva D, Yavelberg L, Jamnik V, Riddell M, et al. Classification of Physical Activity: Information to Artificial Pancreas Control Systems in Real Time. J Diabetes Sci Technol 2015 Oct 06;9(6):1200-1207 [FREE Full text] [doi: 10.1177/1932296815609369] [Medline: 26443291]

117. Jacobs P, Resalat N, El YJ, Reddy R, Branigan D, Preiser N, et al. Incorporating an Exercise Detection, Grading, and Hormone Dosing Algorithm Into the Artificial Pancreas Using Accelerometry and Heart Rate. J Diabetes Sci Technol 2015 Oct 05;9(6):1175-1184 [FREE Full text] [doi: 10.1177/1932296815609371] [Medline: 26438720]

118. Samadi S, Turksoy K, Hajizadeh I, Feng J, Sevil M, Cinar A. Meal Detection and Carbohydrate Estimation Using Continuous Glucose Sensor Data. IEEE J Biomed Health Inform 2017 Dec;21(3):619-627. [doi: 10.1109/JBHI.2017.2677953] [Medline: 28278487]

119. Turksoy K, Samadi S, Feng J, Littlejohn E, Quinn L, Cinar A. Meal Detection in Patients With Type 1 Diabetes: A New Module for the Multivariable Adaptive Artificial Pancreas Control System. IEEE J Biomed Health Inform 2016 Jan;20(1):47-54 [FREE Full text] [doi: 10.1109/JBHI.2015.2446413] [Medline: 26087510]

120. Samadi S, Rashid M, Turksoy K, Feng J, Hajizadeh I, Hobbs N, et al. Automatic Detection and Estimation of Unannounced Meals for Multivariable Artificial Pancreas System. Diabetes Technol Ther 2018 Mar;20(3):235-246. [doi: 10.1089/dia.2017.0364] [Medline: 29406789]

121. Cescon M, DeSalvo D, Ly T, Maahs D, Messer L, Buckingham B, et al. Early Detection of Infusion Set Failure During Insulin Pump Therapy in Type 1 Diabetes. J Diabetes Sci Technol 2016 Dec;10(6):1268-1276 [FREE Full text] [doi: 10.1177/1932296816663962] [Medline: 27621142]

122. Fioravanti A, Fico G, Arredondo M, Leuteritz J. A mobile feedback system for integrated E-health platforms to improve self-care and compliance of diabetes mellitus patients. In: Conf Proc IEEE Eng Med Biol Soc. 2011 Presented at: Annu Int Conf IEEE Eng Med Biol Soc; 2011; Paris p. 3550-3553. [doi: 10.1109/IEMBS.2011.6090591]

123. Fico G, Fioravanti A, Arredondo M, Leuteritz J, Guillén A, Fernandez D. A user centered design approach for patient interfaces to a diabetes IT platform. Conf Proc IEEE Eng Med Biol Soc 2011;2011:1169-1172. [doi: 10.1109/IEMBS.2011.6090274] [Medline: 22254523]

124. Fico G, Arredondo M, Protopappas V, Georgia E, Fotiadis D. Mining data when technology is applied to support patients and professional on the control of chronic diseases: the experience of the METABO platform for diabetes management. Methods Mol Biol 2015;1246:191-216. [doi: 10.1007/978-1-4939-1985-7_13] [Medline: 25417088]

125. Fico G, Fioravanti A, Arredondo M, Ardigó D, Guillén A. A healthy lifestyle coaching-persuasive application for patients with type 2 diabetes. In: Conf Proc IEEE Eng Med Biol Soc. 2010 Presented at: Annu Int Conf IEEE Eng Med Biol; 2010; Paris p. 2221-2224. [doi: 10.1109/IEMBS.2010.5626185]

126. Guillen A, Colas J, Fico G, Guillen S. METABO: A new paradigm towards diabetes disease management. An innovative business model. 2011 Annu Int Conf IEEE Eng Med Biol Soc 2011:3554-3557. [doi: 10.1109/IEMBS.2011.6090592]

127. Fico G, Fioravanti A, Arredondo M, Gorman J, Diazzi C, Arcuri G, et al. Integration of Personalized Healthcare Pathways in an ICT Platform for Diabetes Managements: A Small-Scale Exploratory Study. IEEE J Biomed Health Inform 2016 Jan;20(1):29-38. [doi: 10.1109/JBHI.2014.2367863] [Medline: 25389246]

128. Dagliati A, Sacchi L, Tibollo V, Cogni G, Teliti M, Martinez-Millana A, et al. A dashboard-based system for supporting diabetes care. J Am Med Inform Assoc 2018 May 01;25(5):538-547. [doi: 10.1093/jamia/ocx159] [Medline: 29409033]

129. Hidalgo J, Maqueda E, Risco-Martín J, Cuesta-Infante A, Colmenar J, Nobel J. glUCModel: A monitoring and modeling system for chronic diseases applied to diabetes. J Biomed Inform 2014;48:183-192. [doi: 10.1016/J.JBI.2013.12.015]

130. Everett E, Kane B, Yoo A, Dobs A, Mathioudakis N. A Novel Approach for Fully Automated, Personalized Health Coaching for Adults with Prediabetes: Pilot Clinical Trial. J Med Internet Res 2018 Feb 27;20(2):e72 [FREE Full text] [doi: 10.2196/jmir.9723] [Medline: 29487046]

131. Yom-Tov E, Feraru G, Kozdoba M, Mannor S, Tennenholtz M, Hochberg I. Encouraging Physical Activity in Patients With Diabetes: Intervention Using a Reinforcement Learning System. J Med Internet Res 2017 Oct 10;19(10):e338 [FREE Full text] [doi: 10.2196/jmir.7994] [Medline: 29017988]

132. Hazlehurst B, Lawrence J, Donahoo W, Sherwood N, Kurtz S, Xu S, et al. Automating assessment of lifestyle counseling in electronic health records. Am J Prev Med 2014 May;46(5):457-464 [FREE Full text] [doi: 10.1016/j.amepre.2014.01.001] [Medline: 24745635]

133. Rigla M, Martínez-Sarriegui I, García-Sáez G, Pons B, Hernando M. Gestational Diabetes Management Using Smart Mobile Telemedicine. J Diabetes Sci Technol 2018 Mar;12(2):260-264 [FREE Full text] [doi: 10.1177/1932296817704442] [Medline: 28420257]

134. Caballero-Ruiz E, García-Sáez G, Rigla M, Villaplana M, Pons B, Hernando M. A web-based clinical decision support system for gestational diabetes: Automatic diet prescription and detection of insulin needs. Int J Med Inform 2017 Dec;102:35-49. [doi: 10.1016/j.ijmedinf.2017.02.014] [Medline: 28495347]

135. Cleveringa F, Welsing P, van DDM, Gorter K, Niessen L, Rutten G, et al. Cost-effectiveness of the diabetes care protocol, a multifaceted computerized decision support diabetes management intervention that reduces cardiovascular risk. Diabetes Care 2010 Feb;33(2):258-263 [FREE Full text] [doi: 10.2337/dc09-1232] [Medline: 19933991]

136. Cleveringa F, Minkman M, Gorter K, van den Dock M, Rutten GEHM. Diabetes Care Protocol: effects on patient-important outcomes. A cluster randomized, non-inferiority trial in primary care. Diabet Med 2010 Apr;27(4):442-450. [doi: 10.1111/j.1464-5491.2010.02968.x] [Medline: 20536517]

137. Alotaibi MM, Istepanian R, Philip N. A mobile diabetes management and educational system for type-2 diabetics in Saudi Arabia (SAED). mHealth 2016 Aug 24;2:33-33. [doi: 10.21037/mhealth.2016.08.01] [Medline: 28293606]

138. Fernandez-Llatas C, Martinez-Millana A, Martinez-Romero A, Benedi JM, Traver V. Diabetes care related process modelling using Process Mining techniques. Lessons learned in the application of Interactive Pattern Recognition: coping with the Spaghetti Effect. Lessons learned in the application of Interactive Pattern Recognition; 2015 Presented at: Conf Proc IEEE Eng Med Biol Soc; 25-29/08/2015; Milano p. 2127-2130. [doi: 10.1109/EMBC.2015.7318809]

139. Contreras I, Kiefer S, Vehi J. Adaptive Workflows for Diabetes Management: Self-Management Assistant and Remote Treatment for Diabetes. Stud Health Technol Inform 2017;237:151-156. [Medline: 28479559]

140. Suh M, Woodbridge J, Moin T, Lan M, Alshurafa N, Samy L, et al. Dynamic Task Optimization in Remote Diabetes Monitoring Systems. Proc IEEE Int Conf Healthc Inform Imaging Syst Biol 2012 Sep;2012:3-11 [FREE Full text] [doi: 10.1109/HISB.2012.10] [Medline: 27617297]

141. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. J Med Internet Res 2013 Nov 01;15(11):e239 [FREE Full text] [doi: 10.2196/jmir.2721] [Medline: 24184993]

142. Valdez R, Guterbock T, Fitzgibbon K, Williams I, Wellbeloved-Stone C, Bears J, et al. From loquacious to reticent: understanding patient health information communication to guide consumer health IT design. J Am Med Inform Assoc 2017 Jul 01;24(4):680-696. [doi: 10.1093/jamia/ocw155] [Medline: 28069667]

143. Chen A. Exploring online support spaces: using cluster analysis to examine breast cancer, diabetes and fibromyalgia support groups. Patient Educ Couns 2012 May;87(2):250-257. [doi: 10.1016/j.pec.2011.08.017] [Medline: 21930359]

144. Hamon T, Gagnayre R. Improving knowledge of patient skills thanks to automatic analysis of online discussions. Patient Educ Couns 2013 Aug;92(2):197-204. [doi: 10.1016/j.pec.2013.05.012] [Medline: 23769423]

145. Lo-Ciganic W, Donohue J, Thorpe J, Perera S, Thorpe C, Marcum Z, et al. Using machine learning to examine medication adherence thresholds and risk of hospitalization. Med Care 2015 Aug;53(8):720-728 [FREE Full text] [doi: 10.1097/MLR.0000000000000394] [Medline: 26147866]

146. Fioravanti A, Fico G, Salvi D, García-Betances RI, Arredondo M. Automatic messaging for improving patients engagement in diabetes management: an exploratory study. Med Biol Eng Comput 2015 Dec;53(12):1285-1294. [doi: 10.1007/s11517-014-1237-8] [Medline: 25564181]

147. Heringa M, Siderius H, Floor-Schreudering A, de Smet P, Bouvy M. Lower alert rates by clustering of related drug interaction alerts. J Am Med Inform Assoc 2017 Jan;24(1):54-59. [doi: 10.1093/jamia/ocw049] [Medline: 27107437]

148. Eghbali-Zarch M, Tavakkoli-Moghaddam R, Esfahanian F, Sepehri M, Azaron A. Pharmacological therapy selection of type 2 diabetes based on the SWARA and modified MULTIMOORA methods under a fuzzy environment. Artif Intell Med 2018 Mar 29 (forthcoming). [doi: 10.1016/j.artmed.2018.03.003] [Medline: 29606521]

149. Kurasawa H, Hayashi K, Fujino A, Takasugi K, Haga T, Waki K, et al. Machine-Learning-Based Prediction of a Missed Scheduled Clinical Appointment by Patients With Diabetes. J Diabetes Sci Technol 2016 Dec;10(3):730-736 [FREE Full text] [doi: 10.1177/1932296815614866] [Medline: 26555782]

Abbreviations

AI: artificial intelligence
ANN: artificial neural network
AP: artificial pancreas
BG: blood glucose
CGM: continuous glucose monitoring
CSII: continuous subcutaneous insulin infusion
DSS: decision support system
DT: decision tree
DL: deep learning
EA: evolutionary algorithm
FL: fuzzy logic
GA: genetic algorithm
GDM: gestational diabetes
KDD: knowledge discovery in databases
KNN: k-nearest neighbor
MDI: multiple daily injection
NB: Naïve Bayes
NLP: natural language processing
PH: prediction horizon
RA: regression algorithm
RF: random forest
RMSE: relative mean square error
SVM: support vector machine
T1D: type 1 diabetes
T2D: type 2 diabetes

XSL•FO
RenderX

Original Paper

"It's Totally Okay to Be Sad, but Never Lose Hope": Content Analysis of Infertility-Related Videos on YouTube in Relation to Viewer Preferences

Margot Kelly-Hedrick[1], BA (Psych); Paul H Grunberg[2], BA Hons (Psych); Felicia Brochu[3], BSc (Psych); Phyllis Zelkowitz[1,3,4], EdD

[1]Department of Psychiatry, Jewish General Hospital, Montreal, QC, Canada
[2]Department of Psychology, McGill University, Montreal, QC, Canada
[3]Department of Psychiatry, McGill University, Montreal, QC, Canada
[4]Lady Davis Institute for Medical Research, Montreal, QC, Canada

Corresponding Author:
Phyllis Zelkowitz, EdD
Jewish General Hospital
Department of Psychiatry
4333 Cote Ste. Catherine
Montreal, QC, H3T 1E4
Canada
Phone: 1 514 340 8222 ext 25258
Fax: 1 514 340 8124
Email: phyllis.zelkowitz@mcgill.ca

Abstract

Background: Infertility patients frequently use the internet to find fertility-related information and support from people in similar circumstances. YouTube is increasingly used as a source of health-related information and may influence health decision making. There have been no studies examining the content of infertility-related videos on YouTube.

Objective: The purpose of this study was to (1) describe the content of highly viewed videos on YouTube related to infertility and (2) identify video characteristics that relate to viewer preference.

Methods: Using the search term "infertility," the 80 top-viewed YouTube videos and their viewing statistics (eg, views, likes, and comments) were collected. Videos that were non-English, unrelated to infertility, or had age restrictions were excluded. Content analysis was used to examine videos, employing a coding rubric that measured the presence or absence of video codes related to purpose, tone, and demographic and fertility characteristics (eg, sex, parity, stage of fertility treatment).

Results: A total of 59 videos, with a median of 156,103 views, met the inclusion criteria and were categorized into 35 personal videos (35/59, 59%) and 24 informational-educational videos (24/59, 41%). Personal videos did not differ significantly from informational-educational videos on number of views, dislikes, subscriptions driven, or shares. However, personal videos had significantly more likes *(P<.001)* and comments *(P<.001)* than informational-educational videos. The purposes of the videos were treatment outcomes (33/59, 56%), sharing information (30/59, 51%), emotional aspects of infertility (20/59, 34%), and advice to others (6/59, 10%). The tones of the videos were positive (26/59, 44%), neutral (25/59, 42%), and mixed (8/59, 14%); there were no videos with negative tone. No videos contained only male posters. Videos with a positive tone did not differ from neutral videos in number of views, dislikes, subscriptions driven, or shares; however, positive videos had significantly more likes *(P<.001)* and comments *(P<.001)* than neutral videos. A majority (21/35, 60%) of posters of personal videos shared a pregnancy announcement.

Conclusions: YouTube is a source of both technical and personal experience-based information about infertility. However, videos that include personal experiences may elicit greater viewer engagement. Positive videos and stories of treatment success may provide hope to viewers but could also create and perpetuate unrealistic expectations about the success rates of fertility treatment.

XSL•FO
RenderX

Introduction

Up to 1 in 6 Canadian couples may experience infertility [1], defined as the failure to achieve conception following at least 12 months of unprotected intercourse [2]. Prevalence rates are similar or higher in the United States and European countries, with estimates ranging from 10% to 28% [3-7]. The diagnosis and treatment of infertility may have adverse effects on psychological well-being and quality of life [8-10].

More than 85% of infertility patients search for Web-based infertility-related content [11-13]. They report using the Web to find experience-based information; to learn about others going through similar treatment [14]; and to gain emotional support, reduce isolation, and seek normalcy [15]. The internet is an increasingly common source of health-related information, with 72% of internet users using the Web for health information [16].

Social media allows internet users to connect with others online, facilitating communication and social support regarding health conditions [17]. YouTube is one such social media website that allows users to share video content and engage in discussion. YouTube is the most popular website for video sharing and the second most accessed website with over 1 billion users, reaching a global and diverse audience [18,19]. YouTube is a source of experience-based information on health topics [20,21] and the use of YouTube for social support relating to health topics has been found to relate to higher levels of health care–related empowerment and information engagement [22]. Therefore, examining the content of popular health-related videos on YouTube is of interest. Several studies have examined how YouTube videos characterize chronic health conditions, such as epilepsy [23], Alzheimer's disease [24], and chronic obstructive pulmonary disease [25]. These studies have evaluated level of viewership, discussion, and reliability of information presented in videos on YouTube [23,25,26]. There has not yet been an examination of infertility content on YouTube. As such, this is the first study to assess highly viewed infertility videos on YouTube to document their tone, purpose, and characteristics of posters and fertility-related variables, using content analysis. The coding scheme was based on other studies examining health-related content on YouTube [23,25-28]. The study aimed to (1) describe the video content of the most highly viewed videos and (2) identify video characteristics that relate to viewer favorability (eg, video likes, shares). Doing so may provide a richer understanding of people's infertility experiences and purpose for posting infertility information on the internet. Indications of viewers' responses to videos, such as likes and dislikes, may give an indication of viewer preference.

Methods

Ethical Considerations

YouTube meets the criteria for a public online database as it is free, publicly accessible without requiring registration, and has a large membership size [18,29]. Consistent with past research [29,30], our institutional ethics board deemed ethics approval unnecessary as content relevant to this study was in the public domain. YouTube was accessed without a registered account to ensure accessed videos were publicly available and had no age restrictions.

Coding Rubric Development

The coding rubric for the video analysis was initially developed through a deductive approach. Specifically, we reviewed studies that used content analysis to examine health-related YouTube topics [23,25-28]. Videos were coded as either personal, where the primary focus is on sharing experiences, or informational-educational, where the focus is to provide information. Other variables included (1) tone (positive, negative, neutral, or mixed), (2) purpose(s) of video (sharing information, emotional aspects of infertility, treatment outcomes, and advice to others), (3) whether the video has live people present (character video) or absent (noncharacter video), and (4) source of video, which indicates the type of YouTube channel that posted the video (personal account, organization's channel, news channel, or other type of channel). Before viewing videos, the coding team generated categories hypothesized to be relevant to infertility-related videos, such as the stage of fertility treatment of the poster(s) and parity of the poster(s). Two coders tested the rubric using 7 sets of 3 to 5 videos with low view counts. Over the course of this testing period, the codebook was modified until variables were refined and explicitly defined. Through an inductive process, new variables such as the code "pregnancy announcement" were added if agreed upon by both coders during the training process. Training videos were not included in the final dataset. The training period ended when the 2 coders reached agreement for all codes and code definitions (see Table 1 for codebook). Certain variables (purpose of video, type of informational video, and subtones in videos) were not mutually exclusive, and therefore codes could sum to more than 100%. Video statistics were collected for all videos in the sample. These included number of views, likes, dislikes, comments, date video was posted, transcript (generated by YouTube), subscriptions driven (number of people who subscribed to the channel following viewing of video), number of shares, length of video, number of subscribers to channel, account creation date, and total channel views (sum of views for all videos posted by user). YouTube channels are profiles for users to post their videos and share information about themselves and their videos.

Table 1. Codebook variable definitions and inter-rater reliability for 2 coders assessed by Cohen kappa.

Variable definitions	Kappa[a]
Variables coded for all video types	
Type of video	1.00
Personal**:** Primary focus on sharing experiences, stories, or emotions	
Informational-educational: Primary focus is to provide information	
Character and noncharacter videos	1.00
Character: Contains 1 or more live people	
Noncharacter: Contains no live people in video (eg, animation or slideshow of photos)	
Purpose for posting	
Sharing information: Information, facts, details and descriptions of infertility conditions, causes, and treatments. May include information that poster finds online, in books, or by reading scientific papers	1.00
Emotional aspects of infertility: Discussion of the emotional aspects of infertility process, diagnosis and/or treatment (ie, how individual or couple feels, processing of emotions)	.71
Treatment outcomes: Discussion of treatment outcome(s). May include infertility treatment experiences (ie, reporting positive or negative treatment results)	1.00
Advice to others: Individual or couple in video is giving advice to others; advocating for a treatment, decision, or action	.75
Source of video	.63
Personal account: An individual's (or couple's) channel	
Organization's channel: An institute, organized group, or company's channel	
News channel: A news source's channel	
Other: Type of channel is unclear or does not fit other categories	
Source: Type of organization	.77
Drug company: Channel belongs to a drug or pharmaceutical company	
Doctor or clinic: Channel belongs to a medical clinic or a doctor	
Academic institution: Channel belongs to a university, medical school, or other education-promoting channel	
Other: Channel belongs to organization, but does not fit into another category	
Tone of video	1.00
Positive: More than 50% of video tone is positive, indicated by the dominance of 1 or more of the positive subtones (see below)	
Negative: More than 50% of the video tone is negative, indicated by the dominance of 1 or more of the negative subtones (see below)	
Neutral: Absence of overarching positive or negative tones	
Mixed: The video tone is approximately 50% positive and 50% negative	
Positive subtones in video	
Hope: Overall sentiment that their future holds improvements or success, even in the face of current adversities	.69
Encouraging: Reassurance and hope toward viewers/others; offers statements of inspiration, comfort, and/or hope	.85
Humorous: Humor or jokes when talking about their treatment or throughout the video	.83
Joyful: Laughing, excited comments, and joyous exclamations	.83
Grateful: Thankfulness, gratitude, or feeling blessed for life/situation	.84
Negative subtones in video	
Anger or frustration: Anger or frustration toward self, others, or situation	—[b]
Sadness: Sorrow evidenced by verbal statements and/or frowning, crying, downcast expression	—[b]
Hopelessness: An overall sentiment that things will not get better	—[b]
Thanking viewers	
Thanking viewers for providing support or participation on YouTube (ie, subscriptions, comments)	1.00

XSL•FO
RenderX

Variable definitions	Kappa[a]
Variables coded for informational-educational videos only	
Type of informational-educational video	
Infertility causes or detection: Causes, signs, or detection of infertility	.88
Medical procedures: How a medical procedure works (eg, in vitro fertilization) or how a lab technique works (eg, measuring sperm motility)	.89
Lifestyle: Lifestyle choices (ie, diet, exercise, sleep)	1.00
Psychological: Mental health factors relating to infertility	.63
Variables coded for personal videos only	
Relationship of individual(s) in video	.87
Romantic or unspecified: Explicitly states being in a relationship, but marital status not specified	
Engaged: States being engaged	
Married: States being married	
Single: States being single	
Missing: Does not explicitly state relationship or unclear	
Relationship type	.86
Male and female, female and female, male and male, male only (ie, single male), female only (ie, single female), or unclear from video if individuals in video are partnered	
Sex of people in video	1.00
Going through treatment together and in video: Male and female, female and female, male and male, male only (ie, single), female only (ie, single), or unclear from video	
Parity	.88
Number of children explicated stated	
Stage of fertility treatment	1.00
Pretreatment: Have not started any fertility drugs or treatment	
Undergoing treatment: Currently undergoing treatment, procedures, or taking medications	
Pregnancy: Currently pregnant	
Pregnancy failure: Failed round of treatment and no pregnancy	
Postchildbirth: Child has been born	
Posttreatment: Treatment completed and waiting on results	
Round of treatment	1.00
First round of treatment: Started or undergone first round	
Multiple rounds of treatment: Started or undergone more than 1 round	
Live pregnancy results	1.00
Pregnancy results disclosed to spouse, friends, or family in video	

[a]2 coders, n=14 training videos.

[b]Codes for at least 1 coder were a constant; therefore, we were unable to calculate the kappa statistic.

Sample

In January 2017, the most-viewed infertility YouTube videos were identified using the search term "infertility" on YouTube on Google Chrome browser with the search history and cookies cleared [31]. We chose the term "infertility" to capture the most common, basic query a patient would be likely to search, as people generally use keywords and condition names when searching for health information online [32]. This search produced approximately 314,000 results. Videos were sorted by view count and the 80 most-viewed videos were selected for analysis. Previous research suggests over 90% of users click on results in the first 3 pages of a search [33]; therefore, we elected to include the first 4 pages of results to ensure selection of popular infertility videos most likely to be encountered (20 results per page). Data for each video were collected and stored in Microsoft Excel. When variable information was not provided for a video (eg, number of shares), it was coded as missing. The URL and title of video were saved as a reference during data analysis.

Figure 1. Diagram of selection process of YouTube videos for inclusion in sample.
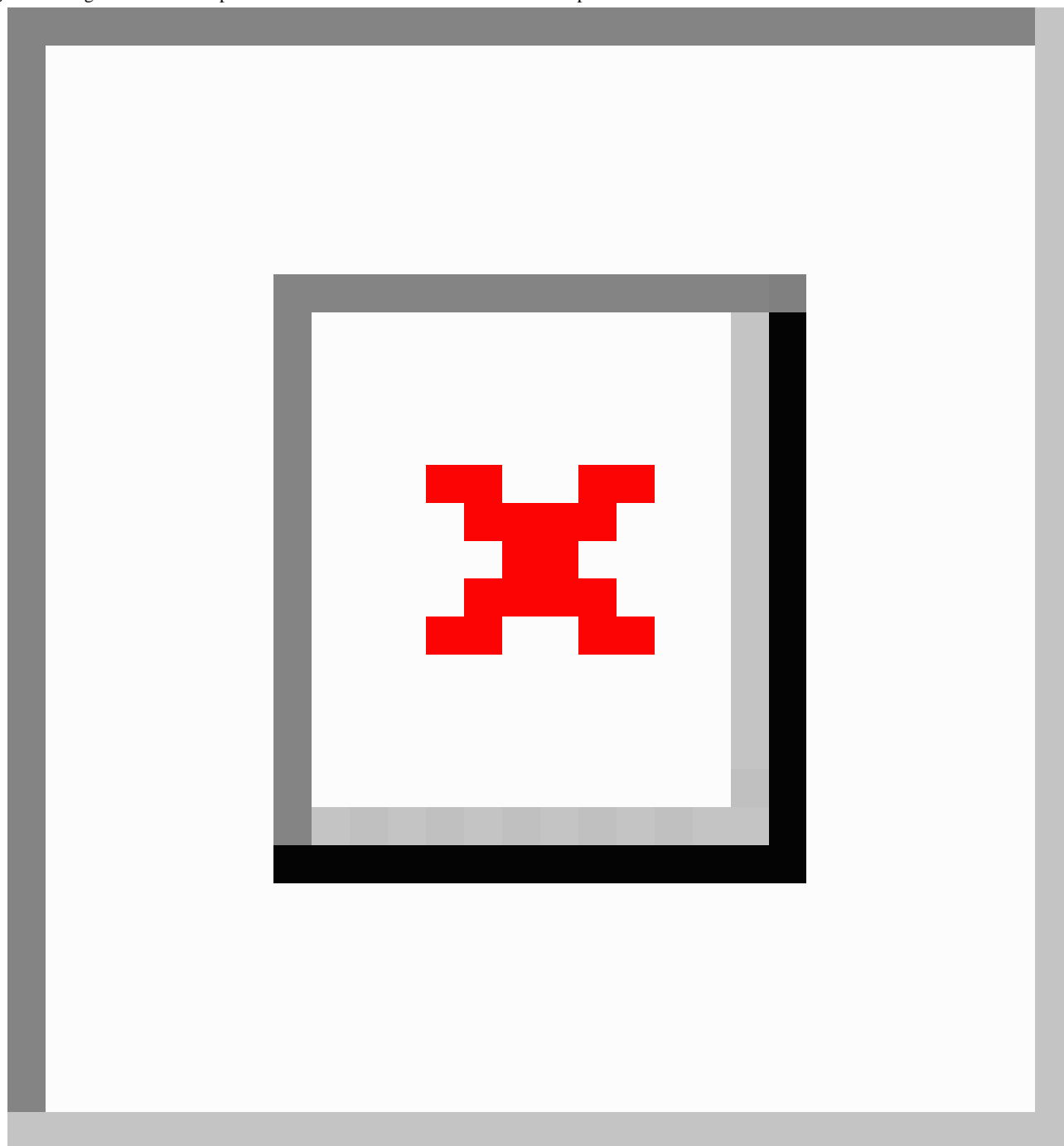
Videos were excluded from the sample if they were non-English (n= 16), unrelated to infertility (n= 3), or had age-restricted access (n=2). This produced a final dataset of 59 videos. Refer to Figure 1 for an inclusion flowchart.

Approximately 20% of the final dataset (14/59, 24%) was independently coded by 2 coders (MKH and FB) to examine inter-rater reliability. Cohen kappa was used to evaluate inter-rater reliability (see Table 1 for kappa coefficients). Kappa coefficients ranged from .638 (substantial agreement) to 1.00 (near perfect agreement), reflecting satisfactory or excellent inter-rater reliability for health-related research [34]. All analysis was performed using SPSS Statistics 23 software (IMB Analytics). Median value and inter-quartile range were used for most video statistics to provide a measure of central tendency,

which tends to be less influenced by skewed data. A similar approach has been used in previous research of health-related topics on YouTube [35-37].

Results

Classification of Videos

The total sample of 59 videos ranged from 50 seconds to 40 min in length (mean 6.6 min, SD 7.10 min). Videos were collectively viewed almost 23 million times (median=156,103 views). Videos received more likes (median=525 likes) than dislikes (median=25.5 dislikes), with a median of 55 comments per video. Only 35 videos indicated number of shares, with a

median of 105 shares per video (see Table 2 for video popularity statistics).

Videos were classified into 4 channel categories—personal accounts (23/59, 39%), organizations' channels (20/59, 34%), news channels (2/59, 3%), and accounts that could not be classified (eg, due to lack of information on channel; 14/59, 24%). The channels had a median of almost 6000 subscribers.

Table 2. Median and interquartile range of video popularity measures in total video sample (n=59).

Video Statistic	n[a]	Median	Interquartile range
Views	59	156103.0	230000.0
Likes	58	525.0	1892.2
Dislikes	58	25.5	47.5
Comments	57	55.0	354.5
Shares	35	105.0	201.0
Subscribers	59	5992.0	189273.0

[a]n varies as video statistics were not available for all videos.

The majority of videos had people in them (character videos, 48/59, 81%), whereas 19% had animation or photos (noncharacter videos, 11/59). More than half of the videos discussed treatment outcomes (33/59, 56%) and half shared information (30/59, 51%). About one-third of videos discussed emotional aspects of infertility (20/59, 34%) and 10% offered advice to others (6/59). The sample contained more personal (35/59, 59%) than informational-educational (24/59, 41%) videos. The characteristics of these 2 categories of videos are discussed below and are summarized in Table 3 (personal videos) and Table 4 (informational-educational videos).

Table 3. Frequency and proportion of codes for personal videos (n=35).

Variable	n (%)
Purpose for posting[a]	
Treatment outcomes	33 (94)
Emotional aspects	18 (51)
Sharing information	6 (17)
Advice to others	4 (11)
Overall tone	
Positive	26 (74)
Mixed	8 (23)
Neutral	1 (3)
Source of video	
Personal account	23 (66)
Organization	1 (3)
Other	9 (26)
News source	2 (6)
Sex of Poster(s)	
Female	12 (34)
Male-female couple	12 (34)
More than 2 individuals in video	10 (29)
Relationship unclear	1 (3)
Relationship type	
Heterosexual	34 (97)
Unspecified	1 (3)
Relationship status	
Married	18 (51)
Unspecified	16 (46)
Unclear	1 (3)
Parity	
No children	9 (26)
One child	6 (17)
More than 1 child	3 (9)
Number of children not stated	17 (48)
Round of treatment	
Unspecified	21 (60)
Multiple rounds	13 (37)
First round	1 (3)

[a]Videos could be coded for more than 1 purpose.

Table 4. Frequency and proportion of codes for informational-educational videos (n=24).

Variable	n (%)
Purpose for posting[a]	
Sharing information	24 (100)
Emotional aspects	2 (8)
Advice to others	2 (8)
Type of informational video[a]	
Infertility causes or detection	16 (67)
Medical procedures	14 (58)
Lifestyle	7 (29)
Psychological	6 (25)
Overall tone	
Neutral	24 (100)
Source of video	
Organization	19 (79)
Doctor/clinic	9 (38)
Other organization	9 (38)
Academic institute	1 (4)
Other	9 (47)

[a]Videos could be coded for more than 1 purpose for posting and more than 1 type of informational video.

Personal Videos

Characteristics of Posters

More than two-thirds of personal videos contained a female (12/35, 34%) or male-female couple (12/35, 34%); none of the videos contained only a male. In all but 1 video, the individual or couple was coded as being in a heterosexual relationship (34/35, 97%). Approximately half of all personal videos included couples who were married (18/35, 51%) and nearly half did not specify their relationship status (16/35, 46%). Similarly, parity was stated in about half of the videos—9 videos (9/35, 26%) reported no children, 6 videos (6/35, 17%) reported 1 child, and 3 videos (3/35, 9%) reported more than 1 child.

Only one-fifth of personal videos mentioned a specific infertility diagnosis, which included polycystic ovary syndrome and low sperm motility (7/35, 20%). Type of infertility treatment was mentioned in about half of personal videos (18/35, 51%), and included both medical (eg, ovulation medication, intrauterine insemination, in vitro fertilization, surgery) and nonmedical treatments (eg, diet, supplements, exercise, psychological treatment, alternative medicine). In almost two-thirds of personal videos (22/35, 63%), the poster was at the pregnancy stage of treatment and the remainder were in pretreatment (1/35, 3%), undergoing treatment (3/35, 9%), posttreatment (2/35, 6%), postchildbirth (6/35, 17%), or not specified (2/35, 6%). One-third of posters disclosed they had undergone multiple rounds of treatment (13/35, 37%), whereas only 1 mentioned it was their first round (1/35, 3%).

Purpose

A total of 33 of the personal videos (33/35, 94%) coded as "treatment outcomes" disclosed the outcome of a diagnosis, test, or treatment. For example, 1 woman filmed herself watching her pregnancy test turn positive and exclaimed:

> My heart is racing, okay, let's see. [looks at stick] Oh my god. Oh my god. It's positive...so it's positive.

Of these 33 videos, 21 (21/33, 64%) were pregnancy announcements that included live results of tests or sharing the results with their spouse, friends, or family.

Videos discussing how the poster felt during treatment were coded as "emotional aspects of infertility" (18/35, 51%). One woman detailed her emotional responses to the successes and failures of treatment:

> I was on ovulation medication for several months and each month was an emotional rollercoaster of hoping we'd get pregnant and not being pregnant. Until finally, one month, we were pregnant...we were just so happy...unfortunately, two weeks after...I miscarried the baby. And it was such a sadness, my heart was broken. I remember crying and the pain and the fear and the emptiness and heartache.

In some cases, videos discussing emotional aspects of infertility mentioned interactions with those not struggling with infertility. One couple, reflecting on their past infertility experiences, shared:

Whatever the situation, it is okay to be sad, but it's also important to be happy for others. And that was sometimes very...hard for me.

Similarly, some videos discussed the anxiety related to waiting for test results or preparing for treatment. One woman made a video the day she was preparing to undergo a hysterosalpingogram test and stated:

I have been having extreme bouts of anxiety and I'm...really nervous...I'm like freaking out...the anticipation kills you, you know.

Videos were coded as "sharing information" if they included information provided by doctors or from online research (6/35, 17%). One individual discussed raising awareness about infertility:

I want to be the person to tell you guys, that from a young age, we're taught how to prevent pregnancy and how to not get pregnant for a very long time until we're successful and have a job and have graduated college...what our parents and teachers and doctors aren't telling us, is when you get to a certain age, it is close to impossible to have a baby without fertility treatments and IVF.

Finally, 4 personal videos (4/35, 11%) were coded as 'advice to others,' reflecting recommendations or advice to viewers about concrete actions or decisions they should make regarding infertility. This advice was often based on personal experience. For example, 1 woman talking about her journey to get pregnant at age 40, suggested lifestyle changes to her viewers:

I really recommend that you do the basal body temperature chart...by eating vegetables and fruits and drinking plenty of water and eliminating caffeine and alcohol, you're putting yourself into a more pH balanced state.

Tone

Of the 35 personal videos, the dominant tones were: positive (26/35, 74%), mixed (8/35, 23%), and neutral (1/35, 3%). No videos were coded as having a negative tone.

Videos with a positive tone were coded for the presence of 5 positive subtones—joy (25/26, 96%), hope (5/26, 19%), encouragement (5/26, 19%), humor (5/26, 19%), and gratitude (3/26, 12%). The majority of positive videos included joy, often indicated by facial expression, joyful crying, and exclamations such as feeling "on cloud nine." Videos with the subtone hope included statements such as:

Such build up...we just hope that we've got something to show for it...we hope they're positive [the results] but it's totally out of our control so fingers crossed for it.

Similarly, another poster discussed the decision to undergo a second round of intrauterine insemination (IUI):

I got inside my own head and I pictured everything being a success...I said, 'no I want to do it [IUI], I feel really confident.

When those posting conveyed hope or support to the viewer, the video was coded as encouraging, exemplified by the following statement:

I think it's totally okay to be sad, but never lose hope. Don't ever get yourself down. I mean, we all have moments where we feel vulnerable and sad and upset and everything but let that just be a small moment of your time. Let happiness fill you.

Videos that incorporated humor included jokes and laughter. A man in 1 video bantered with a relative when disclosing a pregnancy announcement, "this time I'm not joking." Finally, a small number of videos contained statements of gratitude. In 1 video, a couple discussing their child and their second child soon to come, stated:

We say family prayers together and we talk about what we're grateful for...we really don't let a day pass us that we don't...remind ourselves...I cannot believe we have [a child]...I'm grateful to be pregnant.

A total of 11 videos (11/59, 19%) included messages thanking viewers or subscribers for watching, commenting on, liking, subscribing, or otherwise showing support to the poster.

When posters discussed both positive and negative experiences, the tone was coded as mixed. For example, 1 woman describing her infertility journey expressed frustration (ie, negative tone) with the doctors she initially encountered:

He [the doctor] was literally dismissing me because I was not old enough in his eyes to deserve treatment...that was something that was one of the hardest things that I dealt with, with my infertility. It was people not taking us seriously...Whose business was it to tell us we couldn't [have kids yet]? And it wasn't fair and it was frustrating and I was angry because...I felt like my right as a human being, as a woman was taken from me. It was the most frustrating, emotional, devastating thing to go through.

After switching fertility doctors and undergoing treatment, the woman was able to achieve pregnancy, making the positive statement:

I finally saw those two lines [on the pregnancy stick]...It was the greatest moment...It's the greatest thing to achieve getting pregnant when you haven't been able to...I will never take it for granted.

The 1 neutral video included an acupuncturist who shared her personal story on natural pregnancy, explaining the steps she took to conceive.

Informational-Educational Videos

The purpose of all 24 informational-educational videos was coded as sharing information and each was neutral in tone. Emotional aspects of infertility treatment and advice to viewers were each present in 2 videos. Over half of the videos contained information relating to infertility causes or detection (16/24, 67%) and/or medical procedures (14/24, 58%), whereas less

than one-third discussed lifestyle factors (7/24, 29%) and/or psychological factors related to infertility (6/24, 25%).

Characteristics Relating to Video Popularity

Personal videos (mean 2507.66, SD 2663.87) had significantly more likes than informational-educational videos (mean 241.22, SD 338.38; t_{56}=4.04; P<.001). Personal and informational-educational videos did not differ on number of views, dislikes, subscriptions driven, or shares. Similarly, personal videos (mean 347.32, SD 405.63) had significantly more comments than informational-educational videos (mean 33.48, SD 36.48; t_{55}=3.69; P<.001).

Positive (n=26) and neutral (n=25) videos were further analyzed for features relating to video popularity. Personal videos with mixed tone (n=8) were not analyzed due to small sample size. Positive and neutral videos did not differ on number of views, dislikes, subscriptions driven, or shares. Positive videos had significantly more likes (mean 2059.56, SD 2355.67) than neutral videos (mean 254.29, SD 337.08; t_{47}=3.79; P<.001). Positive videos also had significantly more comments (mean 267.32, SD 321.34) than neutral videos (mean 34.38, SD 35.95; t_{46}=3.63; P<.001). As all positive videos (n= 26) were personal videos, it was not possible to determine whether greater engagement was due to videos being positive or personal.

Discussion

Principal Findings

This study was the first to examine infertility-related videos on YouTube using a content analysis method. These videos had high numbers of comments, likes, shares, and subscriptions, indicating that many individuals seek out and engage with infertility-related information on YouTube. The videos were categorized into informational-educational videos and personal videos. Personal and informational-educational videos did not differ in number of views, supporting claims that people use YouTube for both health-related information [38] and social interaction and discussion [38,39]. However, personal videos did elicit more engagement as evidenced by more likes and comments. This suggests that, compared with informational-educational videos, personal videos may resonate more strongly with viewers. One reason may be that personal videos allow people to observe others undergoing similar experiences to seek reassurance, normalization, and niche support [14,15,40]. In addition, these videos may provide a platform for the posters to share and possibly validate their own infertility experiences.

Viewers also showed greater engagement with positive videos, which garnered more likes and comments than neutral videos. This may indicate viewer preference for videos that convey a positive outlook related to fertility. Video messages showing positive fertility outcomes may increase hope and optimism in the viewer, but they may also give a false impression of the success rates of fertility treatment. In actuality, less than half of embryo transfers for in vitro fertilization result in a live birth in the United States and Canada, with rates dropping with increasing maternal age [41,42]. Reliance on YouTube for infertility information may foster unrealistic expectations regarding the success rates of treatment, which may influence treatment decisions [38,43]. As fertility patients frequently access the internet for information [13,14], research is needed to establish whether YouTube content can affect the perceptions of infertility treatment among fertility patients.

In line with research suggesting that men are less likely to share YouTube videos [39], there was an absence of videos containing men only. When men were present, they were always accompanied by their female partner or other people (eg, friends or family). This finding is consistent with research showing that online activity and communication about infertility is more common among women [44-46]. Although we found a lack of male characters in videos, it remains possible that men seek out infertility content through other online avenues, such as websites and discussion forums. Research is needed to determine whether men use YouTube to search for information related to infertility or if they prefer other online platforms.

Limitations and Future Directions

There are several limitations to the study and future directions to consider. As we only examined the content of high-rated videos, the results may not be generalizable to all infertility videos on YouTube. It is possible that less viewed and favored videos may also contain stories of negative experiences and treatment failures, which were mostly absent from highly viewed videos. YouTube videos with negative tones, such as sadness and anger, may be less viewed and preferred by viewers. Our study was a cross-sectional examination of videos and did not explore how video popularity changed over time. Due to the low number of videos where advice was offered to viewers, we were not able to assess quality of the information posted. Further research is needed to undertake such an assessment, as some studies have found health information on YouTube to be false or misleading [28,47,48]. Finally, as we analyzed publicly available content, we were only able to assess metrics available from YouTube and did not ask viewers their motivations for watching videos or the impact of these videos on their attitudes and beliefs. Future research should explore how infertile men and women search for videos, their reasons for doing so, and their reactions to the content of the videos.

Conclusion

This analysis establishes highly viewed YouTube videos as a source of both technical and experience-based information related to infertility. Viewers appeared to prefer and engage more with videos containing personal experiences and positive tones. Although these videos may provide hope to viewers by frequently sharing positive treatment outcomes such as pregnancy announcements, they may also create and perpetuate unrealistic expectations about the success rates of fertility treatments.

XSL•FO

RenderX

Authors' Contributions

MKH, PHG, FB, and PZ were involved in study design and codebook development. MKH and FB independently coded the videos. MKH wrote the manuscript with assistance from PGG and PZ. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Government of Canada. Fertility URL: https://www.canada.ca/en/public-health/services/fertility/fertility.html [accessed 2018-02-15] [WebCite Cache ID 6xG59iMkh]
2. White L, McQuillan J, Greil AL, Johnson DR. Infertility: testing a helpseeking model. Soc Sci Med 2006 Feb;62(4):1031-1041. [doi: 10.1016/j.socscimed.2005.11.012] [Medline: 16360257]
3. Thoma ME, McLain AC, Louis JF, King RB, Trumble AC, Sundaram R, et al. Prevalence of infertility in the United States as estimated by the current duration approach and a traditional constructed approach. Fertil Steril 2013 Apr;99(5):1324-1331.e1 [FREE Full text] [doi: 10.1016/j.fertnstert.2012.11.037] [Medline: 23290741]
4. Datta J, Palmer MJ, Tanton C, Gibson LJ, Jones KG, Macdowall W, et al. Prevalence of infertility and help seeking among 15 000 women and men. Hum Reprod 2016 Sep;31(9):2108-2118 [FREE Full text] [doi: 10.1093/humrep/dew123] [Medline: 27365525]
5. Terävä AN, Gissler M, Hemminki E, Luoto R. Infertility and the use of infertility treatments in Finland: prevalence and socio-demographic determinants 1992-2004. Eur J Obstet Gynecol Reprod Biol 2008 Jan;136(1):61-66. [doi: 10.1016/j.ejogrb.2007.05.009] [Medline: 17640794]
6. Schmidt L. Infertility and assisted reproduction in Denmark. Epidemiology and psychosocial consequences. Dan Med Bull 2006 Nov;53(4):390-417. [Medline: 17150146]
7. Jacob MC, McQuillan J, Greil AL. Psychological distress by type of fertility barrier. Hum Reprod 2007 Mar;22(3):885-894. [doi: 10.1093/humrep/del452] [Medline: 17178745]
8. Dooley M, Dineen T, Sarma K, Nolan A. The psychological impact of infertility and fertility treatment on the male partner. Hum Fertil (Camb) 2014 Sep;17(3):203-209. [doi: 10.3109/14647273.2014.942390] [Medline: 25116275]
9. Valoriani V, Lotti F, Lari D, Miccinesi G, Vaiani S, Vanni C, et al. Differences in psychophysical well-being and signs of depression in couples undergoing their first consultation for assisted reproduction technology (ART): an Italian pilot study. Eur J Obstet Gynecol Reprod Biol 2016 Feb;197:179-185. [doi: 10.1016/j.ejogrb.2015.11.041] [Medline: 26773309]
10. Chachamovich JR, Chachamovich E, Ezer H, Fleck MP, Knauth D, Passos EP. Investigating quality of life and health-related quality of life in infertility: a systematic review. J Psychosom Obstet Gynaecol 2010 Jun;31(2):101-110. [doi: 10.3109/0167482X.2010.481337] [Medline: 20443659]
11. Talarczyk J, Hauke J, Poniewaz M, Serdyńska-Szuster M, Pawelczyk L, Jedrzejczak P. [Internet as a source of information about infertility among infertile patients]. Ginekol Pol 2012 Apr;83(4):250-254. [Medline: 22712255]
12. Kahlor L, Mackert M. Perceptions of infertility information and support sources among female patients who access the Internet. Fertil Steril 2009 Jan;91(1):83-90. [doi: 10.1016/j.fertnstert.2007.11.005] [Medline: 18243181]
13. Zelkowitz P, Robins S, Grunberg P. Searching for Infertility Information Online: Differences Between Men and Women. iproc 2016 Dec 30;2(1):e37. [doi: 10.2196/iproc.6243]
14. Porter M, Bhattacharya S. Helping themselves to get pregnant: a qualitative longitudinal study on the information-seeking behaviour of infertile couples. Hum Reprod 2008 Mar;23(3):567-572. [doi: 10.1093/humrep/dem398] [Medline: 18089553]
15. Hinton L, Kurinczuk JJ, Ziebland S. Infertility; isolation and the Internet: a qualitative interview study. Patient Educ Couns 2010 Dec;81(3):436-441. [doi: 10.1016/j.pec.2010.09.023] [Medline: 21036506]
16. Fox S, Duggan M. PewInternet. Health Online 2013 URL: http://www.pewinternet.org/files/old-media/ /Files/Reports/PIP_HealthOnline.pdf [accessed 2018-02-13] [WebCite Cache ID 6xBspL4Lp]
17. Antheunis ML, Tates K, Nieboer TE. Patients' and health professionals' use of social media in health care: motives, barriers and expectations. Patient Educ Couns 2013 Sep;92(3):426-431. [doi: 10.1016/j.pec.2013.06.020] [Medline: 23899831]
18. Alexa. 2018. Youtube.com Traffic Statistics URL: https://www.alexa.com/siteinfo/youtube.com [accessed 2018-02-15] [WebCite Cache ID 6xG6FuGNV]
19. YouTube. 2018. YouTube for Press URL: https://www.youtube.com/yt/about/press/ [accessed 2018-02-15] [WebCite Cache ID 6xG6NC5tA]
20. Flanders CE, Pragg L, Dobinson C, Logie C. Young sexual minority women's use of the internet and other digital technologies for sexual health information seeking. Can J Hum Sex 2017 Apr;26(1):17-25. [doi: 10.3138/cjhs.261-A2]
21. Vance K, Howe W, Dellavalle RP. Social internet sites as a source of public health information. Dermatol Clin 2009 Apr;27(2):133-6, vi. [doi: 10.1016/j.det.2008.11.010] [Medline: 19254656]
22. Park DY, Goering EM. The health-related uses and gratifications of YouTube: motive, cognitive involvement, online activity, and sense of empowerment. J Consum Health Internet 2016 May 09;20(1-2):52-70. [doi: 10.1080/15398285.2016.1167580]

XSL•FO

RenderX

23. Lo AS, Esser MJ, Gordon KE. YouTube: a gauge of public perception and awareness surrounding epilepsy. Epilepsy Behav 2010 Apr;17(4):541-545. [doi: 10.1016/j.yebeh.2010.02.004] [Medline: 20236867]

24. Tang W, Olscamp K, Choi SK, Friedman DB. Alzheimer's disease in social media: content analysis of YouTube videos. Interact J Med Res 2017 Oct 19;6(2):e19 [FREE Full text] [doi: 10.2196/ijmr.8612] [Medline: 29051137]

25. Stellefson M, Chaney B, Ochipa K, Chaney D, Haider Z, Hanik B, et al. YouTube as a source of chronic obstructive pulmonary disease patient education: a social media content analysis. Chron Respir Dis 2014 May;11(2):61-71 [FREE Full text] [doi: 10.1177/1479972314525058] [Medline: 24659212]

26. Lewis SP, Heath NL, St Denis JM, Noble R. The scope of nonsuicidal self-injury on YouTube. Pediatrics 2011 Mar;127(3):e552-e557. [doi: 10.1542/peds.2010-2317] [Medline: 21339269]

27. Desai T, Shariff A, Dhingra V, Minhas D, Eure M, Kats M. Is content really king? An objective analysis of the public's response to medical videos on YouTube. PLoS One 2013 Dec 18;8(12):e82469 [FREE Full text] [doi: 10.1371/journal.pone.0082469] [Medline: 24367517]

28. Murugiah K, Vallakati A, Rajput K, Sood A, Challa NR. YouTube as a source of information on cardiopulmonary resuscitation. Resuscitation 2011 Mar;82(3):332-334. [doi: 10.1016/j.resuscitation.2010.11.015] [Medline: 21185643]

29. Eysenbach G, Till JE. Ethical issues in qualitative research on internet communities. BMJ 2001 Nov 10;323(7321):1103-1105 [FREE Full text] [Medline: 11701577]

30. Lewis SP, Heath NL, Sornberger MJ, Arbuthnott AE. Helpful or harmful? An examination of viewers' responses to nonsuicidal self-injury videos on YouTube. J Adolesc Health 2012 Oct;51(4):380-385. [doi: 10.1016/j.jadohealth.2012.01.013] [Medline: 22999839]

31. YouTube. 2017. URL: https://www.youtube.com/ [accessed 2017-01-09] [WebCite Cache ID 6xKGAqY8h]

32. Jadhav A, Andrews D, Fiksdal A, Kumbamu A, McCormick JB, Misitano A, et al. Comparative analysis of online health queries originating from personal computers and smart devices on a consumer health information portal. J Med Internet Res 2014 Jul 04;16(7):e160 [FREE Full text] [doi: 10.2196/jmir.3186] [Medline: 25000537]

33. iProspect. 2006. iProspect Search Engine User Behavior Study URL: http://district4.extension.ifas.ufl.edu/Tech/TechPubs/WhitePaper_2006_SearchEngineUserBehavior.pdf [accessed 2018-02-15] [WebCite Cache ID 6xGJQ1KDn]

34. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]

35. Primack BA, Colditz JB, Pang KC, Jackson KM. Portrayal of alcohol intoxication on YouTube. Alcohol Clin Exp Res 2015 Mar;39(3):496-503 [FREE Full text] [doi: 10.1111/acer.12640] [Medline: 25703135]

36. Samuel N, Alotaibi NM, Lozano AM. YouTube as a source of information on neurosurgery. World Neurosurg 2017 Sep;105:394-398. [doi: 10.1016/j.wneu.2017.05.111] [Medline: 28599904]

37. MacLean SA, Basch CH, Reeves R, Basch CE. Portrayal of generalized anxiety disorder in YouTube™ videos. Int J Soc Psychiatry 2017 Dec;63(8):792-795. [doi: 10.1177/0020764017728967] [Medline: 29145793]

38. 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]

39. Khan ML. Social media engagement: What motivates user participation and consumption on YouTube? Comput Human Behav 2017 Jan;66:236-247. [doi: 10.1016/j.chb.2016.09.024]

40. Malik SH, Coulson NS. Computer-mediated infertility support groups: an exploratory study of online experiences. Patient Educ Couns 2008 Oct;73(1):105-113. [doi: 10.1016/j.pec.2008.05.024] [Medline: 18639409]

41. Canadian Fertility and Andrology Society. 2016. Press Release Assisted Reproductive Technologies (ART) in Canada: 2015 Results from the Canadian ART Register (CARTR) URL: https://cfas.ca/public-affairs/canadian-art-register/ [WebCite Cache ID 6xG6cw0UE]

42. Centers for Disease Control and Prevention. ART National Data Success Rates URL: https://www.cdc.gov/art/index.html [accessed 2018-05-09] [WebCite Cache ID 6zHWo0Ytx]

43. Fox S, Jones S. Pew Internet. 2009 Jun. The social life of health information URL: http://www.pewinternet.org/files/old-media/Files/Reports/2009/PIP_Health_2009.pdf [accessed 2018-02-13] [WebCite Cache ID 6xBsM798G]

44. Malik SH, Coulson NS. Coping with infertility online: an examination of self-help mechanisms in an online infertility support group. Patient Educ Couns 2010 Nov;81(2):315-318. [doi: 10.1016/j.pec.2010.01.007] [Medline: 20144521]

45. Malik SH, Coulson NS. A comparison of lurkers and posters within infertility online support groups. Comput Inform Nurs 2011 Oct;29(10):564-573. [doi: 10.1097/NCN.0b013e318224b31d] [Medline: 21709546]

46. Haagen EC, Tuil W, Hendriks J, de Bruijn RP, Braat DD, Kremer JA. Current Internet use and preferences of IVF and ICSI patients. Hum Reprod 2003 Oct;18(10):2073-2078. [Medline: 14507823]

47. Pandey A, Patni N, Singh M, Sood A, Singh G. YouTube as a source of information on the H1N1 influenza pandemic. Am J Prev Med 2010 Mar;38(3):e1-e3. [doi: 10.1016/j.amepre.2009.11.007] [Medline: 20171526]

48. Sood A, Sarangi S, Pandey A, Murugiah K. YouTube as a source of information on kidney stone disease. Urology 2011 Mar;77(3):558-562. [doi: 10.1016/j.urology.2010.07.536] [Medline: 21131027]

Abbreviations

IUI: intrauterine insemination

XSL•FO
RenderX

Original Paper

Spatial Distribution of Partner-Seeking Men Who Have Sex With Men Using Geosocial Networking Apps: Epidemiologic Study

Angel B Algarin[1], MPH; Patrick J Ward[2], MPH; W Jay Christian[2], MPH, PhD; Abby E Rudolph[3], MPH, PhD; Ian W Holloway[4], PhD; April M Young[2,5], MPH, PhD

[1]Robert Stempel College of Public Health & Social Work, Florida International University, Miami, FL, United States

[2]Department of Epidemiology, University of Kentucky, Lexington, KY, United States

[3]Department of Epidemiology, Boston University School of Public Health, Boston, MA, United States

[4]Department of Social Welfare, Luskin School of Public Affairs, University of California Los Angeles, Los Angeles, CA, United States

[5]Center on Drug and Alcohol Research, University of Kentucky, Lexington, KY, United States

Corresponding Author:
April M Young, MPH, PhD
Department of Epidemiology
University of Kentucky
111 Washington Avenue
Lexington, KY,
United States
Phone: 1 859 218 2090
Email: april.young@uky.edu

Abstract

Background: Geosocial networking apps have made sexual partner-seeking easier for men who have sex with men, raising both challenges and opportunities for human immunodeficiency virus and sexually transmitted infection prevention and research. Most studies on men who have sex with men geosocial networking app use have been conducted in large urban areas, despite research indicating similar patterns of online- and app-based sex-seeking among men who have sex with men in rural and midsize cities.

Objective: The goal of our research was to examine the spatial distribution of geosocial networking app usage and characterize areas with increasing numbers of partner-seeking men who have sex with men in a midsize city in the South.

Methods: Data collection points (n=62) were spaced in 2-mile increments along 9 routes (112 miles) covering the county encompassing the city. At each point, staff logged into 3 different geosocial networking apps to record the number of geosocial networking app users within a 1-mile radius. Data were collected separately during weekday daytime (9:00 AM to 4:00 PM) and weekend nighttime (8:00 PM to 12:00 AM) hours. Empirical Bayesian kriging was used to create a raster estimating the number of app users throughout the county. Raster values were summarized for each of the county's 208 Census block groups and used as the outcome measure (ie, geosocial networking app usage). Negative binomial regression and Wilcoxon signed rank sum tests were used to examine Census block group variables (eg, median income, median age) associated with geosocial networking app usage and temporal differences in app usage, respectively.

Results: The number of geosocial networking app users within a 1-mile radius of the data collection points ranged from 0 to 36 during weekday daytime hours and 0 to 39 during weekend nighttime hours. In adjusted analyses, Census block group median income and percent Hispanic ethnicity were negatively associated with geosocial networking app usage for all 3 geosocial networking apps during weekday daytime and weekend nighttime hours. Population density and the presence of businesses were positively associated with geosocial networking app usage for all 3 geosocial networking apps during both times.

Conclusions: In this midsize city, geosocial networking app usage was highest in areas that were more population-dense, were lower income, and had more businesses. This research is an example of how geosocial networking apps' geospatial capabilities can be used to better understand patterns of virtual partner-seeking among men who have sex with men.

XSL•FO

RenderX

Introduction

HIV in the United States continues to disproportionately affect men who have sex with men (MSM) despite ongoing prevention measures taken by public health officials [1]. In 2016, MSM accounted for 67% of all new HIV infections [2]. HIV and sexually transmitted infection research and intervention among MSM increasingly has focused on the social environment where risk behavior occurs [3-5], particularly as more MSM are using Web-based tools or mobile phone geosocial networking (GSN) apps (eg, Grindr, Hornet, Adam4Adam, Scruff, etc) to meet sex partners [6].

GSN apps provide information on geographic proximity between users making sexual partner seeking quick and convenient [7-10]. Research among MSM who use GSN apps to find sex partners has shown mixed results regarding the relationship between GSN app use and risky sexual practices. While some studies found no association between GSN app use for partner seeking and sexual risk behavior [11,12], some research suggests that partner seeking on GSN apps is associated with increased condomless anal intercourse [9,13,14], drug use (ie methamphetamine, Viagra, poppers, painkillers) [10,13,15], number of partners [10,13,16,17], and history of sexually transmitted infection diagnosis [15,16,18-21].

Though many studies have surveyed MSM to examine the use of GSN apps to find sex partners [10-16,18-22], relatively few studies have used the GSN apps' geospatial capabilities to better understand geographical patterns of partner seeking among MSM. Previous research in Atlanta, Georgia, described a methodology for using the geolocation features of a GSN app as a novel approach to calculating the estimated spatial density of GSN app-using MSM. The study collected information on the closest 50 users or, when the total was less than 50, all users within a 2-mile radius of each data collection point. The data were then used to create race-stratified maps that highlighted areas of high spatial densities of MSM [23]. Research projects like the aforementioned show promise in informing geotargeted HIV prevention, treatment, and recruitment strategies. For example, geofencing, a practice widely used in mobile advertising, relies on mobile phone Global Positioning System and radio-frequency identification technology to trigger strategic HIV prevention and treatment messaging when a user enters or exits a specified area [24].

Much of the research on GSN app usage among MSM has been conducted in large urban areas [10,15,22,23,25,26], yet similar patterns of online sex seeking are reported among MSM residing in rural and midsize cities [27,28]. Smaller cities often differ from larger cities in terms of stigma [29,30] and availability of visible gay spaces [31-33]. As others have noted, research on sexual health and app-facilitated sexual behavior among MSM in midsized cities is limited [34-36]. Some research has shown that, compared with online-recruited MSM in larger cities, those in small towns were more likely to report using apps to meet long-term partners and men for sex [37]. Other studies have indicated that rural MSM experiencing hostility, stigma, and social and sexual isolation often use the internet to find sex partners [38,39]. It is important to also acknowledge that these online forums and apps can be central in facilitating positive social connectedness and friendship among MSM [40,41,42]. Taken together, these findings indicate a changing landscape of social connection and risk behavior among MSM in which technology is increasingly relevant yet not well understood, especially in settings outside of major urban centers.

This study is among the first of its kind to use the geospatial capabilities of GSN apps to provide insights on the use of 3 different GSN apps by MSM in a midsize city during weekday daytime and weekend nighttime hours, thus building upon previous research in a large urban area that focused only on 1 app and lacked detail on weekday daytime versus weekend nighttime differences [23]. The purpose of this study was to describe the spatial distribution of GSN app-using MSM in a midsize city in the South and identify geographic and demographic factors associated with areas of high numbers of GSN app users.

Methods

Setting

We collected information on the use of 3 different MSM GSN apps using a geographically systematic sample of points in Fayette County, Kentucky [43]. Fayette County encompasses the city of Lexington, Kentucky, and has a land area of 284 square miles and a population of 295,805 people who are predominantly white (75.7%) and non-Hispanic (93.1%), with 30.7% between the ages of 18 and 34 years [43].

Data Collection

The road network in the county resembles a spoke pattern in which the city is bisected by a series of main roads that extend to the county's perimeter. The bisecting routes that offered the best county coverage were selected using Google Maps [44]. Data collection points (n=62) were spaced in 2-mile increments along 9 routes with a total driving distance of 112 miles (Figure 1).

We visited each point twice: once during weekday daytime (Monday through Friday, 9:00 am to 3:02 pm) and once during weekend nighttime (Friday or Saturday, 8:02 pm to 11:50 pm) hours. While stopped at each point (approximately 5 minutes per stop), we logged into a blank profile created for study purposes on each of the 3 GSN apps. Once we were logged into the profile, the app displayed the number of users within varying distances from the collection point. We recorded the number of users within 1 mile on each app, the time of collection, and the latitude and longitude at that collection point on a paper form and via the Fulcrum data collection app [45]. All apps used in this study have been self-described as providing a space for gay men to look for dates, friends, fun, relationships, and hookups. Apps 1 and 3 are nongroup/niche specific apps that have been described in previous research [46-49], with the latter attracting gay, bisexual, and curious men [50]. App 2 targets a subgroup identified as "bears" [47,48], who have a more traditional masculine style and acceptance of diverse body shapes [49].

Figure 1. Data collection points with 1-mile coverage buffers.
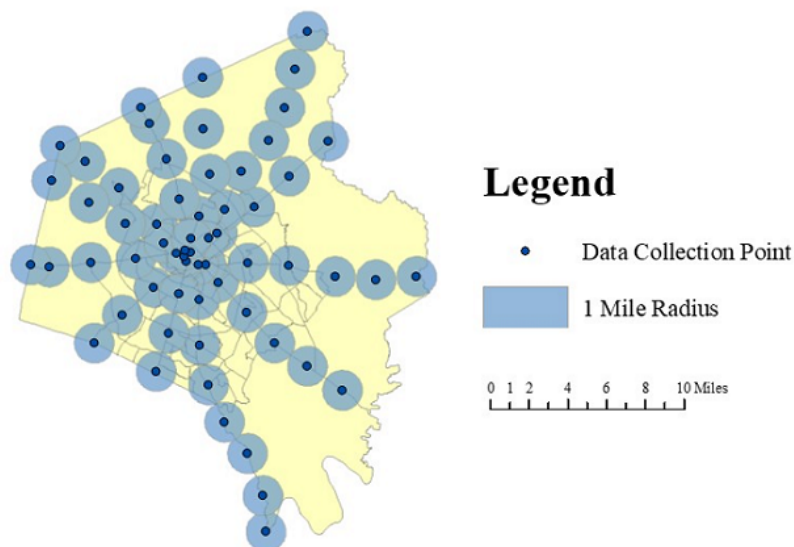
According to preliminary data from interviewer-administered surveys from 57 MSM aged 18 to 35 years from Lexington (data collection ongoing), apps 1, 2, and 3 were in the top 5 most frequently named MSM GSN apps used by respondents to "meet people for sex."

Measures

Variables included in the regression model were derived from the 2011-2015 American Community Survey 5-year estimates [51], the 2010 Census data from the US Census Bureau [43], and the Lexington-Fayette Urban County Government Open Data Web Portal [52]. Independent variables in the model included block group population density (population count per square mile; range 3.3-15,737.5), median age (range 7.5-64.5 years), median annual household income (range $7,905-$201,429), percent white (range 10.8%-100.0%), percent Hispanic (range 0.0%-64.2%), and business zoning amounting to greater (or equal) versus less than 1% of the area of the block group (binary: 40.4% had greater or equal to 1% of the area business zoned). These variables were included in the analyses because past research on GSN app use has found that demographic variables such as age [10,15], race/ethnicity [10,15], and income [22] were associated with app use. Business zoning, the areas specified for business use by the local government, was included to capture the presence of gathering spaces (eg, restaurants, bars, shopping venues, employers).

Analysis

Spatial Analyses

Of the 62 data collection locations, 61 had cell service (ie, LTE, 3G, 4G) to access apps for data collection. The point with no service was excluded from kriging analysis. We conducted spatial analyses using ArcMap version 10.3.1 (Environmental Systems Research Institute Inc) and negative binomial regression and Wilcoxon signed rank sum tests using SAS version 9.4 (SAS Institute Inc).

We constructed maps to illustrate differences in the spatial distribution and number of app-using MSM across apps at different time periods. The Fayette County shapefile used in

the depictions was procured from the Census Bureau [53]. We used empirical Bayesian kriging (EBK) for spatial interpolation as it predicts values for areas where data has not been collected based upon the specific values at each collected observation point and their relative proximity to other points. Past research has employed similar techniques, such as kernel smoothing, and found that kriging offered similar results [23]. The ArcMap EBK tool predicted the average number of app-using MSM for each raster grid (pixel), as show in Figure 2.

We next converted the EBK raster grid cell values to points at their centroids so these values could be assigned to the Census block groups; there were from 8 to 4302 of these points in each block group, depending on area, with a median of 46. This further enabled estimation of the average number of GSN app users within each Census block group. We used this average, rounded to the nearest whole number, in each block group as the dependent variable in a negative binomial regression analysis and to calculate weekday daytime/weekend nighttime differences in the number of GSN app-using MSM by app (Figure 3).

Statistical Analyses

Our dependent variable was the number of app users for a particular app at a particular time period in each Census block group. Given that we examined 3 different apps, each at 2 different times, with possible duplication of users across apps and times (ie, people using more than 1 app at a time and people using the same app at different time windows), we ran 6 independent models. We used Wilcoxon signed rank sum tests to examine differences in the number of people using the app between the 3 apps and between weekday daytime and weekend nighttime data collection periods for the same app because the outcome variables were paired and nonparametric. Negative binomial regression was used to examine geographic and demographic factors associated with areas with increasing numbers of GSN app-using MSM at the Census block group level. Negative binomial regression was used because the counts were overdispersed and therefore not appropriate for a typical Poisson model. We also ran the models with the independent

variables median age, median income, and population density log-transformed to try to force more linear relationships with the outcome. To test for collinearity, we ran the PROC REG collinearity diagnostic collinoint [54] to determine how related each variable was to each other in the presence of all other variables. The analysis for collinearity showed that none of the independent variables was collinear and all could be included in the same final model.

Figure 2. Empirical Bayesian kriging analysis of the estimated spatial distribution of geosocial networking app users by time and app.
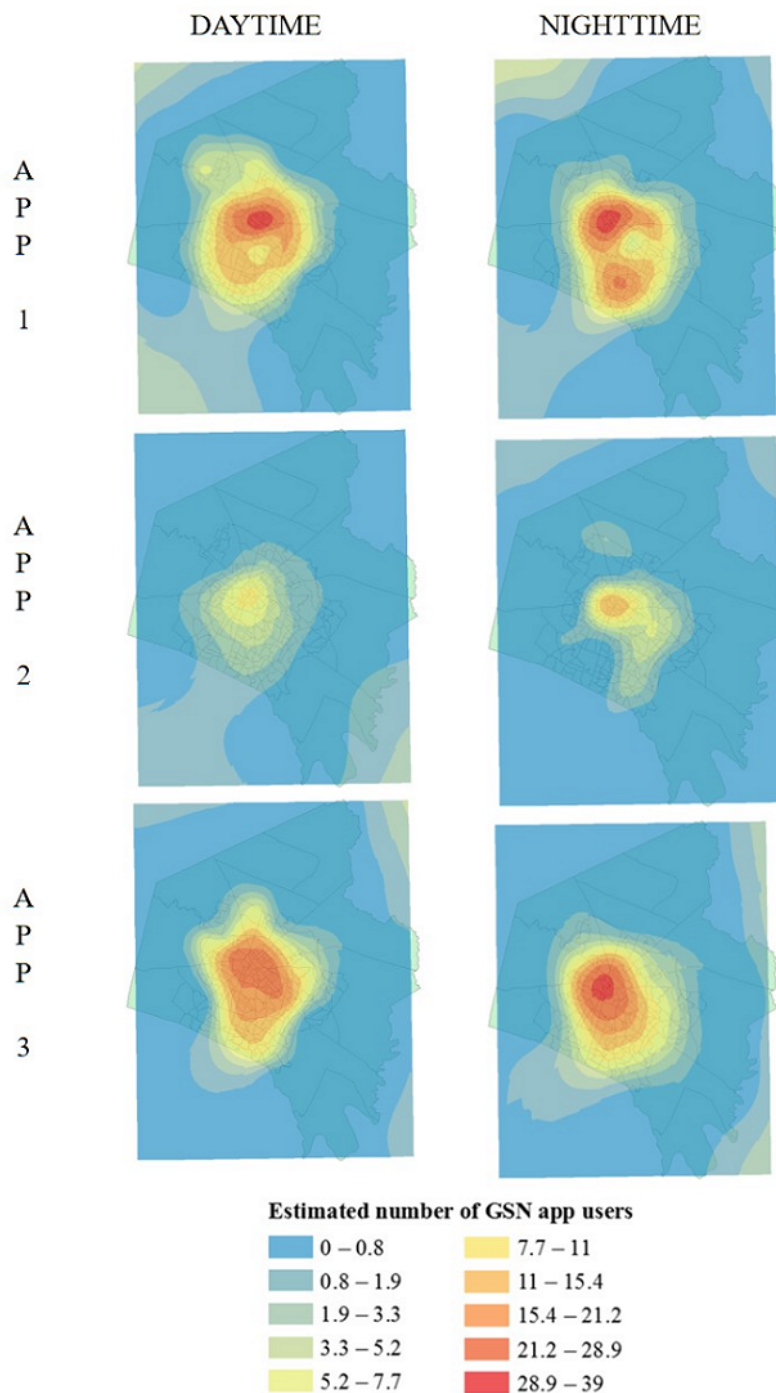
Figure 3. Maps displaying temporal (weekday day-weekend night) differences in the spatial distribution of app users.
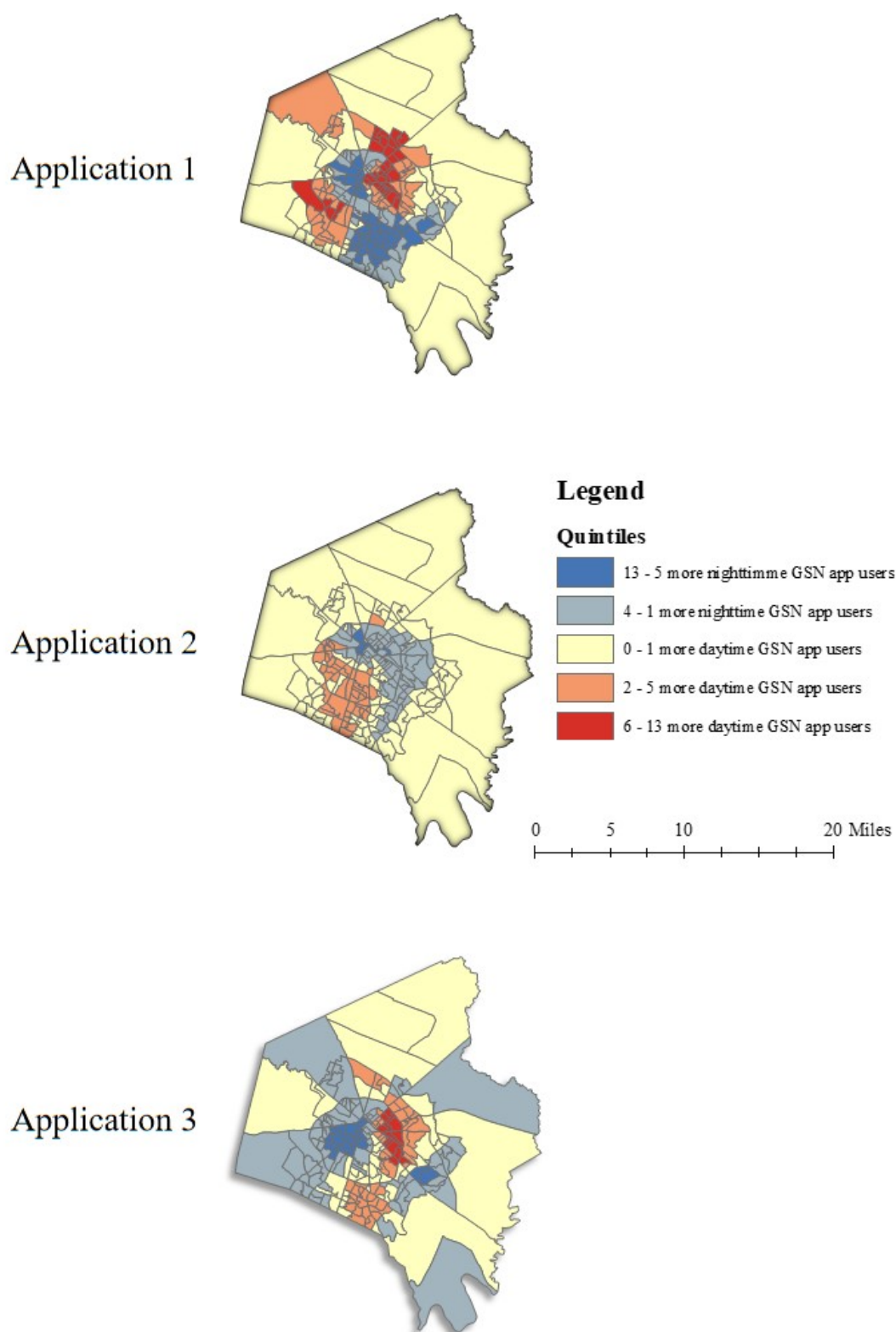
Results

Descriptive

The number of GSN app users within a 1-mile radius of the data collection points ranged from 0 to 36 during weekday daytime hours and 0 to 39 during weekend nighttime hours. The median number of estimated GSN app users in each Census block group in Fayette County during weekday daytime and weekend nighttime varied by app (10.0, 2.0, 10.5, and 10.0, 1.0, 9.0, respectively), but within-app comparisons revealed no significant temporal differences for any of the 3 apps (P=.95, .39, and .65; see Table 1).

Table 1. Descriptive statistics of estimated average number of geosocial networking app users per Census block group.

App and time of day	Weekday day, median (Q1,Q3)	Weekend night, median (Q1,Q3)	*P* value
App 1	10.0 (4.0,15.0)	10.0 (3.0,17.0)	.95
App 2	2.0 (1.0,4.0)	1.0 (0.0,5.0)	.39
App 3	10.5 (3.0,18.5)	9.0 (4.0,18.0)	.65
Day			
App 1 vs App 2 (ref[a])	7.0 (3.0,11.0)	—	<.001[b]
App 2 vs App 3 (ref)	−8.0 (−14.0,−2.0)	—	<.001[b]
App 1 vs App 3 (ref)	0.0 (−3.0,2.0)	—	.46
Night			
App 1 vs App 2 (ref)	—	7.0 (2.0,14.0)	<.001[b]
App 2 vs App 3 (ref)	—	−7.0 (−14.0,−3.0)	<.001[b]
App 1 vs App 3 (ref)	—	−1.0 (−3.0,2.5)	.46

[a]Ref: reference group.

[b]Indicates significant difference between apps.

We created choropleth maps to display differences in the spatial distribution and number of GSN app-using MSM by time of day (Figure 3). From weekday daytime to weekend nighttime, the spatial distribution of app-using MSM varied for 2 of the 3 apps. The use of app 1 was concentrated in downtown Lexington (ie, the center of the county) during weekday daytime and weekend nighttime; however, at weekend nighttime, a second area south of the city emerged as an area of high app usage. For app 3, use was concentrated in the downtown area both during weekday daytime and weekend nighttime but use was more intensely concentrated in the downtown area at weekend nighttime.

Block Group Regression Analysis

Fayette County comprises 208 Census block groups. Based on the dependent variable of estimated number of GSN app users in each Census block group, unadjusted negative binomial regression with each app stratified by time of collection was used to estimate crude risk ratios (Table 2). In all combinations of apps and collection times, the number of GSN app-using MSM was positively associated with population density (*P*<.001) and presence of business zoning of the block group area (*P*<.001), and negatively associated with age (*P*<.001 to *P*=.03) and median income (*P*<.001). For every 100 persons per square mile increase in population density, the number of GSN app users increased by 1.1% to 1.5%. In the presence of business zoning, the number of GSN app users increased by 106% to 188%. For every year increase in median age, the number of GSN app users decreased by 1.5% to 3%. For every $5,000 increase in median income, the number of GSN app users decreased by 5.7% to 8.9%.

Percent Hispanic ethnicity was not significantly associated with the number of GSN app-using MSM in any app or time combinations (*P*=.20 to *P*=.99). The statistical significance of the association between percent white and number of GSN app-using MSM varied between app and collection time (*P*<.001 to *P*=.13); percent white was negatively associated with the number of GSN app users for app 1 and app 3 during the daytime and app 1 and app 2 during the weekend nighttime, but not for app 2 during the weekday daytime or app 3 during weekend nighttime.

The multivariable models are shown in Table 3. In all combinations of app and collection times, number of GSN app-using MSM was positively associated with population density (*P*<.001 to *P*=.045) and business zoning (*P*<.001) and negatively associated with median income (*P*<.001) and percent Hispanic (*P*<.001 to *P*=.045), adjusting for all other variables. The statistical significance of the association of median age (*P*<.001 to *P*=.46) and percent white (*P*=.02 to *P*=.77) with the outcome of interest varied between app and collection time in the final model; median age was positively associated with the number of GSN app users for app 1 during the weekday daytime but not for any other time or app. Percent white was negatively associated with the number of GSN app users for app 1 during the weekday daytime and app 2 during the weekend nighttime but not for any other time or app. In multivariable models with income, age, and population density log-transformed (data not shown in table), the results were similar except that in both of the models for app 2 and app 3, median age became statistically significant and in app 1 day and app 2 night, percent white lost statistical significance.

Table 2. Unadjusted analysis of the association between Census block level characteristics and number of geosocial networking app–using men who have sex with men.

Parameters	Crude incidence rate ratio (95% CI)	P value
App 1: weekday day		
Age, year, median	0.99 (0.97-1.00)	.03
White (%)	0.99 (0.98-0.99)	<.001
Hispanic (%)	1.01 (1.00-1.03)	.20
Median household income (per $5,000)	0.92 (0.91-0.94)	<.001
Population (100 per mi^2)	1.01 (1.01-1.02)	<.001
Business zoning	2.24 (1.82-2.75)	<.001
App 1: weekend night		
Age, year, median	0.97 (0.96-0.99)	<.001
White (%)	0.99 (0.98-1.00)	.01
Hispanic (%)	1.00 (0.99-1.02)	.93
Median household income (per $5,000)	0.93 (0.92-0.95)	<.001
Population (100 per mi^2)	1.02 (1.01-1.02)	<.001
Business zoning	2.06 (1.64-2.58)	<.001
App 2: weekday day		
Age, year, median	0.98 (0.97-0.99)	.003
White (%)	1.00 (0.99-1.00)	.13
Hispanic (%)	1.00 (0.99-1.01)	.99
Median household income (per $5,000)	0.93 (0.91-0.94)	<.001
Population (100 per mi^2)	1.01 (1.01-1.02)	<.001
Business zoning	2.28 (1.89-2.73)	<.001
App 2: weekend night		
Age, year, median	0.97 (0.95-0.99)	.002
White (%)	0.99 (0.98-1.00)	.001
Hispanic (%)	1.00 (0.97-1.08)	.65
Median household income (per $5,000)	0.91 (0.89-0.94)	<.001
Population (100 per mi^2)	1.02 (1.01-1.02)	<.001
Business zoning	2.88 (2.13-3.91)	<.001
App 3: weekday day		
Age, year, median	0.98 (0.97-1.00)	.01
White (%)	0.99 (0.99-1.00)	.049
Hispanic (%)	1.00 (0.99-1.02)	.58
Median household income (per $5,000)	0.94 (0.93-0.96)	<.001
Population (100 per mi^2)	1.01 (1.01-1.02)	<.001
Business zoning	2.12 (1.67-2.68)	<.001
App 3: weekend night		
Age, year, median	0.98 (0.97-0.99)	.001
White (%)	1.00 (0.99-1.00)	.07
Hispanic (%)	1.00 (0.99-1.02)	.61
Median household income (per $5,000)	0.94 (0.93-0.96)	<.001

Parameters	Crude incidence rate ratio (95% CI)	P value
Population (100 per mi^2)	1.01 (1.01-1.02)	<.001
Business zoning	2.26 (1.85-2.76)	<.001

Table 3. Multivariable analysis of the association between Census block level characteristics and number of geosocial networking app–using men who have sex with men.

Parameters	Adjusted incidence rate ratio (95% CI)	*P* value
App 1: weekday day		
Age, year, median	1.03 (1.01-1.04)	<.001
White (%)	0.99 (0.99-1.00)	.03
Hispanic (%)	0.99 (0.98-1.00)	.01
Median household income (per $5,000)	0.93 (0.91-0.95)	.001
Population (100 per mi^2)	1.00 (1.00-1.01)	.04
Business zoning	1.69 (1.43-2.00)	.001
App 1: weekend night		
Age, year, median	1.01 (0.99-1.02)	.46
White (%)	1.00 (0.99-1.00)	.29
Hispanic (%)	0.98 (0.97-0.99)	<.001
Median household income (per $5,000)	0.94 (0.92-0.96)	<.001
Population (100 per mi^2)	1.01 (1.00-1.01)	<.001
Business zoning	1.56 (1.28-1.92)	<.001
App 2: weekday day		
Age, year, median	1.01 (1.00-1.03)	.06
White (%)	1.00 (1.00-1.01)	.47
Hispanic (%)	0.98 (0.97-0.99)	.003
Median household income (per $5,000)	0.94 (0.91-0.96)	<.001
Population (100 per mi^2)	1.01 (1.00-1.01)	.01
Business zoning	1.77 (1.47-2.13)	<.001
App 2: weekend night		
Age, year, median	1.01 (0.99-1.03)	.36
White (%)	0.99 (0.98-1.00)	.02
Hispanic (%)	0.96 (0.95-0.98)	<.001
Median household income (per $5,000)	0.93 (0.90-0.97)	<.001
Population (100 per mi^2)	1.01 (1.00-1.01)	.045
Business zoning	2.22 (1.67-2.93)	<.001
App 3: weekday day		
Age, year, median	1.01 (100-1.03)	.11
White (%)	0.99 (0.99-1.01)	.64
Hispanic (%)	0.99 (0.97-1.00)	.045
Median household income (per $5,000)	0.95 (0.93-1.00)	<.001
Population (100 per mi^2)	1.01 (1.00-1.01)	.01
Business zoning	1.77 (1.41-2.21)	<.001
App 3: weekend night		
Age, year, median	1.01 (0.99-1.02)	.25
White (%)	1.00 (1.00-1.01)	.77
Hispanic (%)	0.99 (0.98-1.00)	.02
Median household income (per $5,000)	0.95 (0.93-1.00)	<.001

XSL•FO
RenderX

Parameters	Adjusted incidence rate ratio (95% CI)	P value
Population (100 per mi^2)	1.01 (1.00-1.01)	.03
Business zoning	1.85 (1.53-2.24)	<.001

Discussion

Principal Findings

This study revealed that the presence of business zoning and population density were positively associated with the number of GSN app-using MSM during both weekday daytime and weekend nighttime for all GSN apps. We also found that median income and percent of the population who were Hispanic were negatively associated with the number of GSN app-using MSM during both weekday daytime and weekend nighttime for all GSN apps, adjusting for other variables in the model. Increased app use in areas with the presence of business zoning could imply that app users may be using these apps in areas of economic activity (eg, bars, restaurants, stores) including areas that cater to a predominantly gay clientele. This could highlight an important overlap between virtual and in-person partner-seeking spaces.

By using Wilcoxon signed rank sum tests and comparing choropleth maps of the differences in the spatial density of GSN app-using MSM, we determined that the total number of users between weekday daytime and weekend nighttime was not significantly different but that specific areas within the county could be experiencing changes in the number of partner-seeking MSM between these 2 time periods. This could imply either that the same users are migrating to different areas over time, different users at different locations are logging in at different times, or a combination of these. Previous research with GSN app-using MSM indicates that more than 50% log on 5 or more times per day with significantly greater percentages logging on during evening hours [55]; however, no studies to our knowledge have tracked use by time and geographic position. This information could be informative to local health departments because instead of using mobile HIV testing units only during weekend nighttime hours at nightlife venues, these units could also be used at specified hotspots during weekday daytime hours. These data may also be able to inform more cost-efficient geotargeted and temporally targeted recruitment strategies for research. However, more research is needed to explore attitudes about the presence of HIV outreach activities such as mobile HIV testing units or research recruitment efforts near daytime hotspots such as neighborhoods and places of business.

The novel methods of data collection used in this study addressed some of the limitations of past GSN app research. For example, prior research using GSN apps for data collection largely relied on a single app [12,14,23,26,41,42,56]. Our study used 3 apps, and the variance in spatial distribution of use and number of users across apps highlights the importance of using multiple apps in future research. Popularity of GSN apps shifts over time, and some apps are more popular among specific subgroups of MSM [57]. Previous research has also aggregated geographic data across time points rather than examining daytime and nighttime use separately [23]. Our study, which examined weekday daytime and weekend nighttime use separately, revealed that spatial patterns of use may vary by time of day.

Finally, data on the use of GSN apps is also novel in the context of midsize cities as most studies to date have focused solely on larger cities [13,15-17,20,22-24,26]. Collecting data on GSN apps for MSM in midsize cities could be of increased importance due to differences in social context [29-33,38,39] and the way in which users interact with the app in contrast with urban GSN app users [37]. This is especially important given the disparities in resources for sexual and gender minority people in rural areas compared to large urban centers. Prior research has demonstrated the importance of technology in the lives of rural sexual and gender minority people in combating social isolation and homophobia in public social spaces [37-41]. In general, more research on the role of technology in fostering resilience among MSM in rural areas is needed. However, given the literature linking GSN app use and sexual risk behavior, a closer examination of how these dynamics play out in rural settings is crucial to developing targeted HIV prevention efforts for rural MSM.

Limitations

Although the maps demonstrated substantial differences in estimated GSN app usage throughout the study area, more data collection points could have improved the accuracy of the EBK estimation. Also, our study compared differences between weekday daytime and weekend nighttime but did not include weekday nighttime hours as a sampling window. The sampling time window was large, more so for weekday daytime hours (6-hour window) than weekend nighttime hours (4-hour window); future studies may benefit from narrowing the time window of data collection; although this may require data collection occurring over more calendar time depending on personnel. Additionally, this study used a multiapp approach for data collection, but there were other GSN apps that were not used for this study. At the outset of the study, we contacted 6 app companies to notify them of our intent to do the study and dropped 3 apps due to the fact that it violated their user agreements.

In respect of app users' privacy, we did not record any identifying information about GSN app users, making it impossible to compare demographic characteristics of users with ecological variables based on census data. This also prevented us from determining the extent to which GSN app-using MSM were using more than 1 app and whether the differences in spatial distribution were due to the temporal migration of the same users to different locations or if other app users in different locations were logging on at different times. Although the choice not to collect identifying data resulted in analytic limitations, we believe that this was an important tradeoff to make in respect of user privacy and app company

policies. Ethical issues surrounding the collection and use of GSN profile data for research are complex, rapidly evolving, and warrant increased attention. In the interim, we believe it is important for researchers to communicate with GSN app companies and take steps to prevent violation of user privacy.

Conclusion

The number of GSN app users within a 1-mile radius of the data collection points ranged from 0 to 36 during weekday daytime hours to 0 to 39 during weekend nighttime hours. In the multivariate analysis, the number of GSN app-using MSM was positively associated with business zoning and population density and negatively associated with median income and percent Hispanic residents. Methodology using geospatial data and area demographic data can lead to insights for tailored and targeted interventions to support better health outcomes in underserved populations.

Conflicts of Interest

None declared.

References

1. Centers for Disease Control and Prevention. HIV among gay and bisexual men URL: https://www.cdc.gov/hiv/group/msm/ [WebCite Cache ID 6t2NFhA6G]
2. Centers for Disease Control and Prevention. 2016. HIV surveillance report URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-report-2016-vol-28.pdf [WebCite Cache ID 6vMiMhqlN]
3. Wang K, Brown K, Shen S, Tucker J. Social network-based interventions to promote condom use: a systematic review. AIDS Behav 2011 Oct;15(7):1298-1308 [FREE Full text] [doi: 10.1007/s10461-011-0020-1] [Medline: 21811843]
4. De P, Cox J, Boivin J, Platt RW, Jolly AM. The importance of social networks in their association to drug equipment sharing among injection drug users: a review. Addiction 2007 Nov;102(11):1730-1739. [doi: 10.1111/j.1360-0443.2007.01936.x] [Medline: 17935581]
5. Bauermeister JA, Connochie D, Eaton L, Demers M, Stephenson R. Geospatial indicators of space and place: a review of multilevel studies of HIV prevention and care outcomes among young men who have sex with men in the United States. J Sex Res 2017;54(4-5):446-464. [doi: 10.1080/00224499.2016.1271862] [Medline: 28135857]
6. Grov C, Rendina HJ, Parsons JT. Comparing three cohorts of MSM sampled via sex parties, bars/clubs, and Craigslist.org: implications for researchers and providers. AIDS Educ Prev 2014 Aug;26(4):362-382 [FREE Full text] [doi: 10.1521/aeap.2014.26.4.362] [Medline: 25068182]
7. Kubicek K, Carpineto J, McDavitt B, Weiss G, Kipke MD. Use and perceptions of the internet for sexual information and partners: a study of young men who have sex with men. Arch Sex Behav 2011 Aug;40(4):803-816 [FREE Full text] [doi: 10.1007/s10508-010-9666-4] [Medline: 20809373]
8. Garofalo R, Herrick A, Mustanski BS, Donenberg GR. Tip of the iceberg: young men who have sex with men, the Internet, and HIV risk. Am J Public Health 2007 Jun;97(6):1113-1117. [doi: 10.2105/AJPH.2005.075630] [Medline: 17463378]
9. Liau A, Millett G, Marks G. Meta-analytic examination of online sex-seeking and sexual risk behavior among men who have sex with men. Sex Transm Dis 2006 Sep;33(9):576-584. [doi: 10.1097/01.olq.0000204710.35332.c5] [Medline: 16540884]
10. Phillips G, Magnus M, Kuo I, Rawls A, Peterson J, Jia Y, et al. Use of geosocial networking (GSN) mobile phone applications to find men for sex by men who have sex with men (MSM) in Washington, DC. AIDS Behav 2014 Sep;18(9):1630-1637. [doi: 10.1007/s10461-014-0760-9] [Medline: 24682866]
11. Chiasson MA, Hirshfield S, Remien RH, Humberstone M, Wong T, Wolitski RJ. A comparison of on-line and off-line sexual risk in men who have sex with men: an event-based on-line survey. J Acquir Immune Defic Syndr 2007 Feb 1;44(2):235-243. [doi: 10.1097/QAI.0b013e31802e298c] [Medline: 17179769]
12. 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):1. [doi: 10.4172/2155-6113.S4-005]
13. Benotsch EG, Kalichman S, Cage M. Men who have met sex partners via the Internet: prevalence, predictors, and implications for HIV prevention. Arch Sex Behav 2002 Apr;31(2):177-183. [Medline: 11974643]
14. Winetrobe H, Rice E, Bauermeister J, Petering R, Holloway IW. Associations of unprotected anal intercourse with Grindr-met partners among Grindr-using young men who have sex with men in Los Angeles. AIDS Care 2014;26(10):1303-1308. [doi: 10.1080/09540121.2014.911811] [Medline: 24754563]

XSL•FO
RenderX

15. Beymer MR, Weiss RE, Bolan RK, Rudy ET, Bourque LB, Rodriguez JP, et al. Sex on demand: geosocial networking phone apps and risk of sexually transmitted infections among a cross-sectional sample of men who have sex with men in Los Angeles County. Sex Transm Infect 2014 Nov;90(7):567-572 [FREE Full text] [doi: 10.1136/sextrans-2013-051494] [Medline: 24926041]

16. McFarlane M, Bull SS, Rietmeijer CA. The Internet as a newly emerging risk environment for sexually transmitted diseases. JAMA 2000 Jul 26;284(4):443-446. [Medline: 10904506]

17. Young SD, Szekeres G, Coates T. The relationship between online social networking and sexual risk behaviors among men who have sex with men (MSM). PLoS One 2013 May;8(5):e62271 [FREE Full text] [doi: 10.1371/journal.pone.0062271] [Medline: 23658716]

18. Lehmiller JJ, Ioerger M. Social networking smartphone applications and sexual health outcomes among men who have sex with men. PLoS One 2014;9(1):e86603 [FREE Full text] [doi: 10.1371/journal.pone.0086603] [Medline: 24466166]

19. Bull SS, McFarlane M, Rietmeijer C. HIV and sexually transmitted infection risk behaviors among men seeking sex with men on-line. Am J Public Health 2001 Jun;91(6):988-989. [Medline: 11392947]

20. Eaton LA, Maksut JL, Gamarel KE, Siembida EJ, Driffin DD, Baldwin R. Online sex partner meeting venues as a risk factor for testing HIV positive among a community-based sample of black men who have sex with men. Sex Transm Dis 2016 Jun;43(6):360-364. [doi: 10.1097/OLQ.0000000000000454] [Medline: 27200520]

21. Rhodes SD, DiClemente RJ, Cecil H, Hergenrather KC, Yee LJ. Risk among men who have sex with men in the United States: a comparison of an Internet sample and a conventional outreach sample. AIDS Educ Prev 2002 Feb;14(1):41-50. [Medline: 11900109]

22. Grosskopf NA, LeVasseur MT, Glaser DB. Use of the Internet and mobile-based apps for sex-seeking among men who have sex with men in New York City. Am J Mens Health 2014 Nov;8(6):510-520. [doi: 10.1177/1557988314527311] [Medline: 24658284]

23. Delaney KP, Kramer MR, Waller LA, Flanders WD, Sullivan PS. Using a geolocation social networking application to calculate the population density of sex-seeking gay men for research and prevention services. J Med Internet Res 2014;16(11):e249 [FREE Full text] [doi: 10.2196/jmir.3523] [Medline: 25406722]

24. Ragsdale A, Rotheram-Borus M. Re-shaping HIV interventions with technology. AIDS Behav 2015;19(2):77-80. [Medline: 26088905]

25. Grov C, Agyemang L, Ventuneac A, Breslow A. Navigating condom use and HIV status disclosure with partners met online: a qualitative pilot study with gay and bisexual men from Craigslist.org. AIDS Educ Prev 2013 Feb;25(1):72-85 [FREE Full text] [doi: 10.1521/aeap.2013.25.1.72] [Medline: 23387953]

26. Rendina HJ, Jimenez RH, Grov C, Ventuneac A, Parsons JT. Patterns of lifetime and recent HIV testing among men who have sex with men in New York City who use Grindr. AIDS Behav 2014 Jan;18(1):41-49 [FREE Full text] [doi: 10.1007/s10461-013-0573-2] [Medline: 23925515]

27. Kakietek J, Sullivan PS, Heffelfinger JD. You've got male: internet use, rural residence, and risky sex in men who have sex with men recruited in 12 U.S. cities. AIDS Educ Prev 2011 Apr;23(2):118-127 [FREE Full text] [doi: 10.1521/aeap.2011.23.2.118] [Medline: 21517661]

28. Rosser BRS, Oakes JM, Horvath KJ, Konstan JA, Danilenko GP, Peterson JL. HIV sexual risk behavior by men who use the Internet to seek sex with men: results of the Men's INTernet Sex Study-II (MINTS-II). AIDS Behav 2009 Jun;13(3):488-498 [FREE Full text] [doi: 10.1007/s10461-009-9524-3] [Medline: 19205866]

29. Preston D, D'augelli A, Cain R, Schulze F. Issues in the development of HIV-preventive interventions for men who have sex with men (MSM) in rural areas. J Primary Prev 2002;23(2):199-214. [doi: 10.1023/a:1019968315959]

30. Preston DB, D'Augelli AR, Kassab CD, Starks MT. The relationship of stigma to the sexual risk behavior of rural men who have sex with men. AIDS Educ Prev 2007 Jun;19(3):218-230. [doi: 10.1521/aeap.2007.19.3.218] [Medline: 17563276]

31. Keene DE, Eldahan AI, White HJM, Pachankis JE. 'The big ole gay express': sexual minority stigma, mobility and health in the small city. Cult Health Sex 2017 Mar;19(3):381-394 [FREE Full text] [doi: 10.1080/13691058.2016.1226386] [Medline: 27604293]

32. White HJM, Pachankis JE, Eldahan AI, Keene DE. You can't just walk down the street and meet someone: the intersection of social-sexual networking technology, stigma, and health among gay and bisexual men in the small city. Am J Mens Health 2017 May;11(3):726-736 [FREE Full text] [doi: 10.1177/1557988316679563] [Medline: 27885147]

33. Preston DB, D'Augelli AR, Kassab CD, Cain RE, Schulze FW, Starks MT. The influence of stigma on the sexual risk behavior of rural men who have sex with men. AIDS Educ Prev 2004 Aug;16(4):291-303. [doi: 10.1521/aeap.16.4.291.40401] [Medline: 15342332]

34. Li MJ, Hubach RD, Dodge B. Social milieu and mediators of loneliness among gay and bisexual men in rural Indiana. J Gay Lesbian Mental Health 2015 Apr 20;19(4):331-346. [doi: 10.1080/19359705.2015.1033798]

35. Dodge B, Reece M, Herbenick D, Fisher C, Satinsky S, Stupiansky N. Relations between sexually transmitted infection diagnosis and sexual compulsivity in a community-based sample of men who have sex with men. Sex Transm Infect 2008 Aug;84(4):324-327. [doi: 10.1136/sti.2007.028696] [Medline: 18096648]

36. Hubach R, Dodge B, Cola T, Battani P, Reece M. Assessing the sexual health needs of men who have sex with men (MSM) residing in rural and mixed rural areas. Health Education Monograph 2014;31(2):37-43 [FREE Full text]

XSL•FO
RenderX

37. Bowen A, Williams M, Horvath K. Using the internet to recruit rural MSM for HIV risk assessment: sampling issues. AIDS Behav 2004 Sep;8(3):311-319 [FREE Full text] [doi: 10.1023/B:AIBE.0000044078.43476.1f] [Medline: 15475678]

38. Williams ML, Bowen AM, Horvath KJ. The social/sexual environment of gay men residing in a rural frontier state: implications for the development of HIV prevention programs. J Rural Health 2005;21(1):48-55 [FREE Full text] [Medline: 15667009]

39. Horvath KJ, Bowen AM, Williams ML. Virtual and physical venues as contexts for HIV risk among rural men who have sex with men. Health Psychol 2006 Mar;25(2):237-242. [doi: 10.1037/0278-6133.25.2.237] [Medline: 16569116]

40. Race K. Party and play: online hook-up devices and the emergence of PNP practices among gay men. Sexualities 2015 May 05;18(3):253-275. [doi: 10.1177/1363460714550913]

41. Van de Wiele C, Tong S. Breaking boundaries: the uses & gratifications of grindr. 2014 Presented at: ACM International Joint Conference on Pervasive and Ubiquitous Computing; 2014; Seattle. [doi: 10.1145/2632048.2636070]

42. Goedel WC, Duncan DT. Geosocial-networking app usage patterns of gay, bisexual, and other men who have sex with men: survey among users of Grindr, a mobile dating app. JMIR Public Health Surveill 2015;1(1):e4 [FREE Full text] [doi: 10.2196/publichealth.4353] [Medline: 27227127]

43. US Census Bureau. 2010 US Census URL: https://www.census.gov/2010census/ [WebCite Cache ID 6t2N2jSyH]

44. Google Maps. URL: https://www.google.com/maps [accessed 2018-04-10] [WebCite Cache ID 6t2MsOp1h]

45. Fulcrum Mobile Location Leveragee. URL: http://www.fulcrumapp.com/ [accessed 2017-08-27] [WebCite Cache ID 6t2MT1RDp]

46. Badal HJ, Stryker JE, DeLuca N, Purcell DW. Swipe right: dating website and app use among men who have sex with men. AIDS Behav 2018 Apr;22(4):1265-1272. [doi: 10.1007/s10461-017-1882-7] [Medline: 28884248]

47. Whitfield DL, Kattari SK, Walls NE, Al-Tayyib A. Grindr, scruff, and on the hunt: predictors of condomless anal sex, Internet use, and mobile application use among men who have sex with men. Am J Mens Health 2017 May;11(3):775-784. [doi: 10.1177/1557988316687843] [Medline: 28134002]

48. Miller B. "Dude, where's your face?" Self-presentation, self-description, and partner preferences on a social networking application for men who have sex with men: a content analysis. Sexuality Culture 2015 Apr 10;19(4):637-658. [doi: 10.1007/s12119-015-9283-4]

49. Manley E, Levitt H, Mosher C. Understanding the bear movement in gay male culture: redefining masculinity. J Homosex 2007 Sep;53(4):89-112. [doi: 10.1080/00918360802103365] [Medline: 18689193]

50. Rodriguez NS, Huemmer J, Blumell LE. Mobile masculinities: an investigation of networked masculinities in gay dating apps. Masculinities Soc Change 2016 Oct 21;5(3):241-267. [doi: 10.17583/MCS.2016.2047]

51. US Census Bureau. American Community Survey 2011-2015 URL: https://www.census.gov/programs-surveys/acs/ [WebCite Cache ID 6t2M0UqlC]

52. Lexington-Fayette urban county government—zoning. URL: https://data.lexingtonky.gov/dataset/zoning [accessed 2017-08-27] [WebCite Cache ID 6t2Lht0oT]

53. TIGER/Line Shapefiles, county, Rowan & Fayette County, KY.: US Census Bureau; 2010. URL: https://www.census.gov/geo/maps-data/data/tiger-line.html [accessed 2017-08-27] [WebCite Cache ID 6t2MD4ODL]

54. SAS/STAT 9.2 User's Guide, Second Edition: Collinearity Diagnostics. Cary: SAS Institute Inc URL: https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_reg_sect038.htm [accessed 2018-04-10] [WebCite Cache ID 6t2LDCETf]

55. Holloway IW, Rice E, Gibbs J, Winetrobe H, Dunlap S, Rhoades H. Acceptability of smartphone application-based HIV prevention among young men who have sex with men. AIDS Behav 2014 Feb;18(2):285-296 [FREE Full text] [doi: 10.1007/s10461-013-0671-1] [Medline: 24292281]

56. Landovitz RJ, Tseng C, Weissman M, Haymer M, Mendenhall B, Rogers K, et al. Epidemiology, sexual risk behavior, and HIV prevention practices of men who have sex with men using GRINDR in Los Angeles, California. J Urban Health 2013 Aug;90(4):729-739 [FREE Full text] [doi: 10.1007/s11524-012-9766-7] [Medline: 22983721]

57. Holloway IW, Winder TJ, Lea CH, Tan D, Boyd D, Novak D. Technology use and preferences for mobile phone-based HIV prevention and treatment among black young men who have sex with men: exploratory research. JMIR Mhealth Uhealth 2017 Dec 13;5(4):e46 [FREE Full text] [doi: 10.2196/mhealth.6436] [Medline: 28408360]

Abbreviations

EBK: empirical Bayesian kriging
GSN: geosocial networking
MSM: men who have sex with men

XSL•FO
RenderX

Original Paper

Identifying National Availability of Abortion Care and Distance From Major US Cities: Systematic Online Search

Alice F Cartwright[1], MPH; Mihiri Karunaratne[2], BA; Jill Barr-Walker[3], MPH, MS; Nicole E Johns[1], MPH; Ushma D Upadhyay[1], MPH, PhD

[1]Advancing New Standards in Reproductive Health, Department of Obstetrics, Gynecology, and Reproductive Sciences, University of California, San Francisco, Oakland, CA, United States

[2]University of California, Berkeley, Berkeley, CA, United States

[3]Zuckerberg San Francisco General Library, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Ushma D Upadhyay, MPH, PhD
Advancing New Standards in Reproductive Health
Department of Obstetrics, Gynecology, and Reproductive Sciences
University of California, San Francisco
1330 Broadway, Suite 1100
Oakland, CA, 94612
United States
Phone: 1 510 986 8946
Email: ushma.upadhyay@ucsf.edu

Abstract

Background: Abortion is a common medical procedure, yet its availability has become more limited across the United States over the past decade. Women who do not know where to go for abortion care may use the internet to find abortion facility information, and there appears to be more online searches for abortion in states with more restrictive abortion laws. While previous studies have examined the distances women must travel to reach an abortion provider, to our knowledge no studies have used a systematic online search to document the geographic locations and services of abortion facilities.

Objective: The objective of our study was to describe abortion facilities and services available in the United States from the perspective of a potential patient searching online and to identify US cities where people must travel the farthest to obtain abortion care.

Methods: In early 2017, we conducted a systematic online search for abortion facilities in every state and the largest cities in each state. We recorded facility locations, types of abortion services available, and facility gestational limits. We then summarized the frequencies by region and state. If the online information was incomplete or unclear, we called the facility using a mystery shopper method, which simulates the perspective of patients calling for services. We also calculated distance to the closest abortion facility from all US cities with populations of 50,000 or more.

Results: We identified 780 facilities through our online search, with the fewest in the Midwest and South. Over 30% (236/780, 30.3%) of all facilities advertised the provision of medication abortion services only; this proportion was close to 40% in the Northeast (89/233, 38.2%) and West (104/262, 39.7%). The lowest gestational limit at which services were provided was 12 weeks in Wyoming; the highest was 28 weeks in New Mexico. People in 27 US cities must travel over 100 miles (160 km) to reach an abortion facility; the state with the largest number of such cities is Texas (n=10).

Conclusions: Online searches can provide detailed information about the location of abortion facilities and the types of services they provide. However, these facilities are not evenly distributed geographically, and many large US cities do not have an abortion facility. Long distances can push women to seek abortion in later gestations when care is even more limited.

KEYWORDS

XSL•FO

RenderX

Introduction

Women's ability to determine if and when they get pregnant and continue that pregnancy is key to their overall well-being. Women who are denied wanted abortions experience some negative outcomes compared with women who were able to obtain abortions, including increased economic insecurity [1] and continued exposure to violence from the man involved in the pregnancy [2]. While abortion rates have declined slightly in recent years, over 926,000 abortions were performed in the United States in 2014 [3]. This rate is equivalent to 1 in 4 women of reproductive age having an abortion within her lifetime [3], which underscores that abortion is common.

The explanations for the decline in abortion rates are varied, but part of this drop can likely be attributed to the decrease in facilities at which women can obtain abortion care across the United States over the past decade. Most abortions (95%) are performed in specialized abortion clinics (rather than private physicians' offices or hospitals), and the number of these clinics declined in half of US states from 2011 to 2014, with some regions experiencing up to a 22% decrease [3]. Because 90% of US counties do not have an abortion provider [3], many women seeking abortion must travel outside their home counties to obtain care. Other geographic disparities have been documented: women living in rural areas, the South and Midwest regions of the United States, and those seeking second-trimester or later abortions are more likely to travel farther for services, often 50 miles (80 km) or more one way [4-7]. These shifts in the availability of abortion-providing facilities indicate that women in underserved areas must travel increasingly far for abortion care.

Somedecline in the number of abortion facilities may be due to the more than 400 state laws regulating abortion that have been adopted between 2011 and 2017 [8], which, among other requirements, mandate that physicians have local hospital admitting privileges, facilities have formal transfer agreements with local hospitals, and facilities become ambulatory surgical centers. These laws have likely led to the closure of facilities that could not meet the financial or administrative requirements imposed by these laws. For example, after these types of laws were passed in Texas in 2013, the number of abortion facilities decreased by 54% over 15 months, requiring women whose nearest clinic had closed to travel 85 miles (137 km) one way to a facility [9]. Additional analyses of trends in abortion rates in Texas from 2012 to 2014 found a relationship between increases in distance to the nearest abortion facility and decreases in the county abortion rate [10]. Another analysis from Louisiana estimated that, if admitting privileges laws were to go into effect, 67% of women of reproductive age would live more than 150 miles (241 km) from the nearest abortion facility, thereby tripling the distance women have to travel to reach the nearest facility for care [11,12]. With distance come increased travel time, increased costs for transportation and childcare, lost wages, the need to take time off of work or school, the need to disclose the abortion to more people than desired, and overall delays in care [13-15]. Ultimately, delays in reaching and obtaining care can push women later into their pregnancies, even up to the point that they might not be able to obtain a

wanted abortion, depending on the gestational limits on abortion in their state [16].

To obtain abortion care in their communities, women who do not know where to go may use the internet to find abortion facility information [15]. Almost half (45%) of women seeking abortion services at clinics in Nebraska located the abortion clinic through an online search [17], and a recent study documented an interest in information on self-abortion among people searching online using the search engine Google [18]. Online searching for abortion information appears to be more prevalent in states with restrictive abortion laws and where abortion availability is limited, suggesting that women with reduced access to abortion are more likely to seek out information on abortion online [19,20].

We were interested in examining the question "What does the current landscape of abortion facilities look like to women searching online for abortion services?" There are no publicly available systematically documented and comprehensive lists of US abortion facilities, which makes it difficult to determine how far women must travel to obtain these services. Considering the trends in increased restrictions and decreasing numbers of abortion-providing facilities, it is important to generate accurate estimates of the distances women must travel to obtain abortion services in order to demonstrate potential impacts of closures. This study aimed to address this question by documenting the location of and abortion services available at abortion facilities identified through a systematic online search in all 50 US states (and the District of Columbia) and then calculating travel distances to these facilities from metropolitan areas with populations of 50,000 or more.

Methods

Data Collection

We conducted a systematic online search for abortion facilities using the Google (Google LLC), Bing (Microsoft Corporation), and Yahoo (Oath Inc) search engines between February 22, 2017 and May 22, 2017. Although Google alone accounts for a substantial portion of the market share in the United States (87.5%), together the 3 search engines comprised 99.1% of the total search engine market share as of February 2017 [21]. We conducted a search with the keywords "Abortion clinic in [state]" (no quotes) for all 50 states and the District of Columbia in each of the 3 search engines. In addition, we searched all cities (n=302) with populations over 100,000 based on 2015 US Census population estimates [22] using the keywords "Abortion clinic in [city]" (no quotes). For states that had fewer than 3 cities with populations over 100,000, we used the 3 most populous cities from the same US Census source. We conducted the keyword searches in Google's Chrome browser on Incognito mode and cleared the complete browsing history, including cookies and other site data and cached images and files, prior to each search. The researcher was logged into a Google account created specifically for this study during the searches. We chose keyword searches to reflect the natural language that women would use to search for local abortion services.

We assessed the first 20 results for each *city* for information on abortion-providing facilities, similar to previous analyses of search engine content [23,24], resulting in a review of a total of 18,120 city results across all 3 search engines. To capture the larger number of facilities expected in statewide searches, we reviewed the first 30 results for each state (for a total review of 4590 state results). For each result, we examined the website for relevant information. If the website belonged to an abortion facility, we included the result in our count of facilities and recorded whether they provided medication abortion, or aspiration or surgical abortion, as well as the facility gestational limit. Some facilities noted on their websites that they offered services beyond the gestational limit on a case-by-case basis; however, we recorded the limit that each facility offered to all patients seeking services. We included hospitals and clinics associated with universities and medical schools through the Ryan Residency Training Program in Abortion and Family Planning [25] in the analysis if they provided information about availability of abortion services on their website, even if they did not come up in our systematic Web searches. If a website did not provide information about a facility where abortion care could be obtained or explicitly stated that they did not provide abortion care, we excluded the facility. If the online information was incomplete or unclear, we called the facility using a mystery shopper method, which simulates the perspective of patients calling for services [26]. With these calls, we verified that it was not a crisis pregnancy center, confirmed that the facility was open and providing abortion services, and obtained additional information on its address, including state and zip code, types of abortion services provided, and gestational limits. Finally, because Planned Parenthood is the health care provider most widely known to provide abortion services in the United States, we reviewed all facilities listed by state on the Planned Parenthood Federation of America website as a validity check against the results from our systematic search. We confirmed that all Planned Parenthood facilities providing abortion had been captured by our searches. The study was approved by the institutional review board of the University of California, San Francisco.

Data Analysis

We described the number of facilities and the proportion of facilities that offered medication abortion only, aspiration or surgical abortion only, and both medication and aspiration or surgical abortion in each state and region. We grouped states by region and subregion based on US Census categories. The latest gestational limit at which facilities offered aspiration or surgical abortions was documented for each state. Using 2015 population estimates taken from 2010 US Census data [22], we determined the number of women of reproductive age (15-49 years) per abortion facility in each state.

To calculate the cities farthest from an abortion-providing facility, we defined cities based on the US Census's data on incorporated places of 50,000 or more [22], which amounted to 758 cities. After removing those cities that had at least one abortion provider, we calculated the distance from each city to all the abortion facilities within the state and in any neighboring or nearby states. For each city, we then took the minimum of

these distances to determine the closest provider. We calculated distances in Stata 14 (StataCorp LLC) using the traveltime3 command, which uses a Google Maps application programming interface to calculate driving distances in miles and time. Mapping was performed in Redivis, a Stanford University-based online visualization platform. Rather than using Euclidean (straight-line) distance, Redivis uses road network information, including road type and corresponding average speed, sourced from OpenStreetMap [27] to implement a cost-distance algorithm to predict distance-access to abortion facilities.

Results

Distribution and Characteristics of Abortion Facilities in the United States

We identified 780 abortion facilities in the United States. The distribution of abortion facilities was not uniform across states. The largest numbers of facilities were in the Northeast and the West. California had the highest number of facilities (n=152), while Kentucky, Mississippi, Missouri, North Dakota, South Dakota, and West Virginia had 1 facility each. The geographic region with the highest ratio of women of reproductive age to facility was the Midwest, with 165,886 women per abortion facility (Table 1). The Northeast had the lowest ratio (55,662:1). While population density is not distributed evenly across all regions, the subregions with the highest ratios were the West South Central and East South Central subregions, with 298,733 and 288,463 women per facility, respectively. The state with the highest ratio of women to facility was Missouri, with 1,365,575 women per facility, and the lowest was in Maine, with 13,905 women per facility.

Most facilities reported providing both medication and aspiration or surgical abortion, although the proportion of facilities that provided different types of abortion care also varied by region and state. Over 30% (236/780, 30.3%) of all the facilities reported on their websites that they only provided medication abortion (Table 2), while 65.4% (510/780) provided both medication and aspiration or surgical abortion. Very few offered just aspiration or surgical abortion.

The South region had the highest proportion of facilities offering both medication abortion and aspiration or surgical abortion (169/193, 87.6%). While the Northeast and West had many more facilities overall, almost 40% of facilities in each of these regions offered medication abortion only (89/233, 38.2%; and 104/262, 39.7%, respectively). When looking at subregion, New England (34/74, 45.9%) and Pacific (87/207, 42.0%) had even greater proportions of facilities offering medication abortion only.

The highest gestational limit advertised by facilities also varied by state, subregion, and region (Table 2). The states with the lowest advertised gestation for abortions were Wyoming (12 weeks) and Indiana and South Dakota (both 13 weeks and 6 days), and the lowest subregions were West North Central in the Midwest and East South Central in the South, with limits of 22 weeks. Among all facilities, 50.9% (397/780) provided abortion services at 14 weeks or later and 26.5% (207/780) provided services at 20 weeks or later.

XSL•FO

RenderX

Table 1. Number of US abortion facilities by region and state, May 2017.

Region and state	Number of facilities	Population of women of reproductive age (ages 15-49 years) per facility
United States	780	95,033
Northeast	233	55,662
New England	74	45,655
Connecticut	20	40,632
Maine	20	13,905
Massachusetts	19	84,973
New Hampshire	6	48,395
Rhode Island	3	82,130
Vermont	6	22,743
Middle Atlantic	159	60,320
New Jersey	50	40,927
New York	92	51,293
Pennsylvania	17	166,210
Midwest	92	165,886
East North Central	68	155,508
Indiana	6	250,997
Illinois	25	120,135
Michigan	23	96,054
Ohio	11	235,016
Wisconsin	3	423,591
West North Central	24	195,289
Iowa	9	75,629
Kansas	4	161,029
Minnesota	5	245,486
Missouri	1	1,365,575
Nebraska	3	140,140
North Dakota	1	167,601
South Dakota	1	181,145
South	193	145,645
South Atlantic	147	98,787
Delaware	3	70,851
District of Columbia	3	66,863
Florida	65	67,757
Georgia	17	145,646
Maryland	25	56,665
North Carolina	15	155,709
South Carolina	3	370,733
Virginia	15	131,439
West Virginia	1	392,351
East South Central	15	288,463
Alabama	5	223,458

Region and state	Number of facilities	Population of women of reproductive age (ages 15-49 years) per facility
Kentucky	1	996,488
Mississippi	1	694,045
Tennessee	8	189,891
West South Central	31	298,733
Arkansas	3	222,577
Louisiana	3	363,228
Oklahoma	4	220,527
Texas	21	315,296
West	262	67,883
Mountain	55	97,547
Arizona	8	190,750
Colorado	21	60,902
Idaho	4	91,376
Montana	5	43,161
Nevada	8	83,522
New Mexico	5	91,243
Utah	2	363,970
Wyoming	2	63,250
Pacific	207	60,002
Alaska	6	27,969
California	152	61,740
Hawaii	3	103,715
Oregon	12	75,968
Washington	34	48,391

XSL•FO
RenderX

Table 2. Types of services offered by abortion care facilities (N=780) by US region and state, May 2017.

Region and state	Facilities offering only aspiration or surgical abortion, n (%)	Facilities offering only medication abortion, n (%)	Facilities offering both aspiration or surgical abortion and medication abortion, n (%)	Latest gestational limit (weeks since LMP[a]) as listed on website
United States	34 (4.4)	236 (30.3)	510 (65.4)	28
Northeast	18 (7.7)	89 (38.2)	126 (54.1)	27
New England	4 (5.4)	34 (45.9)	36 (48.6)	27
Connecticut	0 (0)	13 (65.0)	7 (35.0)	<24
Maine	0 (0)	17 (85.0)	3 (15.0)	<19
Massachusetts	4 (21.1)	0 (0)	15 (78.9)	27
New Hampshire	0 (0)	1 (16.7)	5 (83.3)	<16
Rhode Island	0 (0)	0 (0)	3 (100)	<22
Vermont	0 (0)	3 (50.0)	3 (50.0)	<19
Middle Atlantic	14 (8.8)	55 (34.6)	90 (56.6)	<25
New Jersey	6 (12.0)	22 (44.0)	22 (44.0)	<25
New York	8 (8.7)	29 (31.5)	55 (59.8)	24
Pennsylvania	0 (0)	4 (23.5)	13 (76.5)	23
Midwest	2 (2.2)	23 (25.0)	67 (72.8)	24
East North Central	1 (1.5)	16 (23.5)	51 (75.0)	24
Indiana	0 (0)	1 (16.7)	5 (83.3)	<14
Illinois	0 (0)	9 (36.0)	16 (64.0)	24
Michigan	0 (0)	5 (21.7)	18 (78.3)	24
Ohio	1 (9.1)	1 (9.1)	9 (81.8)	<22
Wisconsin	0 (0)	0 (0)	3 (100)	<23
West North Central	1 (4.2)	7 (29.2)	16 (66.7)	22
Iowa	0 (0)	6 (66.7)	3 (33.3)	<22
Kansas	0 (0)	1 (25.0)	3 (75.0)	<22
Minnesota	1 (20.0)	0 (0)	4 (80.0)	22
Missouri	0 (0)	0 (0)	1 (100)	<22
Nebraska	0 (0)	0 (0)	3 (100)	<22[b]
North Dakota	0 (0)	0 (0)	1 (100)	16
South Dakota	0 (0)	0 (0)	1 (100)	<14
South	4 (2.1)	20 (10.4)	169 (87.6)	26
South Atlantic	2 (1.4)	15 (10.2)	130 (88.4)	26
Delaware	0 (0)	1 (33.3)	2 (66.7)	<16
District of Columbia	1 (33.3)	0 (0)	2 (66.7)	26
Florida	1 (1.5)	5 (7.7)	59 (90.8)	24
Georgia	0 (0)	4 (23.5)	13 (76.5)	24[c]
Maryland	0 (0)	4 (16.0)	21 (84.0)	26
North Carolina	0 (0)	0 (0)	15 (100)	<21[b]
South Carolina	0 (0)	0 (0)	3 (100)	14
Virginia	0 (0)	1 (6.7)	14 (93.3)	21
West Virginia	0 (0)	0 (0)	1 (100)	16

Region and state	Facilities offering only aspiration or surgical abortion, n (%)	Facilities offering only medication abortion, n (%)	Facilities offering both aspiration or surgical abortion and medication abortion, n (%)	Latest gestational limit (weeks since LMP[a]) as listed on website
East South Central	1 (6.7)	2 (13.3)	12 (80.0)	22[b,c]
Alabama	1 (20.0)	0 (0)	4 (80.0)	21
Kentucky	0 (0)	0 (0)	1 (100)	22[b,c]
Mississippi	0 (0)	0 (0)	1 (100)	16
Tennessee	0 (0)	2 (25.0)	6 (75.0)	<18
West South Central	1 (3.2)	3 (9.7)	27 (87.1)	<24[d]
Arkansas	0 (0)	2 (66.7)	1 (33.3)	21
Louisiana	0 (0)	0 (0)	3 (100)	<19[b,c]
Oklahoma	0 (0)	1 (25.0)	3 (75.0)	<22[b]
Texas	1 (4.8)	0 (0)	20 (95.2)	<24[d]
West	10 (3.8)	104 (39.7)	148 (56.5)	28
Mountain	4 (7.3)	17 (30.9)	34 (61.8)	28
Arizona	0 (0)	1 (12.5)	7 (87.5)	24
Colorado	2 (9.5)	8 (38.1)	11 (52.4)	26
Idaho	0 (0)	2 (50.0)	2 (50.0)	<16
Montana	0 (0)	2 (40.0)	3 (60.0)	21
Nevada	2 (25.0)	2 (25.0)	4 (50.0)	24
New Mexico	0 (0)	1 (20.0)	4 (80.0)	28
Utah	0 (0)	0 (0)	2 (100)	<22
Wyoming	0 (0)	1 (50.0)	1 (50.0)	12[b]
Pacific	6 (2.9)	87 (42.0)	114 (55.1)	26
Alaska	0 (0)	1 (16.7)	5 (83.3)	<14
California	5 (3.3)	67 (44.1)	80 (52.6)	24
Hawaii	0 (0)	0 (0)	3 (100)	<24[b]
Oregon	0 (0)	4 (33.3)	8 (66.7)	23
Washington	1 (2.9)	15 (44.1)	18 (52.9)	26

[a]LMP: last menstrual period.

[b]Information on highest gestational limit obtained through phone call to facility.

[c]Information obtained in early 2017, and as of September 22, 2017, may no longer be accurate as a result of state laws or staffing changes.

[d]One facility in Texas listed the gestational limit on their website as <24 weeks; this was likely an error, as at the time of data collection, Texas had restrictions on abortion services after 22 weeks.

Abortion Deserts

We identified 27 "abortion deserts," cities from which people would have to travel over 100 miles (160 km) to reach an abortion facility (Table 3). People living in Rapid City, SD had to travel the farthest, 318 miles (512 km), to reach an abortion facility. Although the most cities in any one state (n=10) were located in Texas, there was a wide geographic diversity, with 15 unique states represented. These states were overwhelmingly in the South and Midwest. Figure 1 shows the geographic distribution of these distances, where large areas of the Midwest and Southwest had no abortion facility.

Table 3. Abortion deserts (cities >100 miles/160 km to closest facility) in the United States, May 2017.

	City and state	2015 population, n	Distance to closest facility, miles (km)	Location of closest facility
1	Rapid City, SD	73,569	318 (512)	Billings, MT
2	Lubbock, TX	249,042	308 (496)	Fort Worth, TX
3	Midland, TX	132,950	293 (472)	Fort Worth, TX
4	Odessa, TX	118,968	282 (454)	El Paso, TX
5	Amarillo, TX	198,645	258 (415)	Warr Acres, OK
6	Casper, WY	60,285	223 (359)	Fort Collins, CO
7	San Angelo, TX	100,450	204 (328)	Austin, TX
8	Bismarck, ND	71,167	196 (315)	Fargo, ND
9	Laredo, TX	255,473	160 (257)	San Antonio, TX
10	Lake Havasu City, AZ	53,553	144 (232)	Henderson, NV
11	Abilene, TX	121,721	144 (232)	Fort Worth, TX
12	Corpus Christi, TX	324,074	139 (224	San Antonio, TX
13	La Crosse, WI	52,306	137 (220)	Madison, WI
14	Springfield, MO	166,810	136 (219)	Fayetteville, AR
15	Lake Charles, LA	76,070	132 (212)	Baton Rouge, LA
16	Wichita Falls, TX	104,710	123 (198)	Fort Worth, TX
17	Columbia, MO	119,108	122 (196)	St Louis, MO
18	Evansville, IN	119,943	122 (196)	Bloomington, IN
19	St George, UT	80,202	121 (195)	Las Vegas, NV
20	Manhattan, KS	56,308	120 (193)	Overland Park, KS
21	Green Bay, WI	105,207	119 (192)	Milwaukee, WI
22	Pocatello, ID	54,441	114 (183)	Twin Falls, ID
23	Victoria, TX	67,574	107 (172)	San Antonio, TX
24	Fort Wayne, IN	260,326	106 (171)	Toledo, OH
25	Owensboro, KY	59,042	105 (169)	Louisville, KY
26	Dothan, AL	68,567	103 (166)	Montgomery, AL
27	Chattanooga, TN	176,588	101 (163)	Marietta, GA

XSL•FO
RenderX

Figure 1. Distance to nearest abortion facility in the contiguous United States, May 2017.
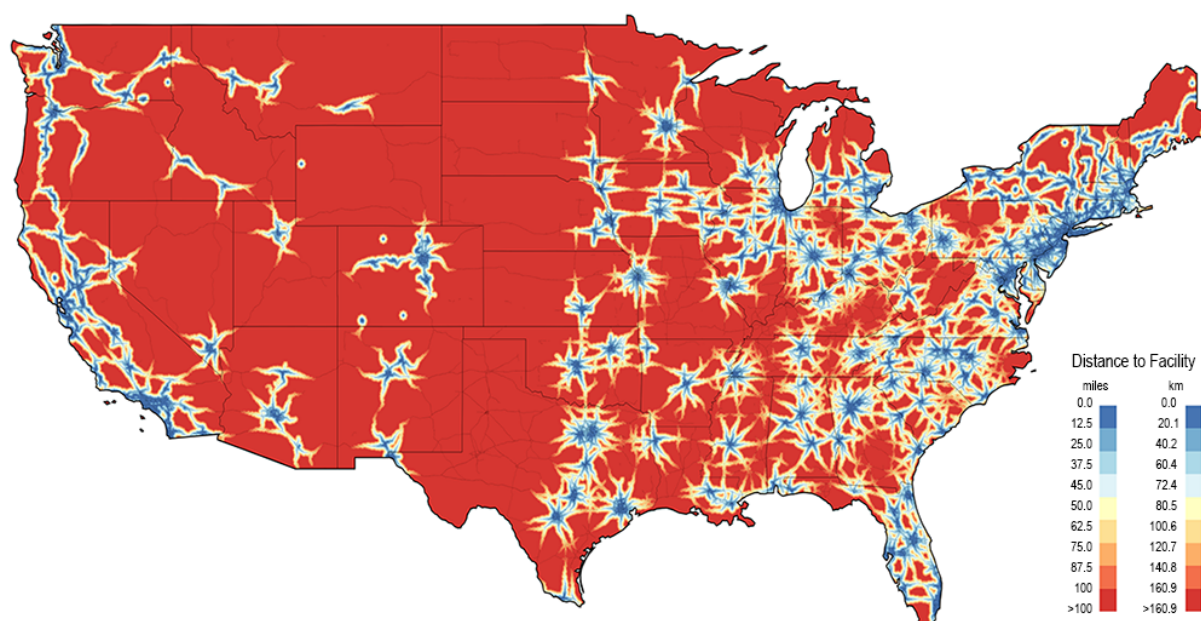
Discussion

Principal Results

Using an online search method, we identified almost 800 abortion facilities in the United States, which is consistent with other estimates of abortion clinics and nonspecialized clinics providing abortion [3]. These facilities were not distributed proportionately by state population. Through our analysis, we also found that 27 US cities, largely in the Midwest and the South, could be characterized as abortion deserts, as they did not have a publicly advertised abortion facility within 100 miles (160 km). These findings are consistent with those published by Bearak and colleagues [6], who found that the US counties where women would have to travel the farthest to reach the nearest abortion clinic were concentrated in the middle of the country, as well as several metropolitan areas in Texas. The lack of access to a common reproductive health service such as abortion is a public health concern in that more women in these cities could be forced to carry unwanted pregnancies to term if they are unable to travel long distances to obtain abortion care.

As states continue to pass, implement, and defend restrictions on abortion [8], it is possible that the number of abortion facilities will continue to decrease in those states with the most restrictions. The 6 states that have only 1 abortion facility have combined populations of almost 4 million women of reproductive age who will be forced to travel out of their home state to access abortion care if those facilities close.

For people seeking abortion services in the cities characterized as abortion deserts and in states with few facilities, reaching a facility for care could be incredibly challenging. Access to transportation is a barrier for people seeking all types of health care, in both urban and rural settings [28]. Lower-income women who are unable to access a car or money for gas may have to travel by bus, train, or other forms of transportation, which also becomes more difficult the farther they have to travel. Delays

in care due to distance or transportation can push women seeking abortion to later gestations [16,29,30] and are likely to disproportionately affect low-income women, who may struggle to cover the cost of transport [11,14]. Delays to abortion care may be particularly crucial to women in Wyoming, Alaska, Indiana, South Dakota, and South Carolina, where the abortion facilities had the lowest gestational limits. We found that 26.5% of identified facilities performed abortions at 20 weeks or later, which is lower than estimates from 2011-2012 [31], perhaps due to an increased number of state restrictions on abortion after 20 weeks since those estimates were published.

It seems likely that the larger number of facilities in the Northeast and West can be attributed to the fact that 40% to 50% of identified facilities in those regions are offering medication abortion only. The high proportion of facilities offering only medication abortion reflects the opportunities provided by medication abortion: the skills required for clinicians to provide it are minimal (compared with aspiration or surgical abortion) and the large majority of abortions in the United States (80.5%) occur at or before 10 weeks' gestation (the current accepted limit by which medication abortion can be provided) [32]. While the proportion of women choosing this method of abortion now accounts for 31% of nonhospital abortions (compared with 6% in 2001) [3], it is difficult to determine what the true demand would be if both medication and aspiration abortion were equally available. However, in states such as California, where fewer barriers to access exist for both types of abortion, medication abortion is now up to 46% of abortions in some populations, such as Medicaid recipients [33]. Additionally, states in the Northeast and West are less likely to have laws that limit the provision of medication abortion to physicians [34] and more likely to have policies that allow nurse practitioners, certified nurse midwives, and physicians assistants to offer medication abortion as part of their scope of practice.

These findings underscore the opportunities to pursue geographic expansion and other innovative models to achieve more equitable access to abortion care [35]. Some states have already managed this through an expansion of medication abortion-only services and increased use of telemedicine, which has been demonstrated to be safe and acceptable to women and to decrease travel for patients [36]. Indeed, in this analysis, Maine had the lowest ratio of women of reproductive age per facility, which was likely the result of an expansion of medication abortion through telemedicine programs offered from the existing Maine facilities [37]. While 19 states (almost exclusively in the South and Midwest) effectively prohibit telemedicine medication abortion [34], a recent Iowa Supreme Court decision could have implications for other states challenging similar restrictions that would allow expansion of medication abortion provision [38]. Given that it is less resource intensive, existing health care providers in the Midwest and the South, particularly in states where there is only 1 abortion provider or those states that contain cities classified as abortion deserts, could consider filling gaps in access by offering medication abortion alone as an entry point into abortion care, especially for primary care providers. Expanding the types of providers who can offer aspiration and medication abortion, such as nurse practitioners, would also increase the number of providers in smaller urban areas, thus expanding access to care [39]. However, it is important to note that in some states in which half or more facilities are only providing medication abortion, such as Idaho and Wyoming, the other facilities in the state offer abortion care up to 16 weeks and 12 weeks, respectively. Simply increasing the availability of medication abortion would not meet the needs of all women seeking abortion, some of whom may prefer aspiration abortion or need later abortion.

Supportive policy related to transportation for reproductive health services could also help reduce the burden on women in abortion deserts who have to travel extended distances to access services. California has recently introduced a Medicaid benefit to provide transportation for reproductive health care services (including abortion) to enrollees [40]. Further research is needed to determine what other policies can be enacted to reduce burdens of transportation and distance.

The internet will likely continue to be a key place for people to obtain the locations of abortion-providing facilities. However, both reduced geographic access and a desire to have more privacy and autonomy around the abortion process may lead women to seek out information on self-abortion [18] and obtain medication abortion pills through online sources, many of which have recently been shown to be selling effective medications with delivery to US-based mailing addresses [41]. There are no accurate estimates of how many women are obtaining abortion pills online, but the existence of online marketplaces and the documented feasibility of ordering from them implies online purchasing is occurring at volume.

Strengths and Limitations

This study is unique in that it systematically documented what people searching for abortion services online would find in search engine listings in early 2017 from a patient-centered perspective. A strength of this study is that it used up-to-date information on facilities of any volume to calculate distances, while other recent studies have been limited to a 2014 iteration of a proprietary database of only high-volume abortion-providing facilities (>400 abortions per year) that the Guttmacher Institute maintains [3,6]. An additional strength of the study is that it did not include abortion providers that offer abortion only to their existing patients or those that do not advertise their services, which would distort an accurate portrayal of the visibility of abortion availability. In addition, this analysis included the maximum gestations at which abortions were provided in each state, regardless of the state laws.

However, this study also has limitations. We used search terms that would enable us to locate abortion facilities in specific cities and states, but someone seeking abortion care might search "abortion clinic near me" and their results could vary from ours based on the location they are searching from. We attempted to eliminate geolocation bias by searching in Incognito mode and clearing both cache and cookies after each search. The information provided here is limited to what women seeking services would encounter—information that facilities chose to make available on their websites and provided through mystery shopper calls. Website information may be inaccurate or updated infrequently. It is possible that, if a woman called a facility to describe her unique situation, the staff could provide her with information about additional services that they do not wish to list on their website.

Distance is not the only barrier that people may face in trying to access abortion services—they may also face abortion stigma, waiting periods [13], and state gestational limits [16,42] as a result of state-level restrictions. In addition, the abortion facility that is closest to where a woman lives may not meet all her needs, particularly if it only provides medication abortion, has low limits on the gestation at which it provides abortion care, or cannot serve women with health conditions who may have higher risks.

Conclusions

Online searches provide information about abortion facilities and their services. The locations of these providers are not distributed equitably geographically across the United States. Having to travel long distances for abortion care can disproportionately affect low-income women and potentially push women to seek abortion at later gestations when care is even less available. Travel burdens may exist in addition to other restrictions on abortion in the state, including waiting periods and gestational limits, which can exacerbate inequities in the ability to access abortion care as part of the full range of reproductive health services.

Authors' Contributions

AFC and UDU led the conceptual design of the study, analyzed and interpreted the data, and drafted the manuscript. MK and AFC led the collection of the data. JBW and NEJ contributed to the analysis and interpretation of the data and reviewed the manuscript for important intellectual content. All authors critically reviewed the paper drafts, gave final approval for the published draft, and are accountable for all aspects of the work.

Conflicts of Interest

None declared.

References

1. Foster DG, Biggs MA, Ralph L, Gerdts C, Roberts S, Glymour MM. Socioeconomic outcomes of women who receive and women who are denied wanted abortions in the United States. Am J Public Health 2018 Mar;108(3):407-413. [doi: 10.2105/AJPH.2017.304247] [Medline: 29345993]

2. Roberts SC, Biggs MA, Chibber KS, Gould H, Rocca CH, Foster DG. Risk of violence from the man involved in the pregnancy after receiving or being denied an abortion. BMC Med 2014 Sep 29;12:144 [FREE Full text] [doi: 10.1186/s12916-014-0144-z] [Medline: 25262880]

3. Jones RK, Jerman J. Abortion incidence and service availability in the United States, 2014. Perspect Sex Reprod Health 2017 Mar;49(1):17-27 [FREE Full text] [doi: 10.1363/psrh.12015] [Medline: 28094905]

4. Johns NE, Foster DG, Upadhyay UD. Distance traveled for Medicaid-covered abortion care in California. BMC Health Serv Res 2017 Apr 19;17(1):287 [FREE Full text] [doi: 10.1186/s12913-017-2241-0] [Medline: 28420438]

5. Jones RK, Jerman J. How far did US women travel for abortion services in 2008? J Womens Health (Larchmt) 2013 Aug;22(8):706-713. [doi: 10.1089/jwh.2013.4283] [Medline: 23863075]

6. Bearak JM, Burke KL, Jones RK. Disparities and change over time in distance women would need to travel to have an abortion in the USA: a spatial analysis. Lancet Public Health 2017 Nov;2(11):e493-e500. [doi: 10.1016/S2468-2667(17)30158-5]

7. Jones RK, Jerman J. Characteristics and circumstances of U.S. women who obtain very early and second-trimester abortions. PLoS One 2017;12(1):e0169969 [FREE Full text] [doi: 10.1371/journal.pone.0169969] [Medline: 28121999]

8. Nash E, Gold RB, Mohammed L, Ansari-Thomas Z, Cappello O. Policy trends in the States, 2017. New York, NY: Guttmacher Institute; 2018. URL: https://www.guttmacher.org/article/2018/01/policy-trends-states-2017 [accessed 2018-04-30] [WebCite Cache ID 6xHYDHerL]

9. Gerdts C, Fuentes L, Grossman D, White K, Keefe-Oates B, Baum SE, et al. Impact of clinic closures on women obtaining abortion services after implementation of a restrictive law in Texas. Am J Public Health 2016 May;106(5):857-864. [doi: 10.2105/AJPH.2016.303134] [Medline: 26985603]

10. Grossman D, White K, Hopkins K, Potter JE. Change in distance to nearest facility and abortion in Texas, 2012 to 2014. JAMA 2017 Dec 24;317(4):437-439. [doi: 10.1001/jama.2016.17026] [Medline: 28114666]

11. Roberts SCM, Fuentes L, Kriz R, Williams V, Upadhyay UD. Implications for women of Louisiana's law requiring abortion providers to have hospital admitting privileges. Contraception 2015 May;91(5):368-372. [doi: 10.1016/j.contraception.2015.02.001] [Medline: 25744615]

12. Roberts SC, Fuentes L, Kriz R, Williams V, Upadhyay UD. Corrigendum to "Implications for women of Louisiana's law requiring abortion providers to have hospital admitting privileges" [Contraception (2015) 91:368-72. doi: 10.1016/j.contraception.2015.02.001]. Contraception 2017 Feb;95(2):221-222. [doi: 10.1016/j.contraception.2016.08.015] [Medline: 27647624]

13. Roberts SCM, Turok DK, Belusa E, Combellick S, Upadhyay UD. Utah's 72-hour waiting period for abortion: experiences among a clinic-based sample of women. Perspect Sex Reprod Health 2016 Dec;48(4):179-187. [doi: 10.1363/48e8216] [Medline: 27010515]

14. Baum SE, White K, Hopkins K, Potter JE, Grossman D. Women's experience obtaining abortion care in Texas after implementation of restrictive abortion laws: a qualitative study. PLoS One 2016;11(10):e0165048 [FREE Full text] [doi: 10.1371/journal.pone.0165048] [Medline: 27783708]

15. Jerman J, Frohwirth L, Kavanaugh ML, Blades N. Barriers to abortion care and their consequences for patients traveling for services: qualitative findings from two states. Perspect Sex Reprod Health 2017 Jun;49(2):95-102. [doi: 10.1363/psrh.12024] [Medline: 28394463]

XSL•FO

RenderX

16. Upadhyay UD, Weitz TA, Jones RK, Barar RE, Foster DG. Denial of abortion because of provider gestational age limits in the United States. Am J Public Health 2014 Sep;104(9):1687-1694. [doi: 10.2105/AJPH.2013.301378] [Medline: 23948000]

17. French V, Anthony R, Souder C, Geistkemper C, Drey E, Steinauer J. Influence of clinician referral on Nebraska women's decision-to-abortion time. Contraception 2016 Mar;93(3):236-243. [doi: 10.1016/j.contraception.2015.10.005] [Medline: 26593095]

18. Jerman J, Onda T, Jones RK. What are people looking for when they google "self-abortion"? Contraception 2018 Feb 22 [FREE Full text] [doi: 10.1016/j.contraception.2018.02.006] [Medline: 29477631]

19. Reis BY, Brownstein JS. Measuring the impact of health policies using internet search patterns: the case of abortion. BMC Public Health 2010;10:514 [FREE Full text] [doi: 10.1186/1471-2458-10-514] [Medline: 20738850]

20. Stephens-Davidowitz S. The New York Times. 2016 Mar 05. The return of the D.I.Y. abortion URL: https://www.nytimes.com/2016/03/06/opinion/sunday/the-return-of-the-diy-abortion.html [accessed 2018-04-26] [WebCite Cache ID 6yyOKd5U9]

21. StatCounter. Search engine market share in United States of America. 2017. URL: http://gs.statcounter.com/search-engine-market-share/all/united-states-of-america [WebCite Cache ID 6vsle7oAp]

22. U.S. Census Bureau Population Division. American Factfinder. 2016. Annual estimates of the resident population for incorporated places of 50,000 or more, ranked by July 1, 2015 population: April 1, 2010 to July 1, 2015 URL: https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk [accessed 2018-04-26] [WebCite Cache ID 6yyOXw6n4]

23. Robins S, Barr HJ, Idelson R, Lambert S, Zelkowitz P. Online health information regarding male infertility: an evaluation of readability, suitability, and quality. Interact J Med Res 2016 Oct 21;5(4):e25 [FREE Full text] [doi: 10.2196/ijmr.6440] [Medline: 27769954]

24. Fast AM, Deibert CM, Hruby GW, Glassberg KI. Evaluating the quality of internet health resources in pediatric urology. J Pediatr Urol 2013 Apr;9(2):151-156. [doi: 10.1016/j.jpurol.2012.01.004] [Medline: 22281281]

25. The Kenneth J. Ryan Residency Training Program in Abortion and Family Planning. Maps and location of Ryan program sites. 2017. URL: https://www.ryanprogram.org/map-and-locations [accessed 2017-12-21] [WebCite Cache ID 6vsm9IDOW]

26. Jacob S, Schiffino N, Biard B. The mystery shopper: a tool to measure public service delivery? Int Rev Adm Sci 2016 Mar 14;84(1):164-184. [doi: 10.1177/0020852315618018]

27. OpenStreetMap. 2017. URL: https://www.openstreetmap.org/ [accessed 2017-12-20] [WebCite Cache ID 6vq4PNy2b]

28. Syed ST, Gerber BS, Sharp LK. Traveling towards disease: transportation barriers to health care access. J Community Health 2013 Oct;38(5):976-993 [FREE Full text] [doi: 10.1007/s10900-013-9681-1] [Medline: 23543372]

29. Grossman D, Baum S, Fuentes L, White K, Hopkins K, Stevenson A, et al. Change in abortion services after implementation of a restrictive law in Texas. Contraception 2014 Nov;90(5):496-501 [FREE Full text] [doi: 10.1016/j.contraception.2014.07.006] [Medline: 25128413]

30. White K, deMartelly V, Grossman D, Turan JM. Experiences accessing abortion care in Alabama among women traveling for services. Womens Health Issues 2016;26(3):298-304. [doi: 10.1016/j.whi.2016.01.003] [Medline: 26897655]

31. Jerman J, Jones RK. Secondary measures of access to abortion services in the United States, 2011 and 2012: gestational age limits, cost, and harassment. Womens Health Issues 2014;24(4):e419-e424 [FREE Full text] [doi: 10.1016/j.whi.2014.05.002] [Medline: 24981401]

32. Jatlaoui TC, Ewing A, Mandel MG, Simmons KB, Suchdev DB, Jamieson DJ, et al. Abortion surveillance - United States, 2013. MMWR Surveill Summ 2016 Nov 25;65(12):1-44. [doi: 10.15585/mmwr.ss6512a1] [Medline: 27880751]

33. Research and Analytics Studies Division, California Department of Health Care Services. Medi-Cal-funded induced abortions, 2014. Sacramento, CA: California Department of Health Care Services; 2016. URL: http://www.dhcs.ca.gov/dataandstats/statistics/Documents/Medi-Cal_Funded_Abortions_2014.pdf, [WebCite Cache ID 6wipOASAK]

34. Guttmacher Institute. State policies in brief: medication abortion. New York, NY: Guttmacher Institute; 2017 Dec. URL: http://www.guttmacher.org/statecenter/spibs/spib_MA.pdf [accessed 2017-12-21] [WebCite Cache ID 6vsqpsBWF]

35. Upadhyay UD. Innovative models are needed for equitable abortion access in the USA. Lancet Public Health 2017 Nov;2(11):e484-e485. [doi: 10.1016/S2468-2667(17)30181-0]

36. Grossman D, Grindlay K. Safety of medical abortion provided through telemedicine compared with in person. Obstet Gynecol 2017 Oct;130(4):778-782. [doi: 10.1097/AOG.0000000000002212] [Medline: 28885427]

37. Farwell J. Bangor Daily News. 2016 Feb 29. Maine Family Planning to offer telemedicine abortions URL: https://bangordailynews.com/2016/02/29/health/maine-family-planning-to-offer-telemedicine-abortions/

38. Yang YT, Kozhimannil KB. Medication abortion through telemedicine: implications of a ruling by the Iowa Supreme Court. Obstet Gynecol 2016 Feb;127(2):313-316 [FREE Full text] [doi: 10.1097/AOG.0000000000001251] [Medline: 26942359]

39. Weitz TA, Taylor D, Desai S, Upadhyay UD, Waldman J, Battistelli MF, et al. Safety of aspiration abortion performed by nurse practitioners, certified nurse midwives, and physician assistants under a California legal waiver. Am J Public Health 2013 Mar;103(3):454-461. [doi: 10.2105/AJPH.2012.301159] [Medline: 23327244]

40.   Chen A. Q&A on non-medical transportation and reproductive health services. Los Angeles, CA: National Health Law Program; 2017 Aug 04. URL: http://www.healthlaw.org/issues/reproductive-health/ q-a-on-non-medical-transportation-and-reproductive-health-services#.WjwVs1WnHIU [WebCite Cache ID 6vsptYDTA]

41.   Murtagh C, Wells E, Raymond EG, Coeytaux F, Winikoff B. Exploring the feasibility of obtaining mifepristone and misoprostol from the internet. Contraception 2018 Apr;97(4):287-291 [FREE Full text] [doi: 10.1016/j.contraception.2017.09.016] [Medline: 29030227]

42.   Roberts SCM, Gould H, Upadhyay UD. Implications of Georgia's 20-week abortion ban. Am J Public Health 2015 Aug;105(8):e77-e82. [doi: 10.2105/AJPH.2015.302728] [Medline: 26066924]

Abbreviations

LMP: last menstrual period

XSL•FO
RenderX

Original Paper

Assessing Competencies Needed to Engage With Digital Health Services: Development of the eHealth Literacy Assessment Toolkit

Astrid Karnoe[1,2], MSc; Dorthe Furstrand[1,3], MSc, MD; Karl Bang Christensen[1], PhD; Ole Norgaard[4,5], MSc; Lars Kayser[1], MD, PhD

[1]Department of Public Health, University of Copenhagen, Copenhagen, Denmark

[2]Danish Multiple Sclerosis Society, Valby, Denmark

[3]Danish Cancer Research Center, Danish Cancer Society, Copenhagen, Denmark

[4]Scias, Copenhagen, Denmark

[5]Steno Diabetes Center Copenhagen, Gentofte, Denmark

Corresponding Author:
Astrid Karnoe, MSc
Department of Public Health
University of Copenhagen
Øster Farimagsgade 5
Copenhagen, 1014
Denmark
Phone: 45 30436701
Fax: 45 35327487
Email: askn@sund.ku.dk

Abstract

Background: To achieve full potential in user-oriented eHealth projects, we need to ensure a match between the eHealth technology and the user's eHealth literacy, described as knowledge and skills. However, there is a lack of multifaceted eHealth literacy assessment tools suitable for screening purposes.

Objective: The objective of our study was to develop and validate an eHealth literacy assessment toolkit (eHLA) that assesses individuals' health literacy and digital literacy using a mix of existing and newly developed scales.

Methods: From 2011 to 2015, scales were continuously tested and developed in an iterative process, which led to 7 tools being included in the validation study. The eHLA validation version consisted of 4 health-related tools (tool 1: "functional health literacy," tool 2: "health literacy self-assessment," tool 3: "familiarity with health and health care," and tool 4: "knowledge of health and disease") and 3 digitally-related tools (tool 5: "technology familiarity," tool 6: "technology confidence," and tool 7: "incentives for engaging with technology") that were tested in 475 respondents from a general population sample and an outpatient clinic. Statistical analyses examined floor and ceiling effects, interitem correlations, item-total correlations, and Cronbach coefficient alpha (CCA). Rasch models (RM) examined the fit of data. Tools were reduced in items to secure robust tools fit for screening purposes. Reductions were made based on psychometrics, face validity, and content validity.

Results: Tool 1 was not reduced in items; it consequently consists of 10 items. The overall fit to the RM was acceptable (Anderson conditional likelihood ratio, CLR=10.8; df=9; P=.29), and CCA was .67. Tool 2 was reduced from 20 to 9 items. The overall fit to a log-linear RM was acceptable (Anderson CLR=78.4, df=45, P=.002), and CCA was .85. Tool 3 was reduced from 23 to 5 items. The final version showed excellent fit to a log-linear RM (Anderson CLR=47.7, df=40, P=.19), and CCA was .90. Tool 4 was reduced from 12 to 6 items. The fit to a log-linear RM was acceptable (Anderson CLR=42.1, df=18, P=.001), and CCA was .59. Tool 5 was reduced from 20 to 6 items. The fit to the RM was acceptable (Anderson CLR=30.3, df=17, P=.02), and CCA was .94. Tool 6 was reduced from 5 to 4 items. The fit to a log-linear RM taking local dependency (LD) into account was acceptable (Anderson CLR=26.1, df=21, P=.20), and CCA was .91. Tool 7 was reduced from 6 to 4 items. The fit to a log-linear RM taking LD and differential item functioning into account was acceptable (Anderson CLR=23.0, df=29, P=.78), and CCA was .90.

Conclusions: The eHLA consists of 7 short, robust scales that assess individual's knowledge and skills related to digital literacy and health literacy.

XSL•FO
RenderX

Introduction

Health care is transforming toward increased patient involvement with the ultimate goal of patients being able to take better care of their own health. This requires that we understand the health-related competencies that patients need to be able to handle information and actively engage in their health condition. Health literacy is one of the key concepts to achieve this. A definition from 1998 states: "Health literacy represents the cognitive and social skills which determine the motivation and ability of individuals to gain access to, understand, and use information in ways which promote and maintain good health" [1].

In accordance with this definition, early measurements of health literacy focused on the patient's ability to read and understand health information [2,3]. This rather narrow understanding has been widened in later multidimensional definitions of health literacy that include, for example, taxonomic levels, navigation in health systems, and social interaction [4-6]. Examples of multidimensional instruments for measuring health literacy include European Health Literacy Survey (HLS-EU) and Health Literacy Questionnaire (HLQ) [4,5].

Simultaneously with the development of health literacy, increased technology use led to the definition of computer literacy or digital literacy as the understanding of necessary skills in technology use and problem-solving [7]. Digital literacy soon became relevant in the health care setting with the emergence of patient-involving digital health care services, for example, email correspondence with general practitioners and looking up health information online.

In 2006, Norman and Skinner addressed this new need to understand the users' digital competencies in a health context. They defined electronic health (eHealth) literacy as "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" [8] and introduced the 8-item eHealth Literacy Scale (eHEALS) instrument for measuring eHealth literacy [9]. In their understanding, a user's eHealth literacy consists of 3 contextual literacies (health literacy, computer literacy, and science literacy) and 3 analytical literacies (traditional literacy, information literacy, and media literacy) [8]. This model is referred to as the lily model, and it represents our initial understanding of the eHealth literacy concept. The model was expanded in 2009 by Chan and Kaufman, who suggested the addition of taxonomic levels for each of the 6 subliteracies, whereas another expansion by Gilstad in 2014 added contextual, cultural, and social dimensions [10-12]. Furthermore, Gilstad introduced a new definition: "eHealth literacy is the ability to identify and define a health problem, to communicate, seek, understand, appraise and apply eHealth information and welfare technologies in the cultural,

social and situational frame and to use the knowledge critically in order to solve the health problem" [12].

Both Chan and Kaufman, and Gilstad elaborated on the existing concept of eHealth literacy by expanding the lily model. However, in 2015, 2 new conceptual understandings of eHealth literacy emerged.

Inspired by the HLS-EU, Bautista performed a systematic keyword clustering on definitions of health literacy, digital literacy, and eHealth literacy [5,13]. This led to a new proposal of a definition of eHealth literacy: "eHealth literacy involves the interplay of individual and social factors in the use of digital technologies to search, acquire, comprehend, appraise, communicate and apply health information in all contexts of health care with the goal of maintaining or improving the quality of life throughout the lifespan" [13]. In the same year, the authors of this study, together with a research group from Deakin University, introduced the eHealth Literacy Framework (eHLF) [14].

eHLF consists of the following 7 dimensions derived from a structured concept mapping process involving both professionals and patients: "ability to process information," "engagement in own health," "ability to actively engage with digital services," "feel safe and in control," "motivated to engage with digital services," "access to digital services that work," and "digital services that suit individual needs," together presenting a multifaceted understanding of eHealth literacy [14]. Subsequently, the eHealth Literacy Questionnaire (eHLQ) was developed as an instrument for measuring eHealth literacy based on the 7 dimensions [15,16].

In this paper, we present the eHealth Literacy Assessment toolkit (eHLA) with a different approach to understand and assess eHealth literacy, compared with the conceptual development. The approach was to combine health literacy, computer and digital literacy, and information literacy in the eHLA. These 3 elements are central subliteracies in the lily model's understanding of eHealth literacy, as well as included in the eHLF's dimensions that describe the individual and the system interaction [8,14]. The combination allows for a deeper understanding of the competencies, knowledge, and skills that a person needs to use and adopt eHealth solutions.

This process was initiated before the development of eHLF and eHLQ, and eHLA has since continued its development in parallel with eHLQ. With eHLA, it has been a goal to develop a toolkit with the combination of test and self-assessment elements and a toolkit suitable for screening purposes in projects involving eHealth solutions.

A similar approach combining skill and self-assessment is seen in the development of the Digital Health Literacy Instrument (DHLI) introduced in January 2017 [17]. DHLI consists of 21 self-assessed items supplemented with 7 performance tasks that

focus on handling digital information, mainly related to navigating the Internet and messaging health professionals [17].

Our approach to the development of eHLA was to combine tools from the more well-established fields of digital/computer literacy and health literacy in a toolkit. This meant to evaluate existing tools and identify those that could be used to improve the understanding of a person's level of eHealth literacy in a screening situation, by being easily administered, automatically evaluated, and independent of reviewer [18].

From 2011 to 2016, the tools went through continuous development and adjustments. Tools were developed to reflect the concept of eHealth literacy, first from the understanding represented in Norman and Skinner's work and later from the understanding represented in the eHLF. Item adjustments were made to ensure that items were robust to changes in technology and different population settings, and that tools could be used together or on their own.

The eHLA was constructed with the purpose to screen and assess the eligibility of an individual's participation in projects involving eHealth solutions. It consists of 7 tools, 4 tools evaluate competencies related to health and 3 tools for evaluating digital aspects. The final toolkit is based on a mixture of an existing scale [19], adapted scales [2,20,21], and scales developed specifically for eHLA in a combination of self-assessment and skill-based methods.

In this study, we describe the development and the validation of the eHLA.

Methods

Development of the Toolkit

The eHLA was developed during the period 2011 to 2016 through continuous testing and adjustments in several iterations with the aim of having a robust and adequate toolkit. Table 1 provides an overview of the 10 tools that were included or developed and tested in the process. The initial tool candidates for eHLA were identified in 2011. A total of 5 tools were evaluated by author ON in a population of 24 diabetes patients. On the basis of these initial results, the second iteration with 4 tools was made in 2012.

Table 1. Development of eHealth literacy assessment toolkit (eHLA). This table illustrates how 10 different tools were adapted or developed through 5 development phases. Seven of the 10 tools were included in the final version of the eHLA.

Tool name	Part of eHLA	Perspective	Development phases					Origin	Type
			2011	Spring 2012	Spring 2014	Spring 2015	Autumn 2015		
Powers	–	Digital	3 items (Transl[a])	–	–	–	–	Power [19]	Self-rated
Functional health literacy	Tool 1	Health	12 items (Transl)	12 items	12 items	12 items	10 items	Parker et al [2]	Performance
eHEALS[b]	–	Health/digital	8 items (Transl)	–	–	–	–	Norman and Skinner [8]	Self-rated
Technology familiarity	Tool 5	Digital	30 items (Transl)	13 items (Adj[c] + red[d])	20 items (Adj)	20 items (Adj)	23 items (Adj)	Hargittai et al [18]	Self-rated
Technology confidence	Tool 6	Digital	13 items (Transl)	5 items (Adj + red)	5 items	5 items	5 items	Cho [17]	Self-rated
COPD[e] tool	–	Health/digital	–	1 item (Dev)	–	–	–	Original by authors	Performance
Health literacy self-assessment/HLS[f]	Tool 2	Health	–	–	7 items (Red)	7 items	20 items (Add[g])	Sørensen et al [5]	Self-rated
Knowledge of health and disease	Tool 4	Health	–	–	6 items (Dev[h])	7 items (Adj)	12 items (Adj + red)	Original by authors	Performance
Incentives for engaging with technology	Tool 7	Digital	–	–	6 items (Dev)	6 items	6 items	Original by authors	Self-rated
Familiarity with health and health care	Tool 3	Health	–	–	–	–	23 items (Dev)	Original by authors	Self-rated

[a]Transl: translation from English to Danish.

[b]eHEALS: eHealth Literacy Scale.

[c]Adj: adjustments and changes to items.

[d]Red: reduction in items.

[e]COPD: chronic obstructive pulmonary disease.

[f]HLS: Health Literacy Survey.

[g]Add: items from original scale added.

[h]Dev: development of new items.

This iteration was pilot tested in a population of 7 patients with chronic obstructive pulmonary disease (COPD; Table 1) [22]. The number of items was reduced for all tools during this evaluation [22].

In 2014, 2 new scales and an existing questionnaire were added to eHLA. This third iteration was evaluated by author DF in cognitive tests (n=4) and a pilot study (n=7) [18,19].

In spring 2015, the fourth iteration with minor adjustments was tested in 8 cognitive interviews and validated in a convenience sample with 187 respondents who completed a digital version of the toolkit [23]. In autumn 2015, 5 cognitive interviews led to minor adjustments of the health-related tools and the addition of a tool exploring familiarity with health and health care.

Cognitive Interviews

Iterations of eHLA were continuously tested using cognitive interviews to make sure that the items in each tool were understood as intended [22,24]. In each cognitive interview,

the respondent was provided with a paper version of eHLA. The respondent was then asked to fill out the questionnaire while the interviewer carefully observed and noted items that caused problems or hesitation in the respondent. After the questionnaire had been completed, the interviewer and the respondent would go through the questionnaire and focus on items difficult to the respondent. The interviewer would ask "What were you thinking about when you were answering that question?" This process elicited the cognitive process behind the answers. A prompt was used if needed: "Why did you select that response option?" The aim of this process was to eliminate items with ambiguity, items with strong relation to prior experience, and items favoring larger population groups such as females or people with heart conditions.

The 7 Tools of eHealth Literacy Assessment Toolkit

A total of 7 tools, 4 health-related and 3 digitally-related, comprise the validation version of eHLA (compare to Table 1). The 4 health-related tools are functional health literacy (tool 1),

XSL•FO
RenderX

self-assessed health literacy (tool 2), familiarity with health and health care (tool 3), and knowledge of health and disease (tool 4). The 3 digitally-related tools are familiarity with technology (tool 5), technology confidence (tool 6), and incentives for engaging with technology (tool 7). eHLA is presented in Danish in Multimedia Appendix 1 and with an English excerpt of items in Multimedia Appendix 2.

Health-Related Tools

Tool 1—Functional Health Literacy

Tool 1 is inspired by Test of Functional Health Literacy in Adults [2]. This tool is a performance test that tests a respondent's ability to read, write, understand, and perform a simple calculation in a health context. The tool is structured as a performance test in which respondents are given a text excerpt from a paracetamol leaflet. In each of the 10 items, a word (8 items) or a number (2 items) has been replaced with a blank. The respondent fills out the blank by underlining 1 of the 4 different response options. One option is correct and 3 are incorrect.

Tool 1 was included in eHLA from first iteration in 2011. Cognitive interviews and pilot tests showed that the items were too easy with almost no incorrect answers. In 2012, as a result, response options for all items were adjusted to increase difficulty levels. In the initial validation in 2015, 2 items yielded no incorrect answers from respondents and contributed to a severe tool ceiling effect. Consequently, these 2 items were removed.

The validation version of tool 1 consisted of 10 items. Results for each participant were calculated as the sum of correct answers, with a minimum score of 0 and a maximum of 10 points.

Tool 2—Health Literacy Self-Assessment

Tool 2 is a shortened version of the HLS-EU-Q47 questionnaire measuring self-reported health literacy [19]. All items were used directly and with no change of wording. Tool 2 aims to give a rough measure of health literacy as defined by the HLS-EU framework matrix, which consists of 3 areas—health care, disease prevention, and health promotion—under which 4 taxonomic levels are defined: find, understand, appraise, and apply. All items can be mapped into one of the cells in the matrix and are formulated as a question of how easy a task is to perform with responses given on a scale from 1 to 4, ranging from very difficult to very easy.

This tool was initially added to eHLA in 2014. To cover the increasing difficulty of the taxonomic levels, items handling the same situation were chosen, covering an emergency situation and a situation of handling information from the media. In total, 7 items out of the 47 from HLS-EU-Q47 were included in the tool.

In 2015, an additional 13 items were added to ensure that the validation version comprised both the items selected by us as well as all items from the 16-item short version of HLS-EU-Q47, the HLS-Q16 [19].

The validation version of eHLA tool 2 consisted of 20 items with a scoring range of 20 to 80 points in the validation version.

Tool 3—Familiarity With Health and Health Care

Tool 3 was created specifically for eHLA by 3 of the authors (DF, LK, and AK). The scale is inspired by the work of Hargittai et al who used the familiarity scale in a digital context [21]. The aim of the self-reported tool is to assess familiarity with the health care system and typical terminology used in the health care. Respondents are asked to rate a number of health care–related terms and concepts on a scale from 1 to 4, ranging from "not at all familiar" to "completely familiar."

Familiarity with health and health care was added to eHLA in autumn 2015 and consequently did not go through prevalidation. In cognitive interviews, the respondents provided positive feedback on the format, which they found easy and accessible.

The validation version consisted of 23 items; thus, scores range from 23 to 92.

Tool 4—Knowledge of Health and Health Care

Tool 4 was developed specifically for eHLA in spring 2014. We wished to include a performance test that tested knowledge of health and health care. The tool was designed as a multiple-choice quiz with questions. For each question, there are 4 response options: 1 correct answer (2 points), 2 incorrect answers (0 points), and one "I would consult with someone else" option (1 point). Thus, partial credit is given to respondents choosing the latter option.

After 2 rounds of cognitive interviews in 2015, thorough changes were made to the items to avoid favoring specific patient groups and to better distinguish lower levels of knowledge from higher levels.

The validation version of tool 4 consisted of 12 items with a possible sum score ranging from 0 to 24 points.

Introduction to the Digitally-Related Tools in eHealth Literacy Assessment Toolkit

The digitally-related tools are meant to cover a range of technologies, but when the word "technology" was used in cognitive interviews, respondents found it difficult to relate to the items. Several tests showed that the best solution was to use the word "computer" in all 3 tools and then add a short introduction that explains that the word computer is used as a term that covers all technologies used in everyday life.

Tool 5—Familiarity With Technology

Tool 5 assesses familiarity with technology based on the work of Hargittai et al, who showed this method to be a valid proxy for digital skills [21]. The concept of the original scale remains, but the items have been changed and adapted to the context of eHLA. New items were selected from the following criteria: (1) they should work in both Danish and English, (2) they should not only relate to a single type of technology, and (3) they should be robust over time. The aim of tool 5 is to estimate familiarity and the respondent's knowledge level.

In 2012, the tool was reduced to 13 items to better reflect technology as a wider term than computers only. In 2014 and 2015, it was adjusted to include a broader difficulty spectrum.

The validation version of tool 5 contained 20 items with varying difficulty. Each item was rated from 1 to 4—1 being "not at all familiar" and 4 being "completely familiar," with a potential range of sum scores from 20 to 80.

Tool 6—Technology Confidence

Tool 6 was inspired by Cho et al [20]. This tool is used to investigate how confident a respondent feels when using technology in general. The tool is based on a self-reported approach in which respondents are asked to rate how confident they feel performing the task stated in the item.

In 2012, the tool was adjusted to better include technology as a wider term than computers only.

The validation version of tool 6 in eHLA consisted of 5 items. The response options were on a scale from 1 to 4, with 1 being very unconfident and 4 being very confident. Sum score ranged from 5 to 20.

Tool 7—Incentives for Engaging With Technology

In 2014, tool 7 was developed specifically for eHLA as a tool to investigate motivation for engaging with technology. The tool was constructed after the introduction of the eHLF to include perspectives on motivation to engage with digital services. The tool is a self-reported questionnaire with items based on statements from the concept mapping process that formed dimension 5 of the eHLF [14].

The validation version of this tool contained 6 items with response options from 1 to 4, with 1 being "completely disagree" and 4 being "completely agree."

Other Tools

A total of 3 other tools were tested but later removed from the eHLA before the validation.

eHEALS [9] was included in the initial development phase as this, at the time, was the only available tool for measuring eHealth literacy. However, it was removed because of limited applicability to the Danish context, primarily because the term "health resources on the Internet" is used in several items. In particular, the word "resources" cannot be translated to a simple concept related to a person's use of the health care system in Denmark.

A tool consisting of 3 questions for identifying patients with limited functional health literacy suggested by Powers et al [25] showed limited ability to discriminate the participants' literacy levels and was subsequently omitted.

A COPD tool was developed and added to the second iteration of eHLA to assess patients' critical health and literacy levels, but it was later removed to avoid disease-specific tools. Furthermore, replies were given as free text, which complicated the analyses [22].

Intertool Correlations

A total of 4 tools are used to assess health-related literacy, and 3 tools assess digitally-related literacy. Intertool correlations were calculated to improve the understanding of how the tools supplement and possibly overlap each other. The expectation was that the 7 tools measured different aspects of eHealth literacy, but with higher correlations within each of the fields for health and digital literacy.

Final Construction of eHealth Literacy Assessment Toolkit

Data Collection for Validation of the eHealth Literacy Assessment Toolkit

Data collection for the validation was conducted from October to December 2015. The eHLA was distributed together with a validation version of the eHLQ [16], and a sociodemographic questionnaire included age, gender, education, self-rated health, other chronic diseases, and nationality. The complete questionnaire consisted of 96 items in eHLA, 58 items in eHLQ, and 6 sociodemographic questions.

Respondents were recruited from an outpatient clinic at Gentofte Hospital, north of Copenhagen, Denmark, and from a general population sample (recruited during visits at, eg, libraries, work places, and sports events). All questionnaires in the outpatient clinic were distributed in paper versions, and respondents who did not have time to finish the questionnaire onsite were given a prepaid envelope they could return the questionnaire in. One of the authors (AK) and 4 student assistants from Institute of Sociology at University of Copenhagen recruited respondents by visiting workplaces, sport events, libraries, and nursing homes. Data collection was carried out as in the outpatient clinic, but with an additional option of filling it out digitally. The recruitment method does not allow for an analysis of those who chose not to respond.

Statistics

Initial data analysis evaluated floor and ceiling effects, interitem correlations, item-total correlations, and calculation of Cronbach coefficient alpha (CCA) [26]. After a potential reduction in the number of items, further analyses examined fit of the data to the Rasch model (RM) [27,28] and evaluated differential item functioning (DIF) with regard to age and gender [29]. We evaluated the overall fit of a subscale to the RM using the Anderson conditional likelihood ratio (CLR) test [30], tested the fit of individual items using comparison of observed and expected item-rest score correlation [31], and evaluated DIF and local dependency (LD; [32]) using the Q3 index [33] log-linear RM tests [34]. In additional analyses, we evaluated the overall fit and the item fit in a log-linear RM where LD was added.

Pearson correlation coefficients were used to estimate correlations between the 7 tools.

Reductions and adjustments to items in this final validation round were done in group session with the authors, and all decisions were made based on statistical results, item content, and information from cognitive interviews.

Ethics

According to Danish law, when survey-based studies are undertaken in accordance with the Helsinki Declaration, specific approval by an ethics committee and written informed consent is not required. Potential respondents were provided with information about the survey and its purpose, including that

participation was voluntary. The completion of the survey by participants was then considered to be implied consent.

Results

Data Collection for Validation of the eHealth Literacy Assessment Toolkit

A total of 100 questionnaires were collected from outpatients (paper and pen) and 375 were collected from the community (328 paper and pen and 47 digitally). The final validation sample consisted of 475 questionnaires. See sociodemographics in Table 2.

The final version of eHLA is presented in its original Danish version in Multimedia Appendix 1, and a translated English excerpt is presented in Multimedia Appendix 2.

Results of the validation are summarized in Table 3.

Table 2. Sociodemographics for the respondents (N=475).

Sociodemographics	n (%)
Age in years	
18-35	147 (30.9)
36-60	174 (36.6)
60+	133 (28.0)
Gender	
Male	213 (44.8)
Female	245 (51.6)
Education	
No education	38 (8.0)
Short education	179 (37.7)
Long education	224 (47.2)
Self-rated health	
Excellent	196 (41.3)
Good	186 (39.2)
Bad	78 (16.4)
Chronic conditions	
Yes	189 (39.8)
No	269 (56.6)

Table 3. Summary of the validation of the 7 eHealth literacy assessment tools.

Tool and item	Item-rest score correlation				
	Observed	Rasch model		Log-linear Rasch model	
		Expected	*P* value	Expected	*P* value
Functional health literacy (tool 1)					
1-1	.65	.76	.39	–	–
1-2	.80	.81	.95	–	–
1-3	.69	.68	.89	–	–
1-4	.46	.75	.02	–	–
1-5	.59	.65	.35	–	–
1-6	.65	.65	.98	–	–
1-7	.72	.65	.30	–	–
1-8	.85	.70	.10	–	–
1-9	.66	.77	.38	–	–
1-10	.74	.72	.80	–	–
Self-assessed health literacy (tool 2)					
2-1	.64	.65	.88	.66	.66
2-2	.64	.62	.61	.65	.79
2-3	.66	.64	.63	.69	.51
2-4	.77	.64	.01	.75	.53
2-5	.63	.63	.99	.64	.86
2-6	.58	.63	.18	.59	.77
2-7	.63	.64	.66	.60	.54
2-8	.64	.64	.99	.60	.32
2-9	.69	.63	.12	.59	.01
Familiarity with health and health care (tool 3)					
3-1	.76	.78	.28	.73	.43
3-2	.73	.78	.02	.73	.84
3-3	.82	.78	.10	.80	.57
3-4	.85	.78	.001	.86	.87
3-5	.80	.78	.46	.80	.82
Knowledge of health care (tool 4)					
4-1	.35	.44	.06	.38	.61
4-2	.66	.36	<.001	.59	.21
4-3	.54	.47	.17	.52	.72
4-4	.42	.56	.008	.51	.11
4-5	.52	.38	.02	.44	.18
4-6	.37	.40	.65	.33	.42
Familiarity with technology (tool 5)					
5-1	.85	.86	.80	–	–
5-2	.85	.87	.39	–	–
5-3	.89	.86	.23	–	–
5-4	.90	.87	.09	–	–

Tool and item	Item-rest score correlation				
	Observed	Rasch model		Log-linear Rasch model	
		Expected	*P* value	Expected	*P* value
5-5	.85	.87	.22	–	–
5-6	.88	.86	.37	–	–
Technology confidence (tool 6)					
6-1	.91	.87	.08	.92	.56
6-2	.90	.88	.22	.90	.85
6-3	.87	.87	.90	.82	.05
6-4	.82	.87	.002	.85	.13
Incentives for engaging with technology (tool 7)					
7-1	.83	.86	.21	.83	.93
7-2	.86	.85	.65	.85	.67
7-3	.87	.85	.33	.86	.81
7-4	.88	.85	.10	.87	.76

Tool 1—Functional Health Literacy

Tool 1 was completed by 404 participants in the validation. The distribution of the score was very skewed. A total of 273 respondents (67.6%) had a maximum score of 10 points, 93 (23.0%) respondents got 9 points, and 28 (9.5%) respondents got 8 points or less.

Interitem correlations ranged from .03 to .39, with item-total correlations being generally low ranging from .40 to .55. No evidence of LD or DIF was disclosed. The overall fit to the RM was acceptable (Anderson CLR=10.8, df=9, *P*=.29) and CCA was .67.

No reduction in items was performed, and consequently, the final version of eHLA tool 1 consists of 10 items.

Tool 2—Health Literacy Self-Assessment

In the first reduction, 4 of the 20 items were removed, 2 because of evidence of DIF. In the second round of reduction, Rasch analyses and considerations about the HLS matrix with the aim of coverage of the framework informed further removal of 7 items. This yielded a 9-item subscale with construct and content validity. The overall fit to a log-linear RM was acceptable (Anderson CLR=78.4, df=45, *P*=.002) and CCA was .85.

Tool 3—Familiarity With Health and Health Care

In the initial analysis, item-total correlations ranged from .54 to .79. In the first round of reductions, 5 items were removed based on a combination of content and data analysis.

In the second round of reductions, a total of 13 items were removed based on evaluation of Rasch item fit statistics, evidence of DIF, and content validity considerations. As an example, the item "emergency room" was found to have the lowest item-total correlation, and several authors expressed concerns that the item and concept of an emergency room will change and that there is a risk that the item may not be meaningful in few years because of organizational changes.

Furthermore, the items "myocardial infarction" and "gastroenteritis" were removed as they fitted the RM poorly. Evidence of LD for the item pairs (1,2), (3,4), and (4,5) was disclosed.

The final version of tool 3 consisted of 5 items and showed excellent fit to a log-linear RM (Anderson CLR=47.7, df=40, *P*=.19) and CCA was .90.

Tool 4—Knowledge of Health and Disease

Five items were removed in the first round of reduction. Reductions were based on a combination of content, item-total correlations, tests of DIF, and Rasch analysis. Interitem correlations ranged from .08 to .45, and item-total correlations ranged from .48 to .70.

In the second round of reductions, one more item was removed. The item "meningitis" showed significant DIF with regard to age and gender and had the lowest item-total correlation. This might indicate that knowledge about meningitis is more extensive among people with children and especially mothers. The Rasch analysis disclosed evidence of LD for the item pairs (2,3) and (2,5).

The final, validated version of tool 4 consists of 6 items. The fit to a log-linear RM was acceptable (Anderson CLR=42.1, df=18, *P*=.001) and CCA was .59.

Tool 5—Familiarity With Technology

The cognitive interviews underlined the importance of displaying the words in each item in both Danish and English, as respondents would sometimes recognize only one of the languages.

In the validation data for tool 5, item-total correlations ranged from .59 to .89. Reductions were performed based on DIF, content, and Rasch analysis. For example, the item "macro" was removed as it contained an ambiguity content-wise, which authors had not been aware of earlier.

Table 4. Pearson correlation coefficients between tools.

eHLA[a] tools	Pearson correlation coefficient						
	Health-related tools				Digitally-related tools		
	Tool 1	Tool 2	Tool 3	Tool 4	Tool 5	Tool 6	Tool 7
Health-related tools							
Tool 1	1.00	–	–	–	–	–	–
Tool 2	.18	1.00	–	–	–	–	–
Tool 3	.21	.41	1.00	–	–	–	–
Tool 4	.40	.30	.30	1.00	–	–	–
Digital tools							
Tool 5	.32	.36	.40	.20	1.00	–	–
Tool 6	.30	.38	.33	.17	.83	1.00	–
Tool 7	.22	.30	.28	.14	.71	.76	1.00

[a]eHLA: eHealth Literacy Assessment toolkit.

This was also reflected in Rasch analysis. The item "turn on the computer" was removed because of a severe ceiling effect, where 387 (87%) respondents replied "completely familiar."

The final version of tool 5 contains 6 items. The fit to the RM was acceptable (Anderson CLR=30.3, df=17, P=.02) and CCA was .94.

Tool 6—Technology Confidence

Validation data showed that 211 (47%) respondents had a sum score of 19 or 20 points, reporting a maximum score. Interitem correlations ranged from .63 to .79. Item-total correlations ranged from .87 to .91. After Rasch analysis and tests for DIF, the item "to open and save a file" was removed.

The Rasch analysis indicated LD between items 6a and 6b in this tool.

The final tool 6 consists of 4 items. The fit to a log-linear RM taking LD into account was acceptable (Anderson CLR=26.1, df=21, P=.20) and CCA was .91.

Tool 7—Incentives for Engaging With Technology

Validation data showed right-skewed data. Interitem correlations were in the range of .39 to .83, and item-total correlations were in the range of .67 to .86. DIF on age and gender showed significant, but not very strong, correlations to 3 of the items.

A total of 2 items were excluded from the tool based on Rasch analyses and content evaluation. The item "computers bring us closer together" performed poorly in item-total correlations, and the content was assessed to be opinion-based and not necessarily an indicator of motivation. The items are intended to explore motivation related to the respondent's own experience. Consequently, the item "computers can be useful in everyday life" was excluded mainly for its ambiguity and lack of relatedness to the respondent's own situation. The Rasch analysis disclosed evidence of LD for the item pairs (1,2) and (3,4) and DIF for item 2 with regard to gender.

The final version of tool 7 consists of 4 items. The fit to a log-linear RM taking LD and DIF into account was acceptable (Anderson CLR=23.0, df=29, P=.78) and CCA was .90.

The validation is summarized in Table 3.

Intertool Correlations

Correlations between all 7 tools were estimated using Pearson correlation coefficients (see Table 4). The health-related tools have correlations ranging from .18 (tools 1 and 2) to .41 (tools 2 and 3). Results from the Pearson correlation coefficients show that the 3 digitally-related tools (tools 5, 6, and 7) are strongly correlated with scores ranging from .70 (tools 5 and 7) to .83 (tools 5 and 6).

Familiarity with health and health care (tool 3) and technology familiarity (tool 5) have the highest intertool correlation (.40) across digitally- and health-related tools.

Discussion

The eHealth Literacy Assessment Toolkit

The eHLA provides the means for gaining insight into people's health-related literacy as well as their confidence, familiarity, and motivation related to digital solutions. This toolkit consists of 7 tools that validly measure constructs with a satisfactory fit to log-linear RMs, thus displaying essential validity and objectivity [35].

Reflection on Method

The development of the eHLA was initiated to be able to swiftly and robustly evaluate a person's eHealth literacy level. The screening format influenced the design and development process in terms of (1) restrictions of total item number and (2) tools should be compatible as self-administration.

Initially, existing tools were used. However, the tested tools conflicted with the screening format, and it instigated the development of new tools. This development process has led to deeper understanding of the aspects of eHealth literacy. However, the development of new tools is an extensive task,

which is not to be recommended where valid and useful tools preexist.

The 7 Tools of eHealth Literacy Assessment Toolkit

Results of the performance-based functional health literacy tool (tool 1) were affected by a low number of respondents with low scores. This may be due to our population sample in the validation, as a study has shown that a Danish validated version of the Test of Functional Health literacy in Adults identified 26% with inadequate levels of functional health literacy in a population of COPD patients (mean age 68.7 years) [36]. This may suggest that eHLA's tool 1 is more relevant to be included in studies among older people with chronic health conditions.

Furthermore, tool 1 (functional health literacy) shows higher correlations to the digitally-related tools than to 2 of the health-related tools (tools 2 and 3). Tool 1 tests the ability to read, write, understand, and calculate in a health context, and these skills may be more similar to information literacy and basic literacy as it is used in relation to technology.

With eHLA's tool 2, comes the introduction of a 9-item short version of the HLS-EU-Q47 questionnaire [19]. The self-rated health literacy tool performed well in statistical tests. The initial validation version consisted of a selection of 20 items from the HLS-EU. In general, of these 20 items, the items in the higher taxonomic levels (appraise and apply) performed worse in statistical tests than the ones in lower levels (find and understand). A study by Neter and Brainin suggests that appraise and apply should be merged into one category [37]. These findings are supported by the results of our analysis, where the final 9 items of tool 2 cover all taxonomic levels in the HLS-EU matrix, when the appraise and apply categories are merged. Consequently, we estimate the 9-item version to be a reliable and validated short version of the HLS-EU-Q47.

Tool 3 and tool 4 were developed for eHLA, and the items in both tools address specific health elements, with tool 3 being self-assessed and tool 4 being a performance test. Tools 1 and 2 assess health competencies related to information processing and navigation in the health services, whereas tools 3 and 4 focus on an individual's familiarity and knowledge of health and health care. The validated versions of tools 3 and 4 are robust measures of health and health care familiarity and knowledge, without items favoring specific patient populations.

It is worth noting that between the health-related tools, the 2 performance tests have higher correlations to each other (tools 1 and 4), and the 2 self-assessment tools have higher correlations to each other. This might indicate a difference between the self-assessment and the performance approach.

The Pearson correlation coefficient test across the 7 tools showed very high correlations between the digitally-related tools (tools 5, 6, and 7) compared with the correlations between the health-related tools. This might indicate that the 3 tools cover some of the same constructs. Consequently, an eHealth project might only need to administer 1 of the 3 digitally-related tools, which can be chosen depending on whether there is a greater need for investigating motivation, the feeling of security, or technology familiarity.

Limitations

Like eHEALS, eHLA does not include the external factors suggested by Gilstad [12], such as social and contextual literacies.

It is a limitation of the toolkit that it does not include a tool to assess the communicative and social aspects of digital literacy in a Web 2.0 context. Existing tools in the area are scarce, and it is still very diffuse what to measure.

Recruitment was conducted in a general population sample as well as in an outpatient clinic. The recruitment was efficient; however, it is a limitation of the study that dropout analyses could not be carried out because of data collection design.

eHealth Literacy Assessment Toolkit Versus eHealth Literacy Scale, Digital Health Literacy Instrument, and eHealth Literacy Questionnaire

Norman and Skinner's eHEALS is based on the lily model with its 6 subliteracies [8,9]. However, eHEALS results are calculated as a sum score, and although there might be 2 or more constructs in the items [37,38], the sum score does not reflect eHealth literacy as a multidimensional construct. eHLA contains elements from all 6 subliteracies in the lily model, and it distinguishes between health and digital literacy. Validation showed that eHLA's tools measure different aspects of digital health. If digital literacy and health literacy are considered the core elements of the lily model, it is possible that tools from eHLA elaborate on key element from the lily model [39,40].

A study by van der Vaart et al showed that eHEALS scores were poorly correlated with actual Internet use [39]. Similar to eHEALS, the eHLA is mainly based on self-assessment with only tools 1 and 4 being health-related tests (yet self-reported tests), and it is to be considered a limitation that no performance tools were identified as a fit for the screening format used in the eHLA. However, with eHLA being a toolkit for screening participants in eHealth projects, we argue that the self-assessed skill will be of importance to the actual use of the technology. A study showed computer use to have a positive effect on computer self-confidence and computer-related attitudes. Furthermore, these 2 factors had a positive effect on computer knowledge [41].

Future studies will have to investigate the correlation between eHLA scores and actual use of eHealth technology by participants in eHealth projects.

DHLI primarily measures digital skills in a health context [17], whereas eHLA assesses digitally- and health-related measures separately. Although there are overlaps between, for example, eHLA's technology familiarity scale and DHLI's operational skills construct, it is possible that tools from eHLA and subscales from DHLI can be combined to suit the need of specific eHealth projects. Further studies would have to be conducted to explore possible combinations.

eHLA and the eHLQ were developed simultaneously. The eHLQ items are based on the eHLF's 7 dimensions, and its 7 scales provide detailed insight on an individual's or a population's eHealth literacy levels. The eHLA tools have more easy-to-understand and content-specific items that make the

toolkit useful in practice settings. The tools in eHLA assess health literacy and digital literacy, and although eHLQ also contain elements of these, eHLA's tools are comprehensive with a combination of self-assessment and test of skills. These advantages make eHLA suitable for use in screening purposes.

eHLA's 7 tools all performed adequately in the statistical validation tests. The construction and validation of eHLA allow future eHealth projects to use the toolkit both as a whole or with a selection of the 7 tools.

Future research should look into validation of eHLA as a tool for screening and preferably test the toolkit in a set of eHealth projects to determine how the tools work as predictors of actual skills and benefit—together, on their own, or in combination with, for example, DHLI. Furthermore, we currently investigate how eHLA correlates to classic sociodemographic factors, as well as how eHealth literacy can supplement assessments of health literacy and empowerment.

Conclusions

eHLA is a validated, robust toolkit consisting of 7 tools that each highlight a specific component of the competencies and resources in eHealth literacy that are necessary for a person to achieve the optimal outcome in projects involving eHealth solutions.

eHLA is suitable for studies that need tools for screening participants' knowledge and skills related to eHealth literacy.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The eHealth Literacy Assessment toolkit in Danish.

[PDF File (Adobe PDF File), 856KB - jmir_v20i5e178_app1.pdf]

Multimedia Appendix 2

The English excerpt of the eHealth Literacy Assessment toolkit.

[PDF File (Adobe PDF File), 809KB - jmir_v20i5e178_app2.pdf]

References

1. Nutbeam D. World Health Organization. 1998. Health Promotion Glossary URL: http://www.who.int/healthpromotion/about/HPR%20Glossary%201998.pdf?ua=1 [accessed 2018-04-04] [WebCite Cache ID 6yQbR9grS]
2. Parker RM, Baker DW, Williams MV, Nurss JR. The test of functional health literacy in adults: a new instrument for measuring patients' literacy skills. J Gen Intern Med 1995 Oct;10(10):537-541. [Medline: 8576769]
3. Murphy PW, Davis TC, Long SW, Jackson RH, Decker BC. Rapid Estimate of Adult Literacy in Medicine (REALM): a quick reading test for patients. J Read 1993 Oct;37(2):124-130 [FREE Full text]
4. 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;13:658 [FREE Full text] [doi: 10.1186/1471-2458-13-658] [Medline: 23855504]
5. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
6. Ishikawa H, Takeuchi T, Yano E. Measuring functional, communicative, and critical health literacy among diabetic patients. Diabetes Care 2008 May;31(5):874-879. [doi: 10.2337/dc07-1932] [Medline: 18299446]
7. Gilster P. Digital Literacy. New York, US: John Wiley & Sons, Inc; 1997.
8. 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]

XSL•FO

RenderX

9. Norman CD, Skinner HA. eHEALS: the eHealth Literacy Scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]

10. Chan CV, Matthews LA, Kaufman DR. A taxonomy characterizing complexity of consumer eHealth Literacy. AMIA Annu Symp Proc 2009 Nov 14;2009:86-90 [FREE Full text] [Medline: 20351828]

11. Chan CV, Kaufman DR. A framework for characterizing eHealth literacy demands and barriers. J Med Internet Res 2011 Nov;13(4):e94 [FREE Full text] [doi: 10.2196/jmir.1750] [Medline: 22094891]

12. Gilstad H. Toward a comprehensive model of eHealth literacy. In: Proceedings of the 2nd European Workshop on Practical Aspects of Health Information. 2014 May 19 Presented at: 2nd European Workshop on Practical Aspects of Health Information; May 19; Trondheim, Norway p. 19-20 URL: http://ceur-ws.org/Vol-1251/paper7.pdf

13. Bautista JR. From solving a health problem to achieving quality of life: redefining eHealth literacy. J Lit Technol 2015 Dec;16(2):33-54 [FREE Full text]

14. Norgaard O, Furstrand 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. Knowl Manag E-Learn 2015;7(4):522-540 [FREE Full text]

15. Adellund HK. 2016. eSundhedskompetence og brugsmønstre af eksisterende e-sundhedstilbud hos ambulante medicinske patienter med kroniske sygdomme-spørgeskemaundersøgelse URL: http://2016.e-sundhedsobservatoriet.dk/forslag/patientrapporterede-oplysninger-pro/esundhedskompetence-og-brugsm-nstre-af-eksisterende-e [accessed 2017-07-04] [WebCite Cache ID 6rhySgK6J]

16. 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]

17. van der Vaart R, Drossaert C. Development of the digital health literacy instrument: measuring a broad spectrum of health 1.0 and health 2.0 Skills. J Med Internet Res 2017 Jan 24;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]

18. Furstrand LD, Kayser L. Development of the eHealth Literacy Assessment Toolkit, eHLA. Stud Health Technol Inform 2015;216:971. [Medline: 26262273]

19. Pelikan JM. 2014. Measuring comprehensive health literacy in general populations validation of instrument, indices and scales of the HLS-EU study (slides) URL: http://www.bumc.bu.edu/healthliteracyconference/files/2014/06/Pelikan-et-al-HARC-2014-fin.pdf [WebCite Cache ID 6rhydMHit]

20. Cho AH, Arar NH, Edelman DE, Hartwell PH, Oddone EZ, Yancy WS. Do diabetic veterans use the Internet? Self-reported usage, skills, and interest in using My HealtheVet Web portal. Telemed J E Health 2010 Jun;16(5):595-602. [doi: 10.1089/tmj.2009.0164] [Medline: 20575727]

21. Hargittai E. An update on survey measures of web-oriented digital literacy. Soc Sci Comput Rev 2008 Apr 07;27(1):130-137. [doi: 10.1177/0894439308318213]

22. Knudsen AK, Vang AR. Fokus på e-health literacy ved hjemmemonitorering af KOL-borgere. Copenhagen, Denmark: University of Copenhagen; 2012.

23. Karnøe KA. Udvikling og validering af redskab til vurdering af eSundhedskompetence. Copenhagen, Denmark: University of Copenhagen; 2015.

24. Drennan J. Cognitive interviewing: verbal data in the design and pretesting of questionnaires. J Adv Nurs 2003 Apr;42(1):57-63. [Medline: 12641812]

25. Powers BJ, Trinh JV, Bosworth HB. Can this patient read and understand written health information? J Am Med Assoc 2010 Jul 7;304(1):76-84. [doi: 10.1001/jama.2010.896] [Medline: 20606152]

26. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951 Sep;16(3):297-334. [doi: 10.1007/BF02310555]

27. Rasch G. Probabilistic Models for Some Intelligence and Achievement Tests. Copenhagen, Denmark: Danish Institute for Educational Research; 1960.

28. Christensen KB, Kreiner S, Mesbah M. Rasch Models in Health. Hoboken, NJ USA: John Wiley & Sons, Inc; 2012.

29. Holland PW, Wainer H, editors. Differential Item Functioning. Hillsdale, NJ: Lawrence Erlbaum Associates; 1993.

30. Andersen EB. A goodness of fit test for the rasch model. Psychometrika 1973 Mar;38(1):123-140 [FREE Full text] [doi: 10.1007/BF02291180]

31. Kreiner S. A Note on Item-Restscore Association in Rasch Models. Appl Psychol Meas 2011 Jul 19;35(7):557-561 [FREE Full text] [doi: 10.1177/0146621611410227]

32. Lazarsfeld PF, Henry NW. Latent Structure Analysis. Boston, MA: Houghton Mifflin; 1968.

33. Christensen KB, Makransky G, Horton M. Critical values for Yen's Q3: identification of local dependence in the Rasch Model using residual correlations. Appl Psychol Meas 2016 Nov 16;41(3):178-194 [FREE Full text] [doi: 10.1177/0146621616677520]

34. Kelderman H. Loglinear Rasch model tests. Psychometrika 1984 Jun;49(2):223-245 [FREE Full text] [doi: 10.1007/BF02294174]

35. Kreiner S, Christensen KB. Validity and objectivity in health-related scales: analysis by graphical loglinear Rasch model. In: Multivariate and Mixture Distribution Rasch Models. Statistics for Social and Behavioral Sciences. New York, NY: Springer-Verlag; 2007:329-346.

36. Emtekaer Haesum LK, Ehlers L, Hejlesen OK. Validation of the test of functional health literacy in adults in a Danish population. Scand J Caring Sci 2015 Sep;29(3):573-581. [doi: 10.1111/scs.12186] [Medline: 25622511]

37. Neter E, Brainin E, Baron-Epel O. The dimensionality of health literacy and eHealth literacy. Eur Health Psychol 2015;17(6):275-280.

38. Soellner R, Huber S, Reder M. The concept of eHealth literacy and its measurement: German translation of the eHEALS. J Media Psychol 2014 Jan;26(1):29-38. [doi: 10.1027/1864-1105/a000104]

39. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]

40. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]

41. Levine T, Donitsa-Schmidt S. Computer use, confidence, attitudes, and knowledge: a causal analysis. Comput Hum Behav 1998 Jan;14(1):125-146. [doi: 10.1016/s0747-5632(97)00036-8]

Abbreviations

CCA: Cronbach coefficient alpha
CLR: conditional likelihood ratio
COPD: chronic obstructive pulmonary disease
DHLI: Digital Health Literacy Instrument
DIF: differential item functioning
eHealth: electronic health
eHEALS: eHealth Literacy Scale
eHLA: eHealth Literacy Assessment toolkit
eHLF: eHealth Literacy Framework
eHLQ: eHealth Literacy Questionnaire
HLS-EU: European Health Literacy Survey
LD: local dependency
RM: Rasch model

<u>Original Paper</u>

Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse

Robert A Verheij[1], PhD; Vasa Curcin[2], PhD; Brendan C Delaney[3], MB, CHB, MD; Mark M McGilchrist[4], PhD

[1]Netherlands Institute for Health Services Research, Utrecht, Netherlands

[2]King's College London, London, United Kingdom

[3]Imperial College London, Imperial College Business School, London, United Kingdom

[4]University of Dundee, Department of Public Health Sciences, Dundee, United Kingdom

Corresponding Author:
Robert A Verheij, PhD
Netherlands Institute for Health Services Research
PO Box 1568
Utrecht, 3500BN
Netherlands
Phone: 31 641242229
Fax: 31 302729729
Email: r.verheij@nivel.nl

Abstract

Background: Enormous amounts of data are recorded routinely in health care as part of the care process, primarily for managing individual patient care. There are significant opportunities to use these data for other purposes, many of which would contribute to establishing a learning health system. This is particularly true for data recorded in primary care settings, as in many countries, these are the first place patients turn to for most health problems.

Objective: In this paper, we discuss whether data that are recorded routinely as part of the health care process in primary care are actually fit to use for other purposes such as research and quality of health care indicators, how the original purpose may affect the extent to which the data are fit for another purpose, and the mechanisms behind these effects. In doing so, we want to identify possible sources of bias that are relevant for the use and reuse of these type of data.

Methods: This paper is based on the authors' experience as users of electronic health records data, as general practitioners, health informatics experts, and health services researchers. It is a product of the discussions they had during the Translational Research and Patient Safety in Europe (TRANSFoRm) project, which was funded by the European Commission and sought to develop, pilot, and evaluate a core information architecture for the learning health system in Europe, based on primary care electronic health records.

Results: We first describe the different stages in the processing of electronic health record data, as well as the different purposes for which these data are used. Given the different data processing steps and purposes, we then discuss the possible mechanisms for each individual data processing step that can generate biased outcomes. We identified 13 possible sources of bias. Four of them are related to the organization of a health care system, whereas some are of a more technical nature.

Conclusions: There are a substantial number of possible sources of bias; very little is known about the size and direction of their impact. However, anyone that uses or reuses data that were recorded as part of the health care process (such as researchers and clinicians) should be aware of the associated data collection process and environmental influences that can affect the quality of the data. Our stepwise, actor- and purpose-oriented approach may help to identify these possible sources of bias. Unless data quality issues are better understood and unless adequate controls are embedded throughout the data lifecycle, data-driven health care will not live up to its expectations. We need a data quality research agenda to devise the appropriate instruments needed to assess the magnitude of each of the possible sources of bias, and then start measuring their impact. The possible sources of bias described in this paper serve as a starting point for this research agenda.

XSL•FO
RenderX

KEYWORDS

electronic health record; data accuracy; data sharing; health information interoperability; health care systems; health information systems; medical informatics

Introduction

Electronic Health Records: A Potential Goldmine

Researchers have long seen the reuse of large-scale, routine health care data as a means of efficiently addressing many research questions of interest. In the United Kingdom, there has been almost 25 years of research using routine primary care data, anonymized at source, through the General Practice Research Database (now CPRD, Clinical Practice Research Datalink [1]), and other data sources, also pooling data from multiple practices and tied to specific electronic health record (EHR) systems (QResearch [2], ResearchOne [3]). A similar development has taken place in the Netherlands, where, in the early 1990s, the Netherlands Institute for Health Services Research (NIVEL) developed its Netherlands Information Network of General Practice [4], now named NIVEL Primary Care Database (NIVEL-PCD) [5,6]. Belgium also has its Intego Network [6,7] and France, until recently, had its l'Observatoire de la médecine générale société [8]. These databases provide valuable information about the use of health services and developments in population health. In the United States, there has not been a tradition of using routine anonymized data, largely because the Health Insurance Portability and Accountability Act (HIPAA) regulations place restrictions on the linkage of health data from different sources without consent [9-11] and because small office practices have not been widely computerized. Instead, the focus has been mainly on secondary care (hospital) data, facilitated by the National Institute of Health's (NIH) Clinical Translational Science Awards (CTSA) [12]. Use or reuse of administrative data for research purposes is becoming more restricted in Europe as well, partly as a consequence of the European General Data Protection Regulation (GDPR) that was established in 2016 [13,14]. In addition, data owners increasingly want control over the use of their data, making it more difficult to construct large centralized databases.

In recent years, new institutions, networks, and informatics tools have appeared, most of them focusing on secondary care and the development of new treatments. For example, the i2b2 platform has proven popular as a means of structuring clinical data, with tools for distributed querying [15]. Networking between the CTSA sites and additional access to primary care health record data have been promoted by the Patient Centered Outcomes Research Institute (PCORI) and its PCORnet distributed data network [11,16] and the US Food and Drug Administration's sentinel database [17].

As more data have become available, so has the funding for research projects to utilize it, such as the Big Data to Knowledge initiative in the United States [18], and the IMI European Medical Informatics Framework [19]. The recently established European institute for Innovation through Health Data (i~HD [20]) also promotes extensive use or reuse of health care data. Increasingly, EHR data are staying where they are, queries are being run across multiple datasets, and large-scale analytics techniques such as data mining or machine learning are being used.

Learning Health Systems and Data Quality

These developments provide a foundation for using routine EHRs in support of a "learning health system" (LHS) [21,22]. An LHS is a system in which knowledge generation and reapplication is a natural product of the health care delivery process and leads to continuous improvement in outcomes and institutional performance [23]. In such a system, routine health data are analyzed and fed back to the health care providers and patients that provided the data, using reports, decision support systems (DSSs), or any other type of feedback method. These data are also used or reused for research that is relevant for clinical practice and/or health policy.

However, it is widely recognized that data collected for one purpose may not be suitable for another and that there are serious issues to be considered in the use or reuse of EHR data [24-28]. There are some strong opinions that data shall be used only for the purpose for which they were collected and that data should not be used if a purpose was not defined before the collection of data [29]. An alternative view, formulated by Juran [30] in 1954 (and reformulated in 2006 by De Lusignan et al [31]), is that: "data are of high quality if they are fit for their intended uses in operations, decision making and planning."

It is this latter definition of data quality that enables the possibility of data use or reuse. Juran's statement is also a warning against the view that sufficiently large and diverse amounts of data will allow us to disregard the quality and provenance of data. More data do not substitute for fit data and fit cannot be judged without knowing the purpose for which the data are to be used. Even inaccurate data can be useful data if the purpose is, for example, to study the quality of data being used by health professionals. Understanding the mechanisms behind variations in data quality is particularly important in the "Big Data" era and for further pursuing the principles of an LHS. The principal aim of this paper was to create awareness among potential and current users of primary care EHR data of the factors that influence the quality of these data and to open the discussion regarding what can be done to deal with these factors. In doing so, we address the following questions:

1. How do EHR data flow from their original source to any form of use or reuse?
2. What are the purposes for which EHR data are used or reused?
3. To what extent may different purposes and the nature of the data flow constitute possible sources of bias?

In this discussion paper, we first describe the steps or stages involved in collecting and processing EHR data. This is followed by a description of the purposes for which the data are and can be used. And finally—given the purposes and the data collection

XSL•FO
RenderX

steps—we identify a number of possible sources of bias involved in the use or reuse of EHR data.

Methods

First, this study is based on the author's discussions during the Translational Research and Patient Safety in Europe (TRANSFoRm) project [32]. The European Commission FP7 sponsored project TRANSFoRm 2010-15 sought to develop, pilot, and evaluate a core information architecture for the LHS in Europe. Second, it is based on the authors' extensive experience in using and reusing EHR data for research (all authors), as cofounder of one of the largest primary care databases in Europe, NIVEL Primary Care Database (RV), as health informatics experts (MM and VC), as well as on their experience as a practicing general practitioner (BD).

One of the objectives of the TRANSFoRm project was to develop tools to assess the quality of EHR data for secondary use. We first assessed the flow of data involved in basically any use or reuse of EHR data, using the privacy and confidentiality framework developed in the project [33], involving the flow of data from a care zone to a database zone, to a research zone, then assessed the different purposes for which these data are and can be used, and finally, we mapped possible sources of bias associated with each of the purposes onto the stages involved in data collection and processing.

Results

Data Flow

In general, data flow from their initial point of generation through one or more systems for processing, ultimately generating information for a desired purpose and creating opportunities for reuse. At any stage in the flow, the data can be wholly characterized in terms of completeness, correctness, and precision relative to purpose.

In terms of the TRANSFoRm Zone Model described by Kuchinke et al [33,34], data move from the care zone to the research zone. The care zone is where health care professionals provide care to their patients, "the area of patient diagnosis and treatment." It is where "personal data are stored and used within the care context by the treating physician." The noncare zone contains "research databases and secondary use databases that have been derived from primary medical care data." In the research zone, "the researcher receives data suitable for processing and analysis in specific research projects, addressing specific research questions […]." [34].

The TRANSFoRm Zone Model was extended with a number of substeps or stages within each of the zones and by naming the different actors involved in each step: health care providers, EHR vendors, data stewards, and researcher/analyst. These stages and the principal actors involved in each of them are depicted in Figure 1.

To avoid redundancy, the distinct stages will be discussed in more detail in the "sources of bias" section.

Purposes

EHRs data can be used and reused for many purposes. An extensive overview is provided by Safran et al [35]. Here, we distinguish 3 broad categories: managing individual patient's care (including also DSSs), management of organizations (including performance indicators), and various types of medical and health services research.

Figure 1. Steps and actors involved in the data flow between the delivery of care and applications reusing the data. EHR: electronic health record.
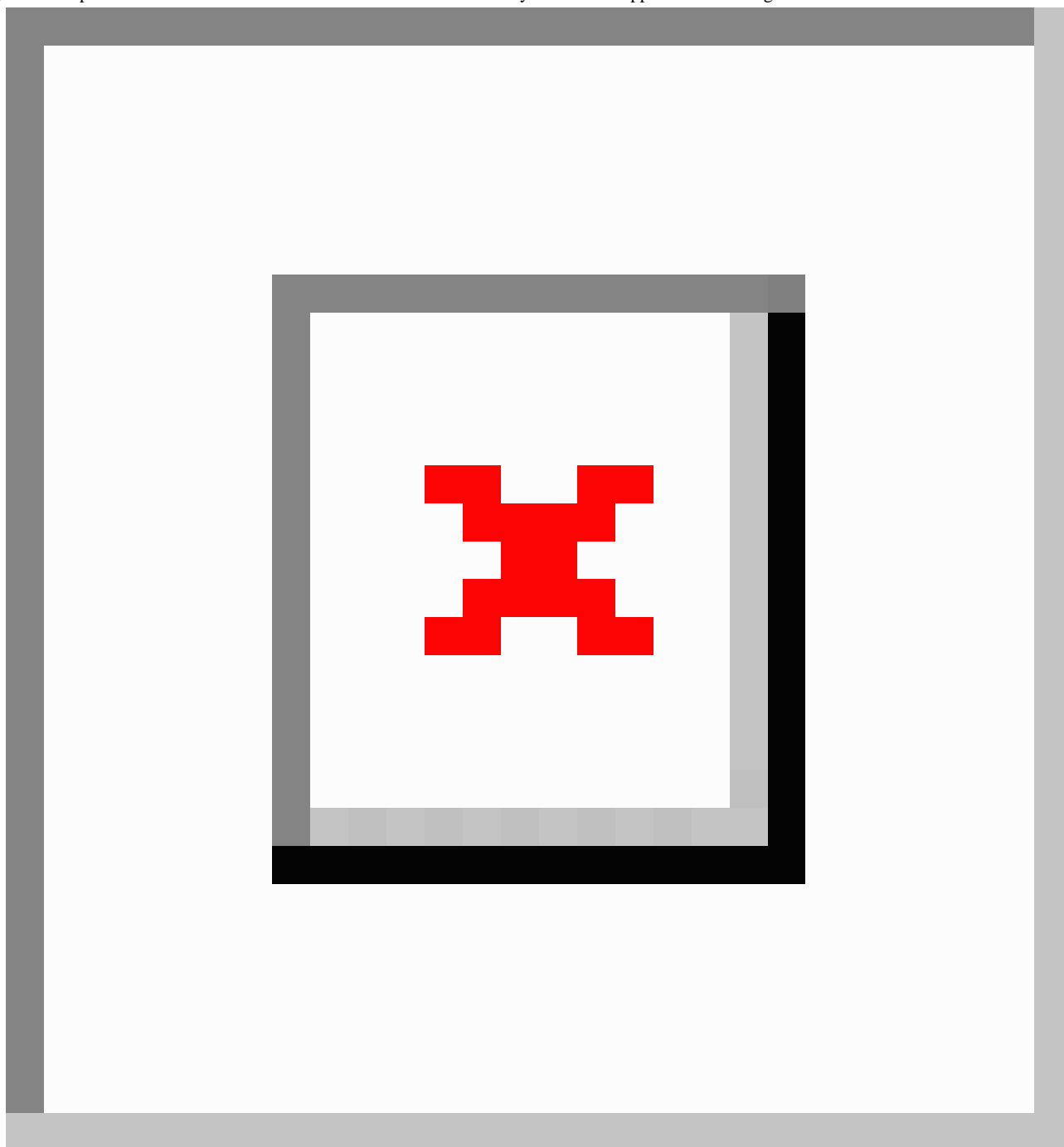
Care for Patients

Electronic health data are primarily recorded to document and facilitate the care for an individual patient. However, many patients receive health care from a variety of health care providers, and sharing relevant information among these health care providers on patients' health problems and treatments is becoming increasingly important. There is an increasing exchange of information between primary care physicians and their nurses within a practice, between primary care and hospital care, pharmacies, out-of-hours services, etc. In the Netherlands, this gave rise to the "national switchboard" initiative that allows health care professionals to see "professional summaries" of a patient's medical history. This project was subsequently voted

down in Parliament, but restarted in 2015 [36]. In the United Kingdom, the NHS National Programme for IT that was to provide a centrally held summary care record (termed "care.data" more recently) was also terminated [37]. In the United Kingdom, summary data for major diagnoses, allergies, test results, and medications are shared nationally, and locality schemes exist for sharing "views" of records between primary care and hospital sites. However, patient access is regarded as a means to empower patients and enhance self-management, and remains high on the political agenda, at least in the Netherlands and the United Kingdom.

To enable useful sharing of EHR data between professionals and patients, the data should be complete, correct, and precise, relative to health care needs. As more use is made of health

XSL•FO
RenderX

data, the more serious the consequences of incomplete, incorrect, or imprecise data, particularly in relation to comorbidity, comedication, allergies, and other intolerances.

EHR data are also increasingly used to enable DSSs [38-41]. For example, almost all Dutch general practitioners (GPs) use an evidence-based electronic prescribing system [42]. EHR data can be used to generate algorithms for DSS and also as a source of data in clinical practice. In either case, a DSS requires stringent data quality to function correctly, especially with respect to diagnosis and prescribing medication.

Management Information

EHR data are also increasingly used to calculate quality-of-care indicators for managers within the health care facility itself, or as a source of information for third-party organizations such as health insurers or governmental bodies. This can be problematic [43]. For example, in the Netherlands, Stirbu-Wagner et al found, in 2008, that it was difficult to retrieve the necessary data from EHR systems. Technically, the data elements could be extracted from the EHR systems, but the quality of the data, in relation to the required purpose, was poor. Similar results were found in more recent Dutch studies [44,45]. However, the situation regarding EHR data quality within primary care in the Netherlands is likely to have changed in recent years. Substantial numbers of practitioners (>90% of the Dutch GPs in 2013 [46]) receive feedback on the quality of their recording, based on the data quality feedback tool developed by NIVEL, as well as the fact that a portion of the reimbursement of GPs was based on the quality of recording [46,47]. Similarly, in the United Kingdom, the Quality and Outcomes Framework (QOF) promoted completeness of recording for agreed data elements within the EHR. These examples suggest that higher-quality data become more available if reimbursement is dependent on it [10]. It also illustrates how the reimbursement system can affect data quality, particularly in regard to systematic distortion of disease prevalence on the basis of the codes entered (eg, coding depression as "low mood" rather than "depression") [48].

Research

Increasingly, EHR data are also used in observational studies, recruitment and follow-up in clinical trials, and health services research. Although there are also distinct disadvantages (one of which is uncertainty about the quality of data; the subject of this paper), in comparison with surveys, EHR data for scientific research have several important advantages, suffering less from systematic errors such as selective nonresponse, response bias (systematic error caused by social desirability or leading questions), and recall bias (systematic error caused by differences in the precision or completeness of the recollections of events or experiences from the past). Moreover, EHR data are generally recorded continuously and routinely rather than periodically.

EHR systems serve as a source of data for monitoring the health of populations, allowing researchers to evaluate, among others, the effects of environmental hazards [49]; the impact of health system reforms [50,51]; how health care systems function; and developments in public health, all at comparatively low cost.

In addition, linking these EHR data to other distinct data sources increases the research possibilities enormously. For example, data from NIVEL's Primary Care Database [5] have been linked to many other data sources providing environmental characteristics [52,53], migration background [54], income, school dropout rates [55], insurance claims [56], and pharmacy data [57]. EHR data are also increasingly used for public health forecasts and surveillance [58,35,59,60]. The research potential of EHR data is also increasingly recognized outside the Western world [61].

EHR data have a distinct advantage over claims data as they are generated as part of the health care process and can potentially be extracted in real time, whereas claims data usually only become available after the treatment and claims processes have been completed. Depending on the health care system, this can take months or even years. The added value of hospital EHR data over claims data was clearly illustrated by Amarasinham [56]. In addition, primary care data have the advantage of containing data from before (and after) hospitalization.

More recently, routine EHRs are increasingly seen as a viable source of data for clinical trials [24,62]. EHR data constitutes a large part of what is called real-world data. By most definitions, real-world data are data that are collected in a usual clinical setting, as opposed to a research clinic [63]. EHR data are increasingly used alongside registry data and patient-recorded data (see for example [64]), all of which can provide contextual information that enriches the data collected directly in controlled trials. Such use of routinely recorded data in the so-called real-world studies aims to address the efficacy-effectiveness gap in drug trials, where a drug performs worse in a real-life context when compared with a trial. Furthermore, EHR data can be used to assess the feasibility of trial criteria and to target sites for recruitment that have relatively high numbers of eligible patients.

Sources of Bias in the Electronic Health Records Data Chain

There are a number of reasons why data may not be fit for a given purpose. To review these reasons, we describe the series of steps that lead from a clinically relevant event that takes place in a health care setting to an application reusing the data. These steps can be regarded as a data food chain. Analogous to a real food chain, any contamination, or "bias" in any of the steps will have consequences for the remaining steps. For each of the steps or stages, the factors that may affect data quality are described below.

Step 1: Delivery of Care (There Must Be an Event That Can Be Recorded)

This step may seem trivial, but (eg) for a blood pressure (BP) reading to be recorded, the measurement must first take place. The actors involved in this step are a health care professional interacting with a patient. The likelihood of such a measurement to take place is partly dependent on factors related to the health care system. Obviously, whether a BP measurement takes place is of course primarily dependent on the GPs professional judgment in relation to this individual patient. BP may be

clinically relevant or necessary to reassure the patient. However, this judgment is dependent on a number of other factors, most of which are strictly medical and related to that individual patient, but there are a number of other factors that may systematically affect the decision to measure a patient's BP as well. For example, as explained below, there are different incentives in the United Kingdom and the Netherlands to record BP. This difference will result in almost complete recordings for the whole population in the United Kingdom, whereas in the Netherlands, there will only be complete recordings for people known to have a chronic disease such as diabetes for which BP readings are relevant. These factors need to be known to anyone using the data in any of the subsequent steps.

First, organizational aspects of the health care system will affect actual medical practice and thereby the opportunity for an event to be recorded. For example, the difference between gatekeeping systems and nongatekeeping systems determines the population, and thereby the denominator, in epidemiological studies. In gate-keeping systems, patients need a referral from a GP before being able to make an appointment with a medical specialist, and usually GPs have a more or less stable patient list [65]. In terms of data quality, such gate-keeping systems have one very important advantage, because they allow for the calculation of an epidemiological denominator. Ideally, prevalence and incidence are expressed per 1000 in the population. This population must therefore be known. Nongatekeeping systems have only the consulting population to report on, whereas in gatekeeping systems, GPs have a more fixed list of patients that can be followed through time [7].

Gatekeeping affects the numerators as well. For example, in a nongatekeeping system, a BP reading may take place outside primary care, resulting in fewer BP readings in primary care settings. Similarly, the existence of a list system, where people are listed as members of the practice population, may not affect the number of BP readings in primary care as a whole, but it will affect the number of BP readings by a particular doctor. Health care system differences such as these have been found to be responsible for international differences in prevalence and incidence of chronic diseases [66,67].

Second, the reimbursement system in one country may stimulate BP readings under certain circumstances, whereas in other countries, it will not. In the Netherlands, prevailing quality of care indicators require BP readings to be scheduled to take place every year for patients with chronic diseases such as diabetes and cardiovascular problems. This is incorporated in the pay for performance part of the GP reimbursement system for these patients in the Netherlands but only for these patient groups. In the United Kingdom on the other hand, the QOF promotes BP readings for the whole population each year [31]. It should be noted that incentives within the health care system may seem to affect completeness of the data, but in this example, it merely reflects differences in medical practice that create data-recording opportunities.

Third, professional guidelines vary across health care systems. If a professional guideline says a BP reading should be done every year in a certain population, it will be more likely that such a measurement takes place (and get recorded).

Fourth, high practice workload may have a negative effect on taking regular BP measurements.

These 4 factors determine whether any intervention takes place in clinical practice, thereby creating a data-recording opportunity. Analysts using data from different health care systems should be aware of these factors. In any of the subsequent steps, differences in data-recording opportunities may be perceived as differences in data quality, but they are not, as they reflect real differences in medical practice. Averaging BP recordings in the United Kingdom and in the Netherlands, using the whole population as the denominator, will render invalid results because the health care system promotes readings in a much larger patient population in the United Kingdom as compared with the Netherlands, where distinct populations of chronically ill patients are targeted.

Step 2: Recording in Electronic Health Record (An Event That Is Not Recorded Will Not Be Present in Any Dataset)

There are 2 actors involved in this step: the health care professional that does the recording and the EHR vendor's software. Whether an event gets recorded is dependent on several factors.

First, there must be a software system actively used by the health care professional. About 99% of practices in the United Kingdom and the Netherlands are today using an EHR system, but this is not the case in the United States and many other countries. In general, functionalities available within the EHR systems may affect the completeness, correctness, and precision of recorded data. Although all software packages in the Netherlands and in the United Kingdom are certified by their respective authorities, considerable differences between packages have been reported in terms of what is actually recorded. For example, considerable differences between primary care EHR software brands were found in the recording of contraindications, episodes of care [68,69], as well as in the quality of prescribing [70]. The most probable factor here is the design and user interface of the software packages involved, but little is known about the actual mechanisms behind these differences. Perhaps, the holistic framework proposed by Van Gemert-Peijnen et al may prove to be useful here [71].

Second, health care professionals may display strategic recording behavior, for example, as a result of monetary incentives. Enhanced reimbursement schemes for chronically ill patients will encourage GPs to diagnose patients with chronic disease. Upcoding has been found to be a risk in relation to diagnosis-related groups used as a basis for reimbursement [72]. In addition, monetary incentives may lead to selective recording habits. For example, Mukherjee et al found that the QOF affected the recording of allergies [73]. This type of strategic behavior may lead to incomplete and incorrect data or both, as incorrectness usually implies incompleteness as well. Moreover, the fact that there are companies providing services to health care facilities to "optimize" their cash flows suggests that there are incentives for strategic recording behavior. As we know that part of the cash flow is dependent on EHR data, it is likely that strategic recording behavior can have an effect on the quality

of the data, especially in systems where billing codes and reimbursement fees are related to recorded diagnoses (as is the case in many countries).

In the United States more than in the EU, health care facilities can get involved in lawsuits with high financial risks. This can result in another form of strategic behavior related to the health care system and lead to differences in quality of the data being recorded either in a positive or negative way.

In addition, awareness of sharing data with other health professionals or patients may have an effect on whether an event gets recorded, and on the way it gets recorded. For example, health care professionals may be more reluctant to record an uncertain diagnosis in situations where this information is shared with colleagues. The size of this effect will be dependent on characteristics of the event involved, on the health professional concerned, and on whether he/she is of the same profession and/or in the same health service organization. A health professional may, for example, be more hesitant to record depression as a diagnosis than diabetes, and this may vary substantially between health professionals. Similarly, GPs may be more hesitant to record a patient's excessive alcohol intake if this information is shared with other professionals. GPs may be less hesitant to share information with GPs than with medical specialists or mental health services.

By facilitating patients' access to EHRs, patient empowerment is part of health policy in many countries [74]. Although very few patients have used this capability thus far, there may be serious consequences in terms of selective or biased recording of information. Quite paradoxically "enforced" sharing of data may lead to incomplete, incorrect, or imprecise data.

Recording behavior will also be dependent on the existence of recording guidelines. In some health care systems, there may be guidelines describing what should be recorded in an EHR system and when [75-77]. In other countries, such guidelines may not exist. Absence of recording guidelines may lead to less precise, less complete, and less correct data.

The available coding systems and thesauruses built into EHR systems determine what will and can be recorded. For example, in the International Classification of Primary Care [78], there are only about 600 codes for diagnoses and symptoms, whereas coding and classification systems such as Read, the Systematised Nomenclature of Medicine, or the various versions of the International Classification of Diseases have many more codes of greater semantic complexity and may prove more difficult to use in primary care settings, resulting in inconsistent recording.

Two other factors at the level of health care professionals will affect adequate use of EHR systems: knowledge and time. Software packages and coding systems may enable health care professionals to do all that is required and recording guidelines may tell them what to do, but if health care professionals are not familiar with these systems and guidelines, there will still be sub-optimal use of the EHR system, leading to incomplete or incorrect data and use of free text where it is not necessary. Parsons et al [79] report a "profound" data quality improvement after providing training and documentation to primary care

services in New York. The effect of feedback on data quality is reported by Van der Bij et al [80]. This feedback makes practitioners aware of the importance of high-quality recording and of the differences among them.

Moreover, the health care professional's workload may play a role. Shortage of time in a consultation will not stimulate proper recording behavior.

Lack of knowledge and time will inhibit appropriate use of the EHR systems and lead to extensive use of free text or no recording at all. The use of free text is generally regarded as problematic and only useful for small-scale studies, unless this free text can be turned into data that can be processed automatically [81]. Within the international context, this difficulty is magnified by the presence of many languages and target coding systems with national variations and varying accuracy. DSSs have an important secondary role in supporting data quality in the EHR if their operation results in more codes being placed in the EHR [82].

Step 3: Extraction From Electronic Health Record (Data Must Be Extracted for Further Analysis or Reporting)

Unless data are only used within the recording practice (the care zone, in terms of the TRANSFoRm Zone Model [34]), it needs to be extracted and transported to another site.

The actors involved in this step include the health care professional in a governance role, the software vendors who are responsible for the necessary software components (receiver as well as sender), and patients.

The database experts together with the software vendors are responsible for the extraction process from a technical point of view. It is the extraction software and associated queries that determine what data elements are extracted and how this is achieved. Different extraction tools, working in combination with different EHR systems, may render different results [83]. This may lead to incomplete and/or incorrect data. Moreover, extraction tools need to be maintained and adapted to changes in the structure and content of the EHR software. Usually—because detailed knowledge of the structure of the EHR software is needed—it is the software vendor/manufacturer that is responsible for the extraction software. How this extraction software actually works is often not explained as the process is protected by intellectual property rights. Those involved in the subsequent steps can only judge the quality of the extraction tools on the basis of the outcomes, if at all.

The third actor involved in the extraction process is the patient. Privacy regulations may allow patients to object to sharing of "their" data with other health care professionals or for research through an opt-out system, or by not giving consent. Similarly, some practices will allow the use of "their" data and others will not. Data governance options may lead to more or less incomplete or incorrect data for some patients.

XSL·FO

RenderX

Step 4: Translation Into Database (Extracted Data Must Be Redatabased as Preparation for Further Analysis or Reporting)

Actors involved in this step include database experts, database staff and domain specialists in the database zone, as the database will be engineered for particular purposes.

First, whether extracted data are actually imported into a database is dependent on the capacity of that database to capture the data that are extracted. This is particularly important in cases where data arrive in multiple formats and coding schemes. These may vary over time, being dependent on, for example, changes in the reimbursement system. The term semantic integration encompasses these issues. When data from different sources are involved, it will almost certainly be necessary to deal with different coding schemes and classifications.

Step 5: Prepare Dataset for Researcher (Generating a Research Data File)

Normally, researchers do not do their analyses on the data within the database, but on a dataset that is derived thereof. Not all variables in a database may be relevant or appropriate for a particular study and may be excluded from the research data file. In fact, the "need to know" principle demands that data that are not needed for a particular research question are not transferred to a researcher.

Determining what data are actually needed for a research question is primarily a responsibility of the researcher together with the database manager. These actors have great impact on the content of the dataset that will be analyzed. For example, quality checks or filters may be employed after data are read into the database (step 4). This means that not all data that are in a repository will go into a data file that is used by a researcher for an agreed purpose.

Furthermore, where data are linked, the resulting database may hold only data on the population common to both sources. This will affect completeness of the data. Complete data will only be available from the population that the 2 (or more) linked datasets have in common.

And finally, a repository may not be able to facilitate all types of research. There may be regulations and steering committees that will or will not grant the possibility to use a certain repository for a certain purpose. This will affect the completeness of the extracted data.

Step 6: Analysis, Outcomes, and Interpretation

These steps are in the research domain in terms of the TRANSFoRm zone Model. Here, we find the end users of exported EHR data. Different researchers will make different choices with respect to the method of analysis and what they report. Different methods may render different results, even with the same data, as was demonstrated by De Vries et al using data from the General Practice Research Database [84]. Moreover, Reeves and coworkers [50] found that different methods for computing quality-of-care scores can lead to different conclusions. This illustrates that research methods have to be based on knowledge about all previous steps and awareness of each of the possible sources of bias in each step mentioned above.

Discussion

In the previous sections, we identified 13 possible sources of bias, associated with different steps in the data chain. Textbox 1 summarizes these possible sources of bias that emerge from the combination of purposes and steps in the data chain.

Awareness and Scope

Awareness of these sources of bias is not self-evident for many that use or reuse EHR data. Where routine electronic health data are readily available, there is a risk of misinterpretation if users are unaware of the different systemic sources of bias and how they interact. It must be emphasized that large volumes of data do not reduce systematic errors, but we do contend that using these data for multiple, distinct purposes is possible, on the condition that users are aware of the risks involved and have strategies for managing them.

This is particularly important when data from different sources and from different countries are being combined in research projects such as the TRANSFoRm [32] project already mentioned, the Electronic Health Record for Clinical Research (EHR4CR) project [85], and the electronic Health Indicator Data (eHID) project [66]. Researchers should be aware of possible sources of bias and take adequate measures to ensure that their research results are not undermined.

This is all the more important because access to data is no longer a privilege of the research community, where individuals are educated and trained to deal with large amounts of data. Academically trained researchers were often the ones that were responsible for the collection of the required data as well as the analyses. Today, this too is no longer the case. Large amounts of data are open and available to the general public, and researchers using the data are very often not the ones who have collected them.

Textbox 1. Possible sources of bias in the use or reuse of electronic health record data that have to be incorporated in the choice of research methods and interpretation of results.

1. Health care system bias, emanating from:

   - Reimbursement system, pay for performance parameters

   - Role of general practitioner in the health care system; gatekeeping/nongatekeeping

   - Professional clinical guidelines

   - Ease of access by patients to their records

   - Data sharing between health care providers

2. Practice workload

3. Variations between electronic health record (EHR) system functionalities and lay-out

4. Coding systems and thesauruses

5. Knowledge and education regarding the use of EHR systems

6. Data extraction tools

7. Data processing—redatabasing

8. Research dataset preparation

9. Research methodologies

The question then arises: is it possible to provide sufficient metadata to prevent mistakes in using these data? Will the users of these data be able to understand and use this information? Will they be able to allocate enough time for that? Is it possible to set requirements for users of a dataset?

The variation in quality found within any body of data when directed at different purposes may slow down the adoption of an LHS by further hindering the formal, large-scale evaluations that have been slow to materialize [86].

The fact that the data are used for so many purposes is not just an issue for researchers, but for anyone using EHRs data not recorded by themselves [87]. Clinicians too must be aware that the patient information they share may not be complete, precise, or current. The same is true for health insurers, who rely on quality-of-care indicators derived from EHRs [44]. The LHS concept allows for greater attention to be paid to the context in which data are recorded in the EHR system, to develop mechanisms for decision support to prospectively address known "information gaps" and to track the provenance of data more thoroughly.

Toward a Data Quality Research Agenda

In this paper, we have considered potential sources of bias in routinely available health data and mapped them onto the steps generally taken in the production and analysis of such data. For each step, we presented an overview of possible sources of bias that might lead to incomparable or invalid analysis results. We

proposed a stepwise, purpose- and actor-oriented approach to understanding these factors and assessing their consequences. The size and direction of the effects from differences in health systems, of access to data by patients, of strategic recording behavior by health care professionals, of the absence or presence of recording guidelines and data quality interventions, and of different EHR systems are all largely unknown and present a huge risk to, potentially inflated, expectations of real-world data.

Unless data quality issues are better understood and unless adequate controls are embedded throughout the data lifecycle, data-driven health care will not live up to its expectations. Understanding these mechanisms is a multidisciplinary task, where medicine, health systems research, health services research, legal experts, and medical informatics have to reach out to each other and understand each other's language.

For now, the factors mentioned summarized in Textbox 1 can be used as a checklist for anyone using or reusing EHR data. However, more targeted research is needed into the actual size of the possible sources of bias described in this paper. In the meantime, it is important for researchers, EHR vendors, and health policy makers to be aware that anything they do may have an effect on the quality of EHR data and the validity of outcomes from these data. We hope this paper will help to establish this awareness and provides input for a data quality research agenda. The possible sources of bias described in this paper can be used as hypotheses for this research agenda.

Conflicts of Interest

None declared.

References

1. Clinical Practice Research Datalink. URL: https://www.cprd.com/home/ [accessed 2017-09-17] [WebCite Cache ID 6tY4lzpvJ]

2. Qresearch. URL: http://www.qresearch.org/SitePages/Home.aspx [accessed 2017-09-17] [WebCite Cache ID 6tY7ULpor]

3. ResearchONE. Transforming data into knowledge URL: http://www.researchone.org/ [accessed 2017-09-17] [WebCite Cache ID 6tY8Gkp97]

4. Verheij RA, van der Zee J. Collecting information in general practice: 'just by pressing a single button'? In: Schellevis FG, Westert G, editors. Morbidity, Performance and Quality in Primary Care: Dutch General Practice on Stage. Oxon: Radcliffe Publishing; 2006:265-272.

5. NIVEL. Utrecht: Netherlands Institute for Health Services Research NIVEL NIVEL Primary Care Database URL: https://www.nivel.nl/en/dossier/nivel-primary-care-database [accessed 2017-09-18] [WebCite Cache ID 6tYDO14QG]

6. Schweikardt C, Verheij RA, Donker GA, Coppieters Y. The historical development of the Dutch Sentinel General Practice Network from a paper-based into a digital primary care monitoring system. J Public Health 2016 Aug 15;24(6):545-562. [doi: 10.1007/s10389-016-0753-4]

7. Bartholomeeusen S, Kim CY, Mertens R, Faes C, Buntinx F. The denominator in general practice, a new approach from the Intego database. Fam Pract 2005 Aug;22(4):442-447. [doi: 10.1093/fampra/cmi054] [Medline: 15964863]

8. Observatoire de la médecine générale. URL: http://omg.sfmg.org/content/reseau/reseau.php [accessed 2017-09-18] [WebCite Cache ID 6tYEO3kIK]

9. HHS.gov. Health Insurance Portability and Accountability Act of 1996 URL: https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html [accessed 2017-09-18] [WebCite Cache ID 6tYEi1XZi]

10. Gliklich RE, Dreyer NA, Leavy MB. PubMed Health. Rockville, MD: Agency for Healthcare Research and Quality; 2014. Registries for Evaluating Patient Outcomes: A User's Guide URL: https://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0077814/ [WebCite Cache ID 6zAOozQpg]

11. Fleurence RL, Beal AC, Sheridan SE, Johnson LB, Selby JV. Patient-powered research networks aim to improve patient care and health research. Health Aff (Millwood) 2014 Jul;33(7):1212-1219. [doi: 10.1377/hlthaff.2014.0113] [Medline: 25006148]

12. Clinical and Translational Science Award. URL: https://ctsacentral.org/ [accessed 2017-09-17] [WebCite Cache ID 6tYFYTpnY]

13. Verschuuren M, Badeyan G, Carnicero J, Gissler M, Asciak RP, Sakkeus L, Work Group on Confidentiality and Data Protection of the Network of Competent Authorities of the Health Information and Knowledge Strand of EU Public Health Programme 2003-08. The European data protection legislation and its consequences for public health monitoring: a plea for action. Eur J Public Health 2008 Dec;18(6):550-551 [FREE Full text] [doi: 10.1093/eurpub/ckn014] [Medline: 19028710]

14. Coppen R, van Veen EB, Groenewegen PP, Hazes JM, de Jong JD, Kievit J, et al. Will the trilogue on the EU Data Protection Regulation recognise the importance of health research? Eur J Public Health 2015 Oct;25(5):757-758 [FREE Full text] [doi: 10.1093/eurpub/ckv149] [Medline: 26265364]

15. Open.med.harvard. Shared Health Research Information Network (SHRINE) URL: https://open.med.harvard.edu/project/shrine/ [accessed 2017-09-17] [WebCite Cache ID 6tYFtHugu]

16. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/ [accessed 2017-09-17] [WebCite Cache ID 6tYG4wuJ0]

17. US Food and Drug Administration. US Food and Drug Administration's Sentinel Initiative URL: https://www.fda.gov/Safety/FDAsSentinelInitiative/ucm2007250.htm [accessed 2017-09-17] [WebCite Cache ID 6tYGEJfQh]

18. National Institute of Health. Big Data to Knowledge URL: https://datascience.nih.gov/bd2k/about [accessed 2017-09-17] [WebCite Cache ID 6tYGXaoVW]

19. European Medical Information Framework. URL: http://www.emif.eu/about [accessed 2017-09-17] [WebCite Cache ID 6tYGfdacL]

20. Kalra D, Stroetmann V, Sundgren M, Dupont D, Schlünder I, Coorevits P, et al. The European Institute for Innovation through Health Data. Learn Health Syst 2017;1(e10008):1-8 [FREE Full text] [doi: 10.1002/lrh2.10008]

21. Friedman C, Rubin J, Brown J, Buntin M, Corn M, Etheredge L, et al. Toward a science of learning systems: a research agenda for the high-functioning Learning Health System. J Am Med Inform Assoc 2015 Jan;22(1):43-50 [FREE Full text] [doi: 10.1136/amiajnl-2014-002977] [Medline: 25342177]

22. Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. Sci Transl Med 2010 Nov 10;2(57):57cm29. [doi: 10.1126/scitranslmed.3001456] [Medline: 21068440]

23. Best Care at Lower Cost: The Path to Continuously Learning Health Care in America Consensus Report. New York: Insitute of Medicine; 2013.

XSL•FO

RenderX

24. Köpcke F, Trinczek B, Majeed RW, Schreiweis B, Wenk J, Leusch T, et al. Evaluation of data completeness in the electronic health record for the purpose of patient recruitment into clinical trials: a retrospective analysis of element presence. BMC Med Inform Decis Mak 2013 Mar 21;13:37 [FREE Full text] [doi: 10.1186/1472-6947-13-37] [Medline: 23514203]

25. Weiskopf NG, Hripcsak G, Swaminathan S, Weng C. Defining and measuring completeness of electronic health records for secondary use. J Biomed Inform 2013 Oct;46(5):830-836 [FREE Full text] [doi: 10.1016/j.jbi.2013.06.010] [Medline: 23820016]

26. Verweij LM, Tra J, Engel J, Verheij RA, de Bruijne MC, Wagner C. Data quality issues impede comparability of hospital treatment delay performance indicators. Neth Heart J 2015 Aug;23(9):420-427 [FREE Full text] [doi: 10.1007/s12471-015-0708-3] [Medline: 26021617]

27. Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PRO, Bernstam EV, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. Med Care 2013 Aug;51(8 Suppl 3):S30-S37 [FREE Full text] [doi: 10.1097/MLR.0b013e31829b1dbd] [Medline: 23774517]

28. Hersh WR, Cimino J, Payne PR, Embi P, Logan J, Weiner M, et al. Recommendations for the use of operational electronic health record data in comparative effectiveness research. EGEMS (Wash DC) 2013;1(1):1018 [FREE Full text] [doi: 10.13063/2327-9214.1018] [Medline: 25848563]

29. van der Lei J. Use and abuse of computer-stored medical records. Methods Inf Med 1991 Apr;30(2):79-80. [Medline: 1857252]

30. Juran JM, Godfrey AB. Juran's Quality Handbook, 5th Edition. New York: McGraw Hill; 1999.

31. de Lusignan S, Mimnagh C. Breaking the first law of informatics: the Quality and Outcomes Framework (QOF) in the dock. Inform Prim Care 2006;14(3):153-156 [FREE Full text] [Medline: 17288700]

32. I-hd.eu. Translational Research and Patient Safety in Europe, the TRANSFoRm project URL: http://www.i-hd.eu/index.cfm/resources/ec-projects-results/transform/ [accessed 2017-09-25] [WebCite Cache ID 6tjkuPiAq]

33. Kuchinke W, Ohmann C, Verheij R, van Veen EB, Delaney BD. Development towards a learning health system? Experiences with the privacy protection model of the TRANSFoRm project. In: Gutwirth S, Leenes R, de Hert P, editors. Data Protection on the Move Current Developments in ICT and Privacy/Data Protection. New York: Springer; 2016.

34. Kuchinke W, Ohmann C, Verheij RA, van Veen EB, Arvanitis TN, Taweel A, et al. A standardised graphic method for describing data privacy frameworks in primary care research using a flexible zone model. Int J Med Inform 2014 Dec;83(12):941-957. [doi: 10.1016/j.ijmedinf.2014.08.009] [Medline: 25241154]

35. 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;14(1):1-9 [FREE Full text] [doi: 10.1197/jamia.M2273] [Medline: 17077452]

36. Pluut B. The Unfolding of Discursive Struggles in the Context of Health Information Exchange (PhD thesis). Utrecht: Utrecht University; 2017. URL: https://dspace.library.uu.nl/handle/1874/347811

37. Justinia T. The UK's National Programme for IT: why was it dismantled? Health Serv Manage Res 2017 Feb;30(1):2-9. [doi: 10.1177/0951484816662492] [Medline: 28166675]

38. Kostopoulou O, Delaney BC, Munro CW. Diagnostic difficulty and error in primary care--a systematic review. Fam Pract 2008 Dec;25(6):400-413. [doi: 10.1093/fampra/cmn071] [Medline: 18842618]

39. Soler JK, Corrigan D, Kazienko P, Kajdanowicz T, Danger R, Kulisiewicz M, et al. Evidence-based rules from family practice to inform family practice; the learning healthcare system case study on urinary tract infections. BMC Fam Pract 2015 May 16;16:63 [FREE Full text] [doi: 10.1186/s12875-015-0271-4] [Medline: 25980623]

40. McGinn T. Putting meaning into meaningful use: a roadmap to successful integration of evidence at the point of care. JMIR Med Inform 2016;4(2):e16 [FREE Full text] [doi: 10.2196/medinform.4553] [Medline: 27199223]

41. Cahan A, Cimino JJ. A learning health care system using computer-aided diagnosis. J Med Internet Res 2017 Mar 08;19(3):e54 [FREE Full text] [doi: 10.2196/jmir.6663] [Medline: 28274905]

42. van Dormaal JE, van der Bemt PM, Zaal RJ, Egberts AC, Lenderink BW, Kosterink JG, et al. The influence that electronic prescribing has on medication errors and preventable adverse drug events: an interrupted time-series study. J Am Med Inform Assoc 2009;16(6):816-825 [FREE Full text] [doi: 10.1197/jamia.M3099] [Medline: 19717798]

43. Chan KS, Fowles JB, Weiner JP. Review: electronic health records and the reliability and validity of quality measures: a review of the literature. Med Care Res Rev 2010 Oct;67(5):503-527. [doi: 10.1177/1077558709359007] [Medline: 20150441]

44. Barkhuysen P, de Grauw W, Akkermans R, Donkers J, Schers H, Biermans M. Is the quality of data in an electronic medical record sufficient for assessing the quality of primary care? J Am Med Inform Assoc 2014;21(4):692-698 [FREE Full text] [doi: 10.1136/amiajnl-2012-001479] [Medline: 24145818]

45. van der Bij S, de Hoon S, de Boer D, Nielen M, Verheij R. Routine zorgdata nog niet geschikt voor kwaliteitsindicatoren [EHR data not yet ready for quality indicators, in Dutch]. Eerstelijns 2016 Mar:22-23 [FREE Full text]

46. van der Bij S, Verheij R. Inzet variabiliseringsgelden 2013 leidt tot belangrijke verbetering EPD. [Pay for performance scheme 2013 led to improvements in EHR data recording. In Dutch]. SynthesHIS 2013 Dec;4(12):16-17.

47. van der Bij S, Khan N, ten Veen P, Roodzant E, Visscher S, Verheij R. De Kwaliteit van Elektronische Verslaglegging Door Huisartsen Gemeten: EPD-Scan Regio Twente, Eindrapport [measuring the quality of EHRs in the Twente region. In Dutch]. Utrecht: NIVEL; 2013.

XSL•FO

RenderX

48.   Dixon A, Khachatryan A, Wallace A, Peckham S, Boyce T, Gillam S. Impact of Quality and Outcomes Framework on Health Inequalities. London: The King's Fund; 2011.

49.   Dorn T, Yzermans JC, Spreeuwenberg PM, Schilder A, van der Zee J. A cohort study of the long-term impact of a fire disaster on the physical and mental health of adolescents. J Trauma Stress 2008 Apr;21(2):239-242. [doi: 10.1002/jts.20328] [Medline: 18404625]

50.   Reeves D, Campbell SM, Adams J, Shekelle PG, Kontopantelis E, Roland MO. Combining multiple indicators of clinical quality: an evaluation of different analytic approaches. Med Care 2007 Jun;45(6):489-496. [doi: 10.1097/MLR.0b013e31803bb479] [Medline: 17515775]

51.   van Dijk C, Verheij RA, Spreeuwenberg P, van den Berg MJ, Groenewegen PP, Braspenning J, et al. Impact of remuneration on guideline adherence: empirical evidence in general practice. Scand J Prim Health Care 2013 Mar;31(1):56-63 [FREE Full text] [doi: 10.3109/02813432.2012.757078] [Medline: 23330604]

52.   Maas J, Verheij RA, de Vries S, Spreeuwenberg P, Schellevis FG, Groenewegen PP. Morbidity is related to a green living environment. J Epidemiol Community Health 2009 Dec;63(12):967-973. [doi: 10.1136/jech.2008.079038] [Medline: 19833605]

53.   Baliatsas C, Van Kamp I, Swart W, Hooiveld M, Yzermans J. Noise sensitivity: symptoms, health status, illness behavior and co-occurring environmental sensitivities. Environ Res 2016 Oct;150:8-13. [doi: 10.1016/j.envres.2016.05.029] [Medline: 27232297]

54.   Fassaert T, Nielen M, Verheij R, Verhoeff A, Dekker J, Beekman A, et al. Quality of care for anxiety and depression in different ethnic groups by family practitioners in urban areas in the Netherlands. Gen Hosp Psychiatry 2010;32(4):368-376. [doi: 10.1016/j.genhosppsych.2010.04.010] [Medline: 20633740]

55.   Uiters E, Maurits E, Droomers M, Zwaanswijk M, Verheij RA, van der Lucht F. The association between adolescents' health and disparities in school career: a longitudinal cohort study. BMC Public Health 2014 Oct 25;14:1104 [FREE Full text] [doi: 10.1186/1471-2458-14-1104] [Medline: 25344832]

56.   Amarasingham R, Velasco F, Xie B, Clark C, Ma Y, Zhang S, et al. Electronic medical record-based multicondition models to predict the risk of 30 day readmission or death among adult medicine patients: validation and comparison to existing models. BMC Med Inform Decis Mak 2015 May 20;15:39 [FREE Full text] [doi: 10.1186/s12911-015-0162-6] [Medline: 25991003]

57.   Florentinus SR, Nielsen MW, van Dijk L, Leufkens HG, Hansen EH, Heerdink ER. Patient characteristics associated with prescribing of a newly introduced drug: the case of rofecoxib. Eur J Clin Pharmacol 2005 Apr;61(2):157-159. [doi: 10.1007/s00228-005-0891-z] [Medline: 15761756]

58.   Hoeymans N, van Loon AJM, Schoemaker CG. Design of the National Public Health Status and Forecast 2014. Bilthoven: Netherlands National Institute for Public Health and the Environment; 2012.

59.   van den Wijngaard C, van Asten L, van Pelt W, Nagelkerke NJ, Verheij R, de Neeling AJ, et al. Validation of syndromic surveillance for respiratory pathogen activity. Emerg Infect Dis 2008 Jun;14(6):917-925 [FREE Full text] [doi: 10.3201/eid1406.071467] [Medline: 18507902]

60.   Hooiveld M, van de Groep T, Verheij TJ, van der Sande MA, Verheij RA, Tacken MA, et al. Prescription of antiviral drugs during the 2009 influenza pandemic: an observational study using electronic medical files of general practitioners in the Netherlands. BMC Pharmacol Toxicol 2013 Oct 21;14:55 [FREE Full text] [doi: 10.1186/2050-6511-14-55] [Medline: 24143932]

61.   Li P, Xie C, Pollard T, Johnson AE, Cao D, Kang H, et al. Promoting secondary analysis of electronic medical records in China: summary of the PLAGH-MIT Critical Data Conference and Health Datathon. JMIR Med Inform 2017 Nov 14;5(4):e43 [FREE Full text] [doi: 10.2196/medinform.7380] [Medline: 29138126]

62.   Prokosch HU, Ganslandt T. Perspectives for medical informatics. Reusing the electronic medical record for clinical research. Methods Inf Med 2009;48(1):38-44. [Medline: 19151882]

63.   Makady A, de Boer A, Hillege H, Klungel O, Goettsch W, (on behalf of GetReal Work Package 1). What is real-world data? A review of definitions based on literature and stakeholder interviews. Value Health 2017;20(7):858-865. [doi: 10.1016/j.jval.2017.03.008] [Medline: 28712614]

64.   Ryu B, Kim N, Heo E, Yoo S, Lee K, Hwang H, et al. Impact of an electronic health record-integrated personal health record on patient participation in health care: development and randomized controlled trial of MyHealthKeeper. J Med Internet Res 2017 Dec 07;19(12):e401 [FREE Full text] [doi: 10.2196/jmir.8867] [Medline: 29217503]

65.   Schäfer WL, Boerma WG, Spreeuwenberg P, Schellevis FG, Groenewegen PP. Two decades of change in European general practice service profiles: conditions associated with the developments in 28 countries between 1993 and 2012. Scand J Prim Health Care 2016;34(1):97-110 [FREE Full text] [doi: 10.3109/02813432.2015.1132887] [Medline: 26862927]

66.   Fleming D, Elliott C, Pringle M, Andersen J, Falcao I, Hebbrecht G, et al. European Commission. 2008 Apr. Electronic Health Indicator Data (eHID): Report from the programme Public Health and Risk Assessment, Health and Consumer Protection. Strand I: Health Information Priority 2.2.5: eHealth Directorate General SANCO European Commission URL: http://ec.europa.eu/health/ph_projects/2003/action1/docs/2003_1_19_frep_en.pdf [accessed 2018-05-01] [WebCite Cache ID 6z6Oiz207]

XSL·FO
RenderX

67. van den Dungen C, Hoeymans N, van den Akker M, Biermans MC, van Boven K, Joosten JH, et al. Do practice characteristics explain differences in morbidity estimates between electronic health record based general practice registration networks? BMC Fam Pract 2014 Oct 30;15:176 [FREE Full text] [doi: 10.1186/s12875-014-0176-7] [Medline: 25358247]

68. van der Bij S, Khan N, ten Veen P, de Bakker DH, Verheij RA. Improving the quality of EHR recording in primary care: a data quality feedback tool. J Am Med Inform Assoc 2017 Jan;24(1):81-87. [doi: 10.1093/jamia/ocw054] [Medline: 27274019]

69. Sollie A, Sijmons RH, Helsper C, Numans ME. Reusability of coded data in the primary care electronic medical record: A dynamic cohort study concerning cancer diagnoses. Int J Med Inform 2017 Mar;99:45-52. [doi: 10.1016/j.ijmedinf.2016.08.004] [Medline: 28118921]

70. Opondo D, Visscher S, Eslami S, Verheij RA, Korevaar JC, Abu-Hanna A. Quality of co-prescribing NSAID and gastroprotective medications for elders in The Netherlands and its association with the electronic medical record. In: PloS one. vol. 10; 2015.

71. 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]

72. Steinbusch PJ, Oostenbrink JB, Zuurbier JJ, Schaepkens FJ. The risk of upcoding in casemix systems: a comparative study. Health Policy 2007 May;81(2-3):289-299. [doi: 10.1016/j.healthpol.2006.06.002] [Medline: 16908086]

73. Mukherjee M, Wyatt JC, Simpson CR, Sheikh A. Usage of allergy codes in primary care electronic health records: a national evaluation in Scotland. Allergy 2016 Nov;71(11):1594-1602. [doi: 10.1111/all.12928] [Medline: 27146325]

74. de Lusignan S, Ross P, Shifrin M, Hercigonja-Szekeres M, Seroussi B. A comparison of approaches to providing patients access to summary care records across old and new europe: an exploration of facilitators and barriers to implementation. Stud Health Technol Inform 2013;192:397-401. [Medline: 23920584]

75. Audit Commission for Local Authorities and the National Health Service in England and Wales. Key Messages From Three Years of Independent Review. Wetherby: Audit Commission Publications; 2004.

76. Nederlands Huisartsengenootschap. Richtlijn adequate dossiervorming met het elektronisch patientendossier ADEPD. [Guideline for adequate record keeping in EHR. In Dutch]. Utrecht: NHG (The Dutch College of General Practiciners); 2009.

77. Verdonck P, Strobbe J, Steenackers J, van Royen P. Het Elektronische Medisch Dossier. Aanbeveling voor goede medische praktijkvoering. [The Electronic Medical Record. Recommendations for correct medical practice. In Dutch]. Huisarts nu: maandblad van de Wetenschappelijke Vereniging van Vlaamse Huisartsen 2004;33(2):58-70.

78. Lamberts H, Wood M. ICPC, International Classification of Primary Care. Oxford: Oxford University Press; 1987.

79. Parsons A, McCullough C, Wang J, Shih S. Validity of electronic health record-derived quality measurement for performance monitoring. J Am Med Inform Assoc 2012;19(4):604-609 [FREE Full text] [doi: 10.1136/amiajnl-2011-000557] [Medline: 22249967]

80. van der Bij S, Khan N, ten Veen P, de Bakker DH, Verheij RA. Improving the quality of EHR recording in primary care: a data quality feedback tool. J Am Med Inform Assoc 2017 Jan;24(1):81-87. [doi: 10.1093/jamia/ocw054] [Medline: 27274019]

81. Wright A, McCoy AB, Henkin S, Kale A, Sittig DF. Use of a support vector machine for categorizing free-text notes: assessment of accuracy across two institutions. J Am Med Inform Assoc 2013;20(5):887-890 [FREE Full text] [doi: 10.1136/amiajnl-2012-001576] [Medline: 23543111]

82. Kostopoulou O, Porat T, Corrigan D, Mahmoud S, Delaney BC. Diagnostic accuracy of GPs when using an early-intervention decision support system: a high-fidelity simulation. Br J Gen Pract 2017 Mar;67(656):e201-e208 [FREE Full text] [doi: 10.3399/bjgp16X688417] [Medline: 28137782]

83. Liaw S, Taggart J, Yu H, de Lusignan S. Data extraction from electronic health records - existing tools may be unreliable and potentially unsafe. Aust Fam Physician 2013 Nov;42(11):820-823 [FREE Full text] [Medline: 24217107]

84. de Vries F, de Vries C, Cooper C, Leufkens B, Van Staa TP. Reanalysis of two studies with contrasting results on the association between statin use and fracture risk: the General Practice Research Database. Int J Epidemiol 2006 Oct;35(5):1301-1308. [doi: 10.1093/ije/dyl147] [Medline: 17053011]

85. Ouagne D, Hussain S, Sadou E, Jaulent MC, Daniel C. The Electronic Healthcare Record for Clinical Research (EHR4CR) information model and terminology. Stud Health Technol Inform 2012;180:534-538. [Medline: 22874248]

86. Budrionis A, Gustave Bellika J. The Learning Healthcare System: where are we now? A systematic review. J Biomed Inform 2016 Dec;64:87-92. [doi: 10.1016/j.jbi.2016.09.018] [Medline: 27693565]

87. Zwaanswijk M, Verheij RA, Wiesman FJ, Friele RD. Benefits and problems of electronic information exchange as perceived by health care professionals: an interview study. BMC Health Serv Res 2011 Oct 07;11:256 [FREE Full text] [doi: 10.1186/1472-6963-11-256] [Medline: 21982395]

Abbreviations

BP: blood pressure

XSL•FO

RenderX

CTSA: Clinical Translational Science Awards
DSSs: decision support systems
EHR: electronic health record
GP: general practitioner
LHS: learning health system
QOF: Quality and Outcomes Framework
TRANSFoRm: Translational Research and Patient Safety in Europe

XSL•FO
RenderX

Original Paper

Differences in Online Consumer Ratings of Health Care Providers Across Medical, Surgical, and Allied Health Specialties: Observational Study of 212,933 Providers

Timothy Daskivich[1,2,3], MSHPM, MD; Michael Luu[3], MPH; Benjamin Noah[2], BS; Garth Fuller[2,4], MSHPM; Jennifer Anger[1], MD; Brennan Spiegel[2,4,5], MD

[1]Division of Urology, Cedars-Sinai Medical Center, Los Angeles, CA, United States

[2]Cedars-Sinai Center for Outcomes Research and Education, Cedars-Sinai Medical Center, Los Angeles, CA, United States

[3]Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, United States

[4]Department of Medicine, Division of Health Services Research, Cedars-Sinai Health System, Los Angeles, CA, United States

[5]Department of Health Policy and Management, UCLA Fielding School of Public Health, Los Angeles, CA, United States

Corresponding Author:
Timothy Daskivich, MSHPM, MD
Division of Urology
Cedars-Sinai Medical Center
8635 West 3rd Street, Suite 1070W
Los Angeles, CA, 90048
United States
Phone: 1 310 423 4700
Fax: 1 310 423 1886
Email: timothy.daskivich@csmc.edu

Abstract

Background: Health care consumers are increasingly using online ratings to select providers, but differences in the distribution of scores across specialties and skew of the data have the potential to mislead consumers about the interpretation of ratings.

Objective: The objective of our study was to determine whether distributions of consumer ratings differ across specialties and to provide specialty-specific data to assist consumers and clinicians in interpreting ratings.

Methods: We sampled 212,933 health care providers rated on the Healthgrades consumer ratings website, representing 29 medical specialties (n=128,678), 15 surgical specialties (n=72,531), and 6 allied health (nonmedical, nonnursing) professions (n=11,724) in the United States. We created boxplots depicting distributions and tested the normality of overall patient satisfaction scores. We then determined the specialty-specific percentile rank for scores across groupings of specialties and individual specialties.

Results: Allied health providers had higher median overall satisfaction scores (4.5, interquartile range [IQR] 4.0-5.0) than physicians in medical specialties (4.0, IQR 3.3-4.5) and surgical specialties (4.2, IQR 3.6-4.6, $P<.001$). Overall satisfaction scores were highly left skewed (normal between –0.5 and 0.5) for all specialties, but skewness was greatest among allied health providers (–1.23, 95% CI –1.280 to –1.181), followed by surgical (–0.77, 95% CI –0.787 to –0.755) and medical specialties (–0.64, 95% CI –0.648 to –0.628). As a result of the skewness, the percentages of overall satisfaction scores less than 4 were only 23% for allied health, 37% for surgical specialties, and 50% for medical specialties. Percentile ranks for overall satisfaction scores varied across specialties; percentile ranks for scores of 2 (0.7%, 2.9%, 0.8%), 3 (5.8%, 16.6%, 8.1%), 4 (23.0%, 50.3%, 37.3%), and 5 (63.9%, 89.5%, 86.8%) differed for allied health, medical specialties, and surgical specialties, respectively.

Conclusions: Online consumer ratings of health care providers are highly left skewed, fall within narrow ranges, and differ by specialty, which precludes meaningful interpretation by health care consumers. Specialty-specific percentile ranks may help consumers to more meaningfully assess online physician ratings.

XSL•FO
RenderX

Introduction

Health care consumers are increasingly using commercial online consumer ratings websites to rate and select medical providers. A recent study of 600 randomly selected physicians from 3 metropolitan areas in the United States revealed that 66% of physicians had at least one rating across several popular online ratings websites, with a median of 7 reviews per physician [1]. Patients also appear to strongly trust these data. Even as early as 2012, a survey found that 59% of US adults believed that online ratings websites were "somewhat important" or "very important" in selecting a physician [2]. And perhaps more strikingly, a survey of 1000 surgical outpatients from the Mayo Clinic in Rochester, MN, USA, found that 75% of patients would choose a physician and 88% would avoid seeing a physician based on ratings data alone [3]. Payers and health systems are also now including consumer ratings in their online tools for patients, which provides tacit endorsement for the ratings' validity in comparing doctors [4,5]. The extent of consumers' use of online ratings suggests that these data have important implications for the use of health services and may even have downstream effects on health.

Yet, despite the public's strong interest and trust in online physician ratings, interpretation of numeric physician ratings is difficult due to the lack of established benchmarks for scoring and the normalization of results for meaningful interpretation [6]. The most popular online consumer ratings websites use a 5-star Likert-type scale to rate providers, often reported as an overall score and sometimes across domains of performance categories. While consumers may assume that higher scores (ie, scores of 4 and 5) indicate above-average performance, this may not be so if ratings are not normally distributed [7]. In fact, the percentile rank for a given star rating may differ drastically based on how scores are distributed, such that a seemingly high score may indicate average or even below-average performance [8]. Furthermore, it is possible that distributions of scores may differ by specialty due to the varying perceptions of performance associated with patients' specific needs and the services provided by different specialties.

In this study, we sought to determine how online provider consumer ratings are distributed across medical, surgical, and allied health professions and whether score distributions differ across individual specialties in the United States. To address this question, we created a novel dataset consisting of over 2.7 million reviews of approximately 830,000 providers reviewed in both the US Centers for Medicare & Medicaid Services (CMS) Physician Compare [9] and the Healthgrades online consumer rating websites [10]. Our objectives were to (1) describe the distribution of quantitative overall satisfaction scores in aggregate and across provider specialties, (2) assess whether these distributions were normal, (3) quantify how overall satisfaction scores related to percentile rank across provider specialties, and (4) provide specialty-specific lookup tables showing percentile rank by overall satisfaction score. We hypothesized that overall satisfaction scores would be strongly left skewed toward higher scores across all specialties, such that seemingly high scores would be associated with a relatively low percentile rank. Lookup tables translating overall

satisfaction scores into specialty-specific percentile ranks would allow for consumer ratings data to be communicated to patients in a more meaningful and accurate manner.

Methods

Data Source and Participants

We sampled online consumer reviews for providers in the United States from the Healthgrades website. Our dataset consisted of all reviews up to March 31, 2017, of 830,308 health care providers. We aggregated data at the provider level to calculate an average rating for each provider across a variety of metrics: overall satisfaction, level of trust in provider's decisions, how well the provider explains medical conditions, how well the provider listens and answers questions, and spending the appropriate amount of time with patients. We collected data on the following office metrics: ease of scheduling urgent appointments, the office environment, staff friendliness and courteousness, and total wait time. We also captured data on the number of reviews per provider. We linked these data to demographic information publicly available on the CMS Physician Compare website [9] using national provider identification numbers to capture medical specialty, region, sex, and year of graduation from medical school. Allied health specialties were defined as health professions distinct from medicine and nursing. We excluded providers with no data on overall patient satisfaction (n=345,862); no data on primary specialty (n=11,762); fewer than 4 reviews (the median number of reviews per provider in the overall dataset; n=255,202); and providers in nursing specialties (n=4549). Our final analytic sample consisted of 212,933 providers.

Variables

Consumer Ratings

The Healthgrades website asks consumers to rate providers on a 5-star Likert-type scale across the domains of patient experience listed above. Individual ratings are quantized at the ordinal level, though average ratings are reported to the 10th decimal place. Average ratings for each domain were aggregated at the provider level.

Covariates

We collected information on US geographical region (New England, Mid-Atlantic, East North Central, West North Central, South Atlantic, East South Atlantic, West South Central, Mountain, and Pacific), sex (male, female), and graduation year (in deciles of graduation year) using linked data from the CMS Physician Compare website.

Statistical Analysis

We first compared our sample characteristics across medical, surgical, and allied health specialties using chi-square analysis for categorical variables and the Wilcoxon-Mann-Whitney test for continuous variables.

To assess whether consumer ratings scores followed a normal distribution, we created histograms showing the distribution of overall patient satisfaction scores across the medical, surgical, and allied health specialties, along with individual specialties.

We then assessed the divergence from normality by determining skewness and kurtosis. Skewness is a measure of symmetry of the distribution of scores, with a negative skew indicating a preponderance of higher scores and a positive skew indicating a preponderance of lower scores; normal distributions generally have skewness values between –0.5 and 0.5. Kurtosis is a measure of the tailedness of the distribution compared with the standard normal distribution; positive kurtosis values indicate a heavier tail and a higher propensity for outliers, while negative values indicate a lighter tail. Normal distributions generally have kurtosis values around 0. We performed bootstrap resampling with 100 replicates to obtain bootstrap confidence intervals for skewness and kurtosis across groupings of specialties using the basic bootstrap method.

We then calculated the percentile rank for overall patient satisfaction scores within individual specialties and visualized them in a scatterplot figure. We used a locally weighted scatterplot smoother to visualize percentile rank by overall patient satisfaction scores across groupings of specialties.

We used $P<.05$ to denote the statistical significance of 2-sided tests. All statistical analyses were performed in R version 3.4.0 (R Foundation for Statistical Computing). The Cedars-Sinai Institutional Review Board approved this study.

Results

Our analytic sample comprised 212,933 providers across 29 medical specialties (n=128,678), 15 surgical specialties (n=72,531), and 6 allied health professions (n=11,724; Table 1). Most providers in our sample were male (156,556/212,933, 73.52%), were from the South region (80,751/212.933, 37.92%), and graduated from medical school after 1985 (146,246/212,933, 68.68%). More of the providers in medical specialties than in surgical specialties or allied health providers were women ($P<.001$). Allied health providers graduated later than those in the medical or surgical specialties ($P<.001$).

Median overall satisfaction scores differed significantly by provider specialty (Figure 1). Allied health providers had higher median overall satisfaction scores (4.5, interquartile range [IQR] 4.0-5.0) than physicians in medical (4.0, IQR 3.3-4.5) and surgical specialties (4.2, IQR 3.6-4.6; $P<.001$). There were also significant differences in median scores across subdomains of physician metrics and office and staff performance metrics by specialty ($P<.001$; Table 1).

Measures of normality also differed by provider specialty. Overall satisfaction scores were highly left skewed for all provider groups, but skewness differed by specialty (Figure 2). Allied health providers had the largest negative skewness (ie, preponderance of higher scores; –1.23, 95% CI –1.280 to –1.181), compared with physicians in the surgical specialties (–0.77, 95% CI –0.787 to –0.755) and medical specialties (–0.64, 95% CI –0.648 to –0.628). Distributions of overall satisfaction scores had variable kurtosis across specialties; allied health providers had the largest positive kurtosis (ie, heavy-tailed with more outliers; 1.30, 95% CI 1.109-1.531), compared with physicians in the surgical specialties (0.26, 95% CI 0.206-0.315) and medical specialties (–0.07, 95% CI –0.101 to –0.041).

To communicate consumer ratings data in a way that accounts for differences in distribution of overall satisfaction scores across specialties, we calculated the percentile rank for overall satisfaction scores by provider specialty. This information allows for translation of a provider's overall satisfaction rating into a percentile ranking compared with others in their specialty. Consistent with the left skew of the data, percentile ranks were low for seemingly high overall satisfaction scores across all specialties. Percentile rank for overall satisfaction varied across allied health, medical specialties, and surgical specialties for scores of 2 (0.7%, 2.9%, 0.8%, respectively), 3 (5.8%, 16.6%, 8.1%), 4 (23.0%, 50.3%, 37.3%), and 5 (63.9%, 89.5%, 86.8%; Figure 3). As a point of reference, if overall satisfaction scores were normally distributed, the 50th percentile would occur at a score of 3. Percentile rank for overall satisfaction scores also differed substantially by individual specialties, reflecting their variable deviation from normality (Figure 4). A Web-based tool for translating overall satisfaction ratings to specialty-specific percentile rankings is available [11].

Table 1. Sample characteristics.

Characteristics	Overall (n=212,933)	Medical specialties (n=128,678)	Allied health providers (n=11,724)	Surgical specialties (n=72,531)	P value[a]
Region, n (%)					.02
Northeast	45,616 (21.42)	27,385 (21.28)	2654 (22.64)	15,577 (21.48)	
Midwest	44,069 (20.70)	26,629 (20.69)	2389 (20.38)	15,051 (20.75)	
South	80,751 (37.92)	48,779 (37.91)	4380 (37.36)	27,592 (38.04)	
West	41,197 (19.35)	25,121 (19.52)	2234 (19.05)	13,842 (19.08)	
Not available	1300 (0.61)	764 (0.59)	67 (0.57)	469 (0.65)	
Division, n (%)					<.001
New England	12,345 (5.80)	7517 (5.84)	615 (5.25)	4213 (5.81)	
Mid-Atlantic	33,271 (15.63)	19,868 (15.40)	2039 (17.39)	11,364 (15.67)	
East North Central	32,797 (15.40)	20,058 (15.59)	1728 (14.74)	11,011 (15.18)	
West North Central	11,272 (5.29)	6571 (5.11)	661 (5.64)	4040 (5.57)	
South Atlantic	45,537 (21.39)	27,829 (21.63)	2570 (21.92)	15,138 (20.87)	
East South Atlantic	11,840 (5.56)	7008 (5.45)	575 (4.90)	4257 (5.87)	
West South Central	23,374 (10.98)	13,942 (10.83)	1235 (10.53)	8197 (11.30)	
Mountain	15,012 (7.05)	8816 (6.85)	1043 (8.90)	5153 (7.10)	
Pacific	26,185 (12.30)	16,305 (12.67)	1191 (10.16)	8689 (11.98)	
Not available	1300 (0.61)	764 (0.59)	67 (0.57)	469 (0.65)	
Sex, n (%)					<.001
Female	56,377 (26.48)	38,293 (29.76)	2592 (22.11)	15,492 (21.36)	
Male	156,556 (73.52)	90,385 (70.24)	9132 (77.89)	57,039 (78.64)	
Graduation year, n (%)					<.001
1945-1954	57 (0.02)	47 (0.04)	2 (0.02)	8 (0.01)	
1955-1964	1579 (0.74)	1025 (0.80)	18 (0.15)	536 (0.74)	
1965-1974	13,475 (6.33)	8360 (6.50)	314 (2.68)	4801 (6.62)	
1975-1984	47,738 (22.42)	29,744 (23.12)	2124 (18.12)	15,870 (21.88)	
1985-1994	64,498 (30.29)	39,094 (30.38)	3128 (26.68)	22,276 (30.71)	
1995-2004	61,338 (28.81)	36,279 (28.19)	3855 (32.88)	21,204 (29.23)	
2005-2014	20,349 (9.56)	11,665 (9.07)	2215 (18.89)	6469 (8.92)	
2015-2016	61 (0.03)	11 (0.01)	40 (0.34)	10 (0.01)	
Not available	3838 (1.80)	2453 (1.91)	28 (0.24)	1357 (1.87)	
Physician performance metrics, median (IQR[b])					
Overall patient satisfaction	4.10 (3.40-4.60)	4.00 (3.30-4.50)	4.50 (4.00-5.00)	4.20 (3.60-4.60)	<.001
Level of trust in provider's decisions	4.20 (3.60-4.60)	4.10 (3.50-4.60)	4.60 (4.10-5.00)	4.30 (3.70-4.70)	<.001
How well provider explains medical condition(s)	4.20 (3.60-4.60)	4.10 (3.50-4.60)	4.60 (4.10-5.00)	4.30 (3.70-4.70)	<.001
How well provider listens and answers questions	4.20 (3.60-4.60)	4.10 (3.50-4.60)	4.60 (4.10-5.00)	4.20 (3.70-4.60)	<.001
Spends appropriate amount of time with patients	4.20 (3.60-4.60)	4.10 (3.50-4.60)	4.60 (4.10-5.00)	4.20 (3.70-4.60)	<.001
Office and staff performance metrics, median (IQR)					
Ease of scheduling urgent appointments	4.20 (3.60-4.60)	4.00 (3.50-4.50)	4.60 (4.20-4.90)	4.30 (3.80-4.60)	<.001

Characteristics	Overall (n=212,933)	Medical specialties (n=128,678)	Allied health providers (n=11,724)	Surgical specialties (n=72,531)	P value[a]
Office environment, cleanliness, comfort	4.30 (3.90-4.70)	4.30 (3.80-4.60)	4.60 (4.20-4.90)	4.40 (4.00-4.70)	<.001
Staff friendliness and courteousness	4.20 (3.70-4.60)	4.10 (3.60-4.50)	4.60 (4.20-4.90)	4.30 (3.90-4.70)	<.001
Total wait time (waiting and exam rooms; minutes), n (%)					<.001
<10	31,177 (14.64)	17,412 (13.53)	5826 (49.69)	7939 (10.95)	
10-15	113,517 (53.31)	69,132 (53.72)	4676 (39.88)	39,709 (54.75)	
16-30	54,412 (25.55)	32,548 (25.29)	1057 (9.02)	20,807 (28.69)	
31-45	12,907 (6.06)	8836 (6.87)	156 (1.33)	3915 (5.40)	
>45	908 (0.43)	740 (0.58)	8 (0.07)	160 (0.22)	

[a]P value calculated by Pearson chi-square test.

[b]IQR: interquartile range.

Figure 1. Boxplots depicting the distribution of mean overall satisfaction ratings by provider specialty. OB/GYN: obstetrics and gynecology.
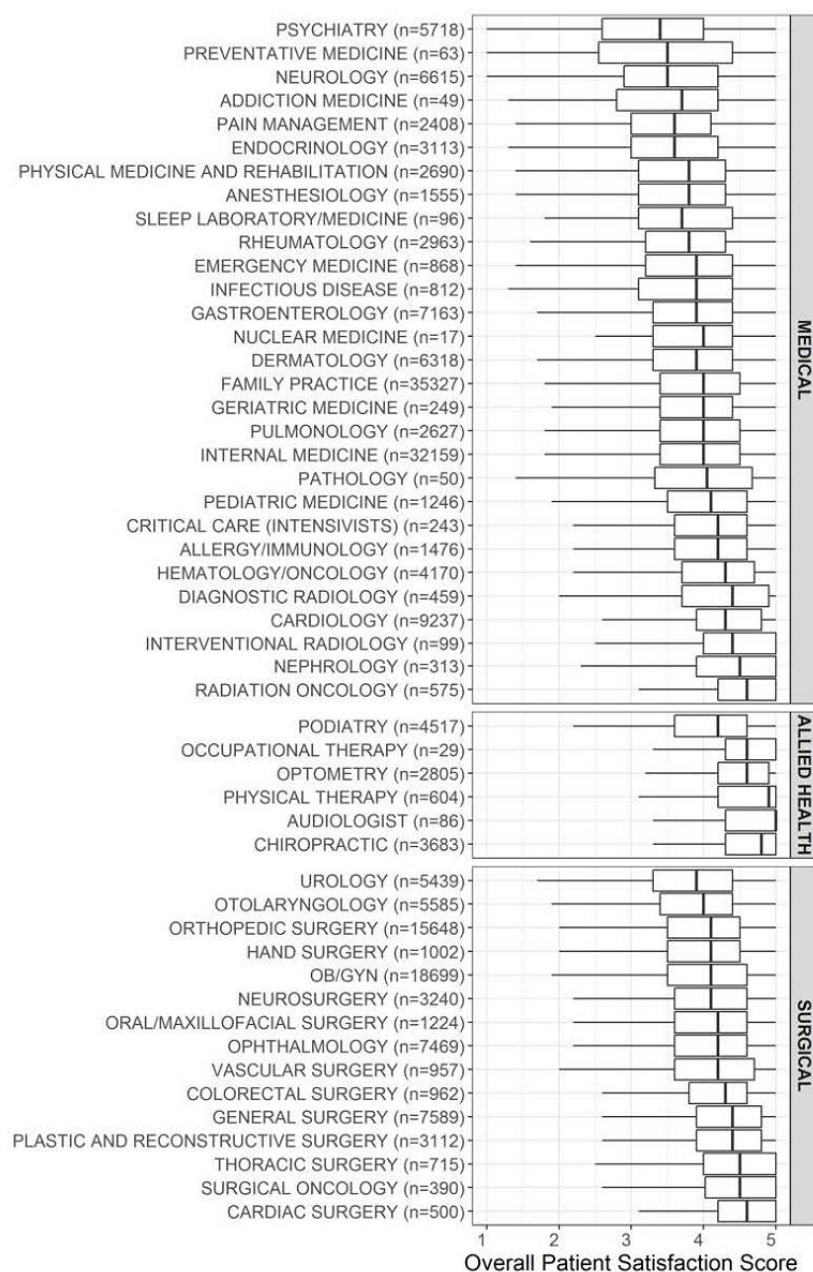
Figure 2. Frequency of mean overall patient satisfaction scores across medical, surgical, and allied health providers.
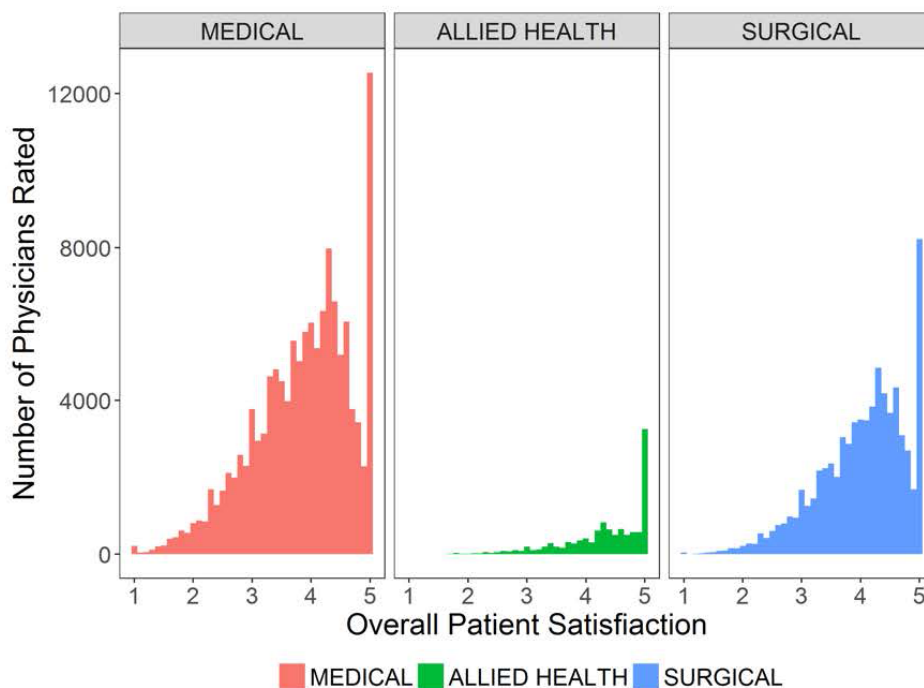
Figure 3. Percentile rank versus mean overall patient satisfaction for allied health, medical specialties, and surgical specialties. Percentile rank associated with overall patient satisfaction was first calculated within individual specialties (eg, internal medicine, podiatry, urology) as represented by scatter dots. Lines represent the locally weighted scatterplot smoothing line smoother best fit for percentile rank among specialty groupings (ie, medical, surgical, allied health). Gray bars around lines represent 95% confidence intervals for percentile rank estimates among specialty groupings.
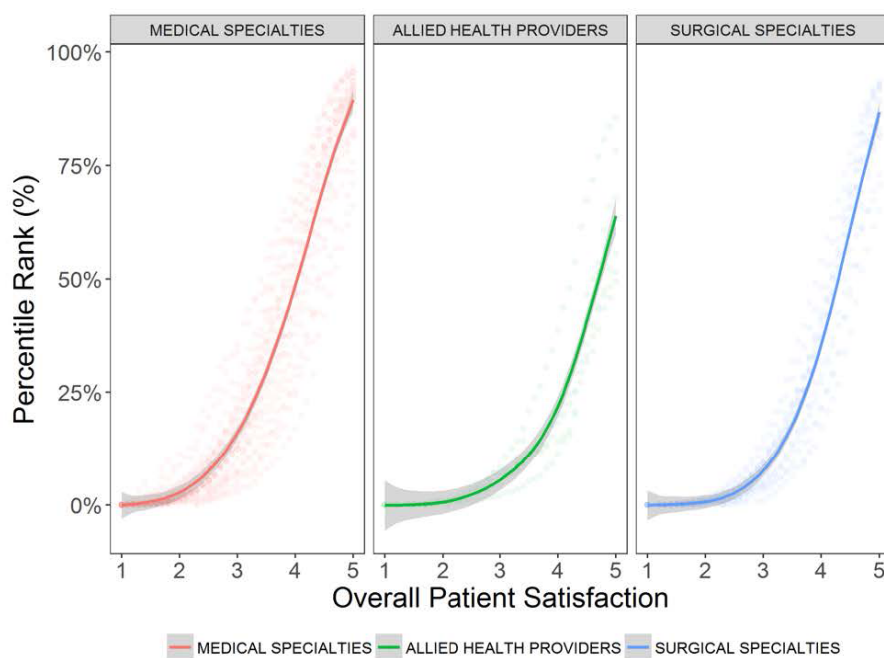
Figure 4. Percentile rank versus mean overall patient satisfaction across individual specialties. Emerg: emergency; Gen: general; Hem/Onc: hematology and oncology; Int: interventional; OB/GYN: obstetrics and gynecology; Occ: occupational; PM&R: physical medicine and rehabilitation; Prev: preventative; Recon: reconstructive.
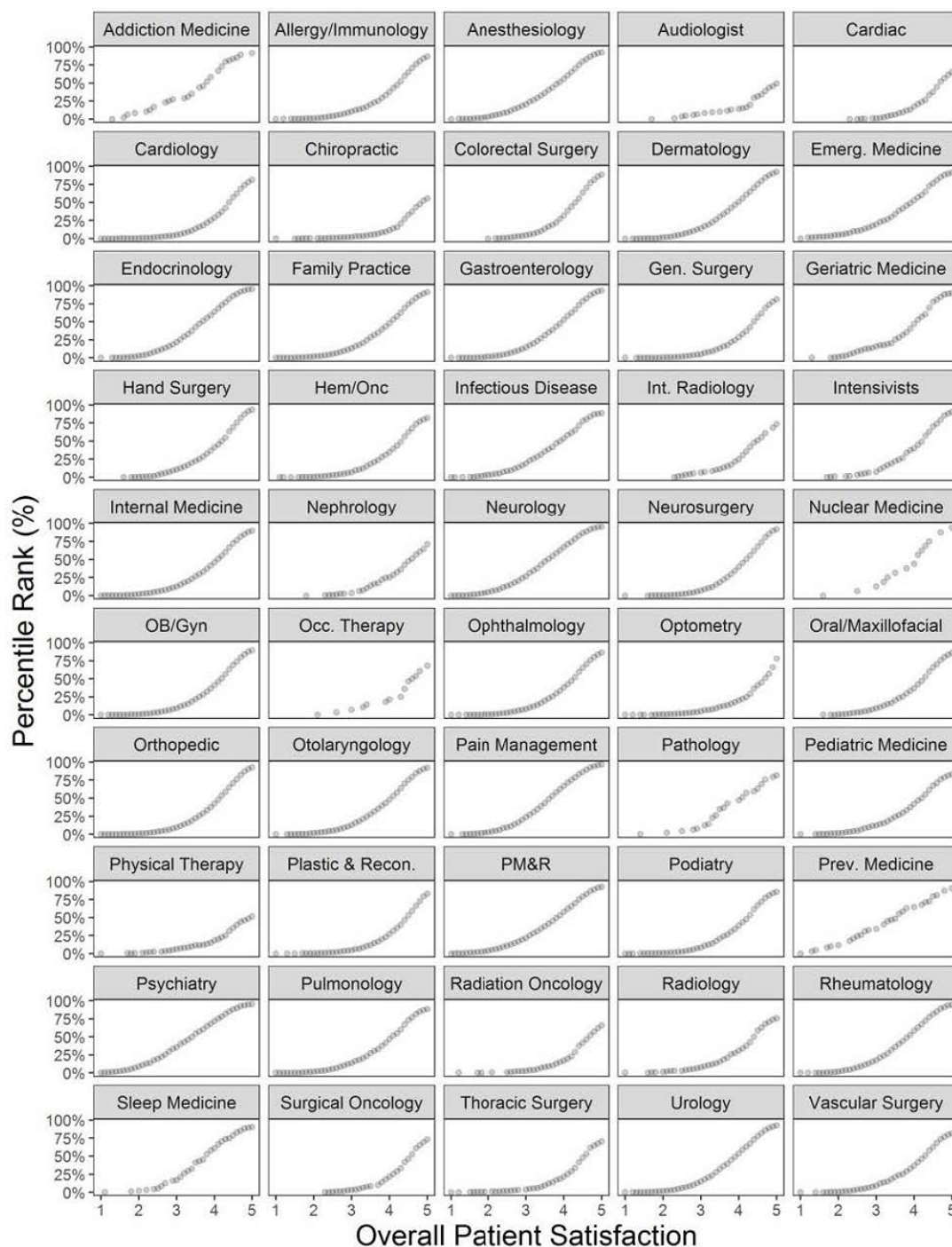
Discussion

Principal Findings

Online consumer ratings of health care providers are playing an increasing role in how consumers perceive and select providers. However, since online ratings lack standardized benchmarks for assessment and because ratings are not normalized, it is unclear how consumers should interpret scores. Our study showed that overall satisfaction scores are consistently left skewed, fall within narrow ranges, and have different distributions across specialties; as a result, scores that appear

high might actually be in the lowest quartile of scores, effectively misleading patients about perceived quality or experience of care. Allied health specialties tend to be the least normally distributed (6/6, 100% of specialties, either moderately or highly skewed—ie, skewness greater than –0.5), followed by surgical specialties (14/15, 93% of specialties), and medical specialties (16/29, 55% of specialties). Overall satisfaction scores also fall within narrow ranges; the average IQR spanned only 1.2 stars for medical specialties and 1.0 for allied health and surgical specialties. Median overall satisfaction scores also varied across specialties, with median values ranging from 3.4

to 4.6 for medical specialties, 3.9 to 4.6 for surgical specialties, and 4.2 to 4.9 for allied health professions.

Deviations from normality and differences in score distributions (ie, median, IQR) across specialties have a substantial impact on how scores should be interpreted by consumers. First, since scores across all specialties were drastically left skewed, consumers should be aware that most scores are high, which falsely implies that most doctors are above average. We found that median values for overall satisfaction scores were 4.0, 4.2 and 4.5, and the 25th percentiles for overall satisfaction scores were 3.4, 3.5, and 4.0 for medical, surgical, and allied health professions, respectively. Given this information, a score of 3—which would be considered average if scores were normally distributed—would be exceedingly low in terms of percentile rank across all medical professions. Second, due to the narrow ranges of scores within professions, consumers should be aware that small differences in scores may represent large differences in percentile rank; for example, a difference of 0.5 stars among a surgical provider may indicate a quartile difference in percentile rank. Third, given the significant differences in median overall satisfaction score distributions across specialties, there may be even more granular differences in how scores should be interpreted for individual specialties. For example, a urologist with 4.6 stars would be at the 80th percentile among his or her peers, whereas a cardiothoracic surgeon with the same star rating would be only at the 50th percentile.

In response to these findings, there are several feasible measures that could improve the interpretability of online physician consumer ratings data. First, data should be reported in a way that accounts for its consistent left skewness and nonnormality. One option would be to report the *median* star rating for each physician as a specialty-specific percentile rank, which would reflect the nonparametric nature of the data, would reduce the impact of outliers, and would be easily interpretable [8]. Another option would be to report the frequency of ratings falling within specialty-specific quartiles of performance, which would accomplish similar goals. Second, data should be reported in a way that accounts for varying distributions across specialties and subspecialties, since our data showed that patients have different benchmarks for scoring for different health care services and types of providers. We believe that our rubric for calculating percentile rank by average overall satisfaction score for individual specialties (available in a user-friendly, Web-based format [11]) may be a useful tool for describing these data to patients in a meaningful way.

While consumer ratings data may seem trivial to health care providers who are often focused on hard end points related to health [12], it is important to note that health care consumers strongly trust these data and choose providers based on them [2,3]. Although studies have shown that numeric online consumer ratings are not related to quality or value of care [13,14], this has not dampened the public's enthusiasm about their use. In fact, numerous surveys have shown that patients

use online consumer ratings as the *sole* determinant of whether or not to see a physician in consultation over three-quarters of the time [3,15]. This underscores the need for physicians to be focused not only on technical execution of their practice but also on providing excellent customer service. If patients believe that customer service (vis-à-vis consumer ratings) is important, we as health care providers should respond by measuring it accurately, describing it meaningfully, and making it a priority in the way we practice, not by ignoring it in favor of what we feel to be more important [12,16,17]. Ultimately, measurements of quality of care and consumer ratings should be provided in tandem to help consumers understand these separate components of the patient experience [5].

Study Limitations

Our study has some limitations. First, it is unclear whether results from the Healthgrades website are generalizable to other consumer ratings platforms, since distributions of scores may differ from platform to platform. Second, our findings may underestimate the degree of nonnormality of physician ratings due to our exclusion of providers with few ratings, since the vast majority of physicians with 1 rating had scores of 5. We decided to exclude physicians with fewer than 4 reviews (the median number of reviews in our overall sample) to ensure that average scores were representative of multiple ratings; sensitivity analyses showed little difference between distributions when we increased the threshold for the number of reviews beyond 4. Third, we cannot account for self-rating of physicians or other practices that may be used to artificially inflate consumer ratings scores; our reported scores represent distributions that would be observed in the real-life setting. Fourth, because we did not weight individual physician ratings scores by number of reviews, our reported results describe the distribution of average scores at the physician level.

Conclusions

Online consumer ratings of physicians are an increasingly important factor in how patients perceive and select physicians. We found that scores were highly left skewed, fell within narrow ranges, and differed by specialty; this may mislead consumers into overestimating providers with seemingly high scores who are actually mediocre or poor when compared with peers in their specialty. We herein provide a Web-based tool for translating an overall satisfaction star rating into a percentile rank comparing the provider across others in his or her specialty, an approach that accounts for the skewness and specialty-specific differences in satisfaction scores. As online consumer ratings grow in popularity, consumers will no doubt demand more detailed forms of information regarding provider service, including comparisons within specialties such as we present here. We hope our work stimulates more research on how to convey consumer ratings data in a clear, fair way, given the degree to which this information affects health care consumers' decisions.

Conflicts of Interest

None declared.

XSL•FO

RenderX

References

1. Lagu T, Metayer K, Moran M, Ortiz L, Priya A, Goff SL, et al. Website characteristics and physician reviews on commercial physician-rating websites. JAMA 2017 Feb 21;317(7):766. [doi: 10.1001/jama.2016.18553] [Medline: 28241346]

2. 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]

3. Burkle CM, Keegan MT. Popularity of internet physician rating sites and their apparent influence on patients' choices of physicians. BMC Health Serv Res 2015;15:416 [FREE Full text] [doi: 10.1186/s12913-015-1099-2] [Medline: 26410383]

4. Physician performance based compensation. Minnetonka, MN: UnitedHealthcare Services, Inc; 2018. URL: https://www.uhcprovider.com/en/reports-quality-programs/physician-perf-based-comp.html?rfid=UHCOContRD [accessed 2018-04-13] [WebCite Cache ID 6yeGq5ys7]

5. Jha A. Harvard Business Review. 2015 Oct 23. Health care providers should publish physician ratings URL: https://hbr.org/2015/10/health-care-providers-should-publish-physician-ratings [accessed 2018-04-23] [WebCite Cache ID 6yt4fwNue]

6. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. J Med Internet Res 2013;15(2):e24 [FREE Full text] [doi: 10.2196/jmir.2360] [Medline: 23372115]

7. 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;13(4):e95 [FREE Full text] [doi: 10.2196/jmir.1960] [Medline: 22088924]

8. 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]

9. Center for Medicare and Medicaid Services. Medicare Physician Compare Website URL: https://www.medicare.gov/physiciancompare/search.html [accessed 2018-04-23] [WebCite Cache ID 6yt4hBoqk]

10. Healthgrades. Denver, CO: Healthgrades Operating Company, Inc; 2018. URL: https://www.healthgrades.com/ [accessed 2018-04-10] [WebCite Cache ID 6yZsNTLKo]

11. Compare My Doc. Los Angeles, CA: Cedars-Sinai CORE; 2017. URL: https://www.comparemydoc.com/ [accessed 2018-04-13] [WebCite Cache ID 6yeHASuKG]

12. Goldman E. How doctors should respond to negative online reviews. New York, NY: Forbes Media LLC; 2013 Nov 21. URL: https://www.forbes.com/sites/ericgoldman/2013/11/21/how-doctors-should-respond-to-negative-online-reviews/ [accessed 2018-04-13] [WebCite Cache ID 6yeGxJFOM]

13. Okike K, Peter-Bibb TK, Xie KC, Okike ON. Association between physician online rating and quality of care. J Med Internet Res 2016;18(12):e324. [doi: 10.2196/jmir.6612] [Medline: 27965191]

14. 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 2017 Sep 08;25(4):401-407. [doi: 10.1093/jamia/ocx083] [Medline: 29025145]

15. Emmert M, Meier F, Pisch F, Sander U. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. J Med Internet Res 2013;15(8):e187 [FREE Full text] [doi: 10.2196/jmir.2702] [Medline: 23985220]

16. Merrell JG, Levy BH, Johnson DA. Patient assessments and online ratings of quality care: a "wake-up call" for providers. Am J Gastroenterol 2013 Nov;108(11):1676-1685. [doi: 10.1038/ajg.2013.112] [Medline: 24192941]

17. Holliday AM, Kachalia A, Meyer GS, Sequist TD. Physician and patient views on public physician rating websites: a cross-sectional study. J Gen Intern Med 2017 Jun;32(6):626-631. [doi: 10.1007/s11606-017-3982-5] [Medline: 28150098]

Abbreviations

CMS: Centers for Medicare & Medicaid Services
OB/GYN: obstetrics and gynecology
IQR: interquartile range

XSL•FO

RenderX

Review

Mapping of Crowdsourcing in Health: Systematic Review

Perrine Créquit[1,2,3], MD, PhD; Ghizlène Mansouri[1], MSc; Mehdi Benchoufi[2], MD; Alexandre Vivot[1,2], MD, PhD; Philippe Ravaud[1,2,3,4], MD, PhD

[1]INSERM UMR1153, Methods Team, Epidemiology and Statistics Sorbonne Paris Cité Research Center, Paris Descartes University, Paris, France

[2]Centre d'Epidémiologie Clinique, Hôpital Hôtel Dieu, Assistance Publique des Hôpitaux de Paris, Paris, France

[3]Cochrane France, Paris, France

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

Corresponding Author:
Perrine Créquit, MD, PhD
INSERM UMR1153, Methods Team
Epidemiology and Statistics Sorbonne Paris Cité Research Center
Paris Descartes University
1 place du Parvis Notre Dame
Paris, 75004
France
Phone: 33 142348932
Email: perrine.crequit@aphp.fr

Abstract

Background: Crowdsourcing involves obtaining ideas, needed services, or content by soliciting Web-based contributions from a crowd. The 4 types of crowdsourced tasks (problem solving, data processing, surveillance or monitoring, and surveying) can be applied in the 3 categories of health (promotion, research, and care).

Objective: This study aimed to map the different applications of crowdsourcing in health to assess the fields of health that are using crowdsourcing and the crowdsourced tasks used. We also describe the logistics of crowdsourcing and the characteristics of crowd workers.

Methods: MEDLINE, EMBASE, and ClinicalTrials.gov were searched for available reports from inception to March 30, 2016, with no restriction on language or publication status.

Results: We identified 202 relevant studies that used crowdsourcing, including 9 randomized controlled trials, of which only one had posted results at ClinicalTrials.gov. Crowdsourcing was used in health promotion (91/202, 45.0%), research (73/202, 36.1%), and care (38/202, 18.8%). The 4 most frequent areas of application were public health (67/202, 33.2%), psychiatry (32/202, 15.8%), surgery (22/202, 10.9%), and oncology (14/202, 6.9%). Half of the reports (99/202, 49.0%) referred to data processing, 34.6% (70/202) referred to surveying, 10.4% (21/202) referred to surveillance or monitoring, and 5.9% (12/202) referred to problem-solving. Labor market platforms (eg, Amazon Mechanical Turk) were used in most studies (190/202, 94%). The crowd workers' characteristics were poorly reported, and crowdsourcing logistics were missing from two-thirds of the reports. When reported, the median size of the crowd was 424 (first and third quartiles: 167-802); crowd workers' median age was 34 years (32-36). Crowd workers were mainly recruited nationally, particularly in the United States. For many studies (58.9%, 119/202), previous experience in crowdsourcing was required, and passing a qualification test or training was seldom needed (11.9% of studies; 24/202). For half of the studies, monetary incentives were mentioned, with mainly less than US $1 to perform the task. The time needed to perform the task was mostly less than 10 min (58.9% of studies; 119/202). Data quality validation was used in 54/202 studies (26.7%), mainly by attention check questions or by replicating the task with several crowd workers.

Conclusions: The use of crowdsourcing, which allows access to a large pool of participants as well as saving time in data collection, lowering costs, and speeding up innovations, is increasing in health promotion, research, and care. However, the description of crowdsourcing logistics and crowd workers' characteristics is frequently missing in study reports and needs to be precisely reported to better interpret the study findings and replicate them.

KEYWORDS

XSL•FO

RenderX

Introduction

Scientific research performed with the involvement of the broader public, the crowd, is attracting increasing attention from scientists and policy makers. Crowdsourcing uses the power of many, using the collective wisdom and resources of the crowd, to complete human intelligence tasks (ie, tasks that cannot be entirely automated and require human intelligence). Crowdsourcing is not a new concept and has often been used in the past as a competition to discover a solution. It originated in 1714 in England, where the British Government proposed £20,000 to anyone who could find a solution for calculating the longitudinal position of a ship [1], and then it was applied in a variety of fields such as astronomy, energy system research, genealogy and genetic research, journalism, linguistics, ornithology, public policy, seismology, and molecular biology [2].

Crowdsourcing currently involves a network of people, the "crowd workers," responding to an open call and completing Web-based tasks of requesters [3]. These crowd workers provide a large wide of activities, especially via the internet, using specific platforms, but have no formal training in the topic of investigation [4]. They have access to the crowdsourcing websites from anywhere at times convenient for them. They carry out tasks posted by requesters, who accept or reject their work and may or not pay them for the work. Crowdsourcing has grown rapidly with the evolution of technology, with 2.3 billion internet users and 6 billion mobile phone subscribers [5]. The main Web platform for crowdsourcing is Amazon Mechanical Turk (MTurk), which was exploited by scientists 5 years ago. In 1 month—May 2016—23,000 people completed 230,000 tasks on their computers in 3.3 million min, corresponding to a total of more than 6 years of effort [6].

Crowdsourcing has several benefits. Crowdsourcing provides easy access to a potentially large pool of participants for a research problem, particularly for increasing the number of respondents for mining crowd data (eg, Web-based surveys) and active crowdsourcing (eg, data processing). It offers important time savings, in that a large number of contributors working in parallel reduces the time required to perform a fixed amount of work, mainly saving the elapsed time to collect data. Project organizers can lower the cost of labor inputs. By soliciting ideas from a large group of people through the internet, crowdsourcing can be used to speed up innovations, particularly those with challenges.

Crowdsourcing has been used primarily in nonmedical fields [7]. The Galaxy Zoo project successfully classified about 900,000 galaxies with the help of hundreds of thousands of Web-based volunteers [8]. The eBird project collected more than 48 million bird observations from more than 35,000 contributors [9]. The fields of research using Amazon MTurk are psychology, marketing, management, business, political science, computer science (improvement of artificial intelligence software, for example, by naming objects to help the computer identify the content of a photograph), and neuroscience [6]. Crowdsourcing is becoming the center of attention of the scientific community and researchers needing to obtain data from any domain.

Crowdsourcing represents a great opportunity in health and medical research. As mentioned by Swan [10], crowdsourced health research studies are the nexus of 3 contemporary trends: "citizen science," crowdsourcing, and Medicine 2.0. Medicine 2.0 or Health 2.0 refers to the active participation of individuals in their health care, particularly using Web 2.0 technologies.

Crowdsourcing is not limited to health research but can also be used in health promotion or health care. Crowdsourcing could be a great way to solve a specific scientific mission that cannot be entirely automated and requires human intelligence in these 3 health categories. However, mapping of crowdsourcing use in health is needed to describe all its applications and to detail specificities, so that health researchers can assess whether they can use this approach in their research.

The aim of the study was to map the different applications of crowdsourcing used in health to outline the fields of health that are using crowdsourcing and the type of crowdsourced tasks involved. We also describe the logistics of crowdsourcing and the characteristics of crowd workers.

Methods

Design

We conducted a systematic review to identify studies using crowdsourcing in health. We uploaded a prespecified protocol to a publicly accessible institutional Website (Multimedia Appendix 1) and followed standard procedures for systematic reviews and reported processes and results according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines [11].

Criteria for Considering Studies for This Review

The inclusion criteria were as follows:

1. Studies reporting on health, considering the definition proposed by Prpic [12], with the activities of the 3 categories of health:
   - Health promotion: disease detection and surveillance, behavioral interventions, health literacy, and health education
   - Health research: pharmaceutical research, clinical trials and health experiment methodology, and improving health care research knowledge
   - Health maintenance (here "health care"): patient- or physician-related, diagnostics, medical practice, and treatment support.

2. Studies conducted with a crowdsourced population: workers are recruited by crowdsourcing (ie, recruited via a website [labor markets such as Amazon MTurk or Crowdflower] or an open call to a large audience using internet-related technologies [eg, scientific games or community challenges with dedicated platforms]) [13]. Studies can refer either to a feasibility study (can crowdsourcing be used for a specific task?) or to the use of crowdsourcing to supply data that support a finding in some research activity.

We excluded studies considering structural and molecular biology (eg, studies reporting Web-based games to manipulate the 3D structures of proteins or moving colored blocks representing different nucleotide sequences).

Search Method for Identification of Studies

We performed an electronic search of MEDLINE via PubMed and EMBASE to identify all reports published from inception to March 30, 2016, with no restriction on date, language, study design, or publication status (published papers or conference abstracts). All databases were searched using both controlled vocabulary (namely, MeSH terms in MEDLINE and Emtree terms in EMBASE) and a wide range of free-text terms. Indeed, crowdsourced health studies may be a blend of crowdsourcing and citizen science (ie, nonprofessionally trained individuals conducting science-related activities); these terms can be used interchangeably and so were included in our search equation. We used different terms referring to crowdsourcing, citizen science, and Web platforms. The search strategy used to search MEDLINE and EMBASE is in Multimedia Appendix 2. We also screened ClinicalTrials.gov (search strategy in Multimedia Appendix 3) and the reference lists of previous systematic reviews [5,10] and selected papers to identify additional studies.

Selection of Studies

Two reviewers (PC and GM) independently examined each title and abstract identified to exclude irrelevant reports. The 2 reviewers then independently examined full-text articles to determine eligibility. Disagreements were discussed to reach consensus. We documented the primary reason for exclusion of full-text articles. For ClinicalTrials.gov, only studies with posted results were included.

Definition of the Crowdsourcing Tasks

We used the classification described by Ranard [5] with 4 tasks of crowdsourcing: (1) problem-solving: to propose empirical solutions to scientific problems; (2) data processing: to perform several human intelligence microtasks to provide in total an analysis of a large amount of data; (3) surveillance or monitoring: to find and collect information into a common location and format such as the creation of collective resources; and (4) surveying: to answer a Web-based survey. Surveillance or monitoring and surveying belong to mining crowd data described by Khare [4] and are defined as data collected and analyzed by crowd workers for the knowledge discovery process. Problem-solving and data processing belong to active crowdsourcing, which refers to crowd workers recruited to solve scientific problems.

Data Extraction and Management

The data were extracted from reports by the two reviewers (PC and GM) who used a standardized data extraction form (provided with the protocol as Multimedia Appendix 1). Disagreements were discussed to reach consensus. From each study, we extracted the following characteristics.

Publication Characteristics of the Study

Publication characteristics of the study were as follows: *Journal Citation Reports* categories (ie, general medicine and health care science, biomedical informatics and technology, or medical specialty journals); impact factor (Clarivate Analytics); average journal impact factor percentile from *Journal Citation Reports* (classified in four categories: >90th percentile, 70th-90th percentile, <70th percentile, and not indexed); and year of publication.

Characteristics of Crowdsourcing Applications in Health

The following characteristics of crowdsourcing applications were extracted:

1. We determined the category of health the study referred to (health promotion, research, or care [12]) and health field (eg, public health, surgery, oncology [details in Multimedia Appendix 4]).
2. We classified the tasks into 1 of the 4 categories of crowdsourcing tasks defined: problem-solving, data processing, surveillance or monitoring, and surveying.
3. We determined whether the study was led by researchers (ie, a traditional study led by institutionally trained researchers) or by participants (ie, studies designed and operated by patients or citizen scientists) [10].

Logistics of Crowdsourcing and Characteristics of Crowd Workers

Considering the logistics of crowdsourcing and characteristics of crowd workers, the following points were extracted:

1. We defined how the crowdsourcing was applied: whether a large task was divided into microtasks and distributed to workers [13] or whether the same task—a high-difficulty task called a megatask, such as a challenge—was given to several groups of workers [14].
2. We extracted the type of platform used (labor markets, scientific games, mobile phone apps, social media, or community challenges with dedicated platforms) [13]; whether monetary incentives were offered and their amount; the time to perform the task; whether a data quality validation was performed; whether the task performed by the crowd workers was compared with that performed by experts (which corresponds to a feasibility study).
3. We extracted the number of crowd workers, the median age, the proportion of women, their status (eg, researchers, physicians, and students), their geographic location, their motivations, whether a skill set was required to perform the task, and whether they had to undergo training and pass a qualification test to be recruited.
4. We also assessed the proportion of studies not reporting all these data.

Analysis

The analysis was descriptive. Data are summarized as number (%) for qualitative variables and median (Q1-Q3) for continuous variables. All analyses involved the use of R v3.0.2 (R Foundation for Statistical Computing, Vienna, Austria) [15].

Results

Systematic Literature Search

The flow of study selection is in Multimedia Appendix 5. Briefly, the electronic search yielded 2354 references; 326 were

selected for further evaluation, and 202 studies were included (182 published papers and 20 conference abstracts [3,16-216]).

More than half of the included studies (108/202, 53.5%) were published during the last 2 years. The median impact factor of the journals of publication was 3.2 (Q1-Q3: 2.1-3.5); for 42/202 studies (20.8%), reports were published in a journal with very high relative impact factor (>90th percentile of journal impact factors averaged across journal categories). Reports for two-thirds of studies (129/202) were published in medical specialty journals and for one-fourth (50/202) in biomedical informatics and technology journals. All these publication characteristics are in Figure 1. A total of 9 studies corresponded to randomized controlled trials, only 1 with results posted on ClinicalTrials.gov.

Mapping of Crowdsourcing Applications in Health

Crowdsourcing applications were more frequent in studies of health promotion (91/202, 45.0%) and health research (72/202, 35.7%) than health care (39/202, 19.3%). More than half of the studies concerned active crowdsourcing (data processing (99/202, 49.0%) and problem-solving (12/202, 5.9%)) and 45% of the studies were about mining crowd data (surveying (70/202, 34.6%) and surveillance or monitoring (21/202, 10.4%)). Examples of crowdsourced tasks by health category are provided in Figure 2.

Almost 50% of the studies related to health promotion used surveys to conduct their research compared with studies related to health care, which used mainly data processing activity. All included studies were led by researchers.

Figure 1. Publication characteristics of included studies. Two-thirds of the studies have been published in one of the 18 medical specialty journals, covering almost all medical fields showing the widespread use of crowdsourcing, and sometimes in a journal with very high relative impact factor.
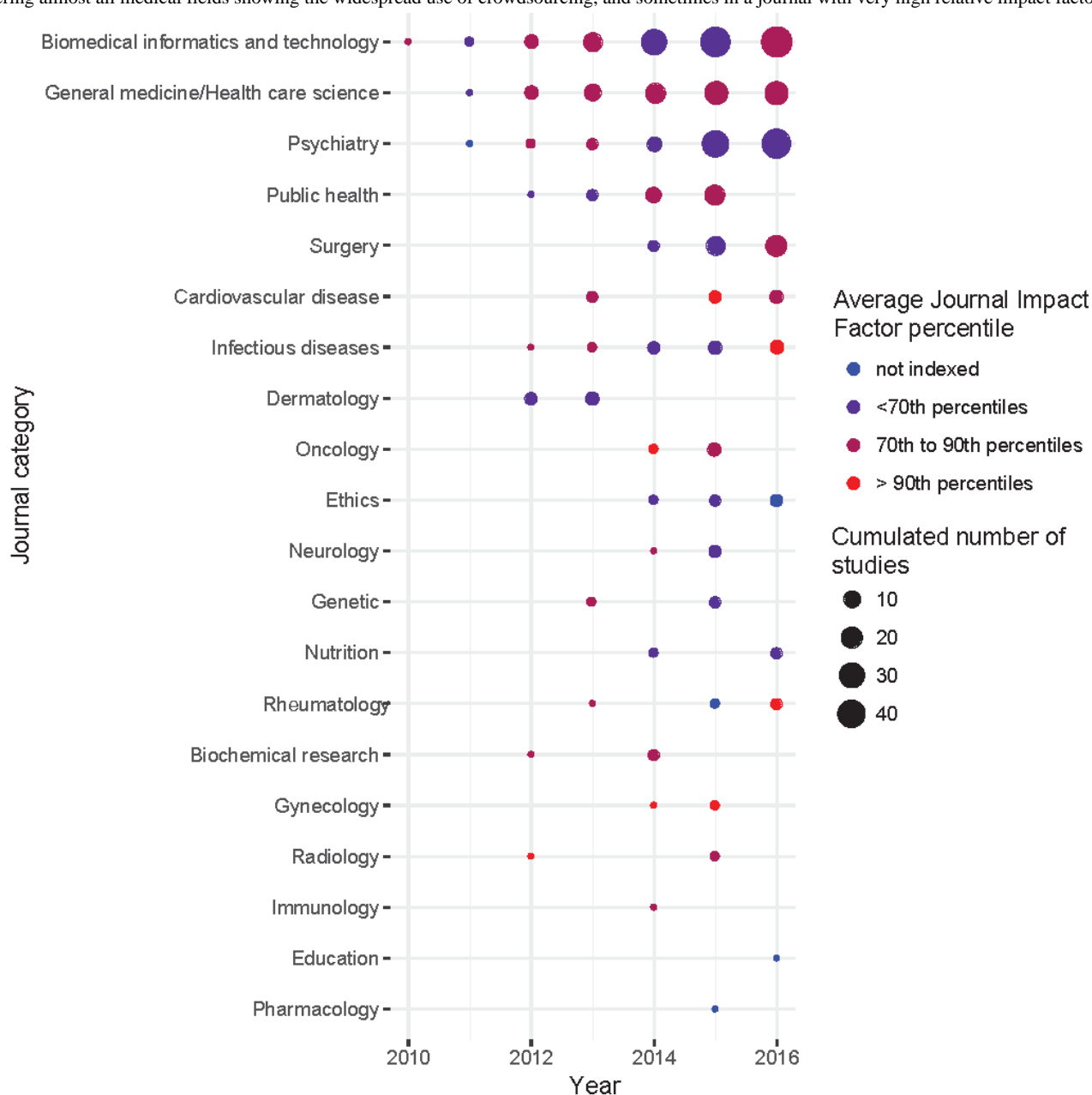
Figure 2. Examples of crowdsourced tasks according to health category. EEG: electroencephalography.

Figure 3. Mapping of crowdsourcing applications in health. Sankey diagram representing the distribution of medical fields applying crowdsourcing for each of the 4 types of task. Width of links is proportional to the number of studies. Medical specialties: anatomopathology (n=3), cardiology (n=5), dermatology (n=5), endocrinology (n=1), gynecology (n=2), infectiology (n=6), nephrology (n=1), neurology (n=7), pediatrics (n=2), pneumology (n=3), radiology (n=2), and rheumatology (n=2).
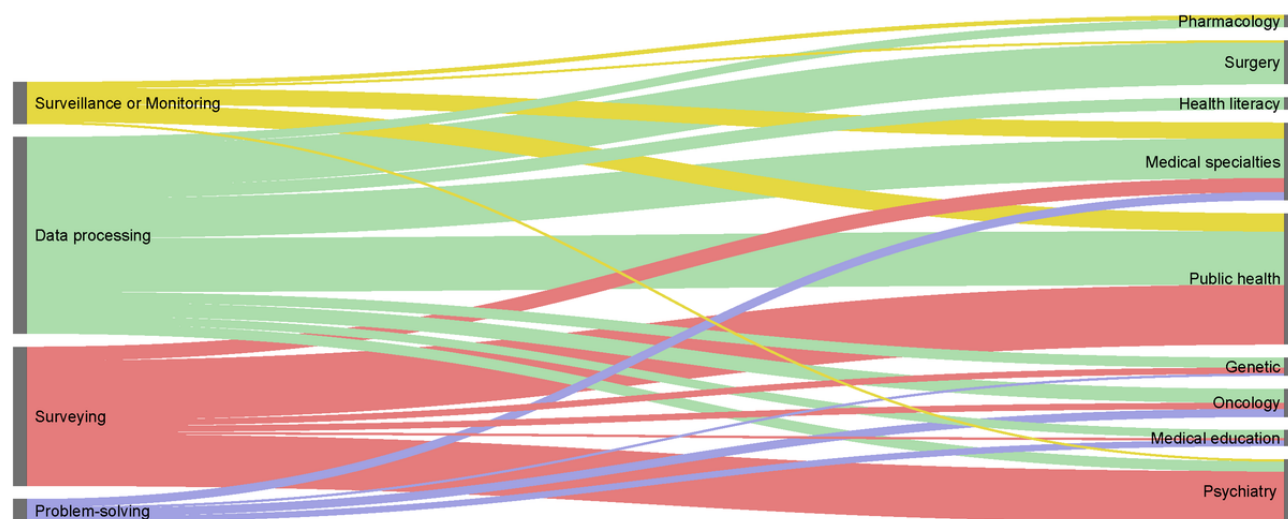
In Figure 3, we provide a mapping of crowdsourcing applications in health, detailing the medical fields applied to each type of task.

Data Processing

One-fourth of studies (27/99) involved public health, one-fifth (21/99) involved surgery, and one-fifth involved medical specialties (20/99). For example, in the Ghani et al study, published in 2016, crowd workers used the Global Evaluative Assessment of Robotic Skills tool to assess surgical skill in a video recording of a nerve-sparing robot-assisted radical prostatectomy [77].

Surveying

A total of 43% of studies (30/70) involved public health, and 37% (26/70) involved psychiatry. In the Stroh et al study, published in 2015, crowd workers completed a questionnaire related to public views on organ donation for people who need transplantation because of alcohol abuse [192]. The survey measured attitudes on liver transplantation in general and early transplantation for this patient population.

Surveillance or Monitoring

A total of 43% of studies (9/21) concerned public health and 24% (5/21) concerned dermatology. In the Merchant et al study, published in 2012, during 2 months, crowd workers had to locate, photograph, and submit the most eligible automated external defibrillator in Philadelphia [168].

Problem-Solving

One-third of studies (4/12) concerned oncology and one-fourth (3/12) concerned medical education. In the Margolin et al study, published in 2013, crowd workers were challenged during 6 months to develop computational models that predict overall survival of breast cancer patients based on clinical information [131].

Reporting of the Logistics of Crowdsourcing and Crowd Workers' Characteristics

For data processing and surveillance or monitoring, a large task was divided into microtasks and distributed to crowd workers. For problem-solving, a megatask was given to several groups of crowd workers. We identified 7 challenges in our sample. A Web platform was used in 190/202 studies (94.1%), of which 133/190 (70.0%) were labor markets (eg, Amazon MTurk; Table 1).

Crowd workers' characteristics and crowdsourcing logistics were poorly reported. Reports for almost one-fourth of studies (47/202) did not mention monetary incentives, and for two-thirds of studies (130/202), the time to perform the task was not mentioned. Crowd workers' characteristics were frequently missing: age and gender were not reported for about 60% of the studies (128/202 and 105/202, respectively), and crowd workers' location was not reported for one-fourth of the studies (50/202).

For 109/202 studies (53.9%), reports mentioned monetary incentives, mainly less than US $1 to perform a task. When reported, the time needed to perform the task was mostly less than 10 min (42/72, 58% of studies). For one-fourth of studies (54/202), reports mentioned using data quality validation, mainly by attention check questions (19/54, 35%) or by replicating the task by several crowd workers (16/54, 30%). About one-fifth of studies (36/202) compared crowd workers' performance with that of experts (corresponding in these cases to a feasibility study), mainly for evaluating surgical skills (15/36, 42%).

The number of crowd workers was reported for 176 studies (87.1%), and the size of the crowd varied from 5 to about 2 million, with median 424 (first and third quartiles Q1-Q3: 167-802; Table 2). When specified, crowd workers' median age was 34 years (Q1-Q3: 32-36) and 55% were men. Crowd workers were recruited nationally in 93/152 studies (61.2%), mainly the United States (83/93, 89%).

Table 1. Logistics of crowdsourcing in systematic review studies.

Characteristics	Statistics (N=202)
Type of platform used, n (%)	
Web platform	190 (94.0)
Labor markets[a]	133 (70.0)
Social media[b]	10 (5.3)
Labor markets and social media	2 (1.0)
Community challenge	7 (3.7)
Scientific games	6 (3.2)
Other websites	32 (16.8)
Mobile apps	12 (6.0)
Monetary incentives to perform a task , n (%)	
Not reported	47 (23.2)
Yes	109 (54.0)
≤US $0.1	15 (13.8)
US $0.2 to 0.5	30 (27.5)
US $0.6 to 1	23 (21.1)
>US $1	23 (21.1)
Amount not specified	18 (16.5)
No	46 (22.8)
Time to perform a task, n (%)	
Not reported	130 (64.4)
Reported	72 (35.6)
≤1 min	17 (24)
2 to 10 min	25 (35)
11 to 30 min	21 (29)
30 min to 1 hr	9 (12)
Data quality validation, n (%)	
Not reported	148 (73.3)
Yes	54 (26.7)
Task replication by several CWs[c]	16 (30)
Attention check questions	19 (35)
Discriminative questions	12 (22)
Limited timing for the task	4 (7)
Not specified	3 (6)
CW performance compared with experts, n (%)	
Yes	36 (17.8)

[a]Amazon MTurk, Crowdflower.

[b]Mainly Facebook, Twitter, LinkedIn, and Curetogether.

[c]CW: crowd worker.

Table 2. Characteristics of crowd workers.

Characteristics	Statistics (N=202)
Size of the crowd	
Median (Q1-Q3)	424 (167-802)
Not reported, n (%)	26 (12.9)
Reported, n (%)	176 (87.1)
<100	20 (11.4)
100-499	78 (44.3)
500-999	41 (23.3)
1000-4999	25 (14.2)
5000-10,000	7 (4.0)
>10,000	5 (2.8)
Geographic location, n (%)	
Not reported	50 (24.8)
Reported	152 (75.2)
International	59 (38.8)
National	93 (61.2)
United States	83 (89.2)
Canada	2 (2.2)
The Netherlands	3 (3.2)
Other[a]	5 (5.4)
Age	
Median (Q1-Q3)	34 (32-36)
Not reported, n (%)	128 (63.4)
Reported, n (%)	74 (36.6)
Gender	
Not reported, n (%)	105 (52)
Mean proportion of men (%)	55.0
Status, n (%)	
Not reported	61 (30.2)
Reported	141 (69.8)
Anyone	51 (36.2)
People graduated from college	35 (24.8)
People with specificities	19 (13.5)
Patients	14 (9.9)
Medical or health care providers	9 (6.4)
Researchers	8 (5.7)
Students	5 (3.5)
Skill set required, n (%)	
Not specified	128 (63.4)
Yes	74 (36.6)
Master qualification[b]	44 (60)
Speak English	14 (19)

Characteristics	Statistics (N=202)
Scientific background	8 (11)
Medical background	8 (10)
Qualification test, n (%)	
Not reported	34 (16.8)
Yes	26 (12.9)
No	142 (70.3)
Training of workers, n (%)	
Not reported	31 (15.3)
Yes	22 (10.9)
No	149 (73.8)

[a]India, Australia, Israel, China, and South Korea.

[b]Defined as "consistently completing *human intelligence tasks* of a certain type with a high degree of accuracy across a variety of requesters."

The motivations of crowd workers were recorded for 5/202 studies (2.5%) and included fun, curiosity, altruism, compensation, contribution to an important cause, personal reasons, research education, and advancing science [82,139,168,183,193]. A skill set was required in 74/202 studies (36.7%); for 60%, this involved previous experience in crowdsourcing. For two-thirds of studies (128/202), a specific skill set required was not specified. For only 12.8% of studies, the Web-based tasks (26/202) required passing a qualification test, and for 10.9%, (22/202), they required training.

Discussion

Principal Findings

In this systematic review of the use of crowdsourcing in studies of health promotion, research, and care, we included 202 studies, mainly published in the last 2 years with for one-fifth of a publication in a journal with very high relative IF. Data processing was the most frequent type of task used (mainly in public health and surgery), followed by surveying (public health and psychiatry), then surveillance or monitoring (public health and dermatology), and finally problem-solving (oncology). Labor market platforms (Amazon MTurk) were mainly used. The description of crowdsourcing logistics and crowd workers' characteristics were frequently missing from reports. When reported, the median size of the crowd was less than 500; crowd workers' median age was around 34 years and 55% were men. Crowd workers were mainly recruited in the United States. A previous experience in crowdsourcing was required in about 60% of the studies, whereas passing a qualification test or training was only needed in about 12%. The time needed to perform the task was mostly less than 10 min for monetary incentives less than US $1. Data quality validation was used in less than one-third of studies.

Our systematic review has advantages over previous ones on the same topic [5,10]. The systematic review conducted by Ranard et al in March 2013 described the scope of crowdsourcing in health and medical research but included only 21 articles [5]. The narrative review conducted by Swan described the use of crowdsourcing in health research studies up to 2011 [10]. Our mapping is more exhaustive—focused on health research but also health promotion and health care—and up-to-date. Many of our studies (80%) were published after the last search date of the Ranard et al's systematic review [5]. This point highlights the increasing use of crowdsourcing in health during the last few years. Indeed, many health fields have since used crowdsourcing, with 20 medical fields identified in our systematic review compared with 8 fields in the Ranard et al's study [5]. Moreover, crowdsourcing use is still growing, as shown by the 11 articles published in *Journal of Medical Internet Research* since our last search date, mainly involving a survey task (9/11, 82%) [217-227]. Our study has some limitations. First, we did not search the gray literature to identify some unpublished studies. However, the EMBASE search allowed us to identify 20 studies (10%) corresponding to conference abstracts. Second, we did not search Google Scholar because of the number of records found (about 30,000). Screening all these references would be extremely time-consuming for only 2 reviewers without using a crowdsourcing process. Third, we did not include studies related to biology, such as studies using the "Fold it" platform to solve protein-folding problems [228]. We did not consider this topic in our definition of health. Finally, we included only crowdsourcing performed via the internet. For example, we did not include studies in which the crowdsourced tasks were performed in a particular workshop without individual data collected online. Therefore, we may have underestimated the number of studies using crowdsourcing in health.

Every health category (promotion, research, and care) has a potential need for human computing power that crowdsourcing could fulfill to accelerate the process. Our systematic review, focusing on peer-reviewed papers, may have not captured some kinds of crowdsourcing. Studies recruiting crowd workers with social media platforms were few in our selection (12/202 studies [5.9%]). This type of recruitment seems less attractive than labor markets, although it is free and easier to use, perhaps because it is considered less reliable or used for purposes other than publication. Another way of exploiting social media data is under development, whereby tweets referring to a specific disease are analyzed as part of a health maintenance approach

XSL•FO
RenderX

(eg, HIV in the Adrover et al's study to identify adverse effects of drug treatment in tweets using crowdsourcing [229]). Considering health research, a fundamental aspect of this crowdsourcing is that it allows research to be performed *with* patients and not only *to* them or *on* them. However, studies with patients as crowd workers represented only 10% of our included studies, perhaps because the primary aim of collecting these data was not to conduct research with the data. Nevertheless, in 2013, the PatientsLikeMe platform [230] had more than 220,000 members sharing health data on more than 2000 diseases and conditions [231]. Using these data and conducting research with the data represent a great future challenge of mining crowd data and a real opportunity to collect large amounts of data on symptoms of diseases, drug efficacy, or adverse events to solve a wide range of health issues with a more real-life approach. Crowdsourcing also has potential in health promotion, especially preventive medicine, by taking it one step further. For example, specific tips in the form of slides or films could be added to the end of a Web-based survey about addiction to conduct a behavioral intervention, in addition to a simple survey. In some cases, data processing tasks may require thinking about a healthier lifestyle, for example, by suggesting healthier alternatives in addition to gathering information on the nutritional characteristics of packaged foods. Such crowdsourced tasks could be expanded to change dietary behaviors, exercise, or adherence to treatment. Finally, the combination of crowdsourcing and mobile health technologies could be the ultimate step in providing an ideal vehicle for behavioral interventions that can reach users in real time, in real life, without being resource-intensive.

Crowdsourcing allows for a large number of crowd workers to be mobilized in record time and at low cost. For instance, in Peabody et al's study [158], experts completed 318 video ratings in 15 days, but crowd workers completed 2531 ratings in 21 hours. These crowdsourced resources might be further harnessed in a world of high health costs. Crowdsourcing also allows for speeding up innovations, when used in the form of collaborative scientific competitions—challenges—to solve diverse and important biomedical problems. Problem-solving was the fourth task we identified in terms of frequency, and only 7 challenges were individualized, perhaps because challenges are an emerging form of crowdsourcing, which should be more prominent in the next few years and lead to more publications [232]. In future, it will be necessary to facilitate and promote the use of this type of crowdsourced tasks in health research, given the amount of data to be considered (big data) and the complexity of medical issues that will require increasingly skilled and qualified individuals to resolve them.

As previously mentioned, crowdsourcing has many advantages: improved cost, speed, quality, flexibility, scalability, and diversity. However, some points that remain controversial include the impact of crowdsourcing on product quality or its unethical aspect. The first remaining potential concern of crowdsourced studies in health is the validity of their results. Some studies have assessed whether we should trust Web-based studies, and it appears that the data provided by internet methods have at least as good quality as those provided by traditional paper-and-pencil methods [233]. In our review, for data

processing tasks, 36/202 feasibility studies (17.8%) compared crowd workers' performance with that of an expert group considered as reference. These studies mainly considered surgical skills evaluation (15/36, 42%) and parasite identification in infectious diseases (4/36, 11%). At each time, the performance of crowd workers was similar to that of the reference group. However, because the participation is anonymous and compensated, participants may provide unsatisfactory quality data. In our review, 54/202 studies (26.7%) reported using data quality validation. Several types of validation techniques were found, from inserting random questions with known answers into the task, to screening for crowd workers who were incorrectly marking answers (31/54, 57%) and to comparing responses among multiple crowd workers to discard outliers (16/54, 30%). The second concern is its unethical aspect: Amazon MTurk is a bargain for researchers but not for crowd workers [234]. Indeed, many MTurk tasks are completed by a small set of workers who spend long hours on the website, many with low income.

A detailed description of the crowdsourcing logistics in the Methods section and all the characteristics of the crowd workers (population of the study) should be provided in high-quality research, even if its importance depends on the type of study. In cases of surveying and surveillance or monitoring studies related to illness, crowd workers' characteristics need to be precisely described to better interpret the study findings and to judge the external validity. In cases of data processing and problem-solving, crowd workers' characteristics also need to be reported to allow reproducibility of studies and to select more quickly and more easily the best population of crowd workers for a future similar study. In our review, the lack of details of crowd workers' characteristics in one-third of the included studies impedes the interpretation of results of these studies. Rather than being a virtually infinite subject pool, crowd workers are far less diverse than was previously thought. As we found, although crowd workers should be recruited from all over the world, 61% were actually recruited nationally, mainly the United States (89%). Previously, crowd workers were mainly young, urban, and single and more often had postsecondary education [6]. In our review, the median age of crowd workers was 34 years, 55% were men, and half reported a high level of education. Therefore, logistics of crowdsourcing and crowd workers' characteristics must be reported, and standardized guidelines on crowdsourcing metrics that needed to be collected and reported could be useful to improve the quality of such studies.

Conclusions

Crowdsourcing appears to be a trendy, efficient, competitive, and useful tool to improve health actions, whether in preventive medicine, research, or care. Its use in health is increasing, particularly in public health, psychiatry, surgery, and oncology. Crowdsourcing allows for access to a large pool of participants, saves time to collect data, lowers costs, and speeds up innovations. Each health field could benefit from some tasks that could be crowdsourced to facilitate advances in research. To optimize the use of crowdsourcing in health, the logistics of crowdsourcing and crowd workers' characteristics must be reported.

Acknowledgments

Authors' Contributions

PC was involved in the study conception, selection of trials, data extraction, data analysis, interpretation of results, and drafting the manuscript and revision. GM was involved in the study conception, the selection of trials, and data extraction. MB was involved in the interpretation of results and drafting the manuscript. AV was involved in the study conception, data analysis, interpretation of results, and drafting the manuscript and revision. PR was involved in the study conception, interpretation of results, and drafting the manuscript and revision. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Protocol of the systematic review.

[PDF File (Adobe PDF File), 12KB - jmir_v20i5e187_app1.pdf]

Multimedia Appendix 2

Search terms for MEDLINE and EMBASE (March 30, 2016).

[PDF File (Adobe PDF File), 241KB - jmir_v20i5e187_app2.pdf]

Multimedia Appendix 3

Search strategy for ClinicalTrials.gov.

[PDF File (Adobe PDF File), 17KB - jmir_v20i5e187_app3.pdf]

Multimedia Appendix 4

Details of the health fields considered.

[PDF File (Adobe PDF File), 31KB - jmir_v20i5e187_app4.pdf]

Multimedia Appendix 5

Flow diagram of selection of studies applying crowdsourcing in health.

[PDF File (Adobe PDF File), 157KB - jmir_v20i5e187_app5.pdf]

References

1. Sobel D. Longitude: The True Story of a Lone Genius Who Solved the Greatest Scientific Problem of His Time. New York: Walker Publishing Company; 1995.
2. Wikipedia. Crowdsourcing URL: https://en.wikipedia.org/wiki/Crowdsourcing [accessed 2018-02-06] [WebCite Cache ID 6x1xeHjIX]
3. Deal SB, Lendvay TS, Haque MI, Brand T, Comstock B, Warren J, et al. Crowd-sourced assessment of technical skills: an opportunity for improvement in the assessment of laparoscopic surgical skills. Am J Surg 2016 Feb;211(2):398-404. [doi: 10.1016/j.amjsurg.2015.09.005] [Medline: 26709011]
4. Khare R, Good BM, Leaman R, Su AI, Lu Z. Crowdsourcing in biomedicine: challenges and opportunities. Brief Bioinform 2016 Jan;17(1):23-32. [doi: 10.1093/bib/bbv021] [Medline: 25888696]
5. Ranard BL, Ha YP, Meisel ZF, Asch DA, Hill SS, Becker LB, et al. Crowdsourcing--harnessing the masses to advance health and medicine, a systematic review. J Gen Intern Med 2014 Jan;29(1):187-203 [FREE Full text] [doi: 10.1007/s11606-013-2536-8] [Medline: 23843021]

6. Bohannon J. Psychology. Mechanical Turk upends social sciences. Science 2016 Jun 10;352(6291):1263-1264. [doi: 10.1126/science.352.6291.1263] [Medline: 27284175]

7. Dawson R, Bynghall S. Getting Results From Crowds: The Definitive Guide to Using Crowdsourcing to Grow Your Business. San Francisco: Advanced Human Technologies; 2012.

8. Lintott C, Schawinski K, Bamford S, Slosar A, Land K, Thomas D, et al. Galaxy Zoo 1: data release of morphological classifications for nearly 900 000 galaxies. Mon Not R Astron Soc 2011 Jan 01;410(1):166-178. [doi: 10.1111/j.1365-2966.2010.17432.x]

9. Marris E. Supercomputing for the birds. Nature 2010 Aug 12;466(7308):807. [doi: 10.1038/466807a] [Medline: 20703280]

10. Swan M. Crowdsourced health research studies: an important emerging complement to clinical trials in the public health research ecosystem. J Med Internet Res 2012;14(2):e46 [FREE Full text] [doi: 10.2196/jmir.1988] [Medline: 22397809]

11. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 2009 Oct;62(10):e1-34 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.06.006] [Medline: 19631507]

12. Prpić PJ. Health Care Crowds: Collective Intelligence in Public Health Internet. Rochester, NY: Social Science Research Network; 2015.

13. Michelucci P, Dickinson JL. Human Computation. The power of crowds. Science 2016 Jan 01;351(6268):32-33. [doi: 10.1126/science.aad6499] [Medline: 26721991]

14. Silberzahn R, Uhlmann EL. Crowdsourced research: many hands make tight work. Nature 2015 Oct 08;526(7572):189-191. [doi: 10.1038/526189a] [Medline: 26450041]

15. R-project. Vienna, Austria: R Foundation for Statistical Computing R: A language and environment for statistical computing URL: https://www.r-project.org/ [accessed 2018-04-05] [WebCite Cache ID 6yS5IavS3]

16. Adams SA. Using patient-reported experiences for pharmacovigilance? Stud Health Technol Inform 2013;194:63-68. [Medline: 23941932]

17. Aghdasi N, Bly R, White LW, Hannaford B, Moe K, Lendvay TS. Crowd-sourced assessment of surgical skills in cricothyrotomy procedure. J Surg Res 2015 Jun 15;196(2):302-306. [doi: 10.1016/j.jss.2015.03.018] [Medline: 25888499]

18. Albarqouni S, Baur C, Achilles F, Belagiannis V, Demirci S, Navab N. AggNet: deep learning from crowds for mitosis detection in breast cancer histology images. IEEE Trans Med Imaging 2016 Dec;35(5):1313-1321. [doi: 10.1109/TMI.2016.2528120] [Medline: 26891484]

19. Kuerbis A, Muench F. Normative feedback for adult drinkers: intervention through developing discrepancy or by validating pre-existing worries? 2015 Presented at: Personalized Feedback Interventions for Individuals with Substance-Related Problems; January 14-18, 2015; New Orleans, LA.

20. Allam A, Schulz PJ, Nakamoto K. The impact of search engine selection and sorting criteria on vaccination beliefs and attitudes: two experiments manipulating Google output. J Med Internet Res 2014 Apr;16(4):e100 [FREE Full text] [doi: 10.2196/jmir.2642] [Medline: 24694866]

21. Alvare G, Gordon R. CT brush and CancerZap!: two video games for computed tomography dose minimization. Theor Biol Med Model 2015 May 12;12(1):7. [doi: 10.1186/s12976-015-0003-4]

22. 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]

23. Andover MS. Non-suicidal self-injury disorder in a community sample of adults. Psychiatry Res 2014 Oct;219(2):305-310. [doi: 10.1016/j.psychres.2014.06.001]

24. Arditte KA, Morabito DM, Shaw AM, Timpano KR. Interpersonal risk for suicide in social anxiety: the roles of shame and depression. Psychiatry Res 2016 May;239:139-144. [doi: 10.1016/j.psychres.2016.03.017]

25. Armstrong AW, Cheeney S, Wu J, Harskamp CT, Schupp CW. Harnessing the power of crowds. Am J Clin Dermatol 2012;13(6):405-416. [doi: 10.2165/11634040-000000000-00000]

26. Armstrong AW, Harskamp CT, Cheeney S, Schupp CW. Crowdsourcing for research data collection in rosacea. Dermatol Online J 2012 Mar 15;18(3):15. [Medline: 22483526]

27. Armstrong AW, Harskamp CT, Cheeney S, Wu J, Schupp CW. Power of crowdsourcing: novel methods of data collection in psoriasis and psoriatic arthritis. J Am Acad Dermatol 2012 Dec;67(6):1273-1281.e9. [doi: 10.1016/j.jaad.2012.05.013] [Medline: 22818792]

28. Armstrong AW, Harskamp CT, Cheeney S, Schupp CW. Crowdsourcing in eczema research: a novel method of data collection. J Drugs Dermatol 2012 Oct;11(10):1153-1155. [Medline: 23285714]

29. Armstrong AW, Wu J, Harskamp CT, Cheeney S, Schupp CW. Crowdsourcing for data collection: a pilot study comparing patient-reported experiences and clinical trial data for the treatment of seborrheic dermatitis. Skin Res Technol 2013 Feb;19(1):55-57. [doi: 10.1111/j.1600-0846.2012.00667.x] [Medline: 22891700]

30. Bahk CY, Goshgarian M, Donahue K, Freifeld CC, Menone CM, Pierce CE, et al. Increasing patient engagement in pharmacovigilance through online community outreach and mobile reporting applications: an analysis of adverse event reporting for the Essure device in the US. Pharmaceut Med 2015;29(6):331-340 [FREE Full text] [doi: 10.1007/s40290-015-0106-6] [Medline: 26635479]

31. Baruch RL, Vishnevsky B, Kalman T. Split-care patients and their caregivers: how collaborative is collaborative care? J Nerv Ment Dis 2015 Jun;203(6):412-417 [FREE Full text] [doi: 10.1097/NMD.0000000000000297] [Medline: 25938507]

32. Bell RA, McGlone MS, Dragojevic M. Bacteria as bullies: effects of linguistic agency assignment in health message. J Health Commun 2014 Sep;19(3):340-358. [doi: 10.1080/10810730.2013.798383] [Medline: 24015807]

33. Bevelander KE, Kaipainen K, Swain R, Dohle S, Bongard JC, Hines PD, et al. Crowdsourcing novel childhood predictors of adult obesity. PLoS One 2014 Feb;9(2):e87756 [FREE Full text] [doi: 10.1371/journal.pone.0087756] [Medline: 24505310]

34. Bickel WK, George WA, Franck CT, Terry ME, Jarmolowicz DP, Koffarnus MN, et al. Using crowdsourcing to compare temporal, social temporal, and probability discounting among obese and non-obese individuals. Appetite 2014 Apr;75:82-89 [FREE Full text] [doi: 10.1016/j.appet.2013.12.018] [Medline: 24380883]

35. Bickel WK, Jarmolowicz DP, Mueller ET, Franck CT, Carrin C, Gatchalian KM. Altruism in time: social temporal discounting differentiates smokers from problem drinkers. Psychopharmacology (Berl) 2012 Nov;224(1):109-120. [doi: 10.1007/s00213-012-2745-6] [Medline: 22644127]

36. Borgida E, Loken B, Williams AL, Vitriol J, Stepanov I, Hatsukami D. Assessing constituent levels in smokeless tobacco products: a new approach to engaging and educating the public. Nicotine Tob Res 2015 Nov;17(11):1354-1361 [FREE Full text] [doi: 10.1093/ntr/ntv007] [Medline: 25634934]

37. Bow HC, Dattilo JR, Jonas AM, Lehmann CU. A crowdsourcing model for creating preclinical medical education study tools. Acad Med 2013 Jun;88(6):766-770. [doi: 10.1097/ACM.0b013e31828f86ef] [Medline: 23619061]

38. Boynton MH, Richman LS. An online daily diary study of alcohol use using Amazon's Mechanical Turk. Drug Alcohol Rev 2014 Jul;33(4):456-461 [FREE Full text] [doi: 10.1111/dar.12163] [Medline: 24893885]

39. Brady CJ, Villanti AC, Pearson JL, Kirchner TR, Gupta OP, Shah CP. Rapid grading of fundus photographs for diabetic retinopathy using crowdsourcing. J Med Internet Res 2014 Oct 30;16(10):e233 [FREE Full text] [doi: 10.2196/jmir.3807] [Medline: 25356929]

40. Brooks SC, Simmons G, Worthington H, Bobrow BJ, Morrison LJ. The PulsePoint Respond mobile device application to crowdsource basic life support for patients with out-of-hospital cardiac arrest: challenges for optimal implementation. Resuscitation 2016 Jan;98:20-26. [doi: 10.1016/j.resuscitation.2015.09.392] [Medline: 26475397]

41. Brown AW, Allison DB. Using crowdsourcing to evaluate published scientific literature: methods and example. PLoS One 2014;9(7):e100647 [FREE Full text] [doi: 10.1371/journal.pone.0100647] [Medline: 24988466]

42. Burger JD, Doughty E, Khare R, Wei CH, Mishra R, Aberdeen J, et al. Hybrid curation of gene-mutation relations combining automated extraction and crowdsourcing. Database (Oxford) 2014;2014: [FREE Full text] [doi: 10.1093/database/bau094] [Medline: 25246425]

43. Cabrera LY, Fitz NS, Reiner PB. Reasons for comfort and discomfort with pharmacological enhancement of cognitive, affective, and social domains. Neuroethics 2014 Oct 10;8(2):93-106. [doi: 10.1007/s12152-014-9222-3]

44. Campisi J, Folan D, Diehl G, Kable T, Rademeyer C. Social media users have different experiences, motivations, and quality of life. Psychiatry Res 2015 Aug 30;228(3):774-780. [doi: 10.1016/j.psychres.2015.04.042] [Medline: 26054935]

45. Candido Dos Reis FJ, Lynn S, Ali HR, Eccles D, Hanby A, Provenzano E, et al. Crowdsourcing the general public for large scale molecular pathology studies in cancer. EBioMedicine 2015 Jul;2(7):681-689 [FREE Full text] [doi: 10.1016/j.ebiom.2015.05.009] [Medline: 26288840]

46. Carter RR, DiFeo A, Bogie K, Zhang GQ, Sun J. Crowdsourcing awareness: exploration of the ovarian cancer knowledge gap through Amazon Mechanical Turk. PLoS One 2014 Jan;9(1):e85508 [FREE Full text] [doi: 10.1371/journal.pone.0085508] [Medline: 24465580]

47. Chaitoff A, Niforatos J, Vega J. Exploring the effects of medical trainee naming: a randomized experiment. Perspect Med Educ 2016 Apr;5(2):114-121 [FREE Full text] [doi: 10.1007/s40037-016-0260-x] [Medline: 26975743]

48. Chan TM, Thoma B, Lin M. Creating, curating, and sharing online faculty development resources: the medical education in cases series experience. Acad Med 2015 Jun;90(6):785-789. [doi: 10.1097/ACM.0000000000000692] [Medline: 25785678]

49. Chen C, White L, Kowalewski T, Aggarwal R, Lintott C, Comstock B, et al. Crowd-sourced assessment of technical skills: a novel method to evaluate surgical performance. J Surg Res 2014 Mar;187(1):65-71. [doi: 10.1016/j.jss.2013.09.024] [Medline: 24555877]

50. Chenette P, Martinez C. Synchronization of women's cycles: a big data and crowdsourcing approach to menstrual cycle analysis. Fertil Steril 2014 Sep;102(3):e250. [doi: 10.1016/j.fertnstert.2014.07.853]

51. Chen SP, Kirsch S, Zlatev DV, Chang T, Comstock B, Lendvay TS, et al. Optical biopsy of bladder cancer using crowd-sourced assessment. J Am Med Assoc Surg 2016 Jan;151(1):90-93. [doi: 10.1001/jamasurg.2015.3121] [Medline: 26422334]

52. Cheung SY, Delfabbro P. Who is a cancer survivor? A study into the lay understanding of cancer identity using a crowdsourced population. Ann Oncol 2015 Dec 19;26(suppl 9):ix122-ix123. [doi: 10.1093/annonc/mdv531.43]

53. Aaron HJ, Deepti A, Bill C, Eyler AA, Pless RB. 2013. Emerging Technologies: Webcams and Crowd-Sourcing to Identify Active Transportation URL: https://openscholarship.wustl.edu/brown_facpubs/3/ [accessed 2018-04-05] [WebCite Cache ID 6yS7MZKXZ]

54. Chunara R, Chhaya V, Bane S, Mekaru SR, Chan EH, Freifeld CC, et al. Online reporting for malaria surveillance using micro-monetary incentives, in urban India 2010-2011. Malar J 2012 Feb 13;11:43 [FREE Full text] [doi: 10.1186/1475-2875-11-43] [Medline: 22330227]

55. Coley HL, Sadasivam RS, Williams JH, Volkman JE, Schoenberger YM, Kohler CL, National Dental PBRN and QUITPRIMO Collaborative Group. Crowdsourced peer- versus expert-written smoking-cessation messages. Am J Prev Med 2013 Nov;45(5):543-550 [FREE Full text] [doi: 10.1016/j.amepre.2013.07.004] [Medline: 24139766]

56. Collier N, Son NT, Nguyen NM. OMG U got flu? Analysis of shared health messages for bio-surveillance. J Biomed Semantics 2011;2 Suppl 5:S9 [FREE Full text] [doi: 10.1186/2041-1480-2-S5-S9] [Medline: 22166368]

57. Corpas M, Valdivia-Granda W, Torres N, Greshake B, Coletta A, Knaus A, et al. Crowdsourced direct-to-consumer genomic analysis of a family quartet. BMC Genomics 2015 Nov 07;16:910 [FREE Full text] [doi: 10.1186/s12864-015-1973-7] [Medline: 26547235]

58. Corrigan PW, Bink AB, Fokuo JK, Schmidt A. The public stigma of mental illness means a difference between you and me. Psychiatry Res 2015 Mar 30;226(1):186-191. [doi: 10.1016/j.psychres.2014.12.047] [Medline: 25660735]

59. Crangle CE, Kart JB. A questions-based investigation of consumer mental-health information. PeerJ 2015;3:e867 [FREE Full text] [doi: 10.7717/peerj.867] [Medline: 25870768]

60. McGregor JA. 2014. Crowdsourced Analysis of GBS Perinatal Disease as a Sexually Transmissible Infection (STI) Underscores Need for GBS Vaccine and Patient Education regarding GBS as an STI to Be Able to Make Well-Informed Sexual Practice Choices URL: https://cdc.confex.com/cdc/std2014/webprogram/Paper34267.html [accessed 2018-04-05] [WebCite Cache ID 6yS7fL0Aa]

61. See C. Crowdsourcing in medical education: a visual example using anatomical henna. FASEB J 2014;28(supplement 1) [FREE Full text]

62. Li L, Adam P, Townsend A, Koehn C, Memetovic J, Esdaile J. Crowdsourcing Priority Setting: A Survey of Canadians Priorities and Views about Using Digital Media in Arthritis Prevention and Treatment. 2015 Jul Presented at: J Rheumatol Conference: 70th Annual Meeting of the Canadian-Rheumatology-Association (CRA); 2015; Quebec, Canada.

63. Sreih A, Aldaghlawi F. 2013. Crowdsourcing Using The Audience Response System To Solve Medical Problems: A Pilot Study URL: http://acrabstracts.org/abstract/crowdsourcing-using-the-audience-response-system-to-solve-medical-problems-a-pilot-study/ [accessed 2018-04-05] [WebCite Cache ID 6yS8BCvMs]

64. Crump MJ, McDonnell JV, Gureckis TM. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. PLoS One 2013 Mar;8(3):e57410 [FREE Full text] [doi: 10.1371/journal.pone.0057410] [Medline: 23516406]

65. Cui L, Carter R, Zhang GQ. Evaluation of a novel conjunctive exploratory navigation Interface for consumer health information: a crowdsourced comparative study. J Med Internet Res 2014 Feb;16(2):e45 [FREE Full text] [doi: 10.2196/jmir.3111] [Medline: 24513593]

66. Dakun T, Rui Z, Jinbo S, Wei Q. Sleep spindle detection using deep learning: a validation study based on crowdsourcing. Conf Proc IEEE Eng Med Biol Soc 2015 Aug;2015:2828-2831. [doi: 10.1109/EMBC.2015.7318980] [Medline: 26736880]

67. Dart RC, Surratt HL, Le Lait M, Stivers Y, Bebarta VS, Freifeld CC, et al. Diversion and illicit sale of extended release tapentadol in the United States. Pain Med 2016 Aug;17(8):1490-1496 [FREE Full text] [doi: 10.1093/pm/pnv032] [Medline: 26814267]

68. Dasgupta N, Freifeld C, Brownstein JS, Menone CM, Surratt HL, Poppish L, et al. Crowdsourcing black market prices for prescription opioids. J Med Internet Res 2013 Aug 16;15(8):e178 [FREE Full text] [doi: 10.2196/jmir.2810] [Medline: 23956042]

69. David SV, Hayden BY. Neurotree: a collaborative, graphical database of the academic genealogy of neuroscience. PLoS One 2012 Oct;7(10):e46608 [FREE Full text] [doi: 10.1371/journal.pone.0046608] [Medline: 23071595]

70. De Weger LA, Hiemstra PS, Op den Buysch E, van Vliet AJ. Spatiotemporal monitoring of allergic rhinitis symptoms in The Netherlands using citizen science. Allergy 2014 Aug;69(8):1085-1091 [FREE Full text] [doi: 10.1111/all.12433] [Medline: 24888457]

71. Donaldson CD, Siegel JT, Crano WD. Nonmedical use of prescription stimulants in college students: attitudes, intentions, and vested interest. Addict Behav 2016 Feb;53:101-107. [doi: 10.1016/j.addbeh.2015.10.007] [Medline: 26476006]

72. Douzgou S, Pollalis YA, Vozikis A, Patrinos GP, Clayton-Smith J. Collaborative crowdsourcing for the diagnosis of rare genetic syndromes: the DYSCERNE experience. Public Health Genomics 2016 Oct;19(1):19-24. [doi: 10.1159/000440710] [Medline: 26447648]

73. Dunford E, Trevena H, Goodsell C, Ng KH, Webster J, Millis A, et al. FoodSwitch: a mobile phone app to enable consumers to make healthier food choices and crowdsourcing of National Food Composition Data. JMIR Mhealth Uhealth 2014 Aug 21;2(3):e37 [FREE Full text] [doi: 10.2196/mhealth.3230] [Medline: 25147135]

74. Dye T, Li D, Demment M, Groth S, Fernandez D, Dozier A, et al. Sociocultural variation in attitudes toward use of genetic information and participation in genetic research by race in the United States: implications for precision medicine. J Am Med Inform Assoc 2016 Jul;23(4):782-786 [FREE Full text] [doi: 10.1093/jamia/ocv214] [Medline: 26984047]

75. Fischetti C, Obidegwu A, Love SM. Crowdsourcing the collateral damage from breast cancer treatment. Ann Epidemiol 2014 Sep;24(9):683. [doi: 10.1016/j.annepidem.2014.06.013]

76.   Gesser-Edelsburg A, Shir-Raz Y, Hayek S, Sassoni-Bar Lev O. What does the public know about Ebola? The public's risk perceptions regarding the current Ebola outbreak in an as-yet unaffected country. Am J Infect Control 2015 Jul 01;43(7):669-675. [doi: 10.1016/j.ajic.2015.03.005] [Medline: 25920703]

77.   Ghani KR, Miller DC, Linsell S, Brachulis A, Lane B, Sarle R, Michigan Urological Surgery Improvement Collaborative. Measuring to improve: peer and crowd-sourced assessments of technical skill with robot-assisted radical prostatectomy. Eur Urol 2016 Apr;69(4):547-550. [doi: 10.1016/j.eururo.2015.11.028] [Medline: 26755338]

78.   Gipson J, Kahane G, Savulescu J. Attitudes of lay people to withdrawal of treatment in brain damaged patients. Neuroethics 2014 Jan;7:1-9 [FREE Full text] [doi: 10.1007/s12152-012-9174-4] [Medline: 24600485]

79.   Golding S, Nadorff MR, Winer ES, Ward KC. Unpacking sleep and suicide in older adults in a combined online sample. J Clin Sleep Med 2015 Dec 15;11(12):1385-1392 [FREE Full text] [doi: 10.5664/jcsm.5270] [Medline: 26194726]

80.   Gonzales L, Davidoff KC, DeLuca JS, Yanos PT. The mental illness microaggressions scale-perpetrator version (MIMS-P): reliability and validity. Psychiatry Res 2015 Sep 30;229(1-2):120-125. [doi: 10.1016/j.psychres.2015.07.061] [Medline: 26233830]

81.   Good BM, Loguercio S, Griffith OL, Nanis M, Wu C, Su AI. The cure: design and evaluation of a crowdsourcing game for gene selection for breast cancer survival prediction. JMIR Serious Games 2014 Jul 29;2(2):e7 [FREE Full text] [doi: 10.2196/games.3350] [Medline: 25654473]

82.   Good BM, Nanis M, Wu C, Su AI. Microtask crowdsourcing for disease mention annotation in PubMed abstracts. Pac Symp Biocomput 2015:282-293 [FREE Full text] [Medline: 25592589]

83.   Gore WL, Widiger TA. Assessment of dependency by the FFDI: comparisons to the PID-5 and maladaptive agreeableness. Personal Ment Health 2015 Nov;9(4):258-276. [doi: 10.1002/pmh.1308] [Medline: 26333624]

84.   Gornick MC, Ryan KA, Kim SY. Impact of non-welfare interests on willingness to donate to biobanks: an experimental survey. J Empir Res Hum Res Ethics 2014 Oct;9(4):22-33 [FREE Full text] [doi: 10.1177/1556264614544277] [Medline: 25747294]

85.   Gottlieb A, Hoehndorf R, Dumontier M, Altman RB. Ranking adverse drug reactions with crowdsourcing. J Med Internet Res 2015 Mar 23;17(3):e80 [FREE Full text] [doi: 10.2196/jmir.3962] [Medline: 25800813]

86.   Grasso KL, Bell RA. Understanding health information seeking: a test of the risk perception attitude framework. J Health Commun 2015 Jul;20(12):1406-1414. [doi: 10.1080/10810730.2015.1018634] [Medline: 26161622]

87.   Greshake B, Bayer PE, Rausch H, Reda J. openSNP--a crowdsourced web resource for personal genomics. PLoS One 2014 Mar;9(3):e89204 [FREE Full text] [doi: 10.1371/journal.pone.0089204] [Medline: 24647222]

88.   Guinney J, Wang T, Bare C, Norman T, Bot B, Shen L, et al. Abstract B204: a DREAM Challenge to improve prognostic models in patients with metastatic castrate-resistant prostate cancer. Mol Cancer Ther 2016 Jan 07;14(12 Supplement 2):B204-B204. [doi: 10.1158/1535-7163.TARG-15-B204]

89.   Hand J, Heil SH, Sigmon SC, Higgins ST. Cigarette smoking and other behavioral risk factors related to unintended pregnancy. Drug Alcohol Depend 2015 Jan;146:e134. [doi: 10.1016/j.drugalcdep.2014.09.282]

90.   Harber P, Leroy G. Assessing work-asthma interaction with Amazon Mechanical Turk. J Occup Environ Med 2015 Apr;57(4):381-385. [doi: 10.1097/JOM.0000000000000360] [Medline: 25851185]

91.   Harris JK, Mart A, Moreland-Russell S, Caburnay CA. Diabetes topics associated with engagement on Twitter. Prev Chronic Dis 2015 May 07;12:E62 [FREE Full text] [doi: 10.5888/pcd12.140402] [Medline: 25950569]

92.   Henshaw EJ. Too sick, not sick enough? Effects of treatment type and timing on depression stigma. J Nerv Ment Dis 2014 Apr;202(4):292-299. [doi: 10.1097/NMD.0000000000000121] [Medline: 24647215]

93.   Hildebrand M, Ahumada C, Watson S. CrowdOutAIDS: crowdsourcing youth perspectives for action. Reprod Health Matters 2013 May 14;21(41):57-68. [doi: 10.1016/S0968-8080(13)41687-7] [Medline: 23684188]

94.   Holst D, Kowalewski TM, White LW, Brand TC, Harper JD, Sorensen MD, et al. Crowd-sourced assessment of technical skills: differentiating animate surgical skill through the wisdom of crowds. J Endourol 2015 Oct;29(10):1183-1188. [doi: 10.1089/end.2015.0104] [Medline: 25867006]

95.   Holst D, Kowalewski TM, White LW, Brand TC, Harper JD, Sorenson MD, et al. Crowd-sourced assessment of technical skills: an adjunct to urology resident surgical simulation training. J Endourol 2015 May;29(5):604-609. [doi: 10.1089/end.2014.0616] [Medline: 25356517]

96.   Hougen HY, Lobo JM, Corey T, Jones R, Rheuban K, Schenkman NS, et al. Optimizing and validating the technical infrastructure of a novel tele-cystoscopy system. J Telemed Telecare 2016 Oct;22(7):397-404. [doi: 10.1177/1357633X15610040] [Medline: 26489430]

97.   Hsi RS, Hotaling JM, Hartzler AL, Holt SK, Walsh TJ. Validity and reliability of a smartphone application for the assessment of penile deformity in Peyronie's disease. J Sex Med 2013 Jul;10(7):1867-1873. [doi: 10.1111/jsm.12136] [Medline: 23551808]

98.   Hughes ML, Lowe DA, Shine HE, Carpenter BD, Balsis S. Using the Alzheimer's Association web site to improve knowledge of Alzheimer's disease in health care providers. Am J Alzheimers Dis Other Demen 2015 Feb;30(1):98-100. [doi: 10.1177/1533317514559827] [Medline: 25425736]

99. Ilakkuvan V, Tacelosky M, Ivey KC, Pearson JL, Cantrell J, Vallone DM, et al. Cameras for public health surveillance: a methods protocol for crowdsourced annotation of point-of-sale photographs. JMIR Res Protoc 2014 Apr 09;3(2):e22 [FREE Full text] [doi: 10.2196/resprot.3277] [Medline: 24717168]

100. Irshad H, Montaser-Kouhsari L, Waltz G, Bucur O, Nowak JA, Dong F, et al. Crowdsourcing image annotation for nucleus detection and segmentation in computational pathology: evaluating experts, automated methods, and the crowd. Pac Symp Biocomput 2015:294-305 [FREE Full text] [Medline: 25592590]

101. Kaczynski AT, Wilhelm Stanis SA, Hipp JA. Point-of-decision prompts for increasing park-based physical activity: a crowdsource analysis. Prev Med 2014 Dec;69:87-89 [FREE Full text] [doi: 10.1016/j.ypmed.2014.08.029] [Medline: 25204987]

102. Dainty KN, Brooks SC. A North American survey of public opinion on the acceptability of crowdsourcing basic life support for out-of-hospital cardiac arrest. Circulation 2015;132:A19542.

103. Kaufman EA, Cundiff JM, Crowell SE. The development, factor structure, and validation of the Self-concept and Identity Measure (SCIM): a self-report assessment of clinical identity disturbance. J Psychopathol Behav Assess 2014 Jun 28;37(1):122-133. [doi: 10.1007/s10862-014-9441-2]

104. Khare R, Burger JD, Aberdeen JS, Tresner-Kirsch DW, Corrales TJ, Hirchman L, et al. Scaling drug indication curation through crowdsourcing. Database (Oxford) 2015 Mar;2015: [FREE Full text] [doi: 10.1093/database/bav016] [Medline: 25797061]

105. Kiefner-Burmeister AE, Hoffmann DA, Meers MR, Koball AM, Musher-Eizenman DR. Food consumption by young children: a function of parental feeding goals and practices. Appetite 2014 Mar;74:6-11. [doi: 10.1016/j.appet.2013.11.011] [Medline: 24275668]

106. Kiefner-Burmeister A, Hoffmann D, Zbur S, Musher-Eizenman D. Implementation of parental feeding practices: does parenting style matter? Public Health Nutr 2016 Sep;19(13):2410-2414. [doi: 10.1017/S1368980016000446] [Medline: 26975423]

107. Kim M, Jung Y, Jung D, Hur C. Investigating the congruence of crowdsourced information with official government data: the case of pediatric clinics. J Med Internet Res 2014 Feb;16(2):e29 [FREE Full text] [doi: 10.2196/jmir.3078] [Medline: 24496094]

108. McKenna MT, Wang S, Nguyen TB, Burns JE, Petrick N, Summers RM. Strategies for improved interpretation of computer-aided detections for CT colonography utilizing distributed human intelligence. Med Image Anal 2012 Aug;16(6):1280-1292 [FREE Full text] [doi: 10.1016/j.media.2012.04.007] [Medline: 22705287]

109. Kowalewski TM, Comstock B, Sweet R, Schaffhausen C, Menhadji A, Averch T, et al. Crowd-sourced assessment of technical skills for validation of basic laparoscopic urologic skills tasks. J Urol 2016 Jun;195(6):1859-1865. [doi: 10.1016/j.juro.2016.01.005] [Medline: 26778711]

110. Kowalewski T, Sweet R, Menhadji A, Averch T, Box G, Brand T, et al. High-volume assessment of surgical videos via crowd-sourcing: the Basic Laparoscopic Urologic Skills (BLUS) Initiative. J Urol 2015 Apr;193(4):e393. [doi: 10.1016/j.juro.2015.02.704]

111. Krauss MJ, Sowles SJ, Moreno M, Zewdie K, Grucza RA, Bierut LJ, et al. Hookah-related Twitter chatter: a content analysis. Prev Chronic Dis 2015 Jul 30;12:E121 [FREE Full text] [doi: 10.5888/pcd12.150140] [Medline: 26226068]

112. Krieke LV, Jeronimus BF, Blaauw FJ, Wanders RB, Emerencia AC, Schenk HM, et al. HowNutsAreTheDutch (HoeGekIsNL): A crowdsourcing study of mental symptoms and strengths. Int J Methods Psychiatr Res 2016 Dec;25(2):123-144. [doi: 10.1002/mpr.1495] [Medline: 26395198]

113. Kristan J, Suffoletto B. Using online crowdsourcing to understand young adult attitudes toward expert-authored messages aimed at reducing hazardous alcohol consumption and to collect peer-authored messages. Transl Behav Med 2015 Mar;5(1):45-52 [FREE Full text] [doi: 10.1007/s13142-014-0298-4] [Medline: 25729452]

114. Kuang J, Argo L, Stoddard G, Bray BE, Zeng-Treitler Q. Assessing pictograph recognition: a comparison of crowdsourcing and traditional survey approaches. J Med Internet Res 2015 Dec;17(12):e281 [FREE Full text] [doi: 10.2196/jmir.4582] [Medline: 26678085]

115. Kudesia R, Chernyak E, McAvey B. Creation & validation of the fertility and infertility treatment knowledge survey (FIT-KS). Fertil Steril 2015 Sep;104(3):e118. [doi: 10.1016/j.fertnstert.2015.07.364]

116. Küffner R, Zach N, Norel R, Hawe J, Schoenfeld D, Wang L, et al. Crowdsourced analysis of clinical trial data to predict amyotrophic lateral sclerosis progression. Nat Biotechnol 2015 Jan;33(1):51-57. [doi: 10.1038/nbt.3051] [Medline: 25362243]

117. Kwitt R, Hegenbart S, Rasiwasia N, Vécsei A, Uhl A. Do we need annotation experts? A case study in celiac disease classification. Med Image Comput Comput Assist Interv 2014;17(Pt 2):454-461. [Medline: 25485411]

118. Lee YO, Jordan JW, Djakaria M, Ling PM. Using peer crowds to segment black youth for smoking intervention. Health Promot Pract 2014 Jul;15(4):530-537 [FREE Full text] [doi: 10.1177/1524839913484470] [Medline: 23628591]

119. Leifman G, Swedish T, Roesch K, Raskar R. Leveraging the crowd for annotation of retinal images. Conf Proc IEEE Eng Med Biol Soc 2015;2015:7736-7739. [doi: 10.1109/EMBC.2015.7320185] [Medline: 26738085]

120. Leiter A, Sablinski T, Diefenbach M, Foster M, Greenberg A, Holland J, et al. Use of crowdsourcing for cancer clinical trial development. J Natl Cancer Inst 2014 Oct;106(10). [doi: 10.1093/jnci/dju258] [Medline: 25217580]

121. Levick N, Dore A, Ralphs S, Arkus B, Nemirovsky D, Deville S, et al. iRescU. 2013. The Use of Social Media and Innovative Technology to Create a Global AED Geolocation Database: The iRescU project URL: https://www.irescu.info/HRS2013finalprint.pdf [accessed 2018-04-05] [WebCite Cache ID 6yS9crXlq]

122. Li S, Feng B, Chen M, Bell RA. Physician review websites: effects of the proportion and position of negative reviews on readers' willingness to choose the doctor. J Health Commun 2015 Apr;20(4):453-461. [doi: 10.1080/10810730.2014.977467] [Medline: 25749406]

123. Lloyd JC, Yen T, Pietrobon R, Wiener JS, Ross SS, Kokorowski PJ, et al. Estimating utility values for vesicoureteral reflux in the general public using an online tool. J Pediatr Urol 2014 Dec;10(6):1026-1031 [FREE Full text] [doi: 10.1016/j.jpurol.2014.02.014] [Medline: 24766856]

124. Luengo-Oroz MA, Arranz A, Frean J. Crowdsourcing malaria parasite quantification: an online game for analyzing images of infected thick blood smears. J Med Internet Res 2012 Nov 29;14(6):e167 [FREE Full text] [doi: 10.2196/jmir.2338] [Medline: 23196001]

125. MacLean DL, Heer J. Identifying medical terms in patient-authored text: a crowdsourcing-based approach. J Am Med Inform Assoc 2013;20(6):1120-1127 [FREE Full text] [doi: 10.1136/amiajnl-2012-001110] [Medline: 23645553]

126. Magid KH, Matlock DD, Thompson JS, McIlvennan CK, Allen LA. The influence of expected risks on decision making for destination therapy left ventricular assist device: An MTurk survey. J Heart Lung Transplant 2015 Jul;34(7):988-990. [doi: 10.1016/j.healun.2015.03.006] [Medline: 25935437]

127. Maier-Hein L, Mersmann S, Kondermann D, Bodenstedt S, Sanchez A, Stock C, et al. Can masses of non-experts train highly accurate image classifiers? A crowdsourcing approach to instrument segmentation in laparoscopic images. Med Image Comput Comput Assist Interv 2014;17(Pt 2):438-445. [Medline: 25485409]

128. Maier-Hein L, Mersmann S, Kondermann D, Stock C, Kenngott H, Sanchez A, et al. Crowdsourcing for reference correspondence generation in endoscopic images. Med Image Comput Comput Assist Interv 2014;17(Pt 2):349-356. [Medline: 25485398]

129. Malpani A, Vedula SS, Chen CC, Hager GD. A study of crowdsourced segment-level surgical skill assessment using pairwise rankings. Int J Comput Assist Radiol Surg 2015 Sep;10(9):1435-1447. [doi: 10.1007/s11548-015-1238-6] [Medline: 26133652]

130. Manuvinakurike R, Velicer WF, Bickmore TW. Automated indexing of Internet stories for health behavior change: weight loss attitude pilot study. J Med Internet Res 2014 Dec 09;16(12):e285 [FREE Full text] [doi: 10.2196/jmir.3702] [Medline: 25491243]

131. Margolin AA, Bilal E, Huang E, Norman TC, Ottestad L, Mecham BH, et al. Systematic analysis of challenge-driven improvements in molecular prognostic models for breast cancer. Sci Transl Med 2013 Apr 17;5(181):181re1 [FREE Full text] [doi: 10.1126/scitranslmed.3006112] [Medline: 23596205]

132. Mathieu E, Barratt A, Carter SM, Jamtvedt G. Internet trials: participant experiences and perspectives. BMC Med Res Methodol 2012 Oct 23;12:162 [FREE Full text] [doi: 10.1186/1471-2288-12-162] [Medline: 23092116]

133. Mavandadi S, Dimitrov S, Feng S, Yu F, Sikora U, Yaglidere O, et al. Distributed medical image analysis and diagnosis through crowd-sourced games: a malaria case study. PLoS One 2012 May;7(5):e37245 [FREE Full text] [doi: 10.1371/journal.pone.0037245] [Medline: 22606353]

134. Mavandadi S, Dimitrov S, Feng S, Yu F, Yu R, Sikora U, et al. Crowd-sourced BioGames: managing the big data problem for next-generation lab-on-a-chip platforms. Lab Chip 2012 Oct 21;12(20):4102-4106 [FREE Full text] [doi: 10.1039/c2lc40614d] [Medline: 22918378]

135. Mavandadi S, Feng S, Yu F, Dimitrov S, Yu R, Ozcan A. BioGames: a platform for crowd-sourced biomedical image analysis and telediagnosis. Games Health J 2012 Oct 01;1(5):373-376 [FREE Full text] [doi: 10.1089/g4h.2012.0054] [Medline: 23724363]

136. McCoy AB, Wright A, Laxmisan A, Ottosen MJ, McCoy JA, Butten D, et al. Development and evaluation of a crowdsourcing methodology for knowledge base construction: identifying relationships between clinical problems and medications. J Am Med Inform Assoc 2012 Sep;19(5):713-718 [FREE Full text] [doi: 10.1136/amiajnl-2012-000852] [Medline: 22582202]

137. McGregor TJ, Batis JC. A novel method for assessing caffeine dependence. J Caffeine Res 2016 Mar;6(1):26-33. [doi: 10.1089/jcr.2015.0017]

138. Yetisgen-Yildiz M, Solti I, Xia F. Using Amazon's Mechanical Turk for annotating medical named entities. AMIA Annu Symp Proc 2010;2010:1316 [FREE Full text] [Medline: 21785667]

139. Merchant RM, Griffis HM, Ha YP, Kiaru AS, Sellers AM, Hershey JC, et al. Hidden in plain sight: a crowdsourced public art contest to make automated external defibrillators more visible. Am J Public Health 2014 Dec;104(12):2306-2312. [doi: 10.2105/AJPH.2014.302211] [Medline: 25320902]

140. Meyer AN, Longhurst CA, Singh H. Crowdsourcing diagnosis for patients with undiagnosed illnesses: an evaluation of CrowdMed. J Med Internet Res 2016 Jan 14;18(1):e12 [FREE Full text] [doi: 10.2196/jmir.4887] [Medline: 26769236]

141. Michl GL, Katz JN, Losina E. Risk and risk perception of knee osteoarthritis in the US: a population-based study. Osteoarthritis Cartilage 2016 Apr;24(4):593-596 [FREE Full text] [doi: 10.1016/j.joca.2015.11.001] [Medline: 26555504]

142. Miller-Graff LE, Scrafford K, Rice C. Conditional and indirect effects of age of first exposure on PTSD symptoms. Child Abuse Negl 2016 Jan;51:303-312. [doi: 10.1016/j.chiabu.2015.09.003] [Medline: 26427886]

143.   Miller JD, Gentile B, Wilson L, Campbell WK. Grandiose and vulnerable narcissism and the DSM-5 pathological personality trait model. J Pers Assess 2013 May;95(3):284-290. [doi: 10.1080/00223891.2012.685907] [Medline: 22594764]

144.   Mitry D, Peto T, Hayat S, Blows P, Morgan J, Khaw KT, et al. Crowdsourcing as a screening tool to detect clinical features of glaucomatous optic neuropathy from digital photography. PLoS One 2015;10(2):e0117401 [FREE Full text] [doi: 10.1371/journal.pone.0117401] [Medline: 25692287]

145.   Mitry D, Peto T, Hayat S, Morgan JE, Khaw K, Foster PJ. Crowdsourcing as a novel technique for retinal fundus photography classification: analysis of images in the EPIC Norfolk cohort on behalf of the UK Biobank Eye and Vision Consortium. PLoS One 2013;8(8):e71154 [FREE Full text] [doi: 10.1371/journal.pone.0071154] [Medline: 23990935]

146.   Mollá D, Santiago-Martínez ME. Creation of a corpus for evidence based medicine summarisation. Australas Med J 2012;5(9):503-506 [FREE Full text] [doi: 10.4066/AMJ.2012.1375] [Medline: 23115585]

147.   Ma P, Businelle MS, Balis DS, Kendzor DE. The influence of perceived neighborhood disorder on smoking cessation among urban safety net hospital patients. Drug Alcohol Depend 2015 Nov 01;156:157-161. [doi: 10.1016/j.drugalcdep.2015.09.004] [Medline: 26386824]

148.   Morris RR, Schueller SM, Picard RW. Efficacy of a Web-based, crowdsourced peer-to-peer cognitive reappraisal platform for depression: randomized controlled trial. J Med Internet Res 2015;17(3):e72 [FREE Full text] [doi: 10.2196/jmir.4167] [Medline: 25835472]

149.   Morse P, Sweeny K, Legg AM. A situational construal approach to healthcare experiences. Soc Sci Med 2015 Aug;138:170-178. [doi: 10.1016/j.socscimed.2015.06.005] [Medline: 26094179]

150.   Mortensen JM, Minty EP, Januszyk M, Sweeney TE, Rector AL, Noy NF, et al. Using the wisdom of the crowds to find critical errors in biomedical ontologies: a study of SNOMED CT. J Am Med Inform Assoc 2015 May;22(3):640-648 [FREE Full text] [doi: 10.1136/amiajnl-2014-002901] [Medline: 25342179]

151.   Mortensen JM, Musen MA, Noy NF. Crowdsourcing the verification of relationships in biomedical ontologies. AMIA Annu Symp Proc 2013;2013:1020-1029 [FREE Full text] [Medline: 24551391]

152.   Muench F, Hayes M, Kuerbis A, Shao S. The independent relationship between trouble controlling Facebook use, time spent on the site and distress. J Behav Addict 2015 Sep;4(3):163-169 [FREE Full text] [doi: 10.1556/2006.4.2015.013] [Medline: 26551906]

153.   Nadorff MR, Nadorff DK, Germain A. Nightmares: under-reported, undetected, and therefore untreated. J Clin Sleep Med 2015 Jul 15;11(7):747-750 [FREE Full text] [doi: 10.5664/jcsm.4850] [Medline: 25845898]

154.   Nguyen TB, Wang S, Anugu V, Rose N, McKenna M, Petrick N, et al. Distributed human intelligence for colonic polyp classification in computer-aided detection for CT colonography. Radiology 2012 Mar;262(3):824-833 [FREE Full text] [doi: 10.1148/radiol.11110938] [Medline: 22274839]

155.   Norr AM, Albanese BJ, Oglesby ME, Allan NP, Schmidt NB. Anxiety sensitivity and intolerance of uncertainty as potential risk factors for cyberchondria. J Affect Disord 2015 Mar 15;174:64-69. [doi: 10.1016/j.jad.2014.11.023] [Medline: 25486275]

156.   Norr AM, Allan NP, Boffa JW, Raines AM, Schmidt NB. Validation of the Cyberchondria Severity Scale (CSS): replication and extension with bifactor modeling. J Anxiety Disord 2015 Apr;31:58-64. [doi: 10.1016/j.janxdis.2015.02.001] [Medline: 25734759]

157.   Norr AM, Oglesby ME, Raines AM, Macatee RJ, Allan NP, Schmidt NB. Relationships between cyberchondria and obsessive-compulsive symptom dimensions. Psychiatry Res 2015 Dec 15;230(2):441-446. [doi: 10.1016/j.psychres.2015.09.034] [Medline: 26429148]

158.   Peabody J, Miller D, Lane B, Sarle R, Brachulis A, Linsell S, et al. Wisdom of the crowds: use of crowdsourcing to assess surgical skill of robot-assisted radical prostatectomy in a statewide surgical collaborative. J Urol 2015 Apr;193(4):e655-e656. [doi: 10.1016/j.juro.2015.02.1826]

159.   Folkestad L, Brodersen JB, Hallas P, Brabrand M. [Laypersons can seek help from their Facebook friends regarding medical diagnosis]. Ugeskr Laeger 2011 Dec 05;173(49):3174-3177. [Medline: 22142603]

160.   Pharoah PD. Cell Slider: Using crowd sourcing for the scoring of molecular pathology. In: Proceedings of the 105th Annual Meeting of the American Association for Cancer Research. 2014 Sep 30 Presented at: 105th Annual Meeting of the American Association for Cancer Research; April 5-9, 2014; San Diego, CA. [doi: 10.1158/1538-7445.AM2014-303]

161.   Platt J, Kardia S. Public trust in health information sharing: implications for biobanking and electronic health record systems. J Pers Med 2015 Feb 03;5(1):3-21 [FREE Full text] [doi: 10.3390/jpm5010003] [Medline: 25654300]

162.   Plenge RM, Greenberg JD, Mangravite LM, Derry JM, Stahl EA, Coenen MJ, International Rheumatoid Arthritis Consortium (INTERACT). Crowdsourcing genetic prediction of clinical utility in the Rheumatoid Arthritis Responder Challenge. Nat Genet 2013 Apr 26;45(5):468-469. [doi: 10.1038/ng.2623] [Medline: 23619782]

163.   Pollert GA, Kauffman AA, Veilleux JC. Symptoms of psychopathology within groups of eating-disordered, restrained eating, and unrestrained eating individuals. J Clin Psychol 2016 Jun;72(6):621-632. [doi: 10.1002/jclp.22283] [Medline: 26990755]

164.   Powers MK, Boonjindasup A, Pinsky M, Dorsey P, Maddox M, Su LM, et al. Crowdsourcing assessment of surgeon dissection of renal artery and vein during robotic partial Nephrectomy: a novel approach for quantitative assessment of surgical performance. J Endourol 2016 Apr;30(4):447-452. [doi: 10.1089/end.2015.0665] [Medline: 26597352]

165. Price M, van Stolk-Cooke K. Examination of the interrelations between the factors of PTSD, major depression, and generalized anxiety disorder in a heterogeneous trauma-exposed sample using DSM 5 criteria. J Affect Disord 2015 Nov 01;186:149-155. [doi: 10.1016/j.jad.2015.06.012] [Medline: 26241663]

166. Pugh J, Kahane G, Maslen H, Savulescu J. Lay attitudes toward deception in medicine: Theoretical considerations and empirical evidence. AJOB Empir Bioeth 2016 Jan 02;7(1):31-38 [FREE Full text] [doi: 10.1080/23294515.2015.1021494] [Medline: 26682239]

167. Quisenberry A, Franck C, Koffarnus MN, Bickel WK. Delay discounting in current, ex, and non- smokers: interactions with gender. Drug Alcohol Depend 2015 Jan;146:e73-e74. [doi: 10.1016/j.drugalcdep.2014.09.569]

168. Merchant RM, Asch DA, Hershey JC, Griffis H, Hill S, Saynisch O, et al. A crowdsourcing, mobile media, challenge to locate automated external defibrillators. Circulation 2012;126:A57 [FREE Full text]

169. Raines AM, Allan NP, Oglesby ME, Short NA, Schmidt NB. Examination of the relations between obsessive-compulsive symptom dimensions and fear and distress disorder symptoms. J Affect Disord 2015 Sep 01;183:253-257. [doi: 10.1016/j.jad.2015.05.013] [Medline: 26042633]

170. Raines AM, Short NA, Sutton CA, Oglesby ME, Allan NP, Schmidt NB. Obsessive-compulsive symptom dimensions and insomnia: the mediating role of anxiety sensitivity cognitive concerns. Psychiatry Res 2015 Aug 30;228(3):368-372. [doi: 10.1016/j.psychres.2015.05.081] [Medline: 26162661]

171. Rass O, Pacek LR, Johnson PS, Johnson MW. Characterizing use patterns and perceptions of relative harm in dual users of electronic and tobacco cigarettes. Exp Clin Psychopharmacol 2015 Dec;23(6):494-503 [FREE Full text] [doi: 10.1037/pha0000050] [Medline: 26389638]

172. Behrend TS, Sharek DJ, Meade AW, Wiebe EN. The viability of crowdsourcing for survey research. Behav Res Methods 2011 Sep;43(3):800-813. [doi: 10.3758/s13428-011-0081-0] [Medline: 21437749]

173. Reese ED, Pollert GA, Veilleux JC. Self-regulatory predictors of eating disorder symptoms: understanding the contributions of action control and willpower beliefs. Eat Behav 2016 Jan;20:64-69. [doi: 10.1016/j.eatbeh.2015.11.005] [Medline: 26643592]

174. Reidy DE, Berke DS, Gentile B, Zeichner A. Masculine discrepancy stress, substance use, assault and injury in a survey of US men. Inj Prev 2016 Oct;22(5):370-374. [doi: 10.1136/injuryprev-2015-041599] [Medline: 26303670]

175. Reidy DE, Brookmeyer KA, Gentile B, Berke DS, Zeichner A. Gender role discrepancy stress, high-risk sexual behavior, and sexually transmitted disease. Arch Sex Behav 2016 Feb;45(2):459-465. [doi: 10.1007/s10508-014-0413-0] [Medline: 25564036]

176. Robillard JM, Roskams-Edris D, Kuzeljevic B, Illes J. Prevailing public perceptions of the ethics of gene therapy. Hum Gene Ther 2014 Aug;25(8):740-746. [doi: 10.1089/hum.2014.030] [Medline: 24773182]

177. Rodrigue JR, Fleishman A, Vishnevsky T, Whiting J, Vella JP, Garrison K, et al. Development and validation of a questionnaire to assess fear of kidney failure following living donation. Transpl Int 2014 Jun;27(6):570-575 [FREE Full text] [doi: 10.1111/tri.12299] [Medline: 24606048]

178. Santiago-Rivas M, Schnur JB, Jandorf L. Sun protection belief clusters: analysis of Amazon Mechanical Turk data. J Cancer Educ 2016 Dec;31(4):673-678. [doi: 10.1007/s13187-015-0882-4] [Medline: 26194777]

179. Schofield CA, Dea Moore C, Hall A, Coles ME. Understanding perceptions of anxiety disorders and their treatment. J Nerv Ment Dis 2016 Feb;204(2):116-122. [doi: 10.1097/NMD.0000000000000433] [Medline: 26669982]

180. Schulte EM, Avena NM, Gearhardt AN. Which foods may be addictive? The roles of processing, fat content, and glycemic load. PLoS ONE 2015 Feb 18;10(2):e0117959. [doi: 10.1371/journal.pone.0117959]

181. Schulte EM, Tuttle HM, Gearhardt AN. Belief in food addiction and obesity-related policy support. PLoS One 2016;11(1):e0147557 [FREE Full text] [doi: 10.1371/journal.pone.0147557] [Medline: 26808427]

182. Seligowski AV, Orcutt HK. Support for the 7-factor hybrid model of PTSD in a community sample. Psychol Trauma 2016 Mar;8(2):218-221. [doi: 10.1037/tra0000104] [Medline: 26689217]

183. Shaer O, Nov O, Okerlund J, Balestra M, Stowell E, Ascher L, et al. Informing the design of direct-to-consumer interactive personal genomics reports. J Med Internet Res 2015 Jun;17(6):e146 [FREE Full text] [doi: 10.2196/jmir.4415] [Medline: 26070951]

184. Shapiro RE, Lipton RB, Reiner PB. Factors influencing stigma towards persons with migraine. J Headache Pain 2014 Sep 18;15(Suppl 1):E36. [doi: 10.1186/1129-2377-15-S1-E36]

185. Siegel JT, Navarro MA, Thomson AL. The impact of overtly listing eligibility requirements on MTurk: An investigation involving organ donation, recruitment scripts, and feelings of elevation. Soc Sci Med 2015 Oct;142:256-260. [doi: 10.1016/j.socscimed.2015.08.020] [Medline: 26322721]

186. Silva I, Behar J, Sameni R, Zhu T, Oster J, Clifford GD, et al. Noninvasive fetal ECG: the PhysioNet/Computing in Cardiology Challenge 2013. Comput Cardiol 2013 Mar;40:149-152. [Medline: 25401167]

187. Sims MH, Bigham J, Kautz H, Halterman MW. Crowdsourcing medical expertise in near real time. J Hosp Med 2014 Jul;9(7):451-456. [doi: 10.1002/jhm.2204] [Medline: 24740747]

188. Sims MH, Fagnano M, Halterman JS, Halterman MW. Provider impressions of the use of a mobile crowdsourcing app in medical practice. Health Informatics J 2016 Jun;22(2):221-231. [doi: 10.1177/1460458214545896] [Medline: 25167866]

189. Smolinski MS, Crawley AW, Baltrusaitis K, Chunara R, Olsen JM, Wójcik O, et al. Flu Near You: crowdsourced symptom reporting spanning 2 influenza seasons. Am J Public Health 2015 Oct;105(10):2124-2130. [doi: 10.2105/AJPH.2015.302696] [Medline: 26270299]

190. Sokol Y, Eisenheim E. The relationship between continuous identity disturbances, negative mood, and suicidal ideation. Prim Care Companion CNS Disord 2016 Jan;18(1): [FREE Full text] [doi: 10.4088/PCC.15m01824] [Medline: 27247841]

191. Strickland JC, Stoops WW. Perceptions of research risk and undue influence: implications for ethics of research conducted with cocaine users. Drug Alcohol Depend 2015 Nov 01;156:304-310. [doi: 10.1016/j.drugalcdep.2015.09.029] [Medline: 26460141]

192. Stroh G, Rosell T, Dong F, Forster J. Early liver transplantation for patients with acute alcoholic hepatitis: public views and the effects on organ donation. Am J Transplant 2015 Jun;15(6):1598-1604 [FREE Full text] [doi: 10.1111/ajt.13176] [Medline: 25707427]

193. Su AI, Good BM, van Wijnen AJ. Gene Wiki reviews: marrying crowdsourcing with traditional peer review. Gene 2013 Dec 01;531(2):125. [doi: 10.1016/j.gene.2013.08.093] [Medline: 24012870]

194. Syme ML, Cohn TJ, Barnack-Tavlaris J. A comparison of actual and perceived sexual risk among older adults. J Sex Res 2017 Feb;54(2):149-160. [doi: 10.1080/00224499.2015.1124379] [Medline: 26813853]

195. Tang W, Han L, Best J, Zhang Y, Mollan K, Kim J, et al. Crowdsourcing HIV test promotion videos: a noninferiority randomized controlled trial in China. Clin Infect Dis 2016 Jun 01;62(11):1436-1442 [FREE Full text] [doi: 10.1093/cid/ciw171] [Medline: 27129465]

196. Taylor S, Conelea CA, McKay D, Crowe KB, Abramowitz JS. Sensory intolerance: latent structure and psychopathologic correlates. Compr Psychiatry 2014 Jul;55(5):1279-1284 [FREE Full text] [doi: 10.1016/j.comppsych.2014.03.007] [Medline: 24703593]

197. IMPROVER project team (in alphabetical order), Boue S, Fields B, Hoeng J, Park J, Peitsch MC, Challenge Best Performers (in alphabetical order), et al. Enhancement of COPD biological networks using a web-based collaboration interface. F1000Res 2015 May;4:32 [FREE Full text] [doi: 10.12688/f1000research.5984.2] [Medline: 25767696]

198. Thomas KB, Lund EM, Bradley AR. Composite trauma and mental health diagnosis as predictors of lifetime nonsuicidal self-injury history in an adult online sample. J Aggress Maltreatment Trauma 2015 Jul 27;24(6):623-635. [doi: 10.1080/10926771.2015.1049766]

199. Turner AM, Kirchhoff K, Capurro D. Using crowdsourcing technology for testing multilingual public health promotion materials. J Med Internet Res 2012 Jun 04;14(3):e79 [FREE Full text] [doi: 10.2196/jmir.2063] [Medline: 22664384]

200. Turner-McGrievy GM, Helander EE, Kaipainen K, Perez-Macias JM, Korhonen I. The use of crowdsourcing for dietary self-monitoring: crowdsourced ratings of food pictures are comparable to ratings by trained observers. J Am Med Inform Assoc 2015 Apr;22(e1):e112-e119. [doi: 10.1136/amiajnl-2014-002636] [Medline: 25092793]

201. Van de Belt TH, Engelen LJL, Berben SA, Teerenstra S, Samsom M, Schoonhoven L. Internet and social media for health-related information and communication in health care: preferences of the Dutch general population. J Med Internet Res 2013 Oct;15(10):e220 [FREE Full text] [doi: 10.2196/jmir.2607] [Medline: 24088272]

202. VanderBroek L, Acker J, Palmer AA, de Wit H, MacKillop J. Interrelationships among parental family history of substance misuse, delay discounting, and personal substance use. Psychopharmacology (Berl) 2016 Jan;233(1):39-48 [FREE Full text] [doi: 10.1007/s00213-015-4074-z] [Medline: 26395990]

203. Vashisht R, Mondal AK, Jain A, Shah A, Vishnoi P, Priyadarshini P, OSDD Consortium, et al. Crowd sourcing a new paradigm for interactome driven drug target identification in Mycobacterium tuberculosis. PLoS One 2012 Jul;7(7):e39808 [FREE Full text] [doi: 10.1371/journal.pone.0039808] [Medline: 22808064]

204. Volpi D, Sarhan MH, Ghotbi R, Navab N, Mateus D, Demirci S. Online tracking of interventional devices for endovascular aortic repair. Int J Comput Assist Radiol Surg 2015 Jun;10(6):773-781. [doi: 10.1007/s11548-015-1217-y] [Medline: 25976832]

205. Wagholikar KB, MacLaughlin KL, Kastner TM, Casey PM, Henry M, Greenes RA, et al. Formative evaluation of the accuracy of a clinical decision support system for cervical cancer screening. J Am Med Inform Assoc 2013;20(4):749-757 [FREE Full text] [doi: 10.1136/amiajnl-2013-001613] [Medline: 23564631]

206. Warby SC, Wendt SL, Welinder P, Munk EG, Carrillo O, Sorensen HB, et al. Sleep-spindle detection: crowdsourcing and evaluating performance of experts, non-experts and automated methods. Nat Methods 2014 Apr;11(4):385-392 [FREE Full text] [doi: 10.1038/nmeth.2855] [Medline: 24562424]

207. Wen X, Higgins ST, Xie C, Epstein LH. Improving public acceptability of using financial incentives for smoking cessation during pregnancy: a randomized controlled experiment. Nicotine Tob Res 2016 May;18(5):913-918. [doi: 10.1093/ntr/ntv204] [Medline: 26385928]

208. Zhang MW, Ward J, Ying JJ, Pan F, Ho RC. The alcohol tracker application: an initial evaluation of user preferences. Br Med J Innov 2016 Jan;2(1):8-13 [FREE Full text] [doi: 10.1136/bmjinnov-2015-000087] [Medline: 27019744]

209. White LW, Kowalewski TM, Dockter RL, Comstock B, Hannaford B, Lendvay TS. Crowd-sourced assessment of technical skill: a valid method for discriminating basic robotic surgery skills. J Endourol 2015 Nov;29(11):1295-1301. [doi: 10.1089/end.2015.0191] [Medline: 26057232]

210. White LW, Lendvay TS, Holst D, Borbely Y, Bekele A, Wright A. Using crowd-assessment to support surgical training in the developing world. J Am Coll Surg 2014 Oct;219(4):e40. [doi: 10.1016/j.jamcollsurg.2014.07.491]

211. Wymbs BT, Dawson AE. Screening Amazon's Mechanical Turk for adults with ADHD. J Atten Disord 2015 Aug 05:. [doi: 10.1177/1087054715597471] [Medline: 26246586]

212. Yang S, Wang SJ, Ji Y. An integrated dose-finding tool for phase I trials in oncology. Contemp Clin Trials 2015 Nov;45(Pt B):426-434. [doi: 10.1016/j.cct.2015.09.019] [Medline: 26419937]

213. Yeh VM, Schnur JB, Margolies L, Montgomery GH. Dense breast tissue notification: impact on women's perceived risk, anxiety, and intentions for future breast cancer screening. J Am Coll Radiol 2015 Mar;12(3):261-266 [FREE Full text] [doi: 10.1016/j.jacr.2014.11.001] [Medline: 25556313]

214. Yin Z, Fabbri D, Rosenbloom ST, Malin B. A scalable framework to detect personal health mentions on Twitter. J Med Internet Res 2015 Jun 05;17(6):e138 [FREE Full text] [doi: 10.2196/jmir.4305] [Medline: 26048075]

215. Yu B, Willis M, Sun P, Wang J. Crowdsourcing participatory evaluation of medical pictograms using Amazon Mechanical Turk. J Med Internet Res 2013 Jun;15(6):e108 [FREE Full text] [doi: 10.2196/jmir.2513] [Medline: 23732572]

216. Zhai H, Lingren T, Deleger L, Li Q, Kaiser M, Stoutenborough L, et al. Web 2.0-based crowdsourcing for high-quality gold standard development in clinical natural language processing. J Med Internet Res 2013 Apr 02;15(4):e73 [FREE Full text] [doi: 10.2196/jmir.2426] [Medline: 23548263]

217. Chang T, Verma BA, Shull T, Moniz MH, Kohatsu L, Plegue MA, et al. Crowdsourcing and the accuracy of online information regarding weight gain in pregnancy: a descriptive study. J Med Internet Res 2016 Apr 07;18(4):e81 [FREE Full text] [doi: 10.2196/jmir.5138] [Medline: 27056465]

218. Juusola JL, Quisel TR, Foschini L, Ladapo JA. The impact of an online crowdsourcing diagnostic tool on health care utilization: a case study using a novel approach to retrospective claims analysis. J Med Internet Res 2016 Dec 01;18(6):e127 [FREE Full text] [doi: 10.2196/jmir.5644] [Medline: 27251384]

219. Zide M, Caswell K, Peterson E, Aberle DR, Bui AA, Arnold CW. Consumers' patient portal preferences and health literacy: a survey using crowdsourcing. JMIR Res Protoc 2016 Jun 08;5(2):e104. [doi: 10.2196/resprot.5122]

220. Bardos J, Friedenthal J, Spiegelman J, Williams Z. Cloud based surveys to assess patient perceptions of health care: 1000 respondents in 3 days for US $300. JMIR Res Protoc 2016 Aug 23;5(3):e166 [FREE Full text] [doi: 10.2196/resprot.5772] [Medline: 27554915]

221. Bock BC, Lantini R, Thind H, Walaska K, Rosen RK, Fava JL, et al. The Mobile Phone Affinity Scale: enhancement and refinement. JMIR Mhealth Uhealth 2016 Dec 15;4(4):e134 [FREE Full text] [doi: 10.2196/mhealth.6705] [Medline: 27979792]

222. Bateman DR, Brady E, Wilkerson D, Yi EH, Karanam Y, Callahan CM. Comparing crowdsourcing and friendsourcing: a social media-based feasibility study to support Alzheimer disease caregivers. JMIR Res Protoc 2017 Apr 10;6(4):e56 [FREE Full text] [doi: 10.2196/resprot.6904] [Medline: 28396304]

223. Aramaki E, Shikata S, Ayaya S, Kumagaya SI. Crowdsourced identification of possible allergy-associated factors: automated hypothesis generation and validation using crowdsourcing services. JMIR Res Protoc 2017 May 16;6(5):e83 [FREE Full text] [doi: 10.2196/resprot.5851] [Medline: 28512079]

224. Brady CJ, Mudie LI, Wang X, Guallar E, Friedman DS. Improving consensus scoring of crowdsourced data using the Rasch model: development and refinement of a diagnostic instrument. J Med Internet Res 2017 Jun 20;19(6):e222 [FREE Full text] [doi: 10.2196/jmir.7984] [Medline: 28634154]

225. DePalma MT, Rizzotti MC, Branneman M. Assessing diabetes-relevant data provided by undergraduate and crowdsourced web-based survey participants for honesty and accuracy. JMIR Diabetes 2017 Jul 12;2(2):e11. [doi: 10.2196/diabetes.7473]

226. Bartek MA, Truitt AR, Widmer-Rodriguez S, Tuia J, Bauer ZA, Comstock BA, et al. The promise and pitfalls of using crowdsourcing in research prioritization for back pain: cross-sectional surveys. J Med Internet Res 2017 Oct 06;19(10):e341 [FREE Full text] [doi: 10.2196/jmir.8821] [Medline: 28986339]

227. Prieto JT, Jara JH, Alvis JP, Furlan LR, Murray CT, Garcia J, et al. Will participatory syndromic surveillance work in Latin America? Piloting a mobile approach to crowdsource influenza-like illness data in Guatemala. JMIR Public Health Surveill 2017 Nov 14;3(4):e87 [FREE Full text] [doi: 10.2196/publichealth.8610] [Medline: 29138128]

228. Cooper S, Khatib F, Treuille A, Barbero J, Lee J, Beenen M, et al. Predicting protein structures with a multiplayer online game. Nature 2010 Aug 05;466(7307):756-760 [FREE Full text] [doi: 10.1038/nature09304] [Medline: 20686574]

229. Adrover C, Bodnar T, Huang Z, Telenti A, Salathé M. Identifying adverse effects of HIV drug treatment and associated sentiments using Twitter. JMIR Public Health Surveill 2015 Jul 27;1(2):e7. [doi: 10.2196/publichealth.4488]

230. PatientsLikeMe. Live better, together! URL: https://www.patientslikeme.com/?format=html [accessed 2018-04-05] [WebCite Cache ID 6ySC9BkFP]

231. Morton CC. Innovating openly: researchers and patients turn to crowdsourcing to collaborate on clinical trials, drug discovery, and more. IEEE Pulse 2014;5(1):63-67. [doi: 10.1109/MPUL.2013.2289467] [Medline: 25296376]

232. Saez-Rodriguez J, Costello JC, Friend SH, Kellen MR, Mangravite L, Meyer P, et al. Crowdsourcing biomedical research: leveraging communities as innovation engines. Nat Rev Genet 2016 Dec 15;17(8):470-486. [doi: 10.1038/nrg.2016.69] [Medline: 27418159]

233. Gosling SD, Vazire S, Srivastava S, John OP. Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. Am Psychol 2004 Mar;59(2):93-104. [doi: 10.1037/0003-066X.59.2.93] [Medline: 14992636]

234. Williamson V. 2016 Feb. Brown Center Chalkboard: Can crowdsourcing be ethical? URL: https://www.brookings.edu/blog/brown-center-chalkboard/2016/02/08/can-crowdsourcing-be-ethical/ [accessed 2018-04-05] [WebCite Cache ID 6ySCUJ4b0]

Abbreviations

MTurk: Amazon Mechanical Turk

XSL•FO
RenderX

Diffusion of the Digital Health Self-Tracking Movement in Canada: Results of a National Survey

Guy Paré[1], PhD; Chad Leaver[2], MSc, MBA; Claire Bourget[3], MBA

[1]Research Chair in Digital Health, HEC Montreal, Montreal, QC, Canada
[2]Canada Health Infoway, Toronto, ON, Canada
[3]CEFRIO, Montreal, QC, Canada

Corresponding Author:
Guy Paré, PhD
Research Chair in Digital Health
HEC Montreal
3000 Cote-Sainte-Catherine Road
Montreal, QC,
Canada
Phone: 1 514 340 6812
Email: guy.pare@hec.ca

Abstract

Background: With the ever-increasing availability of mobile apps, consumer wearables, and smart medical devices, more and more individuals are self-tracking and managing their personal health data.

Objective: The aim of this study was to investigate the diffusion of the digital self-tracking movement in Canada. It provides a comprehensive, yet detailed account of this phenomenon. It examines the profile of digital self-trackers, traditional self-trackers, and nontrackers, further investigating the primary motivations for self-tracking and reasons for nontracking; barriers to adoption of connected care technologies; users' appreciation of their self-tracking devices, including what they perceive to be the main benefits; factors that influence people's intention to continue using connected care technologies in the future; and the reasons for usage discontinuance.

Methods: We conducted an online survey with a sample of 4109 Canadian adults, one of the largest ever. To ensure a representative sample, quota method was used (gender, age), following stratification by region. The maximum margin of error is estimated at 1.6%, 19 times out of 20.

Results: Our findings reveal that 66.20% (2720/4109) of our respondents regularly self-track one or more aspects of their health. About one in 4 respondents (1014/4109, 24.68%) currently owns a wearable or smart medical device, and 57.20% (580/1014) use their devices on a regular basis for self-tracking purposes. Digital self-trackers are typically young or mature adults, healthy, employed, university educated, with an annual family income of over $80,000 CAD. The most popular reported device is the fitness tracker or smartwatch that can capture a range of parameters. Currently, mobile apps and digital self-tracking devices are mainly used to monitor physical activity (856/1669, 51.13%), nutrition (545/1669, 32.65%), sleep patterns (482/1669, 28.88%) and, to a much lesser extent, cardiovascular and pulmonary biomarkers (215/1669, 12.88%), medication intake (126/1669, 7.55%), and glucose level (79/1669, 4.73%). Most users of connected care technologies (481/580, 83.0%) are highly satisfied and 88.2% (511/580) intend to continue using their apps and devices in the future. A majority said smart digital devices have allowed them to maintain or improve their health condition (398/580, 68.5%) and to be better informed about their health in general (387/580, 66.6%). About 33.80% of our sample (1389/4109) is composed of people who do not monitor their health or well-being on a regular basis.

Conclusions: Our study shows an opportunity to advance the health of Canadians through connected care technologies. Our findings can be used to set baseline information for future research on the rise of digital health self-tracking and its impacts. Although the use of mobile apps, consumer wearables, and smart medical devices could potentially benefit the growing population of patients with chronic conditions, the question remains as to whether it will diffuse broadly beyond early adopters and across cost inequities.

Introduction

In *The patient will see you now*, Eric Topol describes mobile phones as the Gutenberg of health care [1]. He argues that small mobile devices with wireless connectivity will prove to be the transformative catalyst for advancing toward the future of medicine. Although the mobile phone remains the device of choice for most individuals, technology manufacturers are creating a future of consumer wearables and smart medical devices that promise to help people live healthier lives [2]. Several such devices are currently available, including physical activity trackers that measure fitness-related metrics such as distance walked or ran; smart forks that vibrate when people eat too fast; smart toothbrushes with 3D motion sensors that monitor brushing performance; and smart clothes (eg, biometric shirts) that measure a person's breathing, pulse, calories, and sleep patterns. Recent estimates predict that 5.2 billion consumer smart devices are in use globally in 2017, setting the stage for an estimated 12.9 billion devices to be deployed by 2020 [3]. Other forecasts indicate that the consumer wearable device market value will reach US $41 billion by 2020 from US $2 billion in 2014 [4].

The need to attach personal numeric data to day-to-day activities such as eating, sleeping, and exercising is called the "quantified-self" movement [5]. Proponents of this movement believe that if they can measure an aspect of their life on a regular basis, they can find a way to improve it [6]. Computers can facilitate self-tracking because of advances in sensor technologies, ubiquity of access to information brought by the Internet, and improvements in user-friendly systems and interfaces [7]. Prior research shows that the measurement of one's daily activities with the assistance of mobile devices provides an advantage with respect to automatic and aggregated data compilation. People have limited memory and capacity to accurately and consistently track computational data about their behaviors such as counting the number of steps throughout the day [7-9]. Another advantage is that data from wearable sensors and smart medical devices can generate automated analytics over time, aggregating personally relevant feedback, which may, in turn, contribute to the sustainable use of digital devices [10,11].

Many experts say the rise of the Internet of Things will bring the next revolution in digital health [12]. Recent surveys on the adoption and impact of consumer digital health technologies reveal important insight about the current state and the trajectory of the purported potential. For instance, a survey of 2225 US adults reveals that the use of wearable self-tracking devices has doubled from 2014 (9%) to 2016 (21%) [13]. This study also indicates that millennials (aged 18-34 years) are the most prevalent users (36%). Another US-based survey (n=2025) on telemedicine, wearables, and postdischarge care found that 27% of adults own a self-tracking device and that 78% would want their doctor to have access to data from their wearables [14]. In Europe, a national survey of 1005 French citizens (aged 15+ years) reveals that 11% possess a health wearable or smart medical device, and 30% of nonusers have a firm intention to buy one in the next 12 months [15]. Finally, a 2016 online survey of more than 20,000 consumers (aged 15+ years) from 16 countries reported that 33% of respondents tracked their physical activity via a mobile app and a fitness band, clip, or smartwatch [16].

Although the abovementioned surveys set relevant baseline information, they do not provide a comprehensive and detailed account of the digital self-tracking movement. Specifically, no prior empirical research has attempted to investigate the sociodemographic and preference profile of digital self-trackers, traditional self-trackers, and nontrackers; the primary reasons for self-tracking and nontracking; the barriers to adoption of smart and connected health devices; users' perceived benefits of these devices; the factors that influence people's intention to continue using connected care technologies in the future; and the reasons for usage discontinuance that remain largely unknown. This study aims to fill this important gap and presents a timely and relevant integration of these issues, which may inform technology manufacturers, health care providers (HCPs), and policy makers' perceptions and future decisions in this area.

Methods

Study Design and Sample

In this section, we report the online survey that was conducted in according with the Checklist for Reporting Results of Internet E-Surveys checklist [17]. We first developed a comprehensive questionnaire instrument to administer with the general Canadian population in 2017. The instrument was based on a review of the extant literature on mobile health (mHealth) and digital self-tracking and was originally designed in French and back translated to English. The questionnaire was pretested during face-to-face interviews with 16 adults representative of the Canadian population in terms of age, gender, and language. Some minor adjustments were made to the questionnaire following this initial step.

The online survey was administered by AC Nielsen Company of Canada. The sample used for this research was the company's proprietary online panel, known as the Harris Panel. This panel is one of the largest, most representative, and best profiled panels in Canada. To begin survey administration, panel members were invited to participate in the study by email. Once participants clicked on the URL provided in the email letter, they were screened for the following eligibility criteria: (1) Canadian resident, (2) aged 18 years or older, and (3) spoke English or French. Those who were eligible read an informed consent form that emphasized the anonymity and confidentiality of respondents and advised that by completing the questionnaire, they are providing their consent to participate. All study procedures were approved by the HEC Montreal's research ethics committee. To ensure a representative sample, the quota method was used (gender, age), following stratification by Canadian geographic regions.

XSL•FO

RenderX

Survey respondents were able to enter the survey at any point during the data collection period, that is, from January 11, 2017 to February 2, 2017. Respondents who partially completed the survey were able to exit the questionnaire and return at a later time to enter additional data. This could be done as many times as necessary. In accessing the online survey, respondents were assigned a unique identifier and passcode that allowed them access their data until the survey was finished. Participants were rewarded points for survey completion. Rewards for completing AC Nielsen surveys range in value from $5 CAD to $75 CAD. Standard options include gift cards and merchandise (eg, Amazon, iTunes, magazine subscriptions, Starbucks, Wal-Mart, and a variety of restaurant gift cards).

Survey Items

Gender, age, region, gross family income, education, occupation, and use of mobile phones and digital tablets were assessed by standard survey items administered in other international surveys [13-16]. Overall health status was obtained by asking participants to self-rate their own health on a scale from 1=poor or fair to 5=very good or excellent. This single-item measure has been used extensively worldwide and represents a valid and acceptable measure [18]. We also asked participants if they had one or several of the following chronic conditions: (1) diabetes, (2) high blood pressure, (3) obesity, (4) cardiovascular disease, (5) lung or respiratory airway disease, (6) cancer, (7) bone or muscular disease, (8) disease of nervous system, (9) mental disorder, (10) chronic infectious disease, and (11) addiction to tobacco or drugs.

Familiarity with connected care technologies was measured by asking "How familiar are you with consumer health wearables and smart medical devices?" using a 5-point Likert scale, where 1=not much at all and 5=extremely. We then asked, "Which of the following devices do you own?" using descriptive nonbrand terms for 13 specific devices commonly listed in the extant literature and available in Canada (see Results section). For each device they own, respondents were then asked how often they use it using a 7-point scale, where 1=once a month or less and 7=many times each day.

Motivations for using digital health self-tracking devices were measured with 10 items developed for this study using 5-point Likert scales, where 1=not at all and 5=very strongly. Items were derived from prior surveys on consumer digital health [12-16]. Examples of items include "know myself better," "give me daily encouragement toward reaching my personal goals," "better follow the treatment plan prescribed by my physician," and "break a bad habit related to my health." Data-sharing behaviors were assessed with a single item asking "Do you ever share with other people the data stored in your device or mobile app?" When answering "yes," respondents were then asked with whom (eg, family members, friends, family doctor, pharmacist, or personal trainer).

Respondents' appreciation of wearables and smart devices were captured with five variables. Measures for perceived usefulness (7 items) and ease of use (4 items) were adapted to the context of this study from Davis [19]. For their part, user satisfaction (3 items), confirmation of initial expectations (3 items), and intention to continue using wearables and smart devices (3

items) were adapted from Bhattacherjee [20] and Hong et al [21]. All five variables were assessed using 5-point Likert scales, where 1=strongly disagree and 5=strongly agree.

Finally, we asked respondents (when applicable) why they did not currently possess health wearables or smart devices. We developed a list of 10 reasons (see Results section), and respondents only checked those that applied to their personal situation. In a similar fashion, we developed a list of 11 items (see next section) that correspond to the reasons why consumers stopped using their wearables and smart devices at some point. Both lists of items were derived from prior surveys on consumer digital health [12-16]. The complete online survey instrument is provided in Multimedia Appendix 1.

Data Analyses

In line with our research objectives, we analyzed the entire sample as well as specific subgroups. General trends regarding ownership and use of connected care technologies are analyzed with descriptive statistics (mean, SD, percentage), comparisons between self-trackers and nontrackers are analyzed with multinomial logistic regression tests, and users' appreciation of digital self-tracking devices is analyzed using Pearson correlation tests and partial least squares (PLS) multiple regression analyses. Analyses are performed using the SPSS (IBM Corp) version 23 software and the SmartPLS (SmartPLS GmbH) version 3.2.7 software.

Results

Profile of the Sample

Our sample is composed of 4109 adults. The maximum margin of error is estimated at 1.6%, 19 times out of 20. Table 1 presents the profile of the sample according to usual sociodemographic variables, in comparison with the total Canadian population. The sample was composed of 2118 men, representing 51.55%. In terms of age, 27.84% of all respondents (1144/4109) were millennials (18-34 years), whereas 35.17% (1445/4109) consisted of baby boomers (55+ years). As expected, the majority of respondents were from the two largest Canadian provinces, namely, Ontario (1575/4109, 38.33%) and Quebec (986/4109, 24.00%). About 1 out of 5 respondents had a gross family annual income of less than $40,000 CAD, whereas 35.58% (1462/4109) had annual family incomes superior to $80,000 CAD. Our survey participants were more educated than the Canadian population according to data from the 2016 national census. Almost half of the respondents had a university degree compared with 28.70% for the whole population, 6 out of 10 respondents were workers (2386/4109), less than 4% (3.68%, 151/4109) were students, and slightly over 23% (22.8%, 937/4109) were retired. Overall, our data indicate that, except for education, the sociodemographic profile of our respondents is representative of the adult population in Canada.

In terms of health status, less than 10% of all respondents (9.78%, 402/4109) perceived themselves to be in poor or fair condition, whereas 50.38% (2070/4109) said they were in good health, and 39.84% (1637/4109) perceived themselves in very good or excellent health.

Table 1. Profile of the sample and comparisons with the Canadian population.

Characteristics	Sample (N=4109), n (%)	Canadian population (N=35,151,730), n (%)
Gender		
Male	2118 (51.55)	17,264,200 (49.11)[a]
Female	1991 (49.45)	17,887,530 (50.89)[a]
Age (years)		
18-34	1144 (27.84)	6,858,075 (25.27)[a]
35-54	1520 (36.99)	9,581,540 (27.28)[a]
55+	1445 (35.17)	10,846,380 (30.86)[a]
Region		
Atlantic provinces	293 (7.13)	2,385,779 (6.58)[a]
Quebec	986 (24.00)	8,321,888 (22.95)[a]
Ontario	1575 (38.33)	13,976,320 (38.54)[a]
Manitoba and Saskatchewan	266 (6.47)	2,466,703 (6.80)[a]
Alberta	437 (10.64)	4,236,376 (11.68)[a]
British Columbia and Northwest Territories	552 (13.43)	4,802,275 (13.24)[a]
Gross family income[b] ($ CAD)		
<$20k	268 (6.52)	8,558,000 (29.88)[a]
≥$20k and <$40k	583 (14.19)	7,014,015 (24.48)[a]
≥$40k and <$60k	614 (14.94)	5,006,820 (17.48)[a]
≥$60k and <$80k	561 (13.65)	2,926,920 (10.22)[a]
≥$80k and <$100k	498 (12.12)	1,716,175 (5.99)[a]
≥$100k	964 (23.46)	2,266,600 (7.91)[a]
Education level		
High school or college	2051 (51.13)	18,730,750 (65.39)[a]
Undergraduate	1300 (32.41)	6,659,615 (23.25)[a]
Graduate	660 (16.45)	1,562,555 (5.45)[a]
Occupation		
Workers	2386 (58.86)	17,230,040 (60.15)[a]
Students	151 (3.72)	19,992,283 (6.99)[a]
Retirees	937 (23.11)	4,912,278 (17.15)[a]
Other	580 (14.31)	4,284,996 (15.96)[a]
Perceived health status		
Bad or average	402 (9.78)	3,443,000 (12.00)[c]
Good	2070 (50.38)	9,561,713 (29.00)
Very good or excellent	1637 (39.84)	18,714,100 (59.00)[c]
Chronic diseases		
Yes	1281 (31.89)	12,053,150 (38.00)[c]
No	2735 (68.11)	19,665,665 (62.00)[c]

Characteristics	Sample (N=4109), n (%)	Canadian population (N=35,151,730), n (%)
Language used to complete the questionnaire		
English	3644 (88.68)	—
French	465 (11.32)	—

[a]Statistics Canada Census 2016.

[b]The median total income in Canada was $80,940 CAD in 2015 according to the Statistics Canada Census 2016.

[c]Health Canada Survey 2014.

Aligned to general population estimates for Canada, the majority of respondents (2735/4109, 68.11%) reported no chronic conditions. The most common self-reported chronic conditions were hypertension, diabetes, and obesity.

Use of Mobile Devices

Our findings indicate that 78.10% of all respondents (3209/4109) owned a mobile phone (eg, Apple iPhone, Samsung Galaxy, Google Nexus, Microsoft Lumia, or Sony Xperia) and used it to download mobile apps, among other things. Our results also show that 56.88% of our respondents (2337/4109) owned a digital tablet (eg, Apple iPad, Samsung Galaxy Tab, Google Nexus Tablet, or Sony Xperia Tablet). These statistics are similar to those reported by the Canadian Radio-television and Telecommunications Commission, which found that in 2016, 73% of Canadian adults owned a mobile phone, and 52% owned a tablet [22]. Overall, 86.01% of our respondents (3534/4109) used either a mobile phone or a tablet, whereas 48.97% (2012/4109) reported owning both devices. As expected, age is negatively associated with use of a mobile device (χ^2_4=389.3; P<.001); 56.24% of millennials (18-34 years) using both devices compared with 37.11% among baby boomers (55+ years).

Self-Tracking Behaviors

We defined three self-tracking profiles for the study. Respondents that regularly track one or more aspect of their health or well-being using connected care technologies, that is, mHealth apps, consumer wearables (eg, fitness trackers), and smart medical devices (eg, blood pressure monitors) were defined as "digital self-trackers." Respondents who regularly monitor one or more aspect of their health using manual tools (other than a mobile app or smart device) such as simply recording the information in writing (on paper, or in a journal or notebook) or by remembering the information were defined as "traditional self-trackers." The remaining respondents reporting that they do not regularly monitor any aspect of their personal health or well-being were defined as "nontrackers."

Table 2 reveals that whether through digital or traditional means, the prevalence of self-trackers in Canada is about two-thirds (2720/4109, 66.20%), a number that is similar to recent estimates from the United States [23]. Our sample is composed of two distinct groups of self-trackers. The first, digital self-trackers comprise 40.61% (1669/4109) of our sample and 61.36% of all self-trackers (1669/2720). This group comprises as many men as women who are, for the most part, active members of the workforce. A majority of digital self-trackers are young or mature adults (18-34 years), highly-educated, and wealthy individuals (average gross family income $80,000 CAD), and people who perceive themselves to be in good or very good health. More than 7 out of 10 respondents in this group (1172/1669) self-report having no chronic diseases.

The second group, traditional self-trackers represents 25.58% (1051/1669) of our sample and 38.64% of all self-trackers (1051/2720). Unlike digital self-trackers, this group comprises slightly more women than men. The majority of traditional trackers are aged 55 years and older, retired, with a gross family income substantially less than digital trackers. Importantly, individuals in this group are more likely to be living with one or several chronic diseases than digital self-trackers (47.07% vs 28.36%, respectively).

The third group, nontrackers, represents 33.80% of our sample (1389/4109). Like traditional trackers, this group comprises slightly less men than women. However, nontrackers are found in all age groups. They are less educated and have lower gross family income on average than the other two groups. One in four nontrackers reported having one or several chronic conditions. The most common reasons given by these respondents for not tracking any aspect of their health were as follows: "the information provided by my physician is sufficient" (335/1389, 24.12%), "no need because I am in excellent health condition" (338/1389, 24.33%), "I am simply not interested" (329/1389, 23.69%), and "I am not disciplined enough" (261/1389, 18.79%). Four out of five nontrackers (1095/1389, 78.83%) possess a mobile phone or a tablet.

A multinomial logistic regression including all sociodemographic and health status variables was performed to calculate odds ratios describing the odds of tracking one's own health using traditional or digital devices compared with the odds of nontracking (reference category). The traditional 0.05 criterion of statistical significance was employed for all tests. Addition of the predictors to a model that contained only the intercept significantly improved the fit between model and data; χ^2_{36} (N=4109)=548.70, Nagelkerke R^2=0.32, P<.001. As indicated in Table 3, our analyses determined no statistically significant differences between groups in terms of gender, region, occupation, and perceived health condition. However, significant differences were observed in terms of age, gross annual income, education level, and chronic condition(s). For instance, millennials (18-34 years) and people in the age range of 35 to 54 years were 3.7 and 1.5 times more likely to be in the digital self-tracking group than baby boomers (55+ years). As another example, compared with people living with no chronic condition, chronic patients were 0.4 times less likely to be in the traditional self-tracking group and 0.6 times less likely to be in the digital self-tracking group.

XSL•FO
RenderX

Table 2. Profile of self-trackers and nontrackers (N=4109).

Characteristics	Nontrackers (N=1389), n (%)	Traditional self-trackers (N=1051), n (%)	Digital self-trackers (N=1669), n (%)
Gender			
Male	721 (51.91)	566 (53.85)	831 (49.79)
Female	668 (48.09)	485 (46.15)	838 (50.21)
Age in years			
18-34	314 (22.61)	147 (13.98)	684 (40.98)
35-54	539 (38.80)	347 (33.02)	633 (37.93)
55+	536 (38.59)	557 (53.00)	352 (21.09)
Region			
Atlantic provinces	106 (7.63)	78 (7.42)	109 (6.53)
Quebec	368 (26.49)	253 (24.07)	365 (21.87)
Ontario	513 (36.93)	414 (39.39)	648 (38.83)
Manitoba and Saskatchewan	94 (6.77)	65 (6.18)	107 (6.41)
Alberta	124 (8.93)	103 (9.80)	211 (12.64)
British Columbia and Terrace	184 (13.25)	138 (13.13)	229 (13.72)
Gross family income ($ CAD)			
<$40k	335 (29.13)	244 (27.23)	272 (19.86)
≥$40k and <$60k	244 (21.22)	171 (19.08)	200 (13.87)
≥$60k and <$80k	190 (16.52)	154 (17.19)	216 (14.98)
≥$80k and <$100k	145 (12.61)	109 (12.16)	244 (16.92)
≥$100k and <$200k	195 (16.96)	192 (21.43)	428 (29.68)
≥$200k	41 (3.56)	26 (2.90)	82 (5.69)
Education level			
High school or college	805 (59.59)	529 (51.30)	717 (44.04)
Undergraduate	376 (27.83)	330 (32.01)	593 (36.43)
Graduate	170 (12.58)	172 (16.68)	318 (19.53)
Occupation			
Workers	752 (54.85)	476 (45.81)	1158 (70.44)
Students	53 (3.87)	23 (2.21)	75 (4.56)
Retirees	347 (25.31)	383 (36.86)	207 (12.59)
Other	219 (15.97)	157 (15.11)	204 (12.41)
Perceived health condition			
Bad or average	118 (8.50)	127 (12.08)	157 (9.41)
Good	712 (51.26)	524 (49.86)	833 (49.91)
Very good or excellent	559 (40.24)	400 (38.06)	679 (40.68)
Chronic diseases			
No	1021 (75.29)	542 (52.93)	1172 (71.64)
Yes	335 (24.71)	482 (47.07)	464 (28.36)

Table 3. Multinomial logistic regression model predicting traditional tracking and e-tracking by patient characteristics. Reference category=nontrackers (N=1389).

Characteristics	Traditional self-trackers (N=1051)		Digital self-trackers (N=1669)	
	Odds ratio (95% CI)	Significance	Odds ratio (95% CI)	Significance
Intercept	—	<.001	—	<.001
Gender				
Female	0.932 (0.765-1.134)	.48	1.170 (0.981-1.394)	.08
Age (years)				
18-34	0.612 (0.434-0.863)	.005	3.732 (2.785-5.002)	<.001
35-54	0.728 (0.555-0.954)	.02	1.552 (1.193-2.018)	<.001
Region				
Atlantic provinces	1.055 (0.682-1.633)	.81	0.921 (0.619-1.370)	.69
Quebec	1.022 (0.739-1.414)	.90	0.858 (0.641-1.147)	.30
Ontario	1.038 (0.764-1.410)	.81	0.963 (0.733-1.263)	.78
Manitoba-Saskatchewan	0.701 (0.434-1.134)	.15	0.859 (0.573-1.287)	.46
Alberta	1.364 (0.900-2.067)	.15	1.335 (0.927-1.922)	.12
Gross family income ($ CAD)				
≤$60K	0.750 (0.552-1.019)	.07	0.429 (0.326-0.566)	<.001
> $60K and ≤ $100K	0.853 (0.673-1.081)	.19	0.550 (0.447-0.675)	<.001
Education level				
High school or college	0.639 (0.477-0.855)	.003	0.623 (0.480-0.808)	<.001
Undergraduate	0.797 (0.590-1.077)	.14	0.832 (0.637-1.086)	.18
Occupation				
Workers	0.808 (0.608-1.075)	.14	1.292 (0.956-1.746)	.01
Students	0.900 (0.442-1.834)	.77	0.680 (0.376-1.229)	.20
Others	0.704 (0.483-1.027)	.07	0.877 (0.603-1.275)	.49
Perceived health condition				
Very poor or poor	0.972 (0.669-1.414)	.88	1.108 (0.789-1.557)	.55
Fair or good	0.923 (0.749-1.137)	.45	1.007 (0.837-1.212)	.94
Chronic disease(s)				
One or several chronic disease(s)	0.403 (0.322-0.503)	<.001	0.548 (0.443-0.677)	<.001

Motivations for Using Digital Self-Tracking Devices

The use of digital self-tracking devices and mobile apps is mainly the result of motivations tied to sustaining individual well-being rather than monitoring or mediating medical problems or illnesses. More precisely, 57.94% of our respondents (967/1669) said they use connected care technologies mainly to know more about their condition and monitor changes in parameters that they consider important for their health. Another common motivation was associated with the day-to-day encouragement that digital health self-tracking technologies provide as people strive to meet their personal goals (883/1669, 52.91%). Importantly, 42.06% (702/1669) said they use digital self-tracking tools to monitor their progress in fitness or athletic training. For their part, motivations such as "follow the treatment plan prescribed by my physician" (447/1669, 26.78%), "improve communication with my physician" (400/1669, 23.97%), and "reduce the number of medical visits" (381/1669, 22.83%) were not the primary drivers of use. Quite conversely, traditional self-trackers tend to monitor specific clinical parameters related to chronic conditions such as weight, heart rate, glucose level, and medication intake (see Table 4).

XSL•FO
RenderX

Table 4. Health aspects monitored by digital and traditional self-trackers.

Dimension and health aspects	Digital self-trackers (N=1669), n (%)	Traditional self-trackers (N=1051), n (%)
Well-being		
Physical activity	856 (51.13)	441 (41.96)
Nutrition and eating habits	545 (32.65)	392 (37.30)
Sleep patterns	482 (28.88)	320 (30.45)
Performance in sports	256 (15.34)	59 (5.61)
Medical		
Weight-related data	483 (28.94)	585 (55.66)
Cardiovascular and respiratory health (eg, heart rate)	215 (12.88)	300 (28.54)
Medication intake	126 (7.55)	339 (32.25)
Glucose level	79 (4.73)	247 (23.50)

Adoption and Use of Wearable Connected Care Technologies

At the time of our survey, 74% of respondents (or 86% of those with a mobile phone or tablet) had already heard of consumer health wearables and smart medical devices. However, the level of familiarity with these remains relatively low, as only 16% of respondents who had already heard of such devices also reported being "very or extremely" familiar with them. We found that the level of familiarity with these tools is negatively correlated with age ($r=-.21$; $P<.001$) and positively correlated with family income ($r=.14$; $P<.001$). Indeed, millennials ($\chi^2_4=50.0$; $P<.001$) and people with annual family income over $80,000 CAD ($\chi^2_4=28.2$; $P=.03$) were more likely to be familiar with digital self-tracking devices than the other groups.

More importantly, our findings indicate that 1014 out of 1669 digital self-trackers (60.75%) own one or several wearables or smart medical devices, representing 24.68% of the entire sample. This ratio is similar to recent statistics from the United States. Indeed, according to two 2016 market reports, between 21% and 27% of American adults owned at least one such device [13,14]. Among our survey respondents, 70.02% (710/1014) said they had one connected wearable device, 20.71% (210/1014) had two, and 9.07% (92/1014) had three or more. On average, Canadian adults were using 1.5 consumer health wearable or smart medical device in early 2017. As these products are relatively new on the market, and our field administration shortly followed the 2016 holiday season, it is not surprising to observe that a majority of owners (549/1014, 54.14%) had been using their devices for less than a year at the time of the survey.

Although 61.11% of digital self-trackers (1020/1669) said they owned one or several wearable connected devices, 34.75% (580/1669) actually use them to self-track one or several aspects of their health. A multinomial logistic regression was performed to model the relationship between the predictors and membership in the two groups (nonusers and users of smart wearables). Addition of the predictors to a model that contained only the intercept significantly improved the fit between model and data; χ^2_{18} (N=4109)=154.82, Nagelkerke R^2=0.23, $P<.05$. As shown in Table 5, our analyses determined no statistically significant differences between users and nonusers of smart wearables in terms of gender, region, occupation, and education level. However, significant differences were observed in terms of age, gross annual income, perceived health status, and chronic condition(s). For one thing, millennials and people in the age range of 35 to 54 years were 2.2 and 1.6 times more likely to use digital devices to self-track their health than baby boomers (55+ years). For their part, people with annual family income inferior to $60,000 CAD were 0.38 times less likely to use smart digital devices than those with annual incomes of $100,000 CAD or more. Finally, our findings indicate that it is people who perceive themselves to be in very good or excellent health condition and those with no chronic condition who are current users of digital self-tracking devices.

In terms of usage, the most popular device is by far the bracelet or smartwatch, which is owned by 87.2% of those who own at least one such device (see Table 6). The main advantage often associated with wrist-wearable trackers is that they can monitor a range of health parameters and align with the common practice of wearing a watch. The bathroom scale and pedometer were the next most common connected devices used by Canadians.

Respondents were also asked how often they use connected self-tracking devices. Answers to this question varied across devices and according to the users' specific needs. For instance, 77.5% of those who have a bracelet or a smartwatch (392/506) use it several times per day. For its part, the bathroom scale is generally used once a day (38/119, 31.9%) or several times per week (41/119, 34.5%), whereas a minority (13/119, 10.9%) use it only a few times per month. As a final example, the blood pressure monitor is used once a month (12/47, 25%), once a day (11/47, 24%), or a few times per day (8/47, 16%) depending on the individual's condition and needs.

Table 5. Multinomial logistic regression model predicting usage of health wearables and smart medical devices by patient characteristics. Reference category=nonusers (N=3529).

Characteristics	Users of health wearables and smart medical devices (N=580)	
	Odds ratio (95% CI)	Significance
Intercept	—	<.001
Gender		
Female	1.041 (0.846-1.282)	.70
Age (years)		
18-34	2.234 (1.577-3.167)	<.001
35-54	1.566 (1.128-2.174)	.007
Region		
Atlantic provinces	0.962 (0.592-1.563)	.88
Quebec	0.752 (0.522-1.083)	.13
Ontario	1.120 (0.811-1.546)	.49
Manitoba-Saskatchewan	0.993 (0.605-1.629)	.98
Alberta	1.242 (0.834-1.850)	.29
Gross family income ($ CAD)		
≤60K	0.381 (0.262-0.554)	<.001
>60K and ≤100K	0.638 (0.511-0.797)	<.001
Education level		
High school or college	0.861 (0.644-1.152)	.31
Undergraduate	1.071 (0.809-1.419)	.63
Occupation		
Workers	1.255 (0.859-1.833)	.24
Students	0.377 (0.146-0.975)	.04
Others	0.780 (0.471-1.292)	.34
Perceived health status		
Very poor or poor	0.428 (0.267-0.685)	<.001
Fair or good	0.689 (0.556-0.854)	<.001
Chronic disease(s)		
One or more chronic condition(s)	0.784 (0.615-0.998)	.049

Data Sharing With Health Care Providers

This study indicates that there are relatively few people who regularly share the data captured with their digital self-tracking devices. In fact, only 34.87% of users (582/1669) reported that they share their personal health data. When they do so, it is primarily with family members (352/582, 60.5%), friends (294/582, 50.5%), and to a much lesser extent, an HCP such as a family doctor (195/582, 33.5%) or a pharmacist (50/582, 8.6%). Although no direct comparisons could be made with other surveys, empirical evidence in the United States shows that 78% of adults who use health wearables would like their doctor to have access to their personal data [14]. Another recent survey conducted in Canada reveals that 67% of users of mobile apps would share their data if their doctor requested it [24].

Users' Appreciation of Connected Care Technologies

As shown in Tables 7 and 8, users of consumer wearables and connected devices claimed to be very satisfied (mean=4.1 on a 5-point Likert scale), perceived their devices to be user-friendly (mean=4.2), and had a firm intention of continuing to use them in the future (mean=4.3). Overall, respondents perceive these devices as highly useful. About 7 out of 10 users (398/580) said that they have maintained or improved their health status by using digital self-tracking connected devices. Importantly, a majority of users report they are more informed or more knowledgeable about their health condition. Close to 6 out of 10 users (435/580) said they feel more confident taking care of their health or more autonomous in the management of their condition. On the other hand, feeling less anxious about one's own health and having more informed discussions with a doctor were not perceived as major benefits digital self-trackers in our study.

Table 6. Types of consumer wearables and smart medical devices among Canadian adults who use at least one such device (N=580).

Types of wearables	n (%)
Bracelet, wristband or smartwatch	506 (87.2)
Bathroom scale	119 (20.5)
Pedometer	76 (13.1)
Blood pressure monitor	47 (8.1)
Intelligent toothbrush	38 (6.6)
Pulse oximeter or spirometer (respiratory functions)	35 (6.0)
Thermometer	33 (5.7)
Glucose monitor	25 (4.3)
Intelligent clothes (eg, pants, shirts, and socks)	20 (3.4)
Spirometer	16 (2.8)
Intelligent pill dispenser	14 (2.4)
Intelligent fork	11 (1.9)

To further investigate users' appreciation of digital self-tracking devices, we tested a research model derived from the works of Bhattacherjee [20] and Hong et al [21], as well as expectation-confirmation theory [25]. To our knowledge, no prior research has investigated the factors influencing the continued usage of these devices. As shown in Figure 1, our model suggests that an individual's intention to continue using health wearables and smart devices is mainly influenced by his or her level of satisfaction. In turn, user satisfaction is influenced by the extent to which initial expectations toward these devices are confirmed, as well as by two factors from the Technology Acceptance Model (TAM) proposed by Davis [19], namely, ease of use and perceived usefulness. Following Hong et al [21], our model also proposes direct links between the TAM constructs and the dependent variable.

The reliability of the measures included in the model was determined with Cronbach alpha. Findings in Table 6 indicate that all the measures, without exception, surpass the 0.70 threshold of statistical significance [26]. This table also demonstrates the validity of the variables included in our research model. In particular, we see that the square root of the variance shared by each variable and its respective items is greater than the intercorrelations between the variables.

PLS regression analyses were performed to test the links in the model. Our findings in Figure 1 supported all relationships, and the model explains 64% of the variance in the dependent variable. Our results indicate that expectations confirmation is strongly related to TAM factors and user satisfaction. This result shows the importance of properly managing consumers' initial expectations to ensure greater adherence and continued usage of health wearables and smart medical devices. Future research on this topic may consider other variables such as information quality and personalization of content that have been recognized as facilitators for adherence (eg, [27]).

Reasons for Abandoning the Use of Digital Self-Tracking Devices

A 2015 study [28] suggested that one-third of consumer wearables end up in a drawer 6 months after purchase or initial use. We observed a slightly reduced observation of this phenomenon, with 25.54% of owners (259/1014) who had stopped using their connected devices at the time of the survey. When asked "Why did you stop using your device(s)?" a majority of respondents (111/259, 42.9%) said they had "lost interest after a while." Other reasons included malfunctioning of the device (51/259, 19.7%), doubts about the reliability of the data (39/259; 15.1%), and "the device was acquired more out of curiosity" (38/259, 14.7%). Most interestingly, we found that while abandoning use of these devices was not associated with gender, age, region, education, or main occupation, it was more prevalent among those who perceive there current health status as "poor or fair" compared with those who self-report their health status as good or excellent (χ^2_4=6.6; P=.048).

Reasons for Not Owning Digital Self-Tracking Devices

Respondents who do not own consumer health wearables or connected devices (n=2035) were asked why. Our results indicate that 46.93% of this segment (955/2035) did not see the interest in owning such tools. Other obstacles to greater diffusion of digital self-tracking devices were related to cost (836/2035, 41.08%), lack of knowledge about the value or benefits associated with the use of these devices (368/2035, 18.08%), and doubts about the reliability of data (341/2035, 16.76%). As many nonowners have limited knowledge of the value proposition for such devices, it is not surprising to observe that intentions to buy and adopt one in the near future were relatively low. Indeed, slightly less than 15% of nonowners (14.99%, 305/2035) reported that they intend to acquire a health wearable or smart medical device in the next 12 months.

Table 7. Users' appreciation of connected care technologies.

Variable and items	Somewhat or strongly disagree, n (%)	Neutral, n (%)	Somewhat or strongly agree, n (%)
Perceived usefulness			
I have maintained or improved my health condition	31 (5.4)	151 (26.1)	398 (68.5)
I am more informed about my health	47 (8.1)	147 (25.1)	387 (66.6)
My knowledge of my health condition has improved	51 (8.8)	179 (30.9)	350 (60.3)
I feel more confident taking care of my health	51 (8.8)	194 (33.5)	435 (57.7)
I am more autonomous in the management of my health	37 (6.4)	215 (37.1)	328 (56.5)
I feel less anxious about my health	81 (14.1)	239 (41.2)	259 (44.8)
I have more informed discussions with my doctor	94 (16.1)[a]	249 (42.9)	238 (41.0)
User friendliness			
I find it easy to use my wearables or smart devices	18 (3.1)	57 (9.8)	506 (87.1)
I find my wearables or smart devices user-friendly	22 (3.9)	58 (9.9)	500 (86.2)
Learning how to use my wearables or smart devices was easy	28 (4.9)	65 (11.3)	486 (83.9)
The information provided stored in the mobile apps is easy to understand and interpret	29 (5.0)	57 (9.9)	493 (85.1)
User satisfaction			
I am satisfied with the use of my wearables or smart devices	28 (4.8)	71 (12.2)	481 (83.0)
I am pleased with the use of my wearables or smart devices	28 (4.8)	71 (12.2)	481 (83.0)
I am delighted with the use of my wearables or smart devices	25 (4.4)	114 (19.6)	441 (76.0)
Confirmation of initial expectations			
My initial expectations concerning my use of wearables or smart devices have been confirmed so far	26 (4.6)	109 (18.7)	445 (76.7)
Using my wearables or smart devices turned out to be easier that I first thought	36 (6.2)	141 (24.3)	404 (69.5)
There are more benefits to using my wearables or smart devices than I first thought	42 (7.3)	150 (25.8)	388 (66.8)
Intention to continue using			
I have every intention of continuing to use wearables or smart devices in the future	23 (4.0)	45 (7.8)	511 (88.2)
I will continue to use wearables or smart devices to monitor different aspects of my health	19 (3.2)	70 (12.0)	492 (84.7)
I have no intention of stopping my use of wearables or smart devices in the future	22 (3.9)	64 (11.1)	493 (85.1)

Table 8. Descriptive statistics and variance shared by the variables (N=580). The ratios in italics on the diagonal represent the square root of the variance shared by each variable and its respective items. The ratios above the diagonal are Pearson correlation coefficients between variables.

Variables	Mean (SD); 1-5	Number of items	Cronbach alpha	Perceived usefulness	Ease of use	Confirmation of initial expectations	User satisfaction	Intention to continue usage
Perceived usefulness	3.6 (0.7)	7	.90	*.80*	.53[a]	.77[a]	.66[a]	.56[a]
Ease of use	4.2 (0.7)	4	.92	—	*.90*	.71[a]	.73[a]	.74[a]
Confirmation of initial expectations	3.9 (0.7)	3	.80	—	—	*.84*	.77[a]	.67[a]
User satisfaction	4.1 (0.8)	3	.89	—	—	—	*.90*	.70[a]
Intention to continue usage	4.3 (0.8)	3	.91	—	—	—	—	*.92*

[a]$P<.001$.

XSL•FO
RenderX

Figure 1. Users' appreciation of smart devices (N=580); ***P<.005; **P<.01; *P<.05.
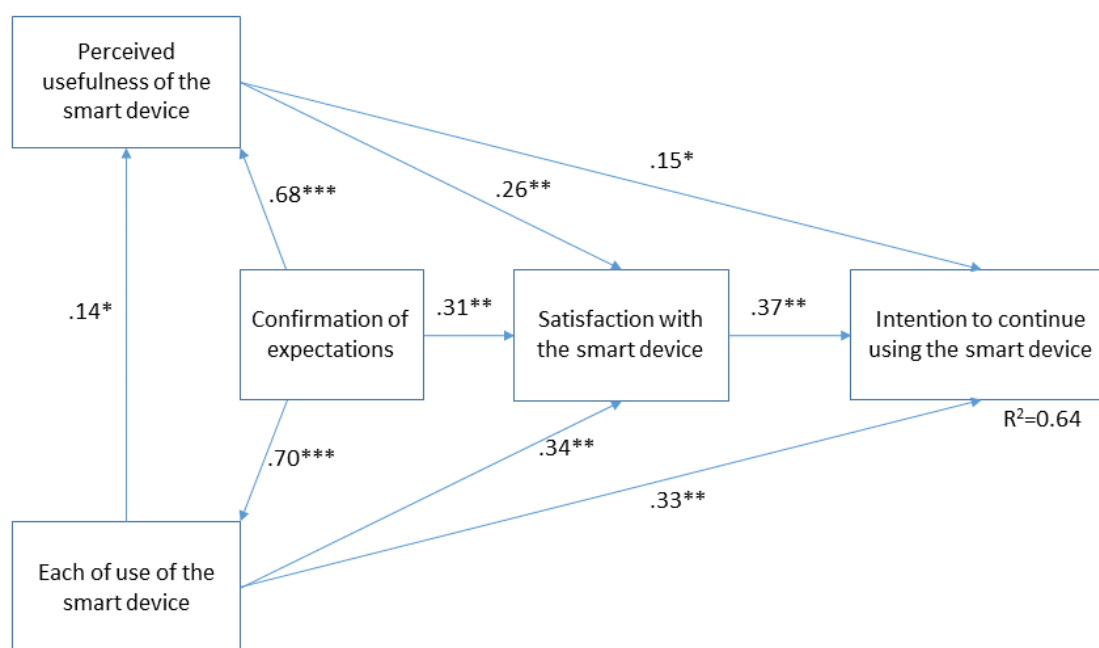
Discussion

Strengths and Limitations

This study investigates Canadian adults' digital health self-tracking behaviors and their use of connected wearables and devices to monitor aspects of their health and well-being. To our knowledge, it is one of the most comprehensive studies on this topic, resulting in highly reliable estimates of findings. Hence, our results set important baseline information that will guide future research on the evolution of the quantified-self phenomenon. Importantly, these findings are relevant to the information and technology industry and mHealth app developers to better understand the current market, segments, and viabilities to achieve behavioral and clinical outcomes. We further contributed to the extant literature by investigating novel, yet important aspects and issues, including the reasons for self-tracking and nontracking, the barriers to adoption of digital devices, consumers' appreciation of wearables and smart devices, the perceived benefits associated with digital self-tracking, and the reasons for usage discontinuance. Hence, this work may inform future policies and efforts in relation to general incorporation of self-tracking digital devices as supportive tools for patient care or reimbursement for technology-enabled quality outcome models of care.

Notwithstanding these strengths and contributions, our results must be interpreted with caution because of some inherent limitations. First, the responses relied on self-report and included only people who participate in Web panels managed by the survey company. Second, this is a cross-sectional survey, and while helpful for examining self-tracking behaviors and use of connected care technologies at one point in time, it is likely that people vary their use patterns and behaviors over time. Third, we did not collect data about race and ethnicity, although these variables might be related to the use of connected health

technologies. Finally, our survey did not include people's health literacy, which may represent an important moderator.

Implications for Practice and Research

Our survey first reveals that the digital health self-tracking movement in Canada is still in an early stage. About one in 4 respondents currently owns a health wearable or smart medical device. Among them, 57.20% (580/1014) use their smart devices on a regular basis for self-tracking purposes. Digital health self-trackers are mainly young, highly educated, and wealthy individuals whose main motivation for use of connected technologies is to monitory or quantify their fitness behaviors or progress on fitness goals. These results indicate an important presence of a significantly health engaged and activated segment of the Canadian population. Indeed, our findings show that many Canadian adults self-track aspects of their health because of the ubiquitous nature of mobile apps for health and consumer wearables and connected devices.

Although the use of connected care technologies could potentially benefit the growing population of patients with chronic conditions [29,30], the question remains as to whether it will diffuse broadly beyond early adopters and across cost inequities. Although technology manufacturers may assume patients with chronic conditions have unlimited enthusiasm for tracking their own health data using self-tracking devices, reality seems to be otherwise. Indeed, our findings show that 29.3% of those with chronic conditions in our sample had abandoned the use of their devices at some point compared with 13.8% for those with no chronic diseases (*P*=.04). A plausible explanation may be that chronic patients often consider it work (ie, a consuming and tiring task) to track their own health data [31]. This would suggest that digital self-tracking devices will successfully spread among chronic patients only if they are highly activated as patients and use is not experienced as a burden on the user but a positive and rewarding user experience.

The use of gamification and positive reinforcement techniques [32,33] may represent effective ways of making the experience more enjoyable and useful for chronic patients. Another explanation may be associated with the fact that medical parameters being tracked by chronic patients can be emotionally charged [31]. Indeed, "bad" data values can be extremely upsetting for many patients, especially when they are perceived to have some link to behavior. A third explanation may be that HCPs, especially physicians, do not seem interested in patients' self-logged data—even data that may be entirely objectively logged [31]. We believe chronic patients, especially those with severe conditions, may need personal coaching and continued support from HCPs to ensure adherence, system continuance, and positive health outcomes. This recommendation is also aligned with the importance of properly managing users' expectations discussed earlier.

To deepen our understanding of usage discontinuance, a prevalent phenomenon identified in these results, we suggest that future research include an approach that borrows a model proposed by Li et al [7]. This model outlines five psychological stages in the process of engaging in digital self-tracking. The first stage, called preparation, concerns people's motivation to collect personal information, how they determine what information to collect, and how they will record it. The next step, collection, is when people collect information about themselves. Integration is the stage where the information collected is prepared, combined, or transformed for the use to reflect on (reflection stage). Finally, action is when people choose what they are going to do with their newfound understanding of themselves (eg, people may tailor their behaviors to match their goals). A key finding of Li et al's study is that individuals have a tendency to focus on a single stage (ie, collecting data on number of steps or hours slept) and to ignore the overall process and intended outcome of self-tracking for health outcomes. This reinforces the importance of providing professional coaching and continued support. Among others, future studies could use this model to identify the barriers that people, especially chronic patients, experience when they self-track using connected care technologies.

Although 70% of digital self-trackers in our sample feel they have maintained or improved their health by using wearables and smart devices, there is little empirical evidence that suggest self-tracking personal health indicators leads to long-term behavioral changes [34]. Among the few studies we found, one trial concludes that the use of pedometers along with nursing consultations increased physical activity among older adults [35]. In another trial, the use of a wearable tracker by overweight and obese adults led to a small increase in moderate-to-vigorous intensity physical activity at follow-up [36]. More research with large samples is definitely needed to determine the effectiveness of wearable connected devices on people's physical activity. It is also unclear whether such devices can motivate adults of all age groups toward other important health mediating behaviors such as adopting a healthy diet, maintaining a healthy weight, adopting good sleep habits, and not smoking. Because healthy behaviors will lead to significant improvements in population health only if they are sustained [37], it will also be important that future (longitudinal) trials investigate whether and under which conditions (eg, health literacy) digital health self-tracking devices can support the creation and maintenance of enduring new lifestyle habits and improve quality of life.

Finally, prior research shows that connected care technologies may create new opportunities for individuals who desire to participate actively in and take responsibility for their personal health. As discussed in Kitsiou et al [38], mobile apps along with consumer wearables and smart medical devices can provide a platform for home telemonitoring programs for chronic patients. Furthermore, physicians can use wearable sensors to monitor acute patients' health in real time, which can aid with diagnosis and treatment decisions [34]. For instance, chronic sleep apnea can be diagnosed with a lightweight wearable that measures heart rate, breathing volume, and snoring instead of a heavy polysomnography assessment [39]. In addition, incorporating wearable and smart device sensors into routine care may improve clinician-patient relationships and increase patient empowerment [40]. It also appears that the widespread integration of these devices into medical practice by clinicians is extremely limited [41]. A recent survey of 989 Canadian HCPs shows that only 30% recommend wearable trackers (eg, smartwatch and bracelet) to their patients, and 25% recommend medical smart devices such as blood pressure monitors and sleep trackers [42]. Several barriers related to patient safety, data accuracy and security, reimbursement policy, and government regulation have been discussed in the extant literature (eg, [29,43]). Future research must continue investigating these important issues for practicing clinicians so that we develop a better understanding of how and under which circumstances the use of connected care technologies can best serve medicine, in general, and prevention and management of chronic conditions, in particular.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey instrument.

[PDF File (Adobe PDF File), 213KB - jmir_v20i5e177_app1.pdf]

References

1. Topol E. The Patient Will See You Now: The Future of Medicine Is in Your Hands. New York, NY: Basic Books; 2015.
2. PwC Health Research Institute. PWC. Health wearables: early days URL: https://www.pwc.com/us/en/health-industries/health-research-institute/publications/health-wearables-early-days.html [accessed 2017-10-26] [WebCite Cache ID 6uViK9t3e]
3. Gartner. 2017. Gartner says 8 billion connected things will be in use in 2017, up 31 percent from 2016 URL: https://www.gartner.com/newsroom/ [accessed 2017-10-27] [WebCite Cache ID 6uWhuGgxF]
4. Soreon Research. 2014. The wearable health revolution URL: http://www.soreonresearch.com/wearable-healthcare-report-2014/ [accessed 2017-10-27] [WebCite Cache ID 6uWfIU59C]
5. Hoy MB. Personal activity trackers and the quantified self. Med Ref Serv Q 2016;35(1):94-100. [doi: 10.1080/02763869.2016.1117300] [Medline: 26794199]
6. Lupton D. The Quantified Self: A Sociology of Self-Tracking. United Kingdom: Cambridge: Polity Press; 2016.
7. Li I, Dey A, Forlizzi J. A stage-based model of personal informatics systems. 2010 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 10-15, 2010; Atlanta, GA.
8. Altschuler A, Picchi T, Nelson M, Rogers JD, Hart J, Sternfeld B. Physical activity questionnaire comprehension: lessons from cognitive interviews. Med Sci Sports Exerc 2009 Feb;41(2):336-343 [FREE Full text] [doi: 10.1249/MSS.0b013e318186b1b1] [Medline: 19127192]
9. Archer E, Pavela G, Lavie CJ. The inadmissibility of what we eat in America and NHANES dietary data in nutrition and obesity research and the scientific formulation of national dietary guidelines. Mayo Clin Proc 2015 Jul;90(7):911-926 [FREE Full text] [doi: 10.1016/j.mayocp.2015.04.009] [Medline: 26071068]
10. 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]
11. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
12. Fox S. Pew Research Center. Washington, DC: Pew Research Center; 2013. The self-tracking data explosion URL: http://www.pewinternet.org/2013/06/04/the-self-tracking-data-explosion/, [WebCite Cache ID 6ibYNkRt6]
13. Accenture Consulting. Accenture. 2016. Patients want a heavy dose of digital URL: https://www.accenture.com/us-en/insight-research-shows-patients-united-states-want-heavy [accessed 2017-10-26] [WebCite Cache ID 6uVitf2ZA]
14. Brown B. Health Tech Insider. 2016 Aug 25. Salesforce's 2016 connected patient report: insights into patient preferences on telemedicine, wearables and post-discharge care, 2016 URL: http://healthtechinsider.com/2016/08/25/2016-connected-patient-report/ [WebCite Cache ID 6uWiWu7XF]
15. IFOP. 2015. Objets connectes et usage des donnees: la perception des Francais URL: http://www.ifop.fr/media/poll/3250-1-study_file.pdf [accessed 2017-10-27] [WebCite Cache ID 6uWiqaOSm]
16. Global GfK. GFK. 2016. Health and fitness tracking URL: http://www.gfk.com/global-studies/global-studies-fitness-tracking/ [accessed 2017-10-26] [WebCite Cache ID 6uVjHS2UI]
17. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
18. Bowling A. Just one question: If one question works, why ask several? J Epidemiol Community Health 2005 May;59(5):342-345 [FREE Full text] [doi: 10.1136/jech.2004.021204] [Medline: 15831678]
19. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989;13(3):319-340.
20. Bhattacherjee A. Understanding information systems continuance: an expectation-confirmation model. MIS Quarterly 2001;25(3):351-370 [FREE Full text]
21. Hong S, Thong JY, Tam KY. Understanding continued information technology usage behaviour: a comparison of three models in the context of mobile internet. Decis Support Syst 2006;42(3):1819-1834. [doi: 10.1016/j.dss.2006.03.009]
22. Canadian Radio-Television and Telecommunications Commission. Crtc.gc. 2016. Communication monitoring report URL: http://crtc.gc.ca/eng/publications/reports/PolicyMonitoring/2016/cmri.htm [WebCite Cache ID 6uWj1GsJK]
23. Fox S, Duggan M. Pewinternet. Washington, DC: Pew Internet & American Life; 2013. Health Online URL: http://www.pewinternet.org/Reports/2013/Health-online/Summary-of-Findings.aspx [accessed 2013-11-20] [WebCite Cache ID 6LH0VgIUW]
24. Canada Health Infoway. 2016. Connecting patients for better health URL: https://www.infoway-inforoute.ca/en/component/edocman/resources/3152-connecting-patients-for-better-health-2016 [accessed 2018-01-27] [WebCite Cache ID 6uVjOPO5P]

25. Oliver RL. A cognitive model of the antecedents and consequences of satisfaction decisions. J Mark Res 1980:460-469 [FREE Full text]

26. Nunnally J. Psychometric Methods. New York, NY: McGraw-Hill; 1978.

27. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behaviour change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]

28. Ledger D, McCaffrey D. Blog.endeavour.partners. Cambridge, MA: Endeavour Partners; 2014. Inside wearables: How the science of human behaviour change offers the secret to long-term engagement, Endeavour Partners: Cambridge, MA URL: https://blog.endeavour.partners/inside-wearable-how-the-science-of-human-behavior-change-offers-the-secret-to-long-term-engagement-a15b3c7d4cf3?gi=b08cfb919c55 [accessed 2018-04-26] [WebCite Cache ID 6yy9e5sj4]

29. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med 2016 Feb;13(2):e1001953 [FREE Full text] [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]

30. Kitsiou S, Paré G, Jaana M. Effects of home telemonitoring interventions on patients with chronic heart failure: an overview of systematic reviews. J Med Internet Res 2015;17(3):e63 [FREE Full text] [doi: 10.2196/jmir.4174] [Medline: 25768664]

31. Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. You get reminded you're a sick person: personal data tracking and patients with multiple chronic conditions. J Med Internet Res 2015;17(8):e202 [FREE Full text] [doi: 10.2196/jmir.4209] [Medline: 26290186]

32. Leahey T, Rosen J. DietBet: a web-based program that uses social gaming and financial incentives to promote weight loss. JMIR Serious Games 2014 Feb 07;2(1):e2 [FREE Full text] [doi: 10.2196/games.2987] [Medline: 25658966]

33. Miller AS, Cafazzo JA, Seto E. A game plan: gamification design principles in mHealth applications for chronic disease management. Health Informatics J 2014 Jul 1;22(2):184-193. [doi: 10.1177/1460458214537511] [Medline: 24986104]

34. Wu Q, Sum K, Nathan-Roberts D. How fitness trackers facilitate health behaviour change. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting.: SAGE Publications; 2016 Presented at: Human Factors and Ergonomics Society Annual Meeting; 2016; Los Angeles, CA URL: http://journals.sagepub.com/doi/pdf/10.1177/1541931213601247

35. Harris T, Kerry SM, Victor CR, Ekelund U, Woodcock A, Iliffe S, et al. A primary care nurse-delivered walking intervention in older adults: PACE (pedometer accelerometer consultation evaluation)-Lift cluster randomised controlled trial. PLoS Med 2015 Feb;12(2):e1001783 [FREE Full text] [doi: 10.1371/journal.pmed.1001783] [Medline: 25689364]

36. Wang JB, Cadmus-Bertram LA, Natarajan L, White MM, Madanat H, Nichols JF, et al. Wearable sensor/device (Fitbit One) and SMS text-messaging prompts to increase physical activity in overweight and obese adults: a randomized controlled trial. Telemed J E Health 2015 Oct;21(10):782-792. [doi: 10.1089/tmj.2014.0176] [Medline: 26431257]

37. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. J Am Med Assoc 2015 Feb 3;313(5):459-460. [doi: 10.1001/jama.2014.14781] [Medline: 25569175]

38. Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: an overview of systematic reviews. PLoS One 2017;12(3):e0173160 [FREE Full text] [doi: 10.1371/journal.pone.0173160] [Medline: 28249025]

39. Harrington J, Schramm PJ, Davies CR, Lee-Chiong TL. An electrocardiogram-based analysis evaluating sleep quality in patients with obstructive sleep apnea. Sleep Breath 2013 Sep;17(3):1071-1078. [doi: 10.1007/s11325-013-0804-9] [Medline: 23354509]

40. Appelboom G, Camacho E, Abraham ME, Bruce SS, Dumont EL, Zacharia BE, et al. Smart wearable body sensors for patient self-assessment and monitoring. Arch Public Health 2014;72(1):28 [FREE Full text] [doi: 10.1186/2049-3258-72-28] [Medline: 25232478]

41. Canada Health Infoway. 2017. Use of Electronic Medical Records among Canadian Physicians URL: https://www.infoway-inforoute.ca/en/what-we-do/news-events/newsroom/2016-news-releases/6752-canadian-physicians-can-improve-patient-care-with-advanced-emr-use [accessed 2017-11-10] [WebCite Cache ID 6us0nF2kj]

42. Canada Health Infoway. 2017. Canadians and Digital Health URL: https://www.infoway-inforoute.ca/en/component/edocman/resources/reports/2815-infographic-report-on-digital-health-canadians-embracing-digital-health [WebCite Cache ID 6ecgdxQQt]

43. West DM. How mobile devices are transforming healthcare issues in technology innovation. Issues Technol Innov 2012;18:14 [FREE Full text]

Abbreviations

HCP: health care provider
mHealth: mobile health
PLS: partial least squares
TAM: Technology Acceptance Model

XSL•FO
RenderX

Experiences of a Motivational Interview Delivered by a Robot: Qualitative Study

Joana Galvão Gomes da Silva[1], MSc; David J Kavanagh[2], PhD; Tony Belpaeme[3], PhD; Lloyd Taylor[1]; Konna Beeson[1], MSc; Jackie Andrade[1], PhD

[1]Cognition Institute, School of Psychology, University of Plymouth, Plymouth, United Kingdom

[2]Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia

[3]Centre for Robotics and Neural Systems, University of Plymouth, Plymouth, United Kingdom

Corresponding Author:
Jackie Andrade, PhD
Cognition Institute
School of Psychology
University of Plymouth
Drake Circus
Plymouth, PL48AA
United Kingdom
Phone: 44 1752584807
Fax: 44 1752584808
Email: jackie.andrade@plymouth.ac.uk

Abstract

Background: Motivational interviewing is an effective intervention for supporting behavior change but traditionally depends on face-to-face dialogue with a human counselor. This study addressed a key challenge for the goal of developing social robotic motivational interviewers: creating an interview protocol, within the constraints of current artificial intelligence, which participants will find engaging and helpful.

Objective: The aim of this study was to explore participants' qualitative experiences of a motivational interview delivered by a social robot, including their evaluation of usability of the robot during the interaction and its impact on their motivation.

Methods: NAO robots are humanoid, child-sized social robots. We programmed a NAO robot with Choregraphe software to deliver a scripted motivational interview focused on increasing physical activity. The interview was designed to be comprehensible even without an empathetic response from the robot. Robot breathing and face-tracking functions were used to give an impression of attentiveness. A total of 20 participants took part in the robot-delivered motivational interview and evaluated it after 1 week by responding to a series of written open-ended questions. Each participant was left alone to speak aloud with the robot, advancing through a series of questions by tapping the robot's head sensor. Evaluations were content-analyzed utilizing Boyatzis' steps: (1) sampling and design, (2) developing themes and codes, and (3) validating and applying the codes.

Results: Themes focused on interaction with the robot, motivation, change in physical activity, and overall evaluation of the intervention. Participants found the instructions clear and the navigation easy to use. Most enjoyed the interaction but also found it was restricted by the lack of individualized response from the robot. Many positively appraised the nonjudgmental aspect of the interview and how it gave space to articulate their motivation for change. Some participants felt that the intervention increased their physical activity levels.

Conclusions: Social robots can achieve a fundamental objective of motivational interviewing, encouraging participants to articulate their goals and dilemmas aloud. Because they are perceived as nonjudgmental, robots may have advantages over more humanoid avatars for delivering virtual support for behavioral change.

Introduction

Background

Lifestyle factors such as physical inactivity impose a considerable burden on society's health care resources and individuals' well-being [1]. Participants in qualitative studies focusing on weight management say they want motivational support to make lifestyle changes [2,3], but public health budgets constrain society's ability to offer face-to-face counseling [4]. Social robots that can deliver effective motivational support could offer a way to increase access and encourage behavior change. This paper reports a study of participants' experiences of a robot-delivered motivational interview to support their goal of becoming more physically active.

Motivational Interviewing

Motivational interviewing (MI) [5] is one of the most effective psychological interventions for supporting behavior change [6,7], including for increasing physical activity (PA) [8,9]. The MI practitioner uses a person-centered counseling style to engage the client in discussion of their current problem and to elicit their own ideas for solutions. This collaborative stance is considered important, because people are likely to react to directive, advice-giving, (doctor-patient) counseling styles by trying to justify their current behavior [10,11]. The aim of MI is to encourage the client to voice their own arguments for change, as hearing oneself arguing for change increases belief that change is important and will happen [12]. Given the focus on personalized dialogue, MI delivered by a robot might seem a distant dream.

Social Robots

Although people may have preconceptions about robots from science fiction films, few have had opportunities to interact with one. Two streams of development dominated early robotics: remote navigation for observing hard-to-reach environments and manipulation for replacing human manual work in industries. Recently, there has been a new focus on humanoid robots as personal assistants or carers in daily life [13,14].

These social robots have been used to provide educational support for children [15] and assistance to elderly individuals [16]. They have proven acceptable and effective for helping children with type 1 diabetes to learn about their condition and how to manage it [17] and are being trialed as therapeutic aids for children with autism spectrum disorders, with results showing therapeutic outcomes similar to those of one-to-one therapy [18,19]. Robots have also become personal trainers, instructing and motivating the completion of exercises such as spinning, rowing, and bodyweights [20] or engaging elderly users in physical exercises [21]. They have served as weight loss coaches, stimulating tracking of calorie consumption and exercise, and being twice as effective as a stand-alone computer or paper log [22]. However, naturalistic dialogue between robots and humans is currently limited by robots' speech processing capabilities and the capacity of artificial intelligence to cope with unconstrained input [23]. The use of robots for therapy has therefore been limited to education and engagement rather than delivery of interventions where dialogue is critical.

Using Technology to Deliver Adaptations of Motivational Interviewing

There have been attempts at mechanizing delivery of MI using text, audio, video, and animations, with some success [24]. For example, Jackson and colleagues used branching logic and a prerecorded Video Doctor to guide pregnant women through a motivational interview. Their trial showed improvements in diet and PA [25] and reductions in smoking [26], although no clear effects on smoking abstinence and weight were observed. There was also evidence that the Video Doctor led to more women discussing partner violence with their health care practitioner [27]. Interfaces have generally relied on participants entering text or selecting preprogrammed options, making the intervention less person-centered than is ideal and removing the benefits central to MI of hearing oneself argue for change.

Social Robots as Motivational Interviewers

Social robots have the potential to engage participants in a motivational interview so that they hear themselves argue for a change. To our knowledge, only one other group has tested robots in this way. Kanaoka and Mutlu [28] used a NAO robot to deliver a motivational interview. They found no benefit of MI compared with a traditional advice condition. They attributed the lack of benefit of MI to a lack of fluency in the dialogue between the robot and the participant, with errors in speech recognition and incongruous nonverbal behaviors destroying the illusion of a meaningful two-way conversation. A complete motivational interview, with personally tailored questions and reflections upon the client's answers, still poses substantial challenges to robot speech recognition and artificial intelligence.

This paper reports the development and assessment of a simpler solution, using a social robot to elicit change talk with a preprogrammed set of questions. In contrast to previous attempts to automate MI, apart from Kanaoka and Mutlu's study, the focus of the interview was on encouraging participants to talk to the robot about their motivation for change, using open questions designed to draw attention to the discrepancy between the participant's current behaviors and core values. Apodaca and Loganbaugh [29] found that change talk and experience of discrepancy are the main mechanisms of change in MI. A preprogrammed set of questions falls short of the person-centered counseling style that is at the heart of MI. However, if this approach succeeds in encouraging participants to talk freely about their concerns and their plans, we contend that it would present a substantial step forward in the use of technology to deliver motivational support.

The aim of the study was to explore participants' experiences of talking to the robot in a dialogue based on MI but constrained by current technology. We specifically wanted to know how people felt about discussing their issues with the robot and whether they felt that the interview affected their motivation.

Methods

Motivational Interview Script

We created an intervention script using manuals developed for face-to-face motivational interviews in clinical trials [30,31] and Miller and Rollnick's book *Motivational Interviewing:*

Preparing People for Change [32] for guidance. Each question needed to make sense, regardless of how the participant answered the question before. To anticipate potential glitches in the flow of the dialogue, we iteratively role-played potential responses to the questions and adjusted the script where necessary.

We shaped participants' expectations by advising them, at the start of the interview, that, "During this interview, sometimes I may ask you questions that you think you've already answered. If that happens, I suggest you use it as an opportunity to think about the issue a bit more." The questions covered MI elements such as advantages and disadvantages of the status quo, optimism about change, intention to change, evocation of ideas about change, hypothetical change, setting goals, and arriving at a plan [32]. As in MI, the interview moved from a general discussion of the pros and cons of change to development of specific plans for change. The questions were designed to encourage participants to articulate their ideas about change and to consider the discrepancy between their current behaviors and core values. In a real motivational interview, the interviewer uses reflection as a tool for amplifying emotions associated with the pros and cons of change; repeating, paraphrasing, or elaborating salient statements made by the participant. Because this personalized reflection is not possible in a prescribed interview, we sought to amplify emotion using open questions to encourage the participant to think deeply about their incentives. For example, the robot asked, "What may happen in the future if you don't change anything?" followed by "Does that worry or concern you?" The script did not refer to any specific goals or behaviors, so that it could be generalized to many situations, but participants knew that the study was about increasing PA. The full script is provided in Multimedia Appendix 1.

To help readers understand the strengths and weaknesses of the robot's script, two of the authors trained in MI characterized it using Shingleton and Palfai's [24] schema for rating technology-delivered adaptations of MI, which was published after we developed the robot interview. Shingleton and Palfai scored features of MI as present or absent. To give a more nuanced picture, we rated the degree to which each quality of MI was present, from 0 (absent) to 3 (fully present; see Table 1). We note that the standard tool for evaluating the quality of MI, the Motivational Interviewing Treatment Integrity manual [33], is not applicable here because it is used for rating the interviewer's interaction with the client.

Programming the NAO Robot

We used a NAO robot (Figure 1) to deliver this intervention. NAO is developed by SoftBank Robotics with speech and movement capabilities. It is brightly colored, 58 cm tall, with large eyes and humanoid appearance. NAO was chosen for this intervention because of its user-friendly software package Choregraphe and because it has been well received by participants in previous research [34]. The robot was programmed with Choregraphe software, which is used to create behaviors, monitor, and control the NAO robot. The instructions were sent wirelessly to NAO so that the experimenters were able to run the programmed script from a computer in a different room. The experimenters could monitor the progress of the interview visually via a live camera in NAO's head.

Much of the programming effort was devoted to determining the intonation and speed of each sentence so that the questions were easy to hear. The robot's voice pitch was kept at the default from the manufacturer. A simpler approach would be to record a human asking the questions and have the robot replay the recording. We felt this option was not viable because it would destroy the illusion that the robot was asking the questions itself. In addition to programming the robot's speech, we incorporated some ready-programmed modules that come with the Choregraphe software package, to give the robot a life-like, animated appearance. These modules included breathing, in which the robot sways gently to emulate breathing and slight fidgeting, and eye color. The robot's eye color changed from blue (question) to green (answer) to indicate whether the participant should listen or answer at each point in the interview.

Table 1. Characterization of the robot-delivered motivational interview using Shingleton and Palfai's criteria for assessing technology-delivered adaptations of motivational interviewing (MI). Examples from the robot script are quoted in italics.

Rating	Components of motivational interviewing
0. Absent	• Roll with resistance • Structure adapted to readiness to change or interest or self-efficacy • Express empathy
1. Present but superficial or inadequate	• Promote autonomy, ask permission: *is it okay if we talk about this now?* • Collaboration: *let's focus on...* • Other MI adherent behaviors: *how does that make you feel?* (amplifying emotion)
2. Present but not optimal	• Develop discrepancy or explore ambivalence: *what may happen in the future if you don't change anything?* • Reflections or summary: summary was used— *I suggest you summarize what you are going to do...* —but reflection is not possible in a pre-scripted interview.
3. Fully present	• Evocation: *why is that important to you now?* • Promote self-efficacy: *what could you do, to make sure you follow your plan over the next week?* • Strengthen commitment to change: *try summarizing the things that are likely to get better if you change your behavior* • Open-ended questions: *what would be the first step?*

The face-tracking mode enabled the robot to follow the participant's face, regardless of his or her movements during the interview, to give a sense that NAO was paying attention.

Setting

For the interview with the robot, we utilized a laboratory (Figure 1), styled as a living room, to create a relaxed atmosphere. The participant was left alone with the robot in this room.

Evaluative Questionnaire

A questionnaire was developed to explore participants' experiences of their interview with the robot and their impressions of its impact on their motivation. We used an anonymous, computerized questionnaire rather than a semistructured interview because we wanted participants to feel as free as possible to give an honest account of their experiences and not feel socially pressured into praising the robot.

The questionnaire included 24 open questions. The questions were designed to address the primary aim of the study: to explore participants' experiences of a motivational interview delivered by a social robot. To ensure that participants considered different aspects of the interview, questions asked about how they felt during the interview (eg, "How was your interaction with the robot?" and "How engaging did you find the interview with the robot?"), how easy they found the robot to use and understand (eg, "How was your understanding of each question? Was the content clear?" and "How did you find the robot's interface? Was it easy or difficult to use?"), how they felt about listening to themselves discussing their goals aloud (because this is a core component of MI), and whether they perceived an impact of the interview on their motivation ("Did this interview with the robot affect your motivation? How?" and "Did you improve your physical activity after the robot interview? How?"). To encourage a balanced appraisal, two questions asked specifically what participants found the best, and worst, aspects of the intervention. The full questionnaire is provided in Multimedia Appendix 2.
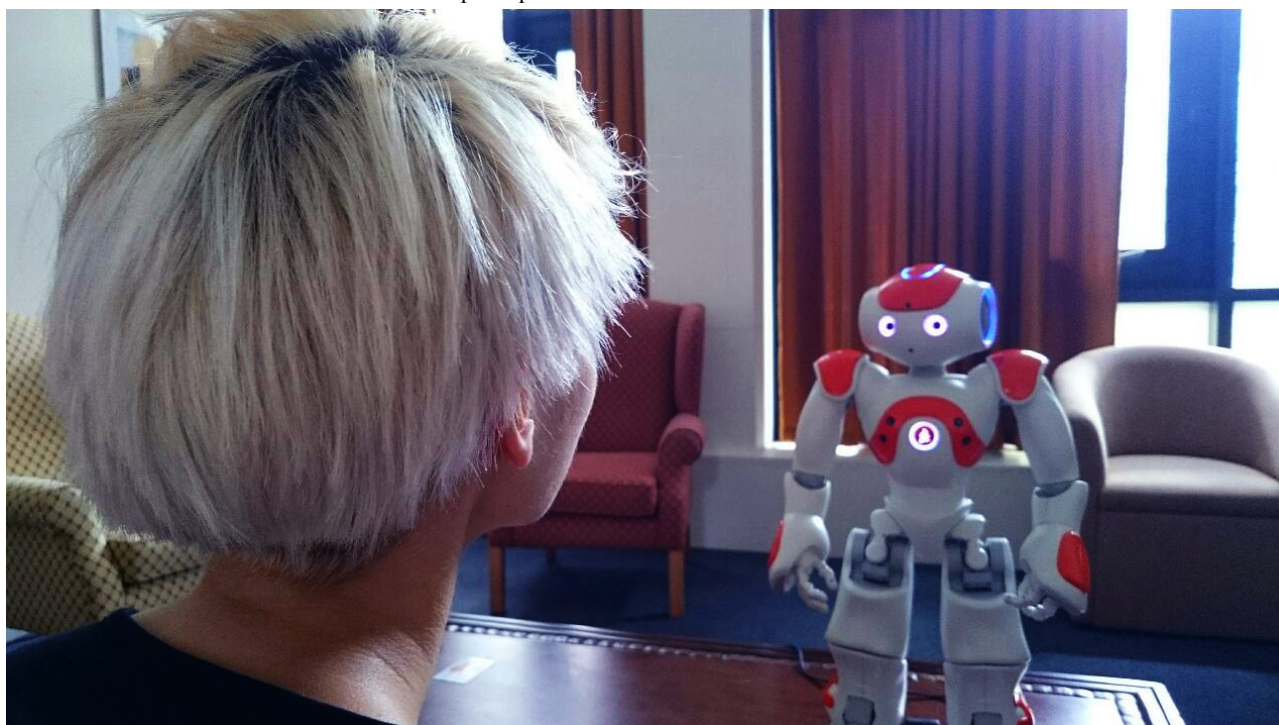
Participants

A total of 20 participants (17 female, 3 male) was recruited from the School of Psychology's pool of research volunteers. Participants were required to be aged 18 years or above and received £8 per hour to participate. Eleven participants were aged 18 to 25 years, 4 participants 26 to 33 years, 1 participant 34 to 42 years, 2 participants 43 to 60 years, and 2 participants above 61 years.

The study advertisement asked for volunteers wishing to increase their PA. Participants were informed that they would take part in a PA intervention over two sessions, which included an interview with a robot. The study was approved by the University of Plymouth Faculty of Health and Human Sciences Research Ethics Committee, and informed consent was signed by participants, who were assured of anonymity and told they could withdraw at any point of the study as per British Psychological Society guidelines. They were told that the interview would take approximately 15 min and would not be recorded. Participants were advised not to take part in the study and to seek medical assessment if they had any concerns about their health or ability to exercise.

Procedure

The robotic intervention was comprised of two phases: lab session I and lab session II. There was a 1-week interval between them to allow time for participants to reflect on any impact of the interview while minimizing forgetting and intervening variables.

Figure 1. An illustration of the interaction between a participant and the NAO robot.

XSL•FO

RenderX

In session I, participants answered the robot's questions out loud in a simulated conversation with the robot, with the participants touching the robot's head sensor to advance to the next question. As previously stated, NAO's eye color changed from question to listening mode when it was the participant's turn to speak. Participants were not alerted to this feature, as we intended it as a subtle turn-taking cue. Interviews were not recorded because participants in pilot work anticipated that they would feel uncomfortable talking to the robot as it was a novel experience and would prefer not to be recorded. We return to this issue in the Discussion.

One week later, in lab session II, participants returned to the lab and evaluated the intervention through a computerized evaluative questionnaire with open-ended questions and typed answers.

Thematic Content Analysis

Participants' answers to the evaluative questionnaire were content-analyzed utilizing a three step methodology recommended by Boyatzis [35]: (1) sampling and design, (2) developing themes and codes, and (3) validating and applying the codes. The first step of the content analysis was to define the set of units of analysis to be investigated further. The full set of responses to the qualitative questionnaire from each participant was delineated as each unit of analysis. There was a total of 20 units of analysis, one set per participant. The coding scheme originated from the text itself, with the main and subthemes being developed based on Boyatzis' [35] steps: (1) generating a code, (2) reviewing and revising the code, and (3) determining the reliability of the code.

Immersive readings of the units of analysis led to the development of a series of potential codes. The text was further analyzed and generated a set of themes and subthemes based on common recurrent topics. Analysis continued until no new themes emerged.

In validating the code, it is important to check that the coding scheme can be applied consistently. Boyatzis recommends having two independent coders rate a subsample separately and computing the interrater reliability (IRR). We did this in two stages. First, two coders directly involved in the study (JGGdS and JA) rated two units of analysis independently. Differences in applying the code were discussed, and the coding scheme was adjusted accordingly. Then, two new raters, with no involvement in the study, applied the adjusted coding scheme to five further randomly selected units of analysis by deciding if each item in the code was mentioned or not. IRR was computed as the percentage of items agreed upon for each unit of analysis.

Results

Thematic Content Analysis

The coding scheme distinguished between experiences of interacting with the robot, participants' own strategies for and barriers to motivation, PA in the week following the intervention, and overall evaluation of the intervention, which included suggestions for improvements. There was some similarity between the themes *interview evaluation* and *overall evaluation*. The interview evaluation theme incorporated answers to most of the questions and covered specific feelings experienced during the interview (for example, feeling relaxed, engaged, or self-conscious) and usability of the interface. The overall evaluation theme covered impressions of the intervention as a whole and suggestions for improvements, particularly but not solely covering responses to the questions about the best and worst aspects of the intervention. The theme on *motivation* covered ideas that participants used spontaneously, whereas the *PA* theme covered impressions of whether the interview affected motivation and activity in the week after the interview. The full coding scheme is shown in Multimedia Appendix 3.

Two experienced raters applied the coding scheme to two sample units of analysis. The IRRs for these units (P2 and P9) were 90% (52/58) and 97% (56/58), respectively. After slight adjustment of the coding scheme, two new raters, who were naïve to the purposes of the study, applied the adjusted coding scheme to five more units of analysis. Their mean IRR was 85.9% (249/290), ranging from 83% (48/58) to 91% (53/58). The coding scheme was assumed reliable.

Participants' Evaluation of the Intervention

Theme 1: Interview Evaluation

Participants' evaluations of the interview clustered around four subthemes: how they felt about the interaction with the robot, their evaluation of the script, usability of the interface, and their experiences of hearing themselves speaking aloud to the robot.

Interaction or Connection With the Robot (1.1)

Most participants found the interaction smooth, felt relaxed or comfortable around the robot, and enjoyed the experience. Others found the experience interesting, unusual, or surreal. Most of the participants had an initial moment of tension followed by a period of relaxation after they became used to the robot. Although the novelty of being in proximity to a robot contributed to the initial awkwardness, it also added to the enjoyment of the experience, as illustrated below:

> *[My experience with the robot was] fine, if not a little awkward. The more time spent with the robot, the more relaxed I felt. [It was] easier to talk to than an actual person.* [P2, age range: 18-25, female]

> *It was a very novel experience as I had never been in such close proximity to a robot before and I certainly found it engaging.* [P3, age range: above 61, female]

> *I enjoyed interacting with the robot. It was like guided self-reflection. I was slightly nervous initially, but this soon passed and it became enjoyable.* [P13, age range: 26-33, male]

> *[It was] fun [talking to the robot]. It made me laugh to see its eyes change colour plus its squeaky voice was a giggle. After a while I forgot about the novelty of it all and just started to answer normally. Occasionally, though, it spouted out too much verbiage and I lost the plot. Over all, a good experience and one which will remain in my mind...It just felt like talking to a fun medical person—without*

the disinterested look of your average GP. [P15, age range: 43-60, female]

For some participants, the lack of a personal response prevented them feeling connected with the robot:

I don't think I interacted as I would have done [with] a human (I tend to look people in the eyes as I talk) as I didn't feel a need to connect with it. [P2, age range: 18-25, female]

However, this participant [P2] later identified advantages of the robot over a human interviewer:

Was easier to talk to than a human so suppose that made the conversation more engaging in that way as I felt able to open up more but really I didn't feel as if the robot was interested in what I had to say, obviously.

Others also drew comparisons with talking to a human, and some preferred it because they felt they could talk without being judged:

Strange, felt like I was talking to a human. I have never experienced an interview with a robot before so it was an unusual experience. [P4, age range: 26-33, female]

Possibly better than talking to a human as I wasn't being judged eg with bored looks, bored body language, cutting words. [P15, age range: 43-60, female]

...allowed you to be more honest as it's not a human so no judgement. [P14, age range: 18-25, female]

Script (1.2)

Most participants found the questions clear and easy to understand. Some had problems with some questions being too vague or ambiguous and having doubts about how to address them, although often they were not able to remember which questions had been problematic. Even though the robot warned at the beginning that it might sometimes ask questions that the participant had already answered, participants sometimes found it disconcerting or frustrating when this happened. This repetition could also be experienced as a positive feature. Some of the participants stated:

Quite engaging, particularly when a question came up that I felt I had already answered, as I would have to think about the topic a bit more in order to add something to my previous answer. [P13, age range: 26-33, male]

At times I was confused as to how deeply the robot wanted me to answer the questions given, and so tended [I think] to delve too deeply as I was asked a few times to repeat what I had just said in another question. Did have one occasion where the automated voice sounded funny and wasn't sure exactly what it had said! [P2, age range: 18-25, female]

The content was clear, each question was clearly spoken...I found it frustrating that a question I may have already answered could be asked. [P5, age range: 18-25, male]

The content was clear. Although I felt should have been more specific to the question. Questions sounded like they could relate to another subject generally so they were too generic and therefore less personal. [P6, age range: 26-33, female]

Interface (1.3)

The instructions were clear and the navigation easy to use, and participants generally found it uncomplicated to touch the robot's head sensor to advance to the next questions. However, some felt that this spoiled the illusion of a natural dialogue:

I felt a little concerned I might press something I should not and muddle up the process but it was fine. [P3, age range: above 61, female]

Once I had stopped giggling at the eye colour change, everything was straightforward. Tapping on the head of the robot for the next question was simple. [P15, age range: 43-60, female]

[The worst aspect of the interview was] not having the immediate response, having to push a button on his head made it feel fake. [P4, age range: 26-33, female]

...It was extremely life like but having to tap it on the head to confirm you had completed your answer broke the rapport slightly. [The interface was] easy to use, it spoke clearly. It was good how his head followed your movement. [P6, age range: 26-33, female]

Listening to Oneself (1.4)

Most of the participants found listening to themselves important. It helped them appreciate the importance of their goals and face the reality of their current behavior and plans for change. Some did not feel comfortable in speaking out loud and found the situation awkward. Some of the participants stated:

[Listening to myself was] very important. It's easy to rationalise unhealthy behaviour in your head but the second you realise how stupid you sound rationalising or how reasonable your reasons are for wanting to do it, your attitude changes. [P2, age range: 18-25, female]

[Listening to myself was] very important. Makes the thoughts hold more weight and actually think about them more than if they are simply passing thoughts. [P9, age range: 18-25, female]

I regularly discuss behaviour with a team mate, so it is something that I consider is generally important. Usually when we discuss our behaviour, we critique errors and try to improve by correcting them. However, the robot also made me talk about times where my behaviour had been positive and this is something I think is very important. [P13, age range: 26-33, male]

Actually, [listening to myself was] really rather important, as I could hear myself suggesting things, then getting a bit doubtful, then more confident as time went on. Hearing myself talking out loud made me feel as if I was chatting to myself and truly sorting

out issues—without anyone else poking their nose in. [P15, age range: 43-60, female]

I actually feel writing it down allows you to express yourself more honestly without the fear of sounding silly. [P6, age range: 26-33, female]

Theme 2: Motivation

Participants' spontaneous strategies (2.1) for supporting motivation included commitment or doing activities with friends or family; flexibility, routine, or planning; focusing on the goals; visualization techniques, mindfulness, or will power; and motivational books:

Using notifications on my phone to remind me when I have to do whatever it is I have to do [helps me staying motivated]. Also using diaries or planners to tick off when I've done it. [P2, age range: 18-25, female]

[What helps me the most in staying motivated] is being mindful of the situation. [P3, age range: above 61, female]

...visualising my goals and setting out steps I can achieve in the short term in order to achieve the long term goal. [P5, age range: 18-25, male]

...Friends making supportive comments. [P15, age range: 43-60, female]

Participants wrote about challenges (2.2) that make it hard to keep themselves motivated, including health problems, bad weather, winter, laziness or being tired, and social distractions:

[The hardest part in keeping myself motivated is] being distracted by something I shouldn't be (like playing a video game for too long or watching another episode of something on Netflix). [P2, age range: 18-25, female]

...tiredness, lack of time due to work. Winter when the days feel shorter. [P6, age range: 26-33, female]

...erratic work shift patterns. Also, not seeing results within a certain timeframe can be demotivating. [P13, age range: 26-33, male]

Theme 3: Engagement in Physical Activity After the Program

There was mixed success in terms of whether participants achieved their goal for the week after the robot interview. Some felt disappointed that they had not done so:

I didn't improve my physical activity. It has been more or less the same as the past weeks. [P12, age range: 18-25, female]

I did go for a run with a friend, as I said I would in the interview, however, this only happened once and so I feel it had not worked as well as maybe I had hoped. [P18, age range: 18-25, female]

Others achieved their goals and occasionally expressed surprise in the way they communicated their success:

I actually carried out my plan... [P9, age range: 18-25, female]

I stretched 3 out of 7 days and practised burlesque on 1, which is way more than I'd done regularly before. [P2, age range: 18-25, female]

I completed at least 20 minutes of additional physical activity every day. [P4, age range: 26-33, female]

Theme 4: Overall Evaluation

Participants' positive appraisal of the intervention focused strongly on their perception that the robot was not judging them, whereas a human might have done. They liked being able to talk without being interrupted and appreciated how the interview gave them space to think about things and voice their goals. One participant described this as *a kind of liberation* [P12, age range: 18-25, female]. Some of the participants stated the following:

[The best aspect of this robotic interview was] being able to talk freely and for as long as I wanted about every aspect of physical activity that concerned me without being judged. [P2, age range: 18-25, female]

...the time to talk without being interrupted. [P4, age range: 26-33, female]

...he didn't interrupt and was not judgemental...I felt more motivated because I talked through my goals without interruption or other people's advice. [P10, age range: 34-42, female]

The robot interview allowed me to reflect on my behaviour in a guided manner. It also encouraged me to focus on positive behaviour from the past and specific changes that I need to make for the future. I felt that this was the best aspect. When reflecting on my behaviour alone, there is a tendency to dwell on things done wrong and this does not always provide a solution. The robot demonstrated that I can reflect on my behaviour without focusing on negative aspects. [P13, age range: 26-33, male]

The best part was the whole idea that I was able to interact with a robot. I think it feels nice to talk and not feel embarrassed by potential judgement. [P18, age range: 18-25, female]

I felt I can talk freely without any judgement which was quite nice. Talking to "a human" is quite daunting as we naturally judge things and people especially people's behaviour. [P19, age range: 18-25, female]

The novelty of the robot was a positive feature for some. One participant explained how the fact that the interview was fun and memorable led her to share her goals with others. The robot may thus have contributed to that participant gaining "support from others," which she cited as something that helps her stay motivated. Participants stated:

The use of a robot made it fun and less pressured which stayed in my mind longer...it played on my mind during the past week and I told others about the robot which made them ask about the goals set during the interview. [P1, age range: 18-25, female]

It was engaging, different and fun...the fact I have thought about it over the past week has been motivational. [P3, age range: above 61, female]

Participants offered insights into how the robotic interview could be improved. Common themes were the problem of not being able to replay a question that had not been understood, needing some time to get used to the robot, and wanting a more natural way of progressing to the next question:

I feel that the interview could be improved by having more off topic questions to begin with allowing the person to get used to the robot. [P5, age range: 18-25, male]

[The robotic interview could be improved by] not having the robot to close, although that is essential to a certain extent, just felt awkward sitting so close—maybe it could be placed more to the side? [P2, age range: 18-25, female]

Having to repeatedly touch the head for the next question was a little off-putting. [P9, age range: 18-25, female]

Maybe it would be useful to have the robot repeating the question. [P12, age range: 18-25, female]

[The worst aspect of this robotic interview was] having no feedback on my responses so I didn't know if I was answering the question correctly. [P14, age range: 18-25, female]

Perhaps a clearer voice. Sometimes, I felt I felt that I might have misunderstood a question due to not understanding the robot as well as I had wanted. [P18, age range: 18-25, female]

Discussion

Principal Findings

We developed a technology-delivered adaptation of MI using a humanoid robot. When MI is translated into technology as a medium, this person-centered counseling technique inevitably loses its full capacity; however, we have developed a script with strong elements of MI, including evocation, promoting self-efficacy, strengthening commitment to change, and asking open questions. Key findings from participants' evaluation of the intervention were that they found it motivating to hear themselves discussing their behavior with the robot; they enjoyed the interaction and found the robot easy to use but wanted longer to get used to it; and they liked the neutrality of the robot. The main drawback was that the robot could not tailor its questions according to the answers already given.

Previous research with technological delivery of MI has typically used text-based responses, for example, Gerbert's work with the Video Doctor [25-27]. In a more ambitious project than ours, Kanaoka and Mutlu [28] used a NAO robot to deliver a motivational interview with personalized responses to the participant's speech. In contrast to their predictions, participants were less motivated after the MI dialogue than after a monologue in which the robot gave advice. Kanaoka and Mutlu attributed this finding to inadequacy of the speech recognition software. They noted that the robot sometimes interrupted

participants and that, when the robot "misheard" them, participants spoke to the microphone rather than to the robot, suggesting a breakdown in the fluency of the interaction. We tried to avoid these problems by using the robot to deliver a series of open questions and requiring the participant to press the robot's head sensor when they had finished talking and were ready to advance to the next question. Participants evaluated this aspect of the interaction positively and negatively. They liked the space to talk freely about themselves, without interruption, and reported that the robot's questions prompted them to think deeply and realistically about their goals and obstacles to achieving them. However, pressing the head sensor broke the flow of the conversation for some. The lack of personalization was frustrating, particularly when the robot asked a question that participants felt they had already answered. We had tried to preempt this problem by having the robot warn participants at the start that it might repeat a question. The interview deliberately asked several questions on one topic before moving to the next, to encourage the participant to think deeply about the issue and why it matters to them. This outcome would normally be achieved in MI through the interviewer reflecting the meaning of the participant's answer back to them. Although participants typically disliked the repetition, one participant found that it helped him feel engaged in the dialogue by encouraging him to add more information to his previous answer.

An important aim of MI is to elicit change talk, where the individual articulates their desire or need to change. The extent to which a motivational interview elicits such talk is positively associated with outcomes [29]. In this study, participants found it motivating to hear themselves argue aloud for change, reporting that it helped them consolidate and take ownership of their plans. Many but not all participants felt that the interview had a positive effect on their behavior in the week that followed.

As most people do not have access to humanoid robots, the interaction with a NAO robot acting as a counselor was a unique experience. Due to the singularity of the situation, participants remembered the interaction and talked about it with other people, reiterating their commitment to change and making a social contract [36,37]. Further research is needed to test whether the effect of the interview fades away once the novelty wears off.

Although participants criticized the interaction for not being as fluent as a conversation with a human interviewer, some benefits of the robot interviewer featured very strongly in participants' evaluations. They felt unhurried because the robot did not interrupt them, and many felt more comfortable discussing issues with the robot than with a human counselor because it would not judge them. A central tenet of MI is that interactions should be collaborative and not judgmental. These findings are an important reminder that, however skilled the interviewer, participants bring their own assumptions and anxieties to the interview, including a fear that the interviewer will judge them. In line with these findings, there is evidence that people will more willingly reveal sensitive information to computers than to humans [27,38,39]. In the drive to develop increasingly naturalistic computer-human interactions, developers must keep sight of the advantages of being perceived as a robot.

Participants spontaneously used a range of strategies to motivate themselves, including setting reminders, engaging peer support, having a routine, and visualizing their goals. As challenges to achieving their goals, they cited competing distractions, tiredness, and lack of time. There is scope for developing the robot interaction further to encourage successful behavior change strategies and reduce counterproductive behaviors. Previous work has shown that people can develop social relationships with robots [34]. Future research could explore the value of the robot for providing social support, which is known to facilitate behavior change, at challenging moments such as those mentioned by participants. This social support could include the ideas suggested by participants, such as reminding them of their plan, providing encouragement, or using imagery to strengthen motivation, for example, by guiding visualization of the goal and how good it will feel to succeed [40].

Participants wanted some time to get used to the robot before starting the interview. Providing a longer introduction before beginning the motivational interview could help address some of the drawbacks identified by participants, including discomfort at being close to the robot and having to touch it, and difficulty understanding its speech.

Limitations and Directions for Future Research

Limitations of this approach include the impossibility of using all MI skills in a prescripted interview. Without more sophisticated speech recognition and branching logic, the robot is unable to reflect the participant's meaning, affirm their choices and autonomy, or summarize what they have said (although we included suggestions, by the robot, that the participant summarize their plan). Even with perfect speech recognition, sophisticated MI skills, such as rolling with resistance and identifying change talk and sustain talk, present a considerable challenge. However, even without these skills, there is evidence that technological adaptations of MI can be beneficial [24]. There is potentially an issue of safety in terms of how a robot might respond to a participant who proposes a dangerous course of action. However, a skilled MI practitioner would elicit the participant's appraisal of their plan, rather than directly advising against it, and this approach could be reproduced in the robot, as we did in this intervention through asking questions that probed an issue deeply before moving to the next. One solution to the problem of participants not knowing a safe solution to their dilemma could be for the robot to ask permission to offer advice or information. The NAO robot's head, hands, and feet sensors also provide opportunities to follow different paths through the prescripted interview—for example, participants could choose information about diet by pressing a hand sensor or about exercise by pressing a foot sensor. Adding limited choice in this way may help to focus the interview on issues that matter most to participants and provide an experience that feels more personal.

The study has several limitations. It focused on participants' experiences of the interaction and impressions of its impact on their behavior. Participants responded to an advertisement for volunteers who wanted to increase their PA, but we did not assess their motivation, baseline activity levels, or changes in

behavior after the interview. Further research should test the robot interview with different populations, including those who wish to start being physically active and those who wish to increase their activity, and measure their pre- and postintervention motivation and behavior. Another limitation is that the people who volunteered for this study were members of a research panel and fairly used to strange experiences in psychology laboratories. They may have been more accepting of the robot, despite meeting it for the first time, than other members of the general population. Having established the acceptability of the intervention using qualitative methods, an important next step is to test its efficacy for changing behavior in a broader sample. Randomized controlled trials are needed to assess quantitative changes in motivation and PA associated with the robot intervention and compare them, in the first instance, against simple information and advice. To maximize the potential for observing benefits over meaningful timescales, we suggest that a series of interactions be designed to incorporate reminders and follow-up sessions so that the robot provides ongoing support for behavior change.

Much could be learned from observing the participant-robot interaction, but interviews were not recorded because participants who helped with the initial development thought they would feel uncomfortable being watched or filmed while talking to the robot. Having completed this study, we are less concerned that this would be an issue. A familiarization phase, with some general conversation between the participant and robot before starting the motivational interview, could help reduce the strangeness of the experience. In a related study evaluating a motivational interview delivered by a human video counselor, analysis of participants' speech showed that the interview successfully elicited change talk (unpublished data, 2017 [41]). Combining analyses of change and sustain talk with quantitative data on behavior change could reveal whether a robot-led motivational interview affected motivation and behavior via the same mechanisms as human-led MI.

A robot-delivered motivational interview may lack elements of an interview with a human counselor, but our findings suggest it could have wide application. Because participants enjoyed the interaction and liked the novelty, a robot-delivered interview may help engage people to discuss sensitive issues and to get a feel for what counseling would be like, encouraging self-help or help-seeking earlier in the time course of a problem. A robot interview could be designed that encouraged people who are not yet contemplating change to consider its pros and cons. The novelty of interacting with a robot could encourage people to engage who might not feel ready to talk to a human counselor. Given that our adult participants were concerned about being judged by another adult, the robot could be particularly important for encouraging children and adolescents to discuss mental health issues, as they may be more susceptible to fears of being judged or misunderstood by an adult. As well as fostering engagement with health care, a robot interviewer could also provide motivational aftercare, ensuring that benefits from a human-led intervention are sustained when the intervention ends. The generic nature of the interview means it can easily be modified for a wide variety of target behaviors, potentially providing motivational support for the very large number of

people who struggle with conditions such as addiction or obesity but do not meet the criteria for accessing professional support.

Conclusions

We have shown, for the first time, that a motivational interview delivered by a social robot can elicit out-loud discussion from participants in an interaction that they perceive as enjoyable, interesting, and helpful. Participants especially found it useful to hear themselves talking about their behavior aloud, giving this new intervention a potential advantage over other technology-delivered adaptations of MI. Concern about being judged by a human interviewer came across strongly in praise for the nonjudgmental nature of the robot, suggesting that robots may be particularly helpful for eliciting talk about sensitive issues.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The motivational interviewing script delivered by the NAO robot.

[PDF File (Adobe PDF File), 17KB - jmir_v20i5e116_app1.pdf]

Multimedia Appendix 2

The questionnaire used to explore participants' experiences of the interview.

[PDF File (Adobe PDF File), 13KB - jmir_v20i5e116_app2.pdf]

Multimedia Appendix 3

The coding scheme developed through the thematic content analysis.

[PDF File (Adobe PDF File), 29KB - jmir_v20i5e116_app3.pdf]

References

1. NHS Digital. 2008. Health Survey for England - 2008: Physical activity and fitness URL: http://content.digital.nhs.uk/catalogue/PUB00430/heal-surv-phys-acti-fitn-eng-2008-rep-v2.pdf [accessed 2017-12-17] [WebCite Cache ID 6mGAY2E1j]
2. Hardcastle S, Hagger MS. "You Can't Do It on Your Own": experiences of a motivational interviewing intervention on physical activity and dietary behaviour. Psychol Sport Exerc 2011 Jun;12(3):314-323. [doi: 10.1016/j.psychsport.2011.01.001]
3. Solbrig L, Jones R, Kavanagh D, May J, Parkin T, Andrade J. People trying to lose weight dislike calorie counting apps and want motivational support to help them achieve their goals. Internet Interv 2017 Mar;7:23-31 [FREE Full text] [doi: 10.1016/j.invent.2016.12.003] [Medline: 28286739]
4. European Commission. 2012. eHealth Action Plan 2012-2020: Innovative healthcare for the 21st century URL: https://ec.europa.eu/digital-single-market/news/ehealth-action-plan-2012-2020-innovative-healthcare-21st-century [accessed 2017-01-18] [WebCite Cache ID 6nbj18LH2]
5. Miller WR, Rollnick S. Motivational Interviewing: Helping People Change. New York, NY: Guilford Press; 2012.
6. Rubak S, Sandbaek A, Lauritzen T, Christensen B. Motivational interviewing: a systematic review and meta-analysis. Br J Gen Pract 2005 Apr;55(513):305-312 [FREE Full text] [Medline: 15826439]
7. Lundahl BW, Kunz C, Brownell C, Tollefson D, Burke BL. A meta-analysis of motivational interviewing: twenty-five years of empirical studies. Res Soc Work Pract 2010 Jan 11;20(2):137-160. [doi: 10.1177/1049731509347850]
8. Bennett JA, Lyons KS, Winters-Stone K, Nail LM, Scherer J. Motivational interviewing to increase physical activity in long-term cancer survivors: a randomized controlled trial. Nurs Res 2007;56(1):18-27. [doi: 10.1097/00006199-200701000-00003]
9. Hardcastle S, Taylor A, Bailey M, Castle R. A randomised controlled trial on the effectiveness of a primary health care based counselling intervention on physical activity, diet and CHD risk factors. Patient Educ Couns 2008 Jan;70(1):31-39. [doi: 10.1016/j.pec.2007.09.014] [Medline: 17997263]
10. Miller WR, Benefield RG, Tonigan JS. Enhancing motivation for change in problem drinking: a controlled comparison of two therapist styles. J Consult Clin Psychol 1993 Jun;61(3):455-461. [Medline: 8326047]

XSL•FO
RenderX

11. Patterson GR, Forgatch MS. Therapist behavior as a determinant for client noncompliance: a paradox for the behavior modifier. J Consult Clin Psychol 1985 Dec;53(6):846-851. [Medline: 4086685]

12. Bem DJ. Self-perception theory. In: Berkowitz L, editor. Advances in Experimental Social Psychology. New York: Academic Press; 1972:1-62.

13. Goeldner M, Herstatt C, Tietze F. The emergence of care robotics–a patient and publication analysis. Technol Forecast Soc Change 2015 Mar;92:115-131. [doi: 10.1016/j.techfore.2014.09.005]

14. Kanda T, Ishiguro H. Human-Robot Interaction in Social Robotics. Boca Raton: CRC Press; 2013.

15. Benitti FB. Exploring the educational potential of robotics in schools: a systematic review. Comput Educ 2012 Apr;58(3):978-988. [doi: 10.1016/j.compedu.2011.10.006]

16. Broekens J, Heerink M, Rosendal H. Assistive social robots in elderly care: a review. Gerontechnology 2009;8(2):94-103 [FREE Full text]

17. Belpaeme T, Baxter P, Read R, Wood R, Cuayáhuitl H, Kiefer B, et al. Multimodal child-robot interaction: Building social bonds. J Hum Robot Interact 2012;1(2):33-53. [doi: 10.5898/jhri.1.2.belpaeme]

18. Scassellati B, Admoni H, Matarić M. Robots for use in autism research. Annu Rev Biomed Eng 2012 Aug;14:275-294. [doi: 10.1146/annurev-bioeng-071811-150036] [Medline: 22577778]

19. Thill S, Pop C, Belpaeme T, Ziemke T, Vanderborght B. Robot-assisted therapy for autism spectrum disorders with (partially) autonomous control: challenges and outlook. Paladyn 2013;3(4):209-217. [doi: 10.2478/s13230-013-0107-7]

20. Schneider S, Goerlich M, Kummert F. A framework for designing socially assistive robot interactions. Cogn Syst Res 2017;43:301-312. [doi: 10.1016/j.cogsys.2016.09.008]

21. Fasola J, Matarić MJ. Using socially assistive human-robot interaction to motivate physical exercise for older adults. In: Xu Y, Qian H, Wu X, editors. Household Service Robotics. Cambridge, MA: Academic Press; 2015:487-516.

22. Kidd CD, Breazeal C. Robots at home: understanding long-term human-robot interaction. 2008 Presented at: IEEE/RSJ International Conference on Intelligent Robots and Systems; September 22-26, 2008; Nice, France. [doi: 10.1109/iros.2008.4651113]

23. Cantrell R, Schermerhorn P, Scheutz M. Learning actions from human-robot dialogues. 2011 Presented at: RO-MAN, IEEE; July 31-Aug 3, 2011; Atlanta, GA. [doi: 10.1109/ROMAN.2011.6005199]

24. Shingleton RM, Palfai TP. Technology-delivered adaptations of motivational interviewing for health-related behaviors: a systematic review of the current research. Patient Educ Couns 2016 Jan;99(1):17-35 [FREE Full text] [doi: 10.1016/j.pec.2015.08.005] [Medline: 26298219]

25. Jackson RA, Stotland NE, Caughey AB, Gerbert B. Improving diet and exercise in pregnancy with Video Doctor counseling: a randomized trial. Patient Educ Couns 2011 May;83(2):203-209. [doi: 10.1016/j.pec.2010.05.019] [Medline: 21459255]

26. Tsoh JY, Kohn MA, Gerbert B. Promoting smoking cessation in pregnancy with Video Doctor plus provider cueing: a randomized trial. Acta Obstet Gynecol Scand 2010;89(4):515-523 [FREE Full text] [doi: 10.3109/00016341003678419] [Medline: 20196678]

27. Humphreys J, Tsoh JY, Kohn MA, Gerbert B. Increasing discussions of intimate partner violence in prenatal care using Video Doctor plus Provider Cueing: a randomized, controlled trial. Womens Health Issues 2011;21(2):136-144 [FREE Full text] [doi: 10.1016/j.whi.2010.09.006] [Medline: 21185737]

28. Kanaoka T, Mutlu B. Designing a motivational agent for behavior change in physical activity. 2015 Presented at: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems; April 18-23, 2011; Seoul, Korea. [doi: 10.1145/2702613.2732924]

29. Apodaca TR, Longabaugh R. Mechanisms of change in motivational interviewing: a review and preliminary evaluation of the evidence. Addiction 2009 May;104(5):705-715 [FREE Full text] [doi: 10.1111/j.1360-0443.2009.02527.x] [Medline: 19413785]

30. Kavanagh D, Graves N, Connor J, Hides L, Gullo M, Andrade J, et al. 2016. A new, low-cost e-health treatment for Alcohol Use Disorder using mental imagery URL: https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=370132 [accessed 2018-03-11] [WebCite Cache ID 6xpdDbkgF]

31. Solbrig LM, Andrade J. ISRCTN. 2016. Personalized weight loss trial comparing motivational interventions for losing weight and being more active URL: https://tinyurl.com/ybw24b8j

32. Miller WR, Rollnick S. Motivational Interviewing: Preparing People for Change. New York, NY: Guilford Press; 2002.

33. Moyers TB, Manuel JK, Ernst D. Motivational Interviewing Treatment Integrity Coding Manual 4.1 URL: https://casaa.unm.edu/download/MITI4_2.pdf [accessed 2018-02-27] [WebCite Cache ID 6xYP8cBam]

34. Baxter P, Ashurst E, Read R, Kennedy J, Belpaeme T. Robot education peers in a situated primary school study: personalisation promotes child learning. PLoS ONE 2017;12(5):e0178126 [FREE Full text] [doi: 10.1371/journal.pone.0178126]

35. Boyatzis RE. Transforming Qualitative Information: Thematic Analysis and Code Development. Thousand Oaks, CA: Sage Publications; 1998.

36. Cavedon L, Sonenberg L. On social commitment, roles and preferred goals. 1998 Presented at: Proceedings of International Conference on Multi Agent Systems, IEEE; July 3-7, 1998; Paris, France. [doi: 10.1109/icmas.1998.699035]

37.   Cudd AE. Commitment as motivation: Amartya Sen's theory of agency and the explanation of behavior. Econ Philos 2014
        Mar 4;30(1):35-56. [doi: 10.1017/s0266267114000030]
38.   Gerbert B, Berg-Smith S, Mancuso M, Caspers N, McPhee S, Null D, et al. Using innovative video doctor technology in
        primary care to deliver brief smoking and alcohol intervention. Health Promot Pract 2003 Jul;4(3):249-261. [doi:
        10.1177/1524839903004003009] [Medline: 14610995]
39.   Pickard M, Roster C, Chen Y. Revealing sensitive information in personal interviews: is self-disclosure easier with humans
        or avatars and under what conditions? Comput Human Behav 2016 Dec;65:23-30. [doi: 10.1016/j.chb.2016.08.004]
40.   Andrade J, Khalil M, Dickson J, May J, Kavanagh DJ. Functional imagery training to reduce snacking: testing a novel
        motivational intervention based on elaborated intrusion theory. Appetite 2016 May 01;100:256-262. [doi:
        10.1016/j.appet.2016.02.015] [Medline: 26877217]
41.   Galvão Gomes da Silva J, Kavanagh DJ, May J, Andrade J. Say it aloud: change talk during a spoken, automated motivational
        interview delivered by video counsellor, and participants' perceptions after one week. Manuscript submitted for publication
        2017.

Abbreviations

IRR: interrater reliability
MI: motivational interviewing
PA: physical activity

XSL•FO
RenderX

Review

Factors Determining the Success and Failure of eHealth Interventions: Systematic Review of the Literature

Conceição Granja[1], PhD; Wouter Janssen[2], MSc; Monika Alise Johansen[1,2], PhD

[1]Future Journal, Norwegian Centre for E-health Research, Tromsø, Norway

[2]Telemedicine and E-health Research Group, University of Tromsø-The Artic University of Norway, Tromsø, Norway

Corresponding Author:
Conceição Granja, PhD
Future Journal
Norwegian Centre for E-health Research
PO Box 35
Tromsø, N-9038 Tromsø
Norway
Phone: 47 90817344
Email: conceicao.granja@ehealthresearch.no

Abstract

Background: eHealth has an enormous potential to improve healthcare cost, effectiveness, and quality of care. However, there seems to be a gap between the foreseen benefits of research and clinical reality.

Objective: Our objective was to systematically review the factors influencing the outcome of eHealth interventions in terms of success and failure.

Methods: We searched the PubMed database for original peer-reviewed studies on implemented eHealth tools that reported on the factors for the success or failure, or both, of the intervention. We conducted the systematic review by following the patient, intervention, comparison, and outcome framework, with 2 of the authors independently reviewing the abstract and full text of the articles. We collected data using standardized forms that reflected the categorization model used in the qualitative analysis of the outcomes reported in the included articles.

Results: Among the 903 identified articles, a total of 221 studies complied with the inclusion criteria. The studies were heterogeneous by country, type of eHealth intervention, method of implementation, and reporting perspectives. The article frequency analysis did not show a significant discrepancy between the number of reports on failure (392/844, 46.5%) and on success (452/844, 53.6%). The qualitative analysis identified 27 categories that represented the factors for success or failure of eHealth interventions. A quantitative analysis of the results revealed the category quality of healthcare (n=55) as the most mentioned as contributing to the success of eHealth interventions, and the category costs (n=42) as the most mentioned as contributing to failure. For the category with the highest unique article frequency, workflow (n=51), we conducted a full-text review. The analysis of the 23 articles that met the inclusion criteria identified 6 barriers related to workflow: workload (n=12), role definition (n=7), undermining of face-to-face communication (n=6), workflow disruption (n=6), alignment with clinical processes (n=2), and staff turnover (n=1).

Conclusions: The reviewed literature suggested that, to increase the likelihood of success of eHealth interventions, future research must ensure a positive impact in the quality of care, with particular attention given to improved diagnosis, clinical management, and patient-centered care. There is a critical need to perform in-depth studies of the workflow(s) that the intervention will support and to perceive the clinical processes involved.

KEYWORDS

Introduction

In the last decades, it has been a challenge for policy makers to ensure access to healthcare to populations living in rural and remote areas [1]. Additionally, global demographic trends, such as the increasing number of elderly people, have been changing healthcare delivery due to a growing demand for long-term care and increasing costs [2-5]. Against this background, eHealth has been presented as a solution [6-8]. In the context of this study, we define eHealth as the use of information and communication technology in healthcare.

A vast amount of eHealth interventions have been reported to fail during clinical implementation [9,10]. Tanriverdi and Iacono [8] demonstrated that a considerable amount of research with promising results did not contribute to clinical practice. Berg [11] stated more specifically that 75% of implemented eHealth should be considered a failure.

According to Bashshur et al [12], the assessment of eHealth interventions rests on three pillars of care: (1) access, (2) quality, and (3) cost containment. They describe these three pillars as the promises that eHealth interventions are required to fulfill to attain a successful outcome and, indeed, that each of these promises must be met. Considering the aforementioned reports on the failure of eHealth interventions [8-11], it appears reasonable to assume that the promises represented by these three pillars are not often accomplished.

To improve the success of eHealth, it is important to identify the factors that can influence, positively or negatively, the outcome of the intervention. Such factors can vary from project-specific to recurring issues, with the three pillars proposed by Bashshur et al [12] expected to have an important role in the success or failure of eHealth interventions. However, as the field of medical informatics is positioned between the fast-changing field of informatics and the rather conservative field of healthcare, organizational and operational aspects can be expected to play an important part in the outcome of eHealth interventions.

The overall aim of this study was to seek, through a systematic review, patterns in the assessment of eHealth intervention outcomes, and through these patterns to identify factors that can help explain why eHealth interventions fail or succeed in clinical practice. Therefore, we systematically searched for original studies that provide data to address the following key questions:

> Key question 1: According to reports in abstracts, why are eHealth interventions failing to achieve the expected results and foreseen benefits? Specifically, (1) What are the major facilitators and barriers contributing to the implementation of eHealth? (2) How are these facilitators and barriers contributing to the adoption of eHealth? (3) Are the perceived facilitators and barriers to eHealth adoption similar among the study participants?

> Key question 2: According to the literature, what is the most relevant factor regarding the possible outcome in terms of success or failure, or both? Specifically, in what manner is this factor affecting the adoption of eHealth?

Methods

This systematic review was guided by the *Cochrane Handbook* [13], and the reporting is based on the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines [14]. We established the review methods before conducting the review, and the reports did not justify any deviations from the protocol.

Search Strategy

We searched the PubMed database in October 2016 for original articles published in English up to this date. The main reason for using only the PubMed search engine was the availability of a vast amount of articles in eHealth research from both a medical and a sociological standpoint. In this way, we expected that we would find most of the relevant clinical outcomes for this study. Since the review focused on implemented eHealth tools, we did not consider the inclusion of articles from a technology perspective to be necessary. As this strategy proved to be useful, we searched no other academic databases.

We performed an initial search using the Medical Subject Heading (MeSH) telemedicine AND challenges, based on prior knowledge obtained from articles that referenced eHealth success or failure, which identified 658 articles. We evaluated the title and abstract of these articles and identified possible search terms. We determined the term "lessons" to be important and used it together with the previous search, identifying 63,299 articles. We analyzed the resulting articles to further specify the search terms and define the search string following the patient, intervention, comparison, and outcome (PICO) framework [15].

We based the search string on three classes of the PICO framework, since the class comparison did not apply to this study. We defined these classes as (P): healthcare; (I): eHealth; and (O): change, failure, or success. We combined all terms in each class with the logical operator OR and linked the classes using the logical operator AND. Over the course of refining the search results, we tried 2 search strings. In the first, we used the search terms extracted from the previous search to create a PICO scheme, constrained to the last 10 years, which identified 11,950 articles. In the second, we removed the time constraint and refined the search terms to focus on eHealth interventions that were used in actual clinical practice. This last search string (Textbox 1) was the one we used to retrieve the articles used in the review process.

Textbox 1. The search string developed according to the patient, intervention, comparisona, and outcome (PICO) framework. Comparison (C) was not applicable to this study.

Healthcare (P)

((clinical practice[Title/Abstract] OR real use[Title/Abstract] OR real practice[Title/Abstract] OR clinical implication OR health care effect[Title/Abstract] OR health care impact[Title/Abstract] OR practical trials[Title/Abstract] OR clinical trials[Title/Abstract] OR practical clinical implementation[Title/Abstract] OR practical clinical trials[Title/Abstract] OR implemented service[Title/Abstract] OR adoption[Title/Abstract] OR adoption rate[Title/Abstract]))

eHealth (I)

(((("telemedicine"[MeSH] OR medical informatics[TIAB] OR eHealth[Title/Abstract] OR ehealth[Title/Abstract] OR telemedicine[Title/Abstract] OR telehealth[Title/Abstract] OR mhealth[Title/Abstract] OR mhealth[Title/Abstract] OR health telematics[Title/Abstract] OR tele-health[Title/Abstract] OR etherapy[Title/Abstract] OR wireless health[Title/Abstract] OR healthcare technology[Title/Abstract] OR telecare[Title/Abstract] OR medical information system[Title/Abstract] OR telemonitoring[Title/Abstract] OR telepresence[Title/Abstract] OR electronic health information[Title/Abstract] OR teleconsultation[Title/Abstract] OR teleintervention[Title/Abstract] OR e-rehabilitation[Title/Abstract])))

Change, failure, or success (O)

((fail*[Title/Abstract] OR succes*[Title/Abstract] OR barrier*[Title/Abstract] OR interoperability[Title/Abstract] OR usability[TIAB] OR lessons learned[Title/Abstract] OR implications[Title/Abstract] OR experiences[Title/Abstract] OR implementation[Title/Abstract])))

Textbox 2. Reading interpretation guidelines for the abstract review.

- Funding from government or tax money equals influence on society.

- Policies have an effect on both organizations and availability of tools for patients.

- Coordination and interoperability problems have consequences for patients, professionals, and systems.

- Extra (or changes in) work is seen as workflow.

- Safety is a relative term, interpreted as compared with traditional ways.

- Workforce problems are interpreted as change of workflow.

- (Un)familiarity with tools is seen as information technology training.

- Paternalism and empowerment is seen as empowerment or engagement.

- Medical (studies) students are seen as health professionals.

- Time is seen as either workflow or costs, depending on the context.

We included no articles based on hand searches of reference lists for the reasons outlined under Section 10.2.2.3 of the *Cochrane Handbook* [13]: "positive studies are more likely to be cited" and "retrieving literature by scanning reference lists may thus produce a biased sample of studies."

Study Selection

We analyzed the titles and abstracts of the articles that resulted from the final search string for inclusion according to the following predefined exclusion criteria: not an original work; unclear or no results; not research; not in English. We defined the exclusion criteria based on the key questions to include original studies in healthcare-related fields, with a focus on success and failure, that reported on a form of eHealth or medical informatics, but which did not have to be the main goal or result but should have been a key component. Textbox 2 lists the reading interpretation guidelines for the abstract review.

Data Collection and Synthesis

A qualitative analysis [16,17] was carried out by 2 of the authors (CG and WJ) to classify the outcomes reported in the articles' abstract according to the following 3 levels.

Category

The category level was evidence that the factor described in the abstract contributed to the success or failure of the eHealth intervention. We defined categories based on the information found in the abstracts (presented in the Results section below). We chose this strategy to minimize the risk of bias, since predetermined categories could have led to a model that merely reflected our opinion.

Success and Failure

This level indicated whether the identified category was described as a success or failure factor reported in the intervention outcome narrative. As success and failure are important concepts of this study, we explain our considerations on the terms here.

We classified the factors in the categories as *success* if they were considered to facilitate the achievement of the study goals. The same category may have been described as success and failure in the same study by different participants. If specific features of an intervention were mentioned to be a success, we attributed these to success, even when the overall project was classified as failure.

XSL•FO
RenderX

Textbox 3. Further exclusion criteria of articles.

- The category is mentioned only in the text.
- No extensive analysis of the category or reporting on the effects on eHealth outcomes is present.
- The category is only identified as relevant to the success or failure of eHealth, and no further considerations are taken.
- The article is not in English.

We classified the factors in the categories as *failure* if they were considered to be barriers to achieving the study goals. Different participants may have described the same category as success and failure in the same study. If specific features of an intervention were mentioned to be a failure factor, we attributed these to failure, even when the overall project was classified as success.

Entity

This level referred to the role of the study participant who reported the identified factor in the categories. The entities could assume the following values.

In this review, *patients* were people who received care. Therefore, we also included clients (ie, people with less-urgent problems) and customers (ie, people who were interested in monitoring their own health). This entity also included people who gave care to patients in a nonprofessional context (ie, parents, family, and friends).

Healthcare professionals comprised all people who provided care services in a professional context. This included physicians, nurses, therapists, mental health workers, and other professional groups trained in providing care. It did not necessarily have to be direct care, but they had to have been providing care to patients.

The *health system* included management and supporting staff, infrastructure, the technological health systems (both software and hardware), and ideological systems such as national health plans and systems.

The *society* value included participants who were described as potential users and were not identified as belonging to the entities described above.

All involved all the above entities, and we classified the category identified in the reports as success and failure.

Full-Text Analysis

In addition to categorizing abstracts for success and failure factors of eHealth interventions, we analyzed the full text for the category with the highest frequency of unique articles according to the categorization results. Such analysis was aimed at gathering the data that provide knowledge related to key question 2.

Articles in which the category with the highest unique article frequency was reported to contribute to the success or failure of the eHealth intervention were eligible for a full-text review. However, in the course of the full-text assessment, we refined

the selection according to the exclusion criteria presented in Textbox 3.

In summary, abstract and full-text reviews were conducted independently by 2 authors (CG and WJ), who extracted data based on the inclusion and exclusion criteria into a structured a Microsoft Excel spreadsheet. The 2 authors analyzed all abstracts a second time to confirm the categorization, identified all relevant factors reported to contribute to the success or failure of eHealth interventions, and noted them in the categorization model. In this manner, every time a new factor was identified, a new category was created. The full-text review was conducted with special attention to the descriptions of how the category with the highest unique article frequency was affecting the success of the eHealth intervention. All disagreements were resolved by consensus discussions.

Results

The search string identified 903 articles, 7 of which we excluded, as they were duplicates or incomplete. The titles and abstracts of the remaining 896 articles were read by 2 of the authors (CG and WJ) for compliance with the exclusion criteria described in the Methods section, and 221 were included in the study.

Figure 1 presents the literature search and selection process based on the PRISMA guidelines [14].

Key Question 1 (Barriers and Facilitators)

We defined the categories included in the categorization model based on the information we found in the abstracts. We identified 27 categories, which are summarized and defined in Table 1 [18].

We classified the abstracts of the 221 articles according to the 3 levels described in the Methods section. Multimedia Appendix 1 [19-239] presents the outcomes.

Furthermore, we analyzed the article frequency, as Multimedia Appendix 2 shows.

The narratives on the category (quality of healthcare) most mentioned as contributing to the success of eHealth interventions reported on improved diagnosis [28,72], better communication with the patient [48,84], and supported patient-centered care [19,48]. Factors less clinically related were also mentioned, such as the diminishment of the care provision gap for patients [22], and the improvement of patients' clinical management [25].

Figure 1. Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) flowchart of the literature search and article selection.
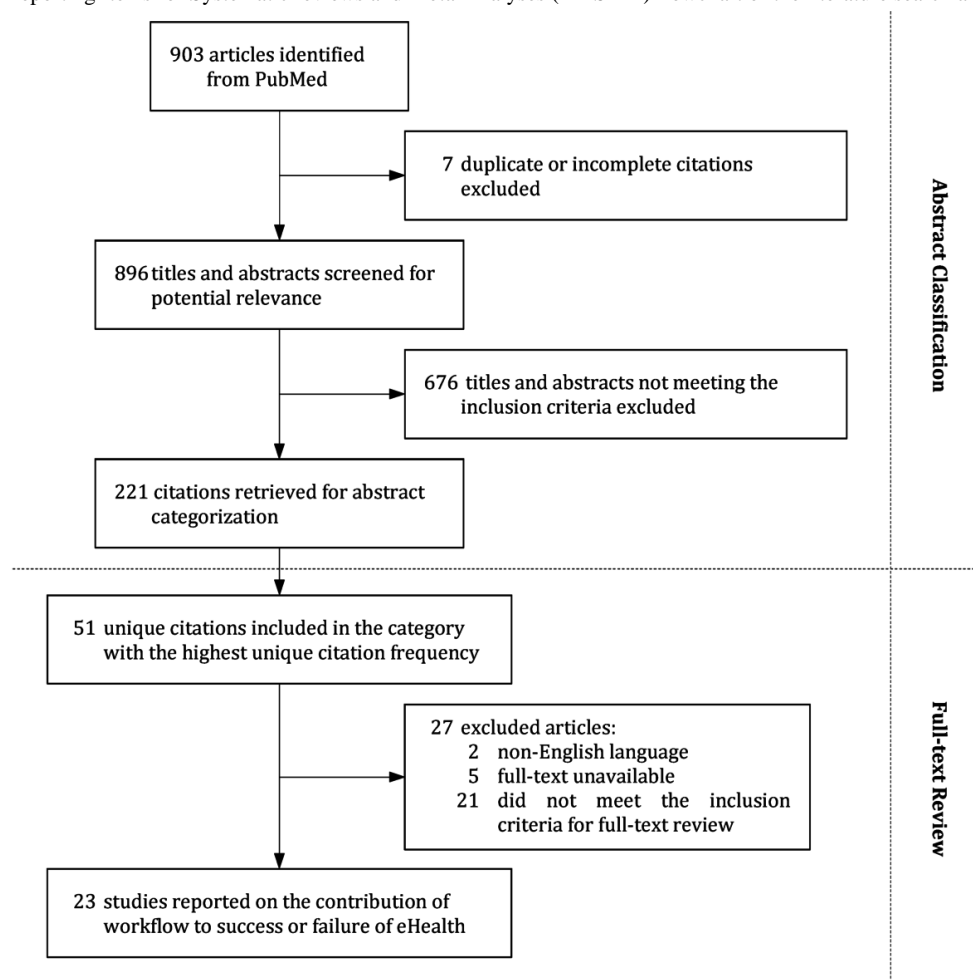
On the other hand, a few articles in the category (costs) most mentioned as contributing to the failure of eHealth interventions established a relationship between the costs and quality of care. As an example, Chan et al [46] pointed out as the main benefit of the eHealth intervention focus of the study that patients in a rural area could be examined by physicians from central hospitals without needing a physical consultation, while at the same time resulting in financial savings. However, the main focus of the articles in this category was on eHealth adoption. Villalba et al [102] identified national investments and funding programs as facilitators in the adoption of 11 eHealth interventions in 8 European countries. O'Toole et al [140], Devriendt et al [45], and Foldy [108] also identified the shortage of financial resources as a common barrier to the adoption and implementation of eHealth. In a similar manner, Ford et al [114] stressed the importance of finding financial mechanisms to support the organizational changes required to adopt eHealth. While acknowledging that a national policy for investment in eHealth interventions is crucial to its adoption, Rozenblum et al [132] argued that financial incentives should be based on patient outcomes that might ensue from the eHealth intervention. DeWorsop et al [129] and Lee and Billings [133] reported on the importance of cost effectiveness to promoting the adoption of eHealth interventions.

To better understand what were, for each entity, the most relevant categories reported to contribute to the success and failure of eHealth interventions, Table 2 summarizes the results from the article frequency analysis. We excluded the society entity from this analysis, as we did not consider the number of articles to be representative.

Key Question 2 (Success/Failure Factors)

The article frequency analysis, presented in Multimedia Appendix 2, revealed that the most representative category was workflow, with 51 unique articles. We assessed all the unique articles according to the exclusion criteria presented in Textbox 2, which resulted in 23 articles remaining for full-text review (Figure 1).

The full-text analysis of the 23 articles, with special attention to the descriptions of how the workflow category affected the success of the eHealth intervention, identified the following 6 barriers.

Workload increased the amount of work and tasks (or time required to perform them) needed to complete a clinical process, when compared with the workflow established before the eHealth intervention [19,22,36,69,73,85,87,93,95,98,99,107].

Workflow disruption resulted in the inability to complete the work process in a linear and smooth manner [19,69,73,84,92,106].

Table 1. Categories and their definitions.

Category	Definition
Usability	Ease of use, learnability (ie, easy for users to learn how to perform basic tasks), and interface intuitiveness are present.
Adoption	Users acknowledge the use of the eHealth tool.
Workflow	Workflow is defined by the way people interact with their work, communication pathways, and other people. It should be noted that different professional groups might have different understanding of workflow. As most of the selected abstracts were based on sociological research, this definition excludes the logistics of information flows.
Costs	This category includes all articles that reported on money, finances, and value in financial contexts.
System architecture	This refers to the fundamental organization of a system embodied in its components, their relationships to each other and to the environment, and the principles guiding its design and evolution [18].
Policies	The policies category is essentially related to governmental policies and rules. It often involves legal and financial aspects based on subsidies to promote the use of eHealth tools.
Interoperability	This refers to the ability of a system to exchange and make use of information from another system.
Patient empowerment and self-management	These are tools or techniques that give patients control over their own health and access to their health data.
Infrastructure	In the context of this study, infrastructure refers to the communication structures required for the operation of the eHealth tool.
Leadership	This refers to all managerial levels and the decisions made by them.
Assessment	This category covers considerations of feasibility, efficiency, effectiveness, operational results or other associated outcomes, and the effects of the implementation of eHealth tools.
Conformity with other healthcare entities	Conformity refers to the usability of information between healthcare providers in regard to clinical processes and the ability to replicate the eHealth tool implementation in different sites.
ICT[a] training	This covers user-focused training and support in the use of the eHealth tool.
Holistic approach	When the focus is on patient care, this approach implies perceiving the subject of study as a whole person, considering mental and social factors, as opposed to just someone who contracted a disease or disability. When the holistic approach focuses on the organization, this is seen as whole rather than separate entities (ie, departments, wards, and different forms of special care).
Reliability of connection and technology	This refers to the stability of communications structures during use. The stability and reliability of eHealth tools, both software and hardware related, are also included.
Standardization	Software and hardware conform to standards.
Culture	This comprises the culture of an organization, country, region, or population group.
ICT vs traditional methods	This category considers the comparability between the use of eHealth tools (ICT), and preestablished methods (traditional) (eg, videoconferencing vs face-to-face consultations).
Privacy and security	Privacy refers to the confidentiality of personal information, usually relating to personal data stored on computer systems. Security refers to the protection of computer systems against information, communications, and physical damage. In the course of classifying the abstracts, reports on security issues were often related to privacy problems in healthcare settings. Therefore, these 2 categories were combined.
Legal	Legal problems relate to legislation issues.
Safety	Safety is considered from a clinical perspective.
Access to healthcare	Access refers to the right of or opportunity for patients to receive, or come in contact with, healthcare organizations or providers.
Education	Education enlightens people about their health (eg, providing information on their disease or disability).
Quality of healthcare	A good quality of healthcare improves the healthcare delivery process and its outcomes, in both an organizational and a clinical context.
Patient-provider relationship	healthcare professionals and patients.
User involvement	This is considered from the human-centered design perspective.
Adherence to treatment	This refers to the patient's compliance with the treatment plan.

[a]ICT: information and communication technology.

XSL•FO

RenderX

Table 2. Categories contributing to the success or failure of eHealth interventions by entity.

Outcome	Entity		
	Patients	Healthcare professionals	Health system
Success	Patient empowerment and self-management	Quality of healthcare	Costs Policies
Failure	Privacy and security	Workflow	Costs

Alignment with clinical processes was a barrier when the eHealth tool did not integrate with or support the existing clinical process [38,85].

Undefined and changed roles resulted when the responsibility for a workflow task was not the same after the eHealth intervention, or new tasks were included in the workflow and no responsibilities were assigned [22,36,87,95,97,105,106].

Undermined face-to-face communication refers to the impact on personal contact with the patient and other healthcare professionals [56,72,86,87,91,106].

Staff turnover refers to the rotation of healthcare professionals between departments, or short-term contracts, that require new learning or training on the eHealth tool [103].

Workload can be classified as the biggest workflow-related concern, since it was overrepresented in the results, being addressed in 12 of the 23 studies. In these studies, healthcare personnel stressed the increase in the amount of work after the implementation of the eHealth tool. eHealth was described as being both time and resource intensive [19,22,36,69,73,85,98,99,107] and [19,87,93,95] indicated discontent about the amount of self-reported and self-recorded health data provided by the tool for assessment.

The second most mentioned barrier, addressed in 7 studies, was the undefined roles and change of work practice of the parties involved in the workflow. Narratives reported, for instance, that the new role was tangential to their role as healthcare professionals [22,36,87]—for example, coaching patients in the use of the technology, analyzing the self-reported data and subsequently answering the patient's questions [87,95], the need for new competences [97], and unresolved attribution of responsibilities [105,106].

Workflow disruption was significantly present in the studies, reflected in narratives describing eHealth as not being fitted to the existing workflow due to time (eg, data provided to the system a priori, and work tasks having to be performed by others) [19,73,84,106] or space [92] constraints, and breaking of traditions [69].

According to Kapadia et al [83], healthcare personnel report a preference for face-to-face communication over digital long-distance systems. Nielsen and Mathiassen [38] mentioned the loss of contact between personnel as a trigger to the reduction of knowledge sharing and collegial relationships. Less mentioned, but still significant, were statements that eHealth is impersonal and, therefore, undermines face-to-face communication [56,72,86,87,91,106], substantiated by the claim that the foundation of good nursing is physical presence, human touch, and the use of all senses.

Less addressed in the studies, but still significant barriers, were the alignment of the eHealth intervention with the clinical workflow [38,85] and staff turnover [103]. These are related to how supportive and well integrated the tool is in the workflow and the need for constant training of staff, respectively.

In the literature, two general workflows were mentioned. The first was the preestablished workflow, defined as the workflow in an organization, or at a specific organizational level (eg, the cardiology ward in a hospital), before the eHealth intervention. The second was the new workflow, which describes the workflow after the eHealth intervention. Different authors have various ideas about this change.

Some [38,84,85,92] argued that eHealth interventions should be adapted into the preestablished workflow in order to succeed. Others [19,72,73 93,97] advocated that the workflow will change, or is necessary to change, in order for the intervention to be successful.

eHealth interventions in healthcare organizations triggered changes in the workflow [73,97]. Such changes were not limited to the directly involved staff, but also had an impact on others within the organization [73]. Therefore, changes in the workflow had the potential to alter the organization both in a negative and in a positive way [36,38,86,88,103]. Additionally, a change that was initially a positive development could lead to a rather negative outcome. Such circumstances are reported by Das et al [86], Davis et al [87], and Chung et al [95], where the overabundance of data led to an inability to use the data due to time constraints.

Professional values and personal feelings should also be considered to be barriers to eHealth interventions, as they may come into conflict with the use of technology [36,38]. In these works, feelings of healthcare staff about the technology resulted in negative thoughts and skepticism, resulting in the technology never fully being integrated into the workflow.

Discussion

General Findings

There is a reasonable amount of original research exploring the effects of eHealth interventions. However, among the 903 articles we identified, only 221 met the inclusion criteria for this systematic review. The studies were heterogeneous by country, type of eHealth intervention, method of implementation, and reporting perspectives (ie, patient, healthcare professionals, health system, and society). The article frequency analysis presented in Multimedia Appendix 2 did not find a significant discrepancy between the number of reports on failure (392/844, 46.5%) and success (452/844, 53.6%),

which encourages definitive conclusions on the key questions that prompted the review.

While evaluating the 903 articles, we realized that the studies could be grouped into three chronological eras, according to their research focus: (1) up to 1999: in this first era, most of the articles addressed the technology with a focus on aspects such as functionality and infrastructure; (2) 2000-2009: in the second era, the focus shifted from the technology to the organization; in this setting, an organization could be a healthcare organization or a community (local, regional, national, or international); (3) 2010 to the present: the third era focused on individuals, where researchers investigated how people work, often from a bottom-up perspective, compared with a top-down scope on the complete organization.

The first two eras defined in this chronological pattern are supported by Nielsen and Mathiassen [38].

When we applied the three chronological eras to the 221 articles selected for abstract classification, it became clear that the search strategy achieved the desired results in identifying studies with relevant clinical outcomes. This is demonstrated by the fact that most of the articles (168 articles) fell into the third era, and only 3 articles were included in the first era, which had a technological focus.

Key Question 1 (Barriers and Facilitators)

The category most mentioned as contributing to the success of eHealth interventions was quality of healthcare. This category was also one of the three pillars of care described by Bashshur et al [12] as being the support of successful eHealth interventions. The authors related quality of care to professional performance standards, related the role and contribution of the intervention to clinical practice, and described how this contribution is achieved [12]. This is in line with the narratives found in the articles assigned to this category.

Most of the unique articles in the category quality of healthcare belonged in the third era, with only 9 being in the second era, and none in the first. Belonging in the third era, the focus was on the individual, which could explain the increase in focus on quality of care.

The articles in the quality of healthcare category were evenly distributed among the entities, showing a significant difference between the article frequency as success (n=55) and failure (n=12), and revealing a positive view among the users, thus positioning this category as an evident facilitator of eHealth interventions.

On the other hand, costs is the category most mentioned as contributing to the failure of eHealth interventions. Similar to quality of healthcare, this category is also one of the three pillars of the promises of eHealth proposed by Bashshur et al [12]. In their work, the cost containments pillar focused on cost reductions for patients and providers in a broad sense, supporting a clear definition of how the eHealth intervention would facilitate the provision of care services at a lower cost without loss of quality of care [12]. However, the main focus of the articles in the costs category was on eHealth adoption, and only

a few articles established a relation between the costs and quality of care in the manner of Bashshur et al.

The classification of unique articles in the costs category according to the aforementioned eras revealed that most, 80% (40 articles), fell in the third era. In line with the focus on the individual that characterized this era, we noted that only 13 of the 50 unique articles included results after the eHealth intervention, and that as many as 37 report on expectations, potentials, and other future possibilities. Most of these articles arose from the social sciences and addressed the actors' expectations of eHealth, revealing the general idea that eHealth interventions bring cost reductions along with the implementation. An important aspect of expectations is that they are based on the actors' opinions or wishes, often before the results are clear. However, the actual costs and financial benefits can be measured only after implementation. Since most research was done during the implementation, often at an early stage, this means that data were being collected at a stage when funding was still available and the full impact of the eHealth intervention on the organization was not yet evident. This being said, the costs category certainly presents as a barrier to eHealth interventions.

Looking at the category's article frequency distribution among the entities, costs was mostly mentioned by the health system entity, being, in fact, considered as the most important to both the success and the failure of eHealth interventions. Such lack of a relevant difference between the contribution of costs to success (18 articles) and failure (19 articles) leads to the conclusion that this category is a major concern for the health system. Thus, and since this entity includes organizations and governmental bodies, we conclude that certain financial conditions are required to distribute eHealth services. Hence, we inferred that, to adopt eHealth in their services, it is not important for the health system to attain financial profit, but rather to not have a loss. It is not surprising to see that, along with costs, the category most cited as contributing to that success of eHealth interventions for the health system entity was policies, as the importance of national policy investments and reimbursement rules was so often mentioned in the articles.

However, such perspective on what is most relevant for the success and failure of eHealth interventions is not shared by the other entities. Considering the narratives related to quality of healthcare, described in the Results section, it is not surprising that this category appeared as the most important for the success of eHealth interventions for healthcare professionals. Following the same line of thought, workflow appeared as the most relevant for the failure of eHealth. We discuss the ways in which each of these issues affected the outcome of eHealth interventions below under Key Question 2. However, we can already assert that what these issues have in common is the deprivation of time to provide care services. Conversely, patients considered of most importance that the eHealth interventions support them in managing their own health independently. This is demonstrated in that the most cited category by the patient entity as contributing to the success of eHealth interventions was patient empowerment and self-management. It seems almost too obvious to state that, as patients want to be provided with the means to manage their own health, their most mentioned

category as leading to the failure of eHealth was privacy and security. This reveals that, even though patients want to be independent and manage their own health remotely, they are aware of the sensitivity of the data that are being shared.

Key Question 2 (Success/Failure Factors)

Kruse et al [240] examined barriers in eHealth research on a quantitative basis for several consecutive years. In their work, they stated that workflow was one of the most mentioned barriers in literature, which is confirmed by the article frequency analysis presented in Multimedia Appendix 2, where workflow is the category with the most unique articles. This positions this category as being of major concern for the entities.

The claim that eHealth interventions should be fitted to the preestablished workflow in order to succeed [38,84,85,92] was also upheld by Gardner [241] and in the systematic review by Kawamoto et al [242]. In opposition, others [19,72,73,93,97] supported that changes in the workflow are inevitable and necessary for the eHealth intervention to be successful, stressing that the adoption of new eHealth tools within the preestablished workflow creates problems during the implementation process [19]. This is corroborated by the findings in the systematic review by Davis et al [243] and by Bowens et al [244], who also point out that the need to reengineer the workflow to integrate eHealth can be a trigger to improve efficiency, distribution of tasks, patient safety, and the quality of the data collected from the patient. Such disparity of opinions leads to 2 questions. (1) What is the importance of eHealth to workflow? (2) How are the design and the outcomes of eHealth related?

What is the Importance of eHealth to Workflow?

Healthcare organizations' resistance to change has identified barriers to eHealth interventions [245,246]. Most organizations were not created to accommodate eHealth and, as noted by Appelbaum and Wohl [247], the inability of healthcare organizations to adapt to changes is a barrier to their own sustainability. The resistance to change is also present at the staff level. Pardo del Val and Martinez Fuentes [248] proposed a framework for the sources to resistance to change, among which are the "relation between change values and organizational values," denoting the difference between what is important for the individual and for the organization, and "cynicism," denoting the negative feelings toward the success of the change [249]. This is in line with the narratives on how professional values and personal feelings hinder the use of eHealth [36,38], found in the articles selected for review.

To maximize the likelihood of a successful eHealth intervention, healthcare professionals must acquaint themselves with the tool [250]. Furthermore, there should not be a possibility to fall back on the old workflow [39,250].

How Are the Design and the Outcomes of eHealth Related?

If eHealth does not meet the expectations or requirements of healthcare personnel, it is possible that the eHealth tool will not be used as anticipated [38]. This is caused by a difference between the prospects, which are often made by management personnel, and the reality of healthcare personnel. In the literature, the technical perspective during the design phase is identified as a significant contributor to the gap between prospect and reality [94,251]. Declerck and Aime [251] used the term technocentrism to classify the importance of technology in the design phase, where other actors, such as healthcare professionals, appear as secondary concerns. In addition, it should be taken into consideration that different professions within healthcare have different needs from eHealth [83], and these needs have to be addressed to make eHealth applications operate successfully in healthcare organizations.

In a veiled manner of recognizing that the technical focus during the design phase was a barrier to eHealth adoption, the technology design mentality shifted to one with a holistic and human focus. There is a widespread perception in human-computer interaction that recognizes that users should be involved in the design in order to create technology that is more relevant, resulting in the wide acceptance of user-centered design and participatory design approaches [252]. Thus, it is legitimate to conclude that the design and the outcomes of eHealth are clearly related, and that user involvement during the design phase is of the most importance for the success eHealth interventions within the domain of workflow.

Conclusions to Key Question 1

With this systematic review, we identified quality of healthcare as the major facilitator of eHealth interventions and costs as the major barrier.

Within the quality of healthcare category, a positive impact in clinical care appears to be relevant to the adoption of eHealth. The impact is assessed with varied metrics, such as improved diagnosis, clinical management, and patient-centered care. The role of the costs category in the adoption of eHealth seems to be more consensual. Most of the studies included in this category identify the shortage of financial resources as a common barrier to the adoption and implementation of eHealth.

The importance of quality of healthcare in the success of eHealth interventions is shared only by the healthcare professionals entity. The concern of healthcare professionals with care services is also reflected by the workflow category being the most mentioned contributor to the failure of eHealth interventions.

The patient entity revealed their wish of controlling their own health by placing the patient empowerment and self-management character of the intervention as the major contributor to success. However, this wish for control does not come without apprehension, as demonstrated by mentioning the privacy and security category as the major contributor to the failure of eHealth.

The health system entity attributed to 1 category, costs, the same relevance to both success and failure of eHealth interventions. This does not appear as a surprise, as this category mostly relates to financial resources. As these financial resources often originate from governmental sources, it is reasonable that the health system entity mentioned the policies category, along with costs, as a major contributor to the success of eHealth interventions.

Conclusions to Key Question 2

According to the reviewed literature, workflow was the most relevant factor to the outcome of eHealth interventions across all entities. A full-text review identified 6 barriers to the adoption of eHealth related to this category. Most of these barriers were reported by the healthcare professionals entity, and are as follows: (1) workload, (2) workflow disruption, (3) alignment with clinical processes, (4) undefined and changed roles, (5) undermined face-to-face communication, and (6) staff turnover.

Limitations

Some limitations in this review must be acknowledged. Despite having searched the PubMed database using standard systematic review protocols, we could have searched further databases. This might have limited the results, as the search process may not have captured relevant studies that were not indexed in the PubMed database. However, we considered that the PubMed search alone yielded a representative overview of the field.

Additionally, many studies were based on data that were gathered either before or during the implementation process when the eHealth intervention was not well established in the workplace. As the reports were collected at such an early stage, they might be overshadowed by emotional resistance and fear of change.

Future Research

In spite of the limitations, this systematic review remains a useful source of information, as it synthesized common challenges in the development of eHealth interventions and in the planning of their implementation. In this regard, considerations for future research include the following:

- Identify potential facilitators and barriers at the earliest possible stage to ensure that full account is taken when defining the development and implementation strategy;
- Determine the impact of eHealth interventions on the quality of care, with particular attention given to improved diagnosis, clinical management, and patient-centered care;
- Evaluate the financial needs and consequences in the short, medium, and long term, to avoid nonadoption due to lack of funding;
- Perform in-depth studies of the workflow(s) that the intervention will support and perceive the clinical processes involved;
- Improve the privacy and security features of eHealth interventions targeted at patients.

Authors' Contributions

All authors contributed substantially to the conception and design of the work; approved the final manuscript; and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. CG and WJ acquired, analyzed, and interpreted the data for the work. CG was the lead writer. WJ and MAJ drafted the manuscript. MAJ analyzed and interpreted the data for the work, and critically revised the manuscript for important intellectual content.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Article abstract classification.

[PDF File (Adobe PDF File), 40KB - jmir_v20i5e10235_app1.pdf]

Multimedia Appendix 2

Article frequencies.

[PDF File (Adobe PDF File), 40KB - jmir_v20i5e10235_app2.pdf]

References

1. World Health Organization. User's guide: the WHO global code of practice on the international recruitment of health personnel. Geneva, Switzerland: WHO; 2011 Jan. URL: http://apps.who.int/iris/bitstream/handle/10665/70525/WHO_HSS_HRH_HMR_2010.2_eng.pdf;jsessionid=A093F16A3CFED3DCB8C2DE38315DCF2D?sequence=1 [accessed 2018-02-12] [WebCite Cache ID 6yv3vyCys]
2. Parker MG, Thorslund M. Health trends in the elderly population: getting better and getting worse. Gerontologist 2007 Apr;47(2):150-158. [Medline: 17440120]

3.   Ward BW, Schiller JS. Prevalence of multiple chronic conditions among US adults: estimates from the National Health Interview Survey, 2010. Prev Chronic Dis 2013 Apr 25;10:E65 [FREE Full text] [doi: 10.5888/pcd10.120203] [Medline: 23618545]

4.   Christensen K, Doblhammer G, Rau R, Vaupel JW. Ageing populations: the challenges ahead. Lancet 2009 Oct 3;374(9696):1196-1208 [FREE Full text] [doi: 10.1016/S0140-6736(09)61460-4] [Medline: 19801098]

5.   Ågren G, Berensson K. Healthy Ageing -- A Challenge for Europe. Stockholm, Sweden: The Swedish National Institute of Public Health; 2006.

6.   Goldberg HS, Morales A, Gottlieb L, Meador L, Safran C. Reinventing patient-centered computing for the twenty-first century. Stud Health Technol Inform 2001;84(Pt 2):1455-1458. [Medline: 11604967]

7.   Hjelm NM. Benefits and drawbacks of telemedicine. J Telemed Telecare 2005 Jan;11(2):60-70. [doi: 10.1258/1357633053499886] [Medline: 15829049]

8.   Tanriverdi H, Iacono C. Knowledge barriers to diffusion of telemedicine. In: Proceedings of the International Conference on Information systems. Atlanta, GA: Association for Information Systems; 1998 Dec Presented at: ICIS; Dec 13-16, 1998; Helsinki, Finland p. 13-16 URL: https://dl.acm.org/citation.cfm?id=353057 [doi: 10.1145/353053.353057]

9.   Jung M, Berthon P. Fulfilling the promise: a model for delivering successful online health care. J Med Marketing 2009 Jul;9(3):243-254. [doi: 10.1057/jmm.2009.26]

10.   Shortliffe EH. Strategic action in health information technology: why the obvious has taken so long. Health Aff (Millwood) 2005;24(5):1222-1233. [doi: 10.1377/hlthaff.24.5.1222] [Medline: 16162567]

11.   Berg M. Patient care information systems and health care work: a sociotechnical approach. Int J Med Inform 1999 Aug;55(2):87-101. [Medline: 10530825]

12.   Bashshur RL, Shannon G, Krupinski EA, Grigsby J. Sustaining and realizing the promise of telemedicine. Telemed J E Health 2013 May;19(5):339-345 [FREE Full text] [doi: 10.1089/tmj.2012.0282] [Medline: 23289907]

13.   Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions. West Sussex, UK: John Wiley & Sons; 2011.

14.   Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009;339:b2535 [FREE Full text] [Medline: 19622551]

15.   Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med Inform Decis Mak 2007;7:16 [FREE Full text] [doi: 10.1186/1472-6947-7-16] [Medline: 17573961]

16.   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]

17.   Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]

18.   ISO/IEC/IEEE. Systems software engineering -- architecture description. In: ISO/IEC/IEEE 42010:2011(E) (Revision of ISO/IEC 42010:2007 and IEEE Std 1471-2000). New York, NY: IEEE; Dec 01, 2011.

19.   Steele GC, Gill A, Khan AI, Hans PK, Kuluski K, Cott C. The electronic patient reported outcome tool: testing usability and feasibility of a mobile app and portal to support care for patients with complex chronic disease and disability in primary care settings. JMIR Mhealth Uhealth 2016 Jun 02;4(2):e58 [FREE Full text] [doi: 10.2196/mhealth.5331] [Medline: 27256035]

20.   Duman-Lubberding S, van Uden-Kraan CF, Jansen F, Witte BI, van der Velden LA, Lacko M, et al. Feasibility of an eHealth application "OncoKompas" to improve personalized survivorship cancer care. Support Care Cancer 2016 May;24(5):2163-2171 [FREE Full text] [doi: 10.1007/s00520-015-3004-2] [Medline: 26563178]

21.   Hoque MR. An empirical study of mHealth adoption in a developing country: the moderating effect of gender concern. BMC Med Inform Decis Mak 2016 Dec 03;16:51 [FREE Full text] [doi: 10.1186/s12911-016-0289-0] [Medline: 27142844]

22.   Hunting G, Shahid N, Sahakyan Y, Fan I, Moneypenny CR, Stanimirovic A, et al. A multi-level qualitative analysis of Telehomecare in Ontario: challenges and opportunities. BMC Health Serv Res 2015 Dec 09;15:544 [FREE Full text] [doi: 10.1186/s12913-015-1196-2] [Medline: 26645639]

23.   Jiménez-Fernández S, de Toledo P, del Pozo F. Usability and interoperability in wireless sensor networks for patient telemonitoring in chronic disease management. IEEE transactions on bio-medical engineering. IEEE Trans Biomed Eng 2013 Dec;60(12):3331-3339. [doi: 10.1109/TBME.2013.2280967] [Medline: 24021636]

24.   Manafò E, Wong S. Promoting eHealth literacy in older adults: key informant perspectives. Can J Diet Pract Res 2013;74(1):37-41. [doi: 10.3148/74.1.2013.37] [Medline: 23449214]

25.   Marzegalli M, Lunati M, Landolina M, Perego GB, Ricci RP, Guenzati G, et al. Remote monitoring of CRT-ICD: the multicenter Italian CareLink evaluation--ease of use, acceptance, and organizational implications. Pacing Clin Electrophysiol 2008 Oct;31(10):1259-1264. [doi: 10.1111/j.1540-8159.2008.01175.x] [Medline: 18811805]

26.   Nikander K, Denyer J, Dodd M, Dyche T, Webb K, Weller P, et al. The Adaptive Aerosol Delivery system in a telehealth setting: patient acceptance, performance and feasibility. J Aerosol Med Pulm Drug Deliv 2010 Apr;23 Suppl 1:S21-S27 [FREE Full text] [doi: 10.1089/jamp.2009.0770] [Medline: 20373906]

27. Nitsch M, Dimopoulos CN, Flaschberger E, Saffran K, Kruger JF, Garlock L, et al. A guided online and mobile self-help program for individuals with eating disorders: an iterative engagement and usability study. J Med Internet Res 2016 Jan 11;18(1):e7 [FREE Full text] [doi: 10.2196/jmir.4972] [Medline: 26753539]

28. Palmier-Claus JE, Rogers A, Ainsworth J, Machin M, Barrowclough C, Laverty L, et al. Integrating mobile-phone based assessment for psychosis into people's everyday lives and clinical care: a qualitative study. BMC Psychiatry 2013;13:34 [FREE Full text] [doi: 10.1186/1471-244X-13-34] [Medline: 23343329]

29. Pelayo S, Ong M. Human factors and ergonomics in the design of health information technology: trends and progress in 2014. Yearb Med Inform 2015 Aug 13;10(1):75-78 [FREE Full text] [doi: 10.15265/IY-2015-033] [Medline: 26293852]

30. Prescher S, Deckwart O, Winkler S, Koehler K, Honold M, Koehler F. Telemedical care: feasibility and perception of the patients and physicians: a survey-based acceptance analysis of the Telemedical Interventional Monitoring in Heart Failure (TIM-HF) trial. Eur J Prev Cardiol 2013 Jun;20(2 Suppl):18-24. [doi: 10.1177/2047487313487483e] [Medline: 23702987]

31. Robben SHM, Perry M, Huisjes M, van Nieuwenhuijzen L, Schers HJ, van Weel C, et al. Implementation of an innovative web-based conference table for community-dwelling frail older people, their informal caregivers and professionals: a process evaluation. BMC Health Serv Res 2012 Aug 15;12:251 [FREE Full text] [doi: 10.1186/1472-6963-12-251] [Medline: 22894654]

32. 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. BMJ 2010;341:c5814 [FREE Full text] [Medline: 21081595]

33. Ehrler F, Lovis C. Supporting elderly homecare with smartwatches: advantages and drawbacks. Stud Health Technol Inform 2014;205:667-671. [Medline: 25160270]

34. Kreps GL, Neuhauser L. New directions in eHealth communication: opportunities and challenges. Patient Educ Couns 2010 Mar;78(3):329-336. [doi: 10.1016/j.pec.2010.01.013] [Medline: 20202779]

35. Radhakrishnan K, Xie B, Jacelon CS. Unsustainable home telehealth: a Texas qualitative study. Gerontologist 2016 Oct;56(5):830-840. [doi: 10.1093/geront/gnv050] [Medline: 26035878]

36. Sølling IK, Carøe P, Mathiesen KS. Development and implementation of IT require focus on user participation, acceptance and workflow. Stud Health Technol Inform 2014;201:219-226. [Medline: 24943547]

37. Zhang X, Yu P, Yan J. Patients' adoption of the e-appointment scheduling service: A case study in primary healthcare. Stud Health Technol Inform 2014;204:176-181. [Medline: 25087546]

38. Nielsen JA, Mathiassen L. Interpretive flexibility in mobile health: lessons from a government-sponsored home care program. J Med Internet Res 2013 Oct 30;15(10):e236 [FREE Full text] [doi: 10.2196/jmir.2816] [Medline: 24172852]

39. Abbott KC, Mann S, DeWitt D, Sales LY, Kennedy S, Poropatich RK. Physician-to-physician consultation via electronic mail: the Walter Reed Army Medical Center Ask a Doc system. Mil Med 2002 Mar;167(3):200-204. [Medline: 11901566]

40. Eason K, Waterson P. The implications of e-health system delivery strategies for integrated healthcare: lessons from England. Int J Med Inform 2013 May;82(5):e96-e106. [doi: 10.1016/j.ijmedinf.2012.11.004] [Medline: 23266062]

41. Fonda SJ, Paulsen CA, Perkins J, Kedziora RJ, Rodbard D, Bursell SE. Usability test of an internet-based informatics tool for diabetes care providers: the comprehensive diabetes management program. Diabetes Technol Ther 2008 Feb;10(1):16-24. [doi: 10.1089/dia.2007.0252] [Medline: 18275359]

42. Landman A, Emani S, Carlile N, Rosenthal DI, Semakov S, Pallin DJ, et al. A mobile app for securely capturing and transferring clinical images to the electronic health record: description and preliminary usability study. JMIR Mhealth Uhealth 2015;3(1):e1 [FREE Full text] [doi: 10.2196/mhealth.3481] [Medline: 25565678]

43. Lodhia V, Karanja S, Lees S, Bastawrous A. Acceptability, usability, and views on deployment of Peek, a mobile phone mHealth intervention for eye care in Kenya: qualitative study. JMIR Mhealth Uhealth 2016 May 09;4(2):e30 [FREE Full text] [doi: 10.2196/mhealth.4746] [Medline: 27160779]

44. Sinclair C, Holloway K, Riley G, Auret K. Online mental health resources in rural Australia: clinician perceptions of acceptability. J Med Internet Res 2013;15(9):e193 [FREE Full text] [doi: 10.2196/jmir.2772] [Medline: 24007949]

45. Devriendt E, Wellens NIH, Flamaing J, Declercq A, Moons P, Boonen S, et al. The interRAI Acute Care instrument incorporated in an eHealth system for standardized and web-based geriatric assessment: strengths, weaknesses, opportunities and threats in the acute hospital setting. BMC Geriatr 2013 Sep 05;13:90 [FREE Full text] [doi: 10.1186/1471-2318-13-90] [Medline: 24007312]

46. Chan FY, Soong B, Watson D, Whitehall J. Realtime fetal ultrasound by telemedicine in Queensland. A successful venture? J Telemed Telecare 2001;7 Suppl 2:7-11. [doi: 10.1258/1357633011937290] [Medline: 11747644]

47. 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 3;5(3):81-85 [FREE Full text] [doi: 10.1016/j.hjdsi.2016.08.004] [Medline: 27720139]

48. George SM, Hamilton A, Baker R. Pre-experience perceptions about telemedicine among African Americans and Latinos in South Central Los Angeles. Telemed J E Health 2009;15(6):525-530 [FREE Full text] [doi: 10.1089/tmj.2008.0152] [Medline: 19566397]

49. Steele GC, Miller D, Kuluski K, Cott C. Tying eHealth tools to patient needs: exploring the use of ehealth for community-dwelling patients with complex chronic disease and disability. JMIR Res Protoc 2014;3(4):e67 [FREE Full text] [doi: 10.2196/resprot.3500] [Medline: 25428028]

XSL·FO
RenderX

50. Islam SMS, Tabassum R. Implementation of information and communication technologies for health in Bangladesh. Bull World Health Organ 2015 Nov 01;93(11):806-809 [FREE Full text] [doi: 10.2471/BLT.15.153684] [Medline: 26549909]

51. Darkins A. Program management of telemental health care services. J Geriatr Psychiatry Neurol 2001;14(2):80-87. [doi: 10.1177/089198870101400207] [Medline: 11419572]

52. El-Mahalli AA, El-Khafif SH, Al-Qahtani MF. Successes and challenges in the implementation and application of telemedicine in the eastern province of Saudi Arabia. Perspect Health Inf Manag 2012;9:1-27 [FREE Full text] [Medline: 23209455]

53. Greenberg A, Kramer S, Welch V, O'Sullivan E, Hall S. Cancer Care Ontario's computerized physician order entry system: a province-wide patient safety innovation. Healthc Q 2006;9 Spec No:108-113 [FREE Full text] [Medline: 17087178]

54. Taylor J, Coates E, Brewster L, Mountain G, Wessels B, Hawley MS. Examining the use of telehealth in community nursing: identifying the factors affecting frontline staff acceptance and telehealth adoption. J Adv Nurs 2015 Feb;71(2):326-337. [doi: 10.1111/jan.12480] [Medline: 25069605]

55. Al-Qirim N. Realizing telemedicine advantages at the national level: cases from the United Arab Emirates. Telemed J E Health 2007 Oct;13(5):545-555. [doi: 10.1089/tmj.2006.0087] [Medline: 17999616]

56. King G, Richards H, Godden D. Adoption of telemedicine in Scottish remote and rural general practices: a qualitative study. J Telemed Telecare 2007;13(8):382-386. [doi: 10.1258/135763307783064430] [Medline: 18078547]

57. Ndetei DM, Jenkins R. The implementation of mental health information systems in developing countries: challenges and opportunities. Epidemiol Psichiatr Soc 2009;18(1):12-16. [Medline: 19378693]

58. Ruas SSM, Assunção A. Teleconsultations by primary care physicians of Belo Horizonte: challenges in the diffusion of innovation. Telemed J E Health 2013 May;19(5):409-414. [doi: 10.1089/tmj.2012.0095] [Medline: 23600411]

59. Hung MC, Jen WY. The adoption of mobile health management services: an empirical study. J Med Syst 2012 Jun;36(3):1381-1388. [doi: 10.1007/s10916-010-9600-2] [Medline: 20878452]

60. Choi W, Park M, Hong E, Kim S, Ahn R, Hong J, et al. Early experiences with mobile electronic health records application in a tertiary hospital in Korea. Healthc Inform Res 2015 Oct;21(4):292-298 [FREE Full text] [doi: 10.4258/hir.2015.21.4.292] [Medline: 26618036]

61. Lin CH, Lin IC, Roan JS, Yeh JS. Critical factors influencing hospitals' adoption of HL7 version 2 standards: an empirical investigation. J Med Syst 2012 Jun;36(3):1183-1192. [doi: 10.1007/s10916-010-9580-2] [Medline: 20827568]

62. Mergel I. The long way from government open data to mobile health apps: overcoming institutional barriers in the US Federal Government. JMIR Mhealth Uhealth 2014;2(4):e58 [FREE Full text] [doi: 10.2196/mhealth.3694] [Medline: 25537314]

63. Schnall R, Odlum M, Gordon P, Bakken S. Barriers to implementation of a Continuity of Care Record (CCR) in HIV/AIDS care. Stud Health Technol Inform 2009;146:248-252 [FREE Full text] [Medline: 19592843]

64. Bull TP, Dewar AR, Malvey DM, Szalma JL. Considerations for the telehealth systems of tomorrow: an analysis of student perceptions of telehealth technologies. JMIR Med Educ 2016 Jul 08;2(2):e11 [FREE Full text] [doi: 10.2196/mededu.5392] [Medline: 27731865]

65. Davis TL, DiClemente R, Prietula M. Taking mHealth forward: examining the core characteristics. JMIR Mhealth Uhealth 2016 Aug 10;4(3):e97 [FREE Full text] [doi: 10.2196/mhealth.5659] [Medline: 27511612]

66. Fico G, Arredondo MT. Use of an holistic approach for effective adoption of user-centred-design techniques in diabetes disease management: experiences in user need elicitation. Conf Proc IEEE Eng Med Biol Soc 2015;2015:2139-2142. [doi: 10.1109/EMBC.2015.7318812] [Medline: 26736712]

67. Finocchario-Kessler S, Odera I, Okoth V, Bawcom C, Gautney B, Khamadi S, et al. Lessons learned from implementing the HIV infant tracking system (HITSystem): a web-based intervention to improve early infant diagnosis in Kenya. Healthc (Amst) 2015 Dec;3(4):190-195 [FREE Full text] [doi: 10.1016/j.hjdsi.2015.07.004] [Medline: 26699342]

68. Follen M, Castaneda R, Mikelson M, Johnson D, Wilson A, Higuchi K. Implementing health information technology to improve the process of health care delivery: a case study. Dis Manag 2007 Aug;10(4):208-215. [doi: 10.1089/dis.2007.104706] [Medline: 17718659]

69. Janols R, Lind T, Göransson B, Sandblad B. Evaluation of user adoption during three module deployments of region-wide electronic patient record systems. Int J Med Inform 2014 Jun;83(6):438-449. [doi: 10.1016/j.ijmedinf.2014.02.003] [Medline: 24630924]

70. Ray KN, Felmet KA, Hamilton MF, Kuza CC, Saladino RA, Schultz BR, et al. Clinician attitudes toward adoption of pediatric emergency telemedicine in rural hospitals. Pediatr Emerg Care 2017 Apr 16;33(4):250--2257. [doi: 10.1097/PEC.0000000000000583] [Medline: 26785087]

71. Saranummi N, Beuscart R, Black N, Maglaveras N, Strano C, Karavidopoulou Y. Adoption space and the idea-to-market process of health technologies. Stud Health Technol Inform 2016;224:9-14. [Medline: 27225546]

72. Praveen D, Patel A, Raghu A, Clifford GD, Maulik PK, Mohammad AA, et al. SMARTHealth India: development and field evaluation of a mobile clinical decision support system for cardiovascular diseases in rural India. JMIR Mhealth Uhealth 2014 Dec 08;2(4):e54 [FREE Full text] [doi: 10.2196/mhealth.3568] [Medline: 25487047]

73. Aas IH. A qualitative study of the organizational consequences of telemedicine. J Telemed Telecare 2001;7(1):18-26. [doi: 10.1258/1357633011936093] [Medline: 11265934]

74. Ancker JS, Witteman HO, Hafeez B, Provencher T, van de Graaf M, Wei E. "You get reminded you're a sick person": personal data tracking and patients with multiple chronic conditions. J Med Internet Res 2015;17(8):e202 [FREE Full text] [doi: 10.2196/jmir.4209] [Medline: 26290186]

75. Bradford NK, Caffery LJ, Smith AC. Awareness, experiences and perceptions of telehealth in a rural Queensland community. BMC Health Serv Res 2015 Sep 28;15:427 [FREE Full text] [doi: 10.1186/s12913-015-1094-7] [Medline: 26416464]

76. Damanpour S, Srivastava D, Nijhawan RI. Self-acquired patient images: the promises and the pitfalls. Semin Cutan Med Surg 2016 Mar;35(1):13-17. [doi: 10.12788/j.sder.2016.013] [Medline: 26963112]

77. De Asmundis C, Ricciardi D, Namdar M, Pappaert G, Rodríguez-Mañero M, Wauters K, et al. ICD function and dysfunction in patients with arrhythmogenic cardiac diseases: the role of home monitoring. Acta Cardiol 2013 Aug;68(4):387-394. [doi: 10.2143/AC.68.4.2988892] [Medline: 24187765]

78. Farzanfar R, Finkelstein J, Friedman RH. Testing the usability of two automated home-based patient-management systems. J Med Syst 2004 Apr;28(2):143-153. [Medline: 15195845]

79. Gabrielian S, Yuan A, Andersen RM, McGuire J, Rubenstein L, Sapir N, et al. Chronic disease management for recently homeless veterans: a clinical practice improvement program to apply home telehealth technology to a vulnerable population. Med Care 2013 Mar;51(3 Suppl 1):S44-S51 [FREE Full text] [doi: 10.1097/MLR.0b013e31827808f6] [Medline: 23407011]

80. Holmström I. Decision aid software programs in telenursing: not used as intended? Experiences of Swedish telenurses. Nurs Health Sci 2007 Mar;9(1):23-28. [doi: 10.1111/j.1442-2018.2007.00299.x] [Medline: 17300541]

81. Cook EJ, Randhawa G, Sharp C, Ali N, Guppy A, Barton G, et al. Exploring the factors that influence the decision to adopt and engage with an integrated assistive telehealth and telecare service in Cambridgeshire, UK: a nested qualitative study of patient 'users' and 'non-users'. BMC Health Serv Res 2016 Apr 19;16:137 [FREE Full text] [doi: 10.1186/s12913-016-1379-5] [Medline: 27095102]

82. Donkin L, Glozier N. Motivators and motivations to persist with online psychological interventions: a qualitative study of treatment completers. J Med Internet Res 2012;14(3):e91 [FREE Full text] [doi: 10.2196/jmir.2100] [Medline: 22743581]

83. Kapadia V, Ariani A, Li J, Ray PK. Emerging ICT implementation issues in aged care. Int J Med Inform 2015 Nov;84(11):892-900. [doi: 10.1016/j.ijmedinf.2015.07.002] [Medline: 26250987]

84. Hao WR, Hsu YH, Chen KC, Li HC, Iqbal U, Nguyen PA, et al. LabPush: a pilot study of providing remote clinics with laboratory results via short message service (SMS) in Swaziland, Africa - a qualitative study. Comput Methods Programs Biomed 2015 Jan;118(1):77-83. [doi: 10.1016/j.cmpb.2014.10.005] [Medline: 25453385]

85. 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]

86. 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]

87. Davis MM, Currey JM, Howk S, DeSordi MR, Boise L, Fagnan LJ, et al. A qualitative study of rural primary care clinician views on remote monitoring technologies. J Rural Health 2014;30(1):69-78 [FREE Full text] [doi: 10.1111/jrh.12027] [Medline: 24383486]

88. Koru G, Alhuwail D, Topaz M, Norcio AF, Mills ME. Investigating the challenges and opportunities in home care to facilitate effective information technology adoption. J Am Med Dir Assoc 2016 Jan;17(1):53-58. [doi: 10.1016/j.jamda.2015.10.008] [Medline: 26612483]

89. Loscertales FR, Rubió FS. [Facilitators in the implantation of telemedicine services. Perspective of professionals involved in its design and implementation]. An Sist Sanit Navar 2011;34(2):235-244 [FREE Full text] [Medline: 21904405]

90. Menachemi N, Brooks RG, Schwalenstocker E, Simpson L. Use of health information technology by children's hospitals in the United States. Pediatrics 2009 Jan;123 Suppl 2:S80-S84. [doi: 10.1542/peds.2008-1755F] [Medline: 19088234]

91. Randell R, Backhouse MR, Nelson EA. Videoconferencing for site initiations in clinical studies: mixed methods evaluation of usability, acceptability, and impact on recruitment. Inform Health Soc Care 2016 Dec;41(4):362-372. [doi: 10.3109/17538157.2015.1064424] [Medline: 26694583]

92. Rossos PG, St-Cyr O, Purdy B, Toenjes C, Masino C, Chmelnitsky D. Hype, harmony and human factors: applying user-centered design to achieve sustainable telehealth program adoption and growth. Stud Health Technol Inform 2015;209:121-127. [Medline: 25980714]

93. MacNeill V, Sanders C, Fitzpatrick R, Hendy J, Barlow J, Knapp M, et al. Experiences of front-line health professionals in the delivery of telehealth: a qualitative study. Br J Gen Pract 2014 Jul;64(624):e401-e407 [FREE Full text] [doi: 10.3399/bjgp14X680485] [Medline: 24982492]

94. Biggs K, Lowe P, Walsh J, Lagios K. Audit of a sexual health website email link for general practitioners. Int J STD AIDS 2010 Jan;21(1):60-62. [doi: 10.1258/ijsa.2009.009192] [Medline: 20029066]

95. Chung C, Cook J, Bales E, Zia J, Munson SA. More than telemonitoring: health provider use and nonuse of life-log data in irritable bowel syndrome and weight management. J Med Internet Res 2015;17(8):e203 [FREE Full text] [doi: 10.2196/jmir.4364] [Medline: 26297627]

96. de Grood C, Eso K, Santana MJ. Physicians' experience adopting the electronic transfer of care communication tool: barriers and opportunities. J Multidiscip Healthc 2015;8:21-31 [FREE Full text] [doi: 10.2147/JMDH.S72953] [Medline: 25609977]

97. Lamothe L, Fortin J, Labbé F, Gagnon M, Messikh D. Impacts of telehomecare on patients, providers, and organizations. Telemed J E Health 2006 Jun;12(3):363-369. [doi: 10.1089/tmj.2006.12.363] [Medline: 16796505]

98. McCullough JM, Zimmerman FJ, Bell DS, Rodriguez HP. Electronic health information exchange in underserved settings: examining initiatives in small physician practices & community health centers. BMC Health Serv Res 2014;14:415 [FREE Full text] [doi: 10.1186/1472-6963-14-415] [Medline: 25240718]

99. Schooley B, Walczak S, Hikmet N, Patel N. Impacts of mobile tablet computing on provider productivity, communications, and the process of care. Int J Med Inform 2016 Apr;88:62-70. [doi: 10.1016/j.ijmedinf.2016.01.010] [Medline: 26878764]

100. Singh R, Mathiassen L, Stachura ME, Astapova EV. Sustainable rural telehealth innovation: a public health case study. Health Serv Res 2010 Aug;45(4):985-1004 [FREE Full text] [doi: 10.1111/j.1475-6773.2010.01116.x] [Medline: 20459449]

101. Smith KL, Kerr DA, Fenner AA, Straker LM. Adolescents just do not know what they want: a qualitative study to describe obese adolescents' experiences of text messaging to support behavior change maintenance post intervention. J Med Internet Res 2014;16(4):e103 [FREE Full text] [doi: 10.2196/jmir.3113] [Medline: 24713407]

102. Villalba E, Casas I, Abadie F, Lluch M. Integrated personal health and care services deployment: experiences in eight European countries. Int J Med Inform 2013 Jul;82(7):626-635. [doi: 10.1016/j.ijmedinf.2013.03.002] [Medline: 23587432]

103. Barrett M, Larson A, Carville K, Ellis I. Challenges faced in implementation of a telehealth enabled chronic wound care system. Rural Remote Health 2009;9(3):1154 [FREE Full text] [Medline: 19705955]

104. Roig F, Saigí F. Barriers to the normalization of telemedicine in a healthcare system model based on purchasing of healthcare services using providers' contracts. Gac Sanit 2011;25(5):397-402 [FREE Full text] [doi: 10.1016/j.gaceta.2011.01.004] [Medline: 21719160]

105. Sørensen T, Dyb K, Rygh E, Salvesen R, Thomassen L. A qualitative description of telemedicine for acute stroke care in Norway: technology is not the issue. BMC Health Serv Res 2014 Dec 19;14:643 [FREE Full text] [doi: 10.1186/s12913-014-0643-9] [Medline: 25523241]

106. Embi PJ, Weir C, Efthimiadis EN, Thielke SM, Hedeen AN, Hammond KW. Computerized provider documentation: findings and implications of a multisite study of clinicians and administrators. J Am Med Inform Assoc 2013;20(4):718-726 [FREE Full text] [doi: 10.1136/amiajnl-2012-000946] [Medline: 23355462]

107. Fairbrother P, Ure J, Hanley J, McCloughan L, Denvir M, Sheikh A, et al. Telemonitoring for chronic heart failure: the views of patients and healthcare professionals - a qualitative study. J Clin Nurs 2014 Jan;23(1-2):132-144. [doi: 10.1111/jocn.12137] [Medline: 23451899]

108. Foldy S. Inventory of electronic health information exchange in Wisconsin, 2006. WMJ 2007 May;106(3):120-125 [FREE Full text] [Medline: 17642349]

109. Levin D. MHealth: promise and pitfalls. Front Health Serv Manage 2012;29(2):33-9; discussion 40. [Medline: 23330301]

110. Park H, Chon Y, Lee J, Choi I, Yoon KH. Service design attributes affecting diabetic patient preferences of telemedicine in South Korea. Telemed J E Health 2011;17(6):442-451 [FREE Full text] [doi: 10.1089/tmj.2010.0201] [Medline: 21631382]

111. Fuji KT, Abbott AA, Galt KA. A qualitative study of how patients with type 2 diabetes use an electronic stand-alone personal health record. Telemed J E Health 2015 Apr;21(4):296-300. [doi: 10.1089/tmj.2014.0084] [Medline: 25614996]

112. Goldberg EM, Levy PD. New approaches to evaluating and monitoring blood pressure. Curr Hypertens Rep 2016 Dec;18(6):49. [doi: 10.1007/s11906-016-0650-9] [Medline: 27137524]

113. Glick G, Druss B, Pina J, Lally C, Conde M. Use of mobile technology in a community mental health setting. J Telemed Telecare 2016 Oct;22(7):430-435. [doi: 10.1177/1357633X15613236] [Medline: 26519378]

114. 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]

115. Jamoom EW, Patel V, Furukawa MF, King J. EHR adopters vs. non-adopters: impacts of, barriers to, and federal initiatives for EHR adoption. Healthc (Amst) 2014 Mar;2(1):33-39 [FREE Full text] [doi: 10.1016/j.hjdsi.2013.12.004] [Medline: 26250087]

116. Minichiello V, Rahman S, Dune T, Scott J, Dowsett G. E-health: potential benefits and challenges in providing and accessing sexual health services. BMC Public Health 2013;13:790 [FREE Full text] [doi: 10.1186/1471-2458-13-790] [Medline: 23987137]

117. Cason J, Behl D, Ringwalt S. Overview of states' use of telehealth for the delivery of early intervention (IDEA Part C) services. Int J Telerehabil 2012;4(2):39-46 [FREE Full text] [doi: 10.5195/IJT.2012.6105] [Medline: 25945202]

118. Lee SI, Park H, Kim JW, Hwang H, Cho EY, Kim Y, et al. Physicians' perceptions and use of a health information exchange: a pilot program in South Korea. Telemed J E Health 2012 Oct;18(8):604-612 [FREE Full text] [doi: 10.1089/tmj.2011.0261] [Medline: 22352898]

119. Leung GM, Yu PL, Wong IO, Johnston JM, Tin KY. Incentives and barriers that influence clinical computerization in Hong Kong: a population-based physician survey. J Am Med Inform Assoc 2003;10(2):201-212 [FREE Full text] [Medline: 12595409]

120. Mattocks K, Lalime K, Tate JP, Giannotti TE, Carr K, Carrabba A, et al. The state of physician office-based health information technology in Connecticut: current use, barriers and future plans. Conn Med 2007 Jan;71(1):27-31. [Medline: 17288104]

121. May CR, Finch TL, Cornford J, Exley C, Gately C, Kirk S, et al. Integrating telecare for chronic disease management in the community: what needs to be done? BMC Health Serv Res 2011;11:131 [FREE Full text] [doi: 10.1186/1472-6963-11-131] [Medline: 21619596]

122. Molfenter T, Boyle M, Holloway D, Zwick J. Trends in telemedicine use in addiction treatment. Addict Sci Clin Pract 2015 May 28;10:14 [FREE Full text] [doi: 10.1186/s13722-015-0035-4] [Medline: 26016484]

123. Woodward A, Fyfe M, Handuleh J, Patel P, Godman B, Leather A, et al. Diffusion of e-health innovations in 'post-conflict' settings: a qualitative study on the personal experiences of health workers. Hum Resour Health 2014 Apr 23;12:22 [FREE Full text] [doi: 10.1186/1478-4491-12-22] [Medline: 24754997]

124. Al-Qirim N. Championing telemedicine adoption and utilization in healthcare organizations in New Zealand. Int J Med Inform 2007 Jan;76(1):42-54. [doi: 10.1016/j.ijmedinf.2006.02.001] [Medline: 16621682]

125. Alaboudi A, Atkins A, Sharp B, Balkhair A, Alzahrani M, Sunbul T. Barriers and challenges in adopting Saudi telemedicine network: the perceptions of decision makers of healthcare facilities in Saudi Arabia. J Infect Public Health 2016;9(6):725-733 [FREE Full text] [doi: 10.1016/j.jiph.2016.09.001] [Medline: 27649882]

126. Chronaki CE, Vardas P. Remote monitoring costs, benefits, and reimbursement: a European perspective. Europace 2013 Jun;15 Suppl 1:i59-i64. [doi: 10.1093/europace/eut110] [Medline: 23737233]

127. Coustasse A, Deslich S, Bailey D, Hairston A, Paul D. A business case for tele-intensive care units. Perm J 2014;18(4):76-84 [FREE Full text] [doi: 10.7812/TPP/14-004] [Medline: 25662529]

128. Cusack CM, Pan E, Hook JM, Vincent A, Kaelber DC, Middleton B. The value proposition in the widespread use of telehealth. J Telemed Telecare 2008;14(4):167-168. [doi: 10.1258/jtt.2007.007043] [Medline: 18534947]

129. DeWorsop D, Creatura G, Bluez G, Thurnauer H, Forselius-Bielen K, Ranganathan M, et al. Feasibility and success of cell-phone assisted remote observation of medication adherence (CAROMA) in clinical trials. Drug Alcohol Depend 2016 Dec 01;163:24-30. [doi: 10.1016/j.drugalcdep.2016.02.045] [Medline: 27068252]

130. Gabriel MH, Jones EB, Samy L, King J. Progress and challenges: implementation and use of health information technology among critical-access hospitals. Health Aff (Millwood) 2014 Jul;33(7):1262-1270. [doi: 10.1377/hlthaff.2014.0279] [Medline: 25006155]

131. Pan E, Cusack C, Hook J, Vincent A, Kaelber DC, Bates DW, et al. The value of provider-to-provider telehealth. Telemed J E Health 2008 Jun;14(5):446-453. [doi: 10.1089/tmj.2008.0017] [Medline: 18578679]

132. Rozenblum R, Jang Y, Zimlichman E, Salzberg C, Tamblyn M, Buckeridge D, et al. A qualitative study of Canada's experience with the implementation of electronic health information technology. CMAJ 2011 Mar 22;183(5):E281-E288 [FREE Full text] [doi: 10.1503/cmaj.100856] [Medline: 21343262]

133. Lee ACW, Billings M. Telehealth implementation in a skilled nursing facility: case report for physical therapist practice in Washington. Phys Ther 2016 Feb;96(2):252-259. [doi: 10.2522/ptj.20150079] [Medline: 26658151]

134. Abramson EL, Silver M, Kaushal R. Meaningful use status and participation in health information exchange among New York State hospitals: a longitudinal assessment. Jt Comm J Qual Patient Saf 2014 Oct;40(10):452-453. [Medline: 26111305]

135. Achey M, Aldred JL, Aljehani N, Bloem BR, Biglan KM, Chan P, International Parkinson and Movement Disorder Society Telemedicine Task Force. The past, present, and future of telemedicine for Parkinson's disease. Mov Disord 2014 Jun;29(7):871-883. [doi: 10.1002/mds.25903] [Medline: 24838316]

136. Al-Qirim NAY. Teledermatology: the case of adoption and diffusion of telemedicine health Waikato in New Zealand. Telemed J E Health 2003;9(2):167-177. [doi: 10.1089/153056203766437507] [Medline: 12855039]

137. Ayatollahi H, Mirani N, Haghani H. Electronic health records: what are the most important barriers? Perspect Health Inf Manag 2014;11:1c [FREE Full text] [Medline: 25593569]

138. Burke BL, Hall RW, Section on Telehealth Care. Telemedicine: pediatric applications. Pediatrics 2015 Jul;136(1):e293-e308 [FREE Full text] [doi: 10.1542/peds.2015-1517] [Medline: 26122813]

139. Driessen J, Castle NG, Handler SM. Perceived benefits, barriers, and drivers of telemedicine from the perspective of skilled nursing facility administrative staff stakeholders. J Appl Gerontol 2018 Jan;37(1):110-120. [doi: 10.1177/0733464816651884] [Medline: 27269289]

140. O'Toole LJ, Slade CP, Brewer GA, Gase LN. Barriers and facilitators to implementing primary stroke center policy in the United States: results from 4 case study states. Am J Public Health 2011 Mar;101(3):561-566. [doi: 10.2105/AJPH.2010.197954] [Medline: 21233430]

141. de Bustos EM, Vuillier F, Chavot D, Moulin T. Telemedicine in stroke: organizing a network--rationale and baseline principles. Cerebrovasc Dis 2009;27 Suppl 4:1-8. [doi: 10.1159/000213052] [Medline: 19546535]

142. Kulcsar M, Gilchrist S, George MG. Improving stroke outcomes in rural areas through telestroke programs: an examination of barriers, facilitators, and state policies. Telemed J E Health 2014 Jan;20(1):3-10. [doi: 10.1089/tmj.2013.0048] [Medline: 24286197]

143. Rogove HJ, McArthur D, Demaerschalk BM, Vespa PM. Barriers to telemedicine: survey of current users in acute care units. Telemed J E Health 2012;18(1):48-53. [doi: 10.1089/tmj.2011.0071] [Medline: 22082107]

XSL•FO

RenderX

144. Cho I, Kim J, Kim JH, Kim HY, Kim Y. Design and implementation of a standards-based interoperable clinical decision support architecture in the context of the Korean EHR. Int J Med Inform 2010 Sep;79(9):611-622. [doi: 10.1016/j.ijmedinf.2010.06.002] [Medline: 20620098]

145. Alajlani M, Clarke M. Effect of culture on acceptance of telemedicine in Middle Eastern countries: case study of Jordan and Syria. Telemed J E Health 2013 Apr;19(4):305-311. [doi: 10.1089/tmj.2012.0106] [Medline: 23540280]

146. Adler-Milstein J, Sarma N, Woskie LR, Jha AK. A comparison of how four countries use health IT to support care for people with chronic conditions. Health Aff (Millwood) 2014 Sep;33(9):1559-1566. [doi: 10.1377/hlthaff.2014.0424] [Medline: 25201660]

147. Mettler T. Anticipating mismatches of HIT investments: Developing a viability-fit model for e-health services. Int J Med Inform 2016 Jan;85(1):104-115. [doi: 10.1016/j.ijmedinf.2015.10.002] [Medline: 26526279]

148. Stevenson P, Finnane AR, Soyer HP. Teledermatology and clinical photography: safeguarding patient privacy and mitigating medico-legal risk. Med J Aust 2016 Mar 21;204(5):198-200e1. [Medline: 26985853]

149. Mozer R, Bradford NK, Caffery LJ, Smith AC. Identifying perceived barriers to videoconferencing by rehabilitation medicine providers. J Telemed Telecare 2015 Dec;21(8):479-484. [doi: 10.1177/1357633X15607136] [Medline: 26556061]

150. Alkhaldi B, Sahama T, Huxley C, Gajanayake R. Barriers to implementing eHealth: a multi-dimensional perspective. Stud Health Technol Inform 2014;205:875-879. [Medline: 25160313]

151. Alsulame K, Khalifa M, Househ M. eHealth in Saudi Arabia: current trends, challenges and recommendations. Stud Health Technol Inform 2015;213:233-236. [Medline: 26153002]

152. Detmer D, Bloomrosen M, Raymond B, Tang P. Integrated personal health records: transformative tools for consumer-centric care. BMC Med Inform Decis Mak 2008;8:45 [FREE Full text] [doi: 10.1186/1472-6947-8-45] [Medline: 18837999]

153. Dixon BE, Simonaitis L, Goldberg HS, Paterno MD, Schaeffer M, Hongsermeier T, et al. A pilot study of distributed knowledge management and clinical decision support in the cloud. Artif Intell Med 2013 Sep;59(1):45-53. [doi: 10.1016/j.artmed.2013.03.004] [Medline: 23545327]

154. May C, Harrison R, Finch T, MacFarlane A, Mair F, Wallace P, Telemedicine Adoption Study Group. Understanding the normalization of telemedicine services through qualitative evaluation. J Am Med Inform Assoc 2003;10(6):596-604 [FREE Full text] [doi: 10.1197/jamia.M1145] [Medline: 12925553]

155. Owens DK. Use of medical informatics to implement and develop clinical practice guidelines. West J Med 1998 Mar;168(3):166-175 [FREE Full text] [Medline: 9549415]

156. Patel RN, Antonarakis GS. Factors influencing the adoption and implementation of teledentistry in the UK, with a focus on orthodontics. Community Dent Oral Epidemiol 2013 Oct;41(5):424-431. [doi: 10.1111/cdoe.12029] [Medline: 23240902]

157. Dinesen B, Nonnecke B, Lindeman D, Toft E, Kidholm K, Jethwani K, et al. Personalized telehealth in the future: a global research agenda. J Med Internet Res 2016 Mar 01;18(3):e53 [FREE Full text] [doi: 10.2196/jmir.5257] [Medline: 26932229]

158. Hawker F, Kavanagh S, Yellowlees P, Kalucy RS. Telepsychiatry in South Australia. J Telemed Telecare 1998;4(4):187-194. [doi: 10.1258/1357633981932181] [Medline: 10505352]

159. Lanzola G, Toffanin C, Di Palma F, Del Favero S, Magni L, Bellazzi R. Designing an artificial pancreas architecture: the AP@home experience. Med Biol Eng Comput 2015 Dec;53(12):1271-1283. [doi: 10.1007/s11517-014-1231-1] [Medline: 25430423]

160. May C. Mobilising modern facts: health technology assessment and the politics of evidence. Sociol Health Illn 2006 Jul;28(5):513-532 [FREE Full text] [doi: 10.1111/j.1467-9566.2006.00505.x] [Medline: 16910945]

161. Zelmer J, Hagens S. Understanding the gap between desire for and use of consumer health solutions. Healthc Pap 2014;13(4):9-21. [Medline: 25148119]

162. Cook DJ, Doolittle GC, Ferguson D, Hanks N, Hood J, Howard M, et al. Explaining the adoption of telemedicine services: an analysis of a paediatric telemedicine service. J Telemed Telecare 2002;8 Suppl 2:106-107. [doi: 10.1177/1357633X020080S248] [Medline: 12217159]

163. Menachemi N, Burke DE, Ayers DJ. Factors affecting the adoption of telemedicine--a multiple adopter perspective. J Med Syst 2004 Dec;28(6):617-632. [Medline: 15615290]

164. Kierkegaard P. Interoperability after deployment: persistent challenges and regional strategies in Denmark. Int J Qual Health Care 2015 Apr;27(2):147-153. [doi: 10.1093/intqhc/mzv009] [Medline: 25699549]

165. Doupi P. Evolving health IT systems evaluation: the convergence of health informatics and HTA. Stud Health Technol Inform 2016;222:220-236. [Medline: 27198105]

166. Gagnon M, Duplantie J, Fortin J, Landry R. Implementing telehealth to support medical practice in rural/remote regions: what are the conditions for success? Implement Sci 2006 Aug 24;1:18 [FREE Full text] [doi: 10.1186/1748-5908-1-18] [Medline: 16930484]

167. LeRouge C, Garfield MJ. Crossing the telemedicine chasm: have the U.S. barriers to widespread adoption of telemedicine been significantly reduced? Int J Environ Res Public Health 2013 Nov 28;10(12):6472-6484 [FREE Full text] [doi: 10.3390/ijerph10126472] [Medline: 24287864]

168. Varsi C, Ekstedt M, Gammon D, Ruland CM. Using the consolidated framework for implementation research to identify barriers and facilitators for the implementation of an internet-based patient-provider communication service in five settings:

a qualitative study. J Med Internet Res 2015 Nov 18;17(11):e262 [FREE Full text] [doi: 10.2196/jmir.5091] [Medline: 26582138]

169. Charani E, Kyratsis Y, Lawson W, Wickens H, Brannigan ET, Moore LSP, et al. An analysis of the development and implementation of a smartphone application for the delivery of antimicrobial prescribing policy: lessons learnt. J Antimicrob Chemother 2013 Apr;68(4):960-967 [FREE Full text] [doi: 10.1093/jac/dks492] [Medline: 23258314]

170. Bouamrane M, Mair FS. Implementation of an integrated preoperative care pathway and regional electronic clinical portal for preoperative assessment. BMC Med Inform Decis Mak 2014 Nov 19;14:93 [FREE Full text] [doi: 10.1186/1472-6947-14-93] [Medline: 25407812]

171. Bullard MJ, Emond SD, Graham TAD, Ho K, Holroyd BR. Informatics and knowledge translation. Acad Emerg Med 2007 Nov;14(11):996-1002 [FREE Full text] [doi: 10.1197/j.aem.2007.06.032] [Medline: 17967961]

172. Umali E, McCool J, Whittaker R. Possibilities and expectations for mHealth in the Pacific Islands: insights from key informants. JMIR Mhealth Uhealth 2016 Jan;4(1):e9 [FREE Full text] [doi: 10.2196/mhealth.4626] [Medline: 26792386]

173. Hernández V, Blanquer I. The Grid as a healthcare provision tool. Methods Inf Med 2005;44(2):144-148. [Medline: 15924164]

174. Sweet LE, Moulaison HL. Electronic health records data and metadata: challenges for big data in the United States. Big Data 2013 Dec;1(4):245-251. [doi: 10.1089/big.2013.0023] [Medline: 27447257]

175. Wake DJ, He J, Czesak AM, Mughal F, Cunningham SG. MyDiabetesMyWay: an evolving national data driven diabetes self-management platform. J Diabetes Sci Technol 2016 Dec;10(5):1050-1058 [FREE Full text] [doi: 10.1177/1932296816648168] [Medline: 27162192]

176. Bond CS. Telehealth as a tool for independent self-management by people living with long term conditions. Stud Health Technol Inform 2014;206:1-6. [Medline: 25365666]

177. Bond CS, Worswick L. Self management and telehealth: lessons learnt from the evaluation of a Dorset telehealth program. Patient 2015 Aug;8(4):311-316. [doi: 10.1007/s40271-014-0091-y] [Medline: 25315192]

178. Camerini L, Camerini A, Schulz PJ. Do participation and personalization matter? A model-driven evaluation of an Internet-based patient education intervention for fibromyalgia patients. Patient Educ Couns 2013 Aug;92(2):229-234. [doi: 10.1016/j.pec.2013.04.007] [Medline: 23664234]

179. Fergie G, Hunt K, Hilton S. Social media as a space for support: young adults' perspectives on producing and consuming user-generated content about diabetes and mental health. Soc Sci Med 2016 Dec;170:46-54 [FREE Full text] [doi: 10.1016/j.socscimed.2016.10.006] [Medline: 27750067]

180. Griebel L, Sedlmayr B, Prokosch H, Criegee-Rieck M, Sedlmayr M. Key factors for a successful implementation of personalized e-health services. Stud Health Technol Inform 2013;192:965. [Medline: 23920739]

181. Schwarz F, Ward J, Willcock S. E-Health readiness in outback communities: an exploratory study. Rural Remote Health 2014;14(3):2871 [FREE Full text] [Medline: 25190566]

182. Tran BQ, Buckley KM, Bertera EM, Gonzales PL. Benefits & barriers to adoption of health IT in an elderly low-income, minority community-based environment. Conf Proc IEEE Eng Med Biol Soc 2009;2009:6441-6443. [doi: 10.1109/IEMBS.2009.5333537] [Medline: 19964428]

183. Duman-Lubberding S, van Uden-Kraan CF, Peek N, Cuijpers P, Leemans CR, Verdonck-de Leeuw IM. An eHealth application in head and neck cancer survivorship care: health care professionals' perspectives. J Med Internet Res 2015 Oct 21;17(10):e235 [FREE Full text] [doi: 10.2196/jmir.4870] [Medline: 26489918]

184. Peng W, Yuan S, Holtz BE. Exploring the challenges and opportunities of health mobile apps for individuals with type 2 diabetes living in rural communities. Telemed J E Health 2016 Sep;22(9):733-738. [doi: 10.1089/tmj.2015.0180] [Medline: 26982017]

185. Ronda MCM, Dijkhorst-Oei L, Rutten GEHM. Reasons and barriers for using a patient portal: survey among patients with diabetes mellitus. J Med Internet Res 2014;16(11):e263 [FREE Full text] [doi: 10.2196/jmir.3457] [Medline: 25424228]

186. Sanders C, Rogers A, Bowen R, Bower P, Hirani S, Cartwright M, et al. Exploring barriers to participation and adoption of telehealth and telecare within the Whole System Demonstrator trial: a qualitative study. BMC Health Serv Res 2012;12:220 [FREE Full text] [doi: 10.1186/1472-6963-12-220] [Medline: 22834978]

187. Vriend I, Coehoorn I, Verhagen E. Implementation of an app-based neuromuscular training programme to prevent ankle sprains: a process evaluation using the RE-AIM Framework. Br J Sports Med 2015 Apr;49(7):484-488. [doi: 10.1136/bjsports-2013-092896] [Medline: 24470587]

188. Hanna L, May C, Fairhurst K. Non-face-to-face consultations and communications in primary care: the role and perspective of general practice managers in Scotland. Inform Prim Care 2011;19(1):17-24 [FREE Full text] [Medline: 22118332]

189. 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;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]

190. Gagnon M, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]

191. Harvey J, Dopson S, McManus RJ, Powell J. Factors influencing the adoption of self-management solutions: an interpretive synthesis of the literature on stakeholder experiences. Implement Sci 2015 Nov 13;10:159 [FREE Full text] [doi: 10.1186/s13012-015-0350-x] [Medline: 26566623]

192. Zhang X, Yu P, Yan J, Ton AMSI. Using diffusion of innovation theory to understand the factors impacting patient acceptance and use of consumer e-health innovations: a case study in a primary care clinic. BMC Health Serv Res 2015 Feb 21;15:71 [FREE Full text] [doi: 10.1186/s12913-015-0726-2] [Medline: 25885110]

193. Lober WB, Quiles C, Wagner S, Cassagnol R, Lamothes R, Alexis DRP, et al. Three years experience with the implementation of a networked electronic medical record in Haiti. AMIA Annu Symp Proc 2008 Nov 06:434-438 [FREE Full text] [Medline: 18999283]

194. Mitchell JG. The uneven diffusion of telemedicine services in Australia. J Telemed Telecare 1999;5 Suppl 1:S45-S47. [Medline: 10534838]

195. Taylor J, Coates L. Caring from a distance: the role of telehealth. Nurs Times 2015;111(28-29):18-20. [Medline: 26477231]

196. Moxham C, Chambers N, Girling J, Garg S, Jelfs E, Bremner J. Perspectives on the enablers of e-heath adoption: an international interview study of leading practitioners. Health Serv Manage Res 2012 Aug;25(3):129-137. [doi: 10.1258/hsmr.2012.012018] [Medline: 23135887]

197. Paré G, Trudel M. Knowledge barriers to PACS adoption and implementation in hospitals. Int J Med Inform 2007 Jan;76(1):22-33. [doi: 10.1016/j.ijmedinf.2006.01.004] [Medline: 16478675]

198. Vimarlund V, Le Rouge C. Barriers and opportunities to the widespread adoption of telemedicine: a bi-country evaluation. Stud Health Technol Inform 2013;192:933. [Medline: 23920707]

199. Nicholas J, Boydell K, Christensen H. mHealth in psychiatry: time for methodological change. Evid Based Ment Health 2016 Dec;19(2):33-34. [doi: 10.1136/eb-2015-102278] [Medline: 27044849]

200. Meyer R. A bio-medical informatics perspective on human factors: how human factors influence adoption of healthcare information technology. Yearb Med Inform 2010:21-24. [Medline: 20938565]

201. Murray E. Web-based interventions for behavior change and self-management: potential, pitfalls, and progress. Med 2 0 2012;1(2):e3 [FREE Full text] [doi: 10.2196/med20.1741] [Medline: 25075231]

202. Cimperman M, Brenčič MM, Trkman P, Stanonik MDL. Older adults' perceptions of home telehealth services. Telemed J E Health 2013 Oct;19(10):786-790 [FREE Full text] [doi: 10.1089/tmj.2012.0272] [Medline: 23931702]

203. de Tolly KM, Constant D. Integrating mobile phones into medical abortion provision: intervention development, use, and lessons learned from a randomized controlled trial. JMIR Mhealth Uhealth 2014;2(1):e5 [FREE Full text] [doi: 10.2196/mhealth.3165] [Medline: 25098569]

204. van Deursen AJ, van Dijk JA. Internet skills performance tests: are people ready for eHealth? J Med Internet Res 2011 Apr;13(2):e35 [FREE Full text] [doi: 10.2196/jmir.1581] [Medline: 21531690]

205. Bielli E, Carminati F, La Capra S, Lina M, Brunelli C, Tamburini M. A wireless health outcomes monitoring system (WHOMS): development and field testing with cancer patients using mobile phones. BMC Med Inform Decis Mak 2004 Jun 15;4:7 [FREE Full text] [doi: 10.1186/1472-6947-4-7] [Medline: 15196308]

206. Young R, Willis E, Cameron G, Geana M. "Willing but unwilling": attitudinal barriers to adoption of home-based health information technology among older adults. Health Informatics J 2014 Jun;20(2):127-135. [doi: 10.1177/1460458213486906] [Medline: 24056750]

207. Armer JM, Harris K, Dusold JM. Application of the concerns-based adoption model to the installation of telemedicine in a rural Missouri nursing home. J Nurses Staff Dev 2004;20(1):42-49. [Medline: 15076128]

208. Plaete J, Crombez G, DeSmet A, Deveugele M, Verloigne M, De Bourdeaudhuij I. What do general practitioners think about an online self-regulation programme for health promotion? Focus group interviews. BMC Fam Pract 2015;16(1):3 [FREE Full text] [doi: 10.1186/s12875-014-0214-5] [Medline: 25608851]

209. Villalba-Mora E, Casas I, Lupiañez-Villanueva F, Maghiros I. Adoption of health information technologies by physicians for clinical practice: the Andalusian case. Int J Med Inform 2015 Jul;84(7):477-485. [doi: 10.1016/j.ijmedinf.2015.03.002] [Medline: 25823578]

210. DeLuca JM, Enmark R. E-health: the changing model of healthcare. Front Health Serv Manage 2000;17(1):3-15. [Medline: 11184427]

211. Baumel A, Muench F. Heuristic evaluation of Ehealth interventions: establishing standards that relate to the therapeutic process perspective. JMIR Ment Health 2016 Jan 13;3(1):e5 [FREE Full text] [doi: 10.2196/mental.4563] [Medline: 26764209]

212. Hsieh H, Tsai C, Chih W, Lin H. Factors affecting success of an integrated community-based telehealth system. Technol Health Care 2015;23 Suppl 2:S189-S196. [doi: 10.3233/THC-150953] [Medline: 26410483]

213. Quinlivan JA, Lyons S, Petersen RW. Attitudes of pregnant women towards personally controlled electronic, hospital-held, and patient-held medical record systems: a survey study. Telemed J E Health 2014 Sep;20(9):810-815. [doi: 10.1089/tmj.2013.0342] [Medline: 25046543]

214. Bruns NE, Irtan S, Rothenberg SS, Bogen EM, Kotobi H, Ponsky TA. Trans-Atlantic telementoring with pediatric surgeons: technical considerations and lessons learned. J Laparoendosc Adv Surg Tech A 2016 Jan;26(1):75-78. [doi: 10.1089/lap.2015.0131] [Medline: 26698191]

215. Mars M. Health capacity development through telemedicine in Africa. Yearb Med Inform 2010:87-93. [Medline: 20938578]

216. Ruxwana N, Herselman M, Pottas D. A Generic Quality Assurance Model (GQAM) for successful e-health implementation in rural hospitals in South Africa. Health Inf Manag 2014;43(1):26-36. [Medline: 27010686]

217. Raisinghani MS, Young E. Personal health records: key adoption issues and implications for management. Int J Electron Healthc 2008;4(1):67-77. [doi: 10.1504/IJEH.2008.018921] [Medline: 18583296]

218. George S, Hamilton A, Baker RS. How do low-income urban African Americans and Latinos feel about telemedicine? A diffusion of innovation analysis. Int J Telemed Appl 2012;2012:715194 [FREE Full text] [doi: 10.1155/2012/715194] [Medline: 22997511]

219. Anvari M. Robot-assisted remote telepresence surgery. Semin Laparosc Surg 2004 Jun;11(2):123-128. [Medline: 15254651]

220. Gorini A, Gaggioli A, Vigna C, Riva G. A second life for eHealth: prospects for the use of 3-D virtual worlds in clinical psychology. J Med Internet Res 2008;10(3):e21 [FREE Full text] [doi: 10.2196/jmir.1029] [Medline: 18678557]

221. Obstfelder A, Engeseth KH, Wynn R. Characteristics of successfully implemented telemedical applications. Implement Sci 2007 Jul 27;2:25 [FREE Full text] [doi: 10.1186/1748-5908-2-25] [Medline: 17662134]

222. 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]

223. Dadd M, Doyle B, Wilson L, Gunaratnam M. Lessons learned from the Hospital Without Walls project. J Telemed Telecare 2002;8 Suppl 3:S3:11-S3:14. [Medline: 12661607]

224. Cocosila M, Archer N. Perceptions of chronically ill and healthy consumers about electronic personal health records: a comparative empirical investigation. BMJ Open 2014 Jul 23;4(7):e005304 [FREE Full text] [doi: 10.1136/bmjopen-2014-005304] [Medline: 25056975]

225. Courtney KL. Privacy and senior willingness to adopt smart home information technology in residential care facilities. Methods Inf Med 2008;47(1):76-81. [Medline: 18213432]

226. Barbieri JS, Nelson CA, Bream KD, Kovarik CL. Primary care providers' perceptions of mobile store-and-forward teledermatology. Dermatol Online J 2015 Aug 15;21(8). [Medline: 26437165]

227. Courtney KL, Demiris G, Rantz M, Skubic M. Needing smart home technologies: the perspectives of older adults in continuing care retirement communities. Inform Prim Care 2008;16(3):195-201 [FREE Full text] [Medline: 19094406]

228. Wilkowska W, Ziefle M. Privacy and data security in E-health: requirements from the user's perspective. Health Informatics J 2012 Sep;18(3):191-201. [doi: 10.1177/1460458212442933] [Medline: 23011814]

229. Driessen J, Bonhomme A, Chang W, Nace DA, Kavalieratos D, Perera S, et al. Nursing home provider perceptions of telemedicine for reducing potentially avoidable hospitalizations. J Am Med Dir Assoc 2016 Jun 01;17(6):519-524 [FREE Full text] [doi: 10.1016/j.jamda.2016.02.004] [Medline: 26969534]

230. McGraw D, Dempsey JX, Harris L, Goldman J. Privacy as an enabler, not an impediment: building trust into health information exchange. Health Aff (Millwood) 2009;28(2):416-427 [FREE Full text] [doi: 10.1377/hlthaff.28.2.416] [Medline: 19275998]

231. Zanaboni P, Knarvik U, Wootton R. Adoption of routine telemedicine in Norway: the current picture. Glob Health Action 2014;7:22801 [FREE Full text] [Medline: 24433942]

232. Martin AB, Probst JC, Shah K, Chen Z, Garr D. Differences in readiness between rural hospitals and primary care providers for telemedicine adoption and implementation: findings from a statewide telemedicine survey. J Rural Health 2012 Jan;28(1):8-15. [doi: 10.1111/j.1748-0361.2011.00369.x] [Medline: 22236310]

233. Williams G, Doughty K, Bradley DA. Safety and risk issues in using telecare. J Telemed Telecare 2000;6(5):249-262. [doi: 10.1258/1357633001935833] [Medline: 11070585]

234. Lewis T, Synowiec C, Lagomarsino G, Schweitzer J. E-health in low- and middle-income countries: findings from the Center for Health Market Innovations. Bull World Health Organ 2012 May 1;90(5):332-340 [FREE Full text] [doi: 10.2471/BLT.11.099820] [Medline: 22589566]

235. De Simone A, Leoni L, Luzi M, Amellone C, Stabile G, La Rocca V, et al. Remote monitoring improves outcome after ICD implantation: the clinical efficacy in the management of heart failure (EFFECT) study. Europace 2015 Aug;17(8):1267-1275. [doi: 10.1093/europace/euu318] [Medline: 25842271]

236. Lau AY, Arguel A, Dennis S, Liaw S, Coiera E. "Why Didn't it Work?" Lessons from a randomized controlled trial of a web-based personally controlled health management system for adults with asthma. J Med Internet Res 2015 Dec;17(12):e283 [FREE Full text] [doi: 10.2196/jmir.4734] [Medline: 26678294]

237. Hardisty AR, Peirce SC, Preece A, Bolton CE, Conley EC, Gray WA, et al. Bridging two translation gaps: a new informatics research agenda for telemonitoring of chronic disease. Int J Med Inform 2011 Oct;80(10):734-744. [doi: 10.1016/j.ijmedinf.2011.07.002] [Medline: 21890403]

238. Naik AD, Lawrence B, Kiefer L, Ramos K, Utech A, Masozera N, et al. Building a primary care/research partnership: lessons learned from a telehealth intervention for diabetes and depression. Fam Pract 2015 Apr;32(2):216-223. [doi: 10.1093/fampra/cmu084] [Medline: 25552674]

239. Poole ES. HCI and mobile health interventions: how human-computer interaction can contribute to successful mobile health interventions. Transl Behav Med 2013 Dec;3(4):402-405 [FREE Full text] [doi: 10.1007/s13142-013-0214-3] [Medline: 24294328]

240. Kruse CS, Regier V, Rheinboldt KT. Barriers over time to full implementation of health information exchange in the United States. JMIR Med Inform 2014;2(2):e26 [FREE Full text] [doi: 10.2196/medinform.3625] [Medline: 25600635]

241. Gardner RM. Computerized clinical decision-support in respiratory care. Respir Care 2004 Apr;49(4):378-86; discussion 386 [FREE Full text] [Medline: 15030611]

242. 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 2;330(7494):765 [FREE Full text] [doi: 10.1136/bmj.38398.500764.8F] [Medline: 15767266]

243. Davis MM, Freeman M, Kaye J, Vuckovic N, Buckley DI. A systematic review of clinician and staff views on the acceptability of incorporating remote monitoring technology into primary care. Telemed J E Health 2014 May;20(5):428-438 [FREE Full text] [doi: 10.1089/tmj.2013.0166] [Medline: 24731239]

244. Bowens FM, Frye PA, Jones WA. Health information technology: integration of clinical workflow into meaningful use of electronic health records. Perspect Health Inf Manag 2010 Oct 01;7:1d [FREE Full text] [Medline: 21063545]

245. LeTourneau B. Managing physician resistance to change. J Healthc Manag 2004;49(5):286-288. [Medline: 15499802]

246. Narine L, Persaud DD. Gaining and maintaining commitment to large-scale change in healthcare organizations. Health Serv Manage Res 2003 Aug;16(3):179-187. [doi: 10.1258/095148403322167933] [Medline: 12908992]

247. Appelbaum SH, Wohl L. Transformation or change: some prescriptions for health care organizations. Manag Serv Q 2000 Oct;10(5):279-298. [doi: 10.1108/09604520010345768]

248. Pardo del Val M, Martínez Fuentes C. Resistance to change: a literature review and empirical study. Manag Decis 2003 Mar;41(2):148-155. [doi: 10.1108/00251740310457597]

249. Landaeta RE, Mun JH, Rabadi G, Levin D. Identifying sources of resistance to change in healthcare. Int J Healthc Technol Manag 2008;9(1):74. [doi: 10.1504/IJHTM.2008.016849]

250. Hickson R, Talbert J, Thornbury WC, Perin NR, Goodin AJ. Online medical care: the current state of "eVisits" in acute primary care delivery. Telemed J E Health 2015 Feb;21(2):90-96. [doi: 10.1089/tmj.2014.0022] [Medline: 25474083]

251. Declerck G, Aimé X. Reasons (not) to spend a few billions more on EHRs: how human factors research can help. Yearb Med Inform 2014 Aug 15;9:90-96 [FREE Full text] [doi: 10.15265/IY-2014-0033] [Medline: 25123727]

252. Frauenberger C, Good J, Fitzpatrick G, Iversen OS. In pursuit of rigour and accountability in participatory design. Int J Hum Comput Stud 2015 Feb;74:93-106 [FREE Full text] [doi: 10.1016/j.ijhcs.2014.09.004] [Medline: 26109833]

Abbreviations

MeSH: Medical Subject Heading
PICO: patient, intervention, comparison, and outcome
PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses

XSL•FO

RenderX

<u>Short Paper</u>

Effects of Contributor Experience on the Quality of Health-Related Wikipedia Articles

Peter Holtz[1], Dipl Psych, Dr Phil; Besnik Fetahu[2], MSc, Dr rer nat; Joachim Kimmerle[1], Dipl Psych, Dr rer nat

[1]IWM Leibniz-Institut für Wissensmedien (Knowledge Media Research Center), Knowledge Construction Lab, Tübingen, Germany
[2]L3S Research Center, Leibniz University of Hannover, Hannover, Germany

Corresponding Author:
Peter Holtz, Dipl Psych, Dr Phil
IWM Leibniz-Institut für Wissensmedien (Knowledge Media Research Center)
Knowledge Construction Lab
Schleichstraße 6
Tübingen, 72076
Germany
Phone: 49 7071979303
Fax: 49 7071979105
Email: p.holtz@iwm-tuebingen.de

Abstract

Background: Consulting the Internet for health-related information is a common and widespread phenomenon, and Wikipedia is arguably one of the most important resources for health-related information. Therefore, it is relevant to identify factors that have an impact on the quality of health-related Wikipedia articles.

Objective: In our study we have hypothesized a positive effect of contributor experience on the quality of health-related Wikipedia articles.

Methods: We mined the edit history of all (as of February 2017) 18,805 articles that were listed in the categories on the portal health & fitness in the English language version of Wikipedia. We identified tags within the articles' edit histories, which indicated potential issues with regard to the respective article's quality or neutrality. Of all of the sampled articles, 99 (99/18,805, 0.53%) articles had at some point received at least one such tag. In our analysis we only considered those articles with a minimum of 10 edits (10,265 articles in total; 96 tagged articles, 0.94%). Additionally, to test our hypothesis, we constructed contributor profiles, where a profile consisted of all the articles edited by a contributor and the corresponding number of edits contributed. We did not differentiate between rollbacks and edits with novel content.

Results: Nonparametric Mann-Whitney U-tests indicated a higher number of previously edited articles for editors of the nontagged articles (mean rank tagged 2348.23, mean rank nontagged 5159.29; U=9.25, *P*<.001). However, we did not find a significant difference for the contributors' total number of edits (mean rank tagged 4872.85, mean rank nontagged 5135.48; U=0.87, *P*=.39). Using logistic regression analysis with the respective article's number of edits and number of editors as covariates, only the number of edited articles yielded a significant effect on the article's status as tagged versus nontagged (dummy-coded; Nagelkerke R^2 for the full model=.17; *B* [SE *B*]=-0.001 [0.00]; Wald c^2 [1]=19.70; *P*<.001), whereas we again found no significant effect for the mere number of edits (Nagelkerke R^2 for the full model=.15; *B* [SE *B*]=0.000 [0.01]; Wald c^2 [1]=0.01; *P*=.94).

Conclusions: Our findings indicate an effect of contributor experience on the quality of health-related Wikipedia articles. However, only the number of previously edited articles was a predictor of the articles' quality but not the mere volume of edits. More research is needed to disentangle the different aspects of contributor experience. We have discussed the implications of our findings with respect to ensuring the quality of health-related information in collaborative knowledge-building platforms.

KEYWORDS

XSL•FO
RenderX

Introduction

Health Information Online

Discussions regarding the quality of health-related information on the internet go back as far as the late 1990s [1,2] and have continued recently [3-5]. Consulting the internet for health-related information has undoubtedly become a common and widespread phenomenon [6,7]. Over the last several years, Wikipedia has emerged as one of the most important knowledge resources for health-related information on the Web [4,8,9]. In this paper we (1) describe potential quality issues as indicated by community-applied tags in health-related Wikipedia articles, and (2) analyze the importance of contributor experience for the quality of health-related Wikipedia articles.

Wikipedia as a Resource

Wikipedia relies heavily on peer review to ensure the quality of its collaboratively constructed knowledge corpus, and Wikipedia contributors are expected and invited to correct other contributors' mistakes [9]. Controversial and conflictual issues are to be debated by the contributors until consensus is reached [10,11]. One way of instigating such a discussion is to apply one of several Wikipedia template messages (or *tags*) to an article, which indicate quality-related problems, such as articles being biased, misleading, or factually wrong. In our study, we used the occurrence of at least one quality-issue tag in an article's edit history as a proxy for potential quality problems.

Although Wikipedia is not governed by a body of experts, the quality of health-related Wikipedia articles is not necessarily worse than that of expert-generated internet content [12]. There have been repeated calls for experienced medical professionals to get more actively involved in improving the accuracy of health-related Wikipedia articles [13,14]. Nevertheless, there is empirical evidence that, in spite of all efforts to ensure knowledge quality, Wikipedia articles can be biased; for example, as a consequence of predominantly male Wikipedia authors underreporting or belittling notable achievements of women (*gender bias*) [15]. Another form of bias is the presentation of information regarding Wikipedia authors' own national group appearing in a more positive way than information about other groups (*ingroup bias*) [16]. In line with previous research [17,18], we hypothesize that a certain proportion of such distortions is attributable to a lack of experience on the side of the respective article's contributors.

Methods

In this study we mined the complete edit history (as of February 2017) of the 18,805 articles that were listed within the categories of the Wikipedia portal *health & fitness* [19] in the English language version of Wikipedia. We developed our own publicly available code to mine the data [20]. We first identified the occurrence of tags or template messages within the articles' edit histories that indicated quality issues. These tags were *neutral point of view policy violation* (42 cases in our sample), *contradictory content* (13), *unbalanced content* (12), *confusing content* (17), and *inaccurate content* (23). Tags in the articles' respective talk pages were not taken into account. Tags such as

neutral point of view policy violation and *unbalanced content* refer to violations of Wikipedia's standards of objectivity, which may be caused by social biases such as an ingroup bias or a gender bias, while the other tags are related to quality issues in general. Of all the sampled articles, 99 (99/18,805, 0.53%) articles had at some point received at least one of the aforementioned tags. For the comparison of tagged versus nontagged articles, we only used articles that were comprised of at least 10 edits, although it is imaginable that some elaborated Wikipedia articles result from relatively few comparatively substantial edits. This limitation reduced the total number of articles to 10,265, whereas the total number of occurrences of tagged articles was reduced from 99 to 96 (96/10,265, 0.94%).

For every contributor that authored at least one edit within the sampled articles, we constructed the contributor's profile, consisting of the total number of edits in all Wikipedia articles and the total number of articles that the contributor had edited up to that point. We did not make any distinction with respect to the quality of edits; hence, we treated rollbacks and edits with novel content in the same way. For all statistical analyses, we used the SPSS 22 software package. All reported significance tests are two-sided and we set the significance level at P=.01.

Results

The Content of the Tagged Articles

The 99 tagged articles were manually assigned to one of four different content categories that had emerged in the analysis: *legislation & politics* (41 articles; examples included "abortion in Iran," "free market healthcare," and "smoking ban"), *medicine-related topics* (21 articles; examples included "antimicrobial resistance," "obesity," and "zidovudine"), *alternative medicine-related topics* (19 articles; examples included "astrology and health," "chiropractic," and "siddha medicine"), *and places, people, and events* (18 articles; examples included "2009 flu pandemic in Mexico," "Bethlem Royal Hospital," and "Arnold Schwarzenegger"). For details see Multimedia Appendix 1.

The average number of total edits for these 99 articles was 940.60 (SD 1458.67), and the articles were authored by an average of 186.27 (SD 251.97) individual contributors. In comparison to the 18,706 nontagged articles (mean edits 49.67, SD 194.56; mean editors 16.16, SD 38.26), the tagged articles were comprised of a significantly higher number of edits ($t_{98.02}$=6.10; P<.001; d=1.08), and they were authored by a significantly higher number of contributors ($t_{98.02}$=6.72; P<.001; d=1.17). Part of these differences can be explained by the fact that a substantial number of the nontagged articles were "stubs" which featured nothing more than a mere article title. Such stubs (fewer than ten edits) were omitted from all further analyses. As a consequence, the differences between the tagged (mean edits 969.75, SD 1472.23; mean editors 191.96, SD 253.81) and the nontagged articles (mean edits 88.13, SD 257.66; mean editors 27.30, SD 49.16) comprising 10 or more edits with regard to the average total numbers of edits ($t_{95.06}$=5.87, P<.001, d=1.07) and editors ($t_{95.07}$=6.36, P<.001, d=1.09) could be

reduced to some extent, but the difference still remained significant.

Effects of Contributor Experience

The 10,265 remaining articles had an average of 100.12 edits (SD 311.00) and were authored by an average of 29.79 (SD 57.97) editors. According to their user profiles, the editors of these articles had made on average 32,031.05 (SD 27,513.01) edits in 1,033.42 (SD 648.71) Wikipedia articles. The number of the editors' total edits, as well as the number of edited articles, were positively skewed and were not normally distributed according to Kolmogorov-Smirnov tests (P values <.001). Hence, we used Mann-Whitney U-tests to analyze differences between the authors of the 96 articles that were comprised of at least 10 edits, and that received at least one of the tags indicating quality issues, and those of the nontagged articles with regard to the editors' previous editing activities. We found a significant difference with regard to the total number of edited articles (mean rank tagged 2348.23, mean rank nontagged 5159.29; U=9.25, P<.001) whereas the difference in terms of the total number of edits did not reach statistical significance (mean rank tagged 4872.85, mean rank nontagged 5135.48; U=0.87, P=.39).

To account for the significant differences between the tagged and the nontagged articles comprising 10 or more edits with regard to the average total numbers of edits and editors (see above), we further tested our initial findings using logistic regression analyses (stepwise) with the tagged versus nontagged status (dummy coded) of the articles as the dependent variable, the respective article's numbers of total editors and total edits as control variables, and the contributor's total number of edits and edited articles (respectively) as independent predictors. This approach controlled for a possible obfuscating linear effect of the number of an article's edits or editors. Again we found that the total number of articles that were edited by the contributors significantly predicted the articles' statuses (Nagelkerke R^2 for the full model=.17; B [SE B]=-0.001 [0.00]; Wald c^2[1]=19.70; P<.001), whereas the total number of the contributors' edits did not yield a significant effect (Nagelkerke R^2 for the full model=.15; B [SE B]=0.000 [0.01]; Wald c^2[1]=0.01; P=.94).

Discussion

Principal Results

The largest category of health-related Wikipedia articles that had at one point in their edit history received a user-applied tag indicating quality issues focused on topics about political and legislative issues. The other articles were related to alternative medicine, generic medical topics, and specific events and people.

The authors of the tagged articles had (on average) edited less Wikipedia articles than the authors of the nontagged articles. However, we did not find a significant difference for the mere number of the contributor's previous edits. Assuming that there is a relationship between the authors' competence and the probability that a Wikipedia article receives a tag indicating quality issues, this could indicate that the mere volume of activity is not indicative of a Wikipedia author's competence, but rather a certain breadth of experience. This finding, if corroborated, could have implications for Wikipedia's (and other platforms') editing system as well as for the quality management of collaborative knowledge construction platforms. For example, articles that were authored by editors with a comparatively narrow range of previously edited articles could be automatically identified and marked for further quality checks as a means of ensuring and improving the quality of health-related articles. However, more research is needed to disentangle the effects of the different facets of Wikipedia contributor activities on the quality of Wikipedia articles.

Limitations

A major limiting factor for our study was that relatively few articles received at least one of the tags indicating quality issues. One way of addressing this issue in future studies, to replicate our findings, would be to use quality metrics that are based on article features such as length, the number of paragraphs, and the number of pictures [21]. It should also be noted that the authors of both the tagged and the nontagged articles had (on average) made several thousand edits in several hundred articles, and hence were relatively experienced Wikipedia contributors. Further research is needed to analyze the effects of different levels of editor experience (or inexperience) on the quality of Wikipedia articles. The articles that were sampled for this study only constitute a part of the medical content that is available at Wikipedia [22]. Future studies are needed to replicate our findings for a wider range of health-related Wikipedia articles and articles in languages other than English.

Conclusion

Consistent with previous studies [4,13,21], our findings highlight the potential of Wikipedia as a valuable resource for health-related information. However, the quality of Wikipedia articles relies on the willingness of experienced and knowledgeable contributors to take on the unpaid labor of editing and improving Wikipedia articles. One way of encouraging experienced professionals to further engage with Wikipedia content would be to provide incentives for such activities (eg, in the form of continuing medical education credits) [22].

Conflicts of Interest

None declared.

Multimedia Appendix 1

Overview of the tagged articles.

[PDF File (Adobe PDF File), 53KB - jmir_v20i5e171_app1.pdf]

References

1. Eysenbach G, Diepgen T. Towards quality management of medical information on the Internet: evaluation, labelling, and filtering of information. BMJ 1998 Nov 28;317(7171):1496-1502. [doi: 10.1136/bmj.317.7171.1496]

2. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: caveant lector et viewor--let the reader and viewer beware. JAMA 1997 Apr 16;277(15):1244-1245. [Medline: 9103351]

3. 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. [doi: 10.1097/JCN.0000000000000324] [Medline: 26938508]

4. Fahy E, Hardikar R, Fox A, Mackay S. Quality of patient health information on the Internet: reviewing a complex and evolving landscape. Australas Med J 2014;7(1):24-28 [FREE Full text] [doi: 10.4066/AMJ.2014.1900] [Medline: 24567763]

5. Scholten MR, Kelders SM, Van GJE. Self-guided Web-based interventions: scoping review on user needs and the potential of embodied conversational agents to address them. J Med Internet Res 2017 Nov 16;19(11):e383 [FREE Full text] [doi: 10.2196/jmir.7351] [Medline: 29146567]

6. Song H, Omori K, Kim J, Tenzek KE, Morey HJ, Lin W, et al. Trusting social media as a source of health information: online surveys comparing the United States, Korea, and Hong Kong. J Med Internet Res 2016 Mar 14;18(3):e25 [FREE Full text] [doi: 10.2196/jmir.4193] [Medline: 26976273]

7. Fox S, Duggan M. Pew Internet & American Life Project. Washington, DC; 2013. Health online 2013 URL: http://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2018-04-16] [WebCite Cache ID 6O18Dgbo0]

8. Laurent MR, Vickers TJ. Seeking health information online: does Wikipedia matter? J Am Med Inform Assoc 2009;16(4):471-479 [FREE Full text] [doi: 10.1197/jamia.M3059] [Medline: 19390105]

9. Jirschitzka J, Kimmerle J, Halatchliyski I, Hancke J, Meurers D, Cress U. A productive clash of perspectives? The interplay between articles' and authors' perspectives and their impact on Wikipedia edits in a controversial domain. PLoS One 2017;12(6):e0178985 [FREE Full text] [doi: 10.1371/journal.pone.0178985] [Medline: 28575077]

10. Oeberst A, Halatchliyski I, Kimmerle J, Cress U. Knowledge construction in Wikipedia: a systemic-constructivist analysis. J Learn Sci 2014 Apr 30;23(2):149-176 [FREE Full text] [doi: 10.1080/10508406.2014.888352]

11. Greving H, Oeberst A, Kimmerle J, Cress U. Emotional content in Wikipedia articles on negative man-made and nature-made events. J Lang Soc Psychol 2017 Jun 29:0261927X1771756. [doi: 10.1177/0261927X17717568]

12. Rajagopalan MS, Khanna V, Leiter Y, Stott M, Showalter T, Dicker A, et al. Patient-oriented cancer information on the Internet: a comparison of wikipedia and a professionally maintained database. J Oncol Pract 2011 Sep;7(5):319-323 [FREE Full text] [doi: 10.1200/JOP.2010.000209] [Medline: 22211130]

13. Heilman JM, Kemmann E, Bonert M, Chatterjee A, Ragar B, Beards GM, et al. Wikipedia: a key tool for global public health promotion. J Med Internet Res 2011;13(1):e14 [FREE Full text] [doi: 10.2196/jmir.1589] [Medline: 21282098]

14. Masukume G, Kipersztok L, Das D, Shafee TMA, Laurent MR, Heilman JM. Medical journals and Wikipedia: a global health matter. Lancet Glob Health 2016 Nov;4(11):e791 [FREE Full text] [doi: 10.1016/S2214-109X(16)30254-6] [Medline: 27765289]

15. Wagner C, Garcia D, Jadidi M, Strohmaier M. It's a man's Wikipedia? Assessing gender inequality in an online encyclopedia. 2015 Apr 21 Presented at: 9th International AAAI Conference on Web and Social Media; April 21, 2015; Oxford, UK p. 454-463 URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10585

16. Oeberst A, Cress U, Back M, Nestler S. Individual versus collaborative information processing: the case of biases in Wikipedia. In: Cress U, Moskaliuk J, Jeong H, editors. Mass Collaboration and Education. Cham, CH: Springer International Publishing; 2016:165-185.

17. Kane GC. A multimethod study of information quality in wiki collaboration. ACM Trans Manage Inf Syst 2011 Mar 01;2(1):1-16. [doi: 10.1145/1929916.1929920]

18. Dang QV, Ignat CL. Measuring quality of collaboratively edited documents: the case of Wikipedia. 2016 Presented at: 2016 IEEE 2nd International Conference on Collaboration and Internet Computing (CIC); Nov 2016; Pittsburgh, PA URL: http://dx.doi.org/10.1109/cic.2016.044 [doi: 10.1109/cic.2016.044]

19. Wikipedia. 2018. Portal: health and fitness URL: https://en.wikipedia.org/wiki/Portal:Health_and_fitness [accessed 2017-11-30] [WebCite Cache ID 6vMEBRCC0]

20. Fetahu B. Github repository of our Wikipedia article parsing functionalities. 2018. URL: https://github.com/bfetahu/utils [accessed 2018-02-07] [WebCite Cache ID 6x3cauBJ2]

XSL•FO

RenderX

21.   Blumenstock JE. Size matters. 2008 Apr Presented at: WWW 2008: 17th international conference World Wide Web
       conference; April 2008; Beijing, China URL: http://dx.doi.org/10.1145/1367497.1367673 [doi: 10.1145/1367497.1367673]

22.   Heilman JM, West AG. Wikipedia and medicine: quantifying readership, editors, and the significance of natural language.
       J Med Internet Res 2015;17(3):e62 [FREE Full text] [doi: 10.2196/jmir.4069] [Medline: 25739399]

XSL·FO
RenderX

Original Paper

Cybersecurity in Hospitals: A Systematic, Organizational Perspective

Mohammad S Jalali[1], MSc, PhD; Jessica P Kaiser[1], MBA

MIT Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, United States

Corresponding Author:
Mohammad S Jalali, MSc, PhD
MIT Sloan School of Management
Massachusetts Institute of Technology
30 Memorial Drive
Cambridge, MA, 02142
United States
Phone: 1 6172532659
Email: jalali@mit.edu

Abstract

Background: Cybersecurity incidents are a growing threat to the health care industry in general and hospitals in particular. The health care industry has lagged behind other industries in protecting its main stakeholder (ie, patients), and now hospitals must invest considerable capital and effort in protecting their systems. However, this is easier said than done because hospitals are extraordinarily technology-saturated, complex organizations with high end point complexity, internal politics, and regulatory pressures.

Objective: The purpose of this study was to develop a systematic and organizational perspective for studying (1) the dynamics of cybersecurity capability development at hospitals and (2) how these internal organizational dynamics interact to form a system of hospital cybersecurity in the United States.

Methods: We conducted interviews with hospital chief information officers, chief information security officers, and health care cybersecurity experts; analyzed the interview data; and developed a system dynamics model that unravels the mechanisms by which hospitals build cybersecurity capabilities. We then use simulation analysis to examine how changes to variables within the model affect the likelihood of cyberattacks across both individual hospitals and a system of hospitals.

Results: We discuss several key mechanisms that hospitals use to reduce the likelihood of cybercriminal activity. The variable that most influences the risk of cyberattack in a hospital is end point complexity, followed by internal stakeholder alignment. Although resource availability is important in fueling efforts to close cybersecurity capability gaps, low levels of resources could be compensated for by setting a high target level of cybersecurity.

Conclusions: To enhance cybersecurity capabilities at hospitals, the main focus of chief information officers and chief information security officers should be on reducing end point complexity and improving internal stakeholder alignment. These strategies can solve cybersecurity problems more effectively than blindly pursuing more resources. On a macro level, the cyber vulnerability of a country's hospital infrastructure is affected by the vulnerabilities of all individual hospitals. In this large system, reducing variation in resource availability makes the whole system less vulnerable—a few hospitals with low resources for cybersecurity threaten the entire infrastructure of health care. In other words, hospitals need to move forward together to make the industry less attractive to cybercriminals. Moreover, although compliance is essential, it does not equal security. Hospitals should set their target level of cybersecurity beyond the requirements of current regulations and policies. As of today, policies mostly address data privacy, not data security. Thus, policy makers need to introduce policies that not only raise the target level of cybersecurity capabilities but also reduce the variability in resource availability across the entire health care system.

KEYWORDS

cybersecurity; hospitals; organizational models; computer simulation

Introduction

Health care data breaches are a growing threat to the health care industry, causing not only data loss and monetary theft but also attacks on medical devices and infrastructure [1]. Hospital data security breaches in particular have the potential to cost a single hospital as much as US $7 million, including fines, litigation, and damaged reputation [2]. A data breach has a combined estimated effect on the health care industry of about US $6 billion [3]. Meanwhile, the health care industry lags behind other industries in securing its data, and in response, health care organizations must invest considerable capital and effort in protecting their systems [4].

However, this is easier said than done, given the complexity of health care organizations. Hospitals are extraordinarily complex organizations with many typical organizational characteristics dialed up or down to extremes [5] such as

- Technology saturated environment: similar to other organizations, they struggle to manage an array of devices ranging from legacy information technology (IT) to connected medical devices; unlike other organizations, they have orders of magnitude more of them, procured not by a single IT department but purchased ad hoc by clinicians, or given for free by medical device companies [6].
- Internal politics: they deal with the same internal politics that other large organizations do but complicated by the complexity of functions contained within the organization: finance, IT, and human resources, just like other organizations; unlike other organizations, they also must support radiology, cardiology, and pediatrics among others [5]. The degree of specialization is high. Each department requires totally different equipment, caters to different patient needs, has different workflows, and employs a highly specialized labor force that requires years to train.
- Regulatory pressures: similar to other organizations, they must abide by the regulations imposed on them by state and federal government; but in the United States, health care data is considered to be particularly sensitive, and thus, is protected under additional specific data protection laws [7].
- Patient-centered care: like all organizations in the United States, hospitals care about their ability to generate positive net revenue for survival, but unlike other organizations, their first mission is to care for their patients, even when they are for-profit [8].

It is interesting to consider what the systemic effect of these characteristics might be on a single hospital's ability to remain robust to cyber breaches. But now consider the range of possible differences among these entities, eg, a rural community hospital has dramatically different priorities than a large, urban research hospital. Specific to IT, outsourcing services is more common in smaller or more rural hospitals, with transcription services being the most commonly outsourced function [9,10]. The decision to outsource interacts with the tendency of these hospitals to make symbolic rather than substantive IT security investments—see Angst and Kelley [11] for more discussion.

Furthermore, significant variability in cybersecurity as a priority has been observed throughout the hospital industry—in the United States, 70% of hospital boards include cybersecurity in their risk management oversight, and only 37% of hospitals perform annual incident response exercises [12]. Similar vulnerabilities in hospitals are also observed in other countries [13-16]. Specifically, pressure from the board of directors appears to be essential in creating substantive cyber resiliency, as research shows that hospital management support is essential for user compliance with information security policies, which in turn are written by health care IT security professionals [17,18].

The importance and complexity of cybersecurity capability development at hospitals raise critical questions: how do the inter- and intradynamics of hospitals interact to form a system of hospital cybersecurity in the United States? Does this leave the health care infrastructure of the United States vulnerable as a whole? As data interoperability becomes an imperative, driven by Affordable Care Act requirements and payment reform, will hospitals with lower cyber capabilities leave all patients vulnerable?

To answer these questions, we interviewed chief information officers (CIOs), chief information security officers (CISOs), and health care cybersecurity experts at hospitals and developed a system dynamics model to study the dynamics of implementation and maintenance of cybersecurity capabilities in hospitals.

This study helps health care leaders reduce hospital vulnerabilities by detailing the outcomes resulting from strategic decisions of cybersecurity development. It also aids cybersecurity professionals in understanding the complexities of cybersecurity capability development in hospitals.

Methods

To develop our model, we conducted semistructured interviews with 19 cybersecurity professionals, primarily in the United States. Interview subjects were contacted by email and volunteered to participate in a 45-min interview on cybersecurity capabilities at their hospitals, noting that all interviews would be stripped of identifying details. Given previous research showing significant differences in capability development driven by size, orientation, and urbanicity, we strove to obtain diverse candidates that represented dimensions previously shown to matter in capability development. To minimize biases (eg, interviewer bias and confirmation bias), we followed standards for semistructured interviews [19], specifically, we ensured that (1) Interviewers were educated to maintain a neutral attitude and avoid suggesting answers or making judgments and (2) The predefined questions had a neutral language and did not include suggestive words. Interviewees included (1) C-level executives who actively participated in the strategies of hospitals, (2) Operationally focused information security professionals, who typically had titles such as information security specialist in hospitals, and (3) Software vendors and consultants with a privacy and security focus, who specialized in the health care industry (see Multimedia Appendix 1 for more information).

Following standards for inductive and generative coding [20], the interview data were coded to extract variables and common themes related to cybersecurity capability development in health care organizations. Coding of the interviews was conducted by the authors—any disagreements or concerns about the extracted data were discussed among the authors until consensus was reached. Coding helped in learning the mechanisms of capability development through identifying key variables and relationships among the variables. For example, "efforts for filling out gaps" (between the actual and desired level of cybersecurity) and "internal stakeholder alignment and resource availability" are two variables extracted from the interviews, and the relationship between these two variables was that "internal stakeholder alignment and resource availability" had a positive causality effect on "efforts for filling out gaps."

Following best practices for analyzing qualitative data to build system dynamics models [21], the emerging relationships among the extracted variables were then integrated into an evolving causal loop diagram, which embedded the key relevant mechanisms important for understanding how the capabilities were built and then eroded—see the next section for the description of the model. We used system dynamics modeling, which is a potential tool to understand the complexity of a sociotechnical system [22-25]; in this case, the system of cybersecurity capabilities in hospitals.

Results

Main Findings

Cybersecurity capabilities include a variety of programs, behaviors, and technologies that a hospital employs to improve cyber resiliency. Not all of these capabilities are self-sustaining, and they may erode over time. If properly adopted, implemented, and maintained, however, they will have a positive impact on the hospital's ability to be resilient to cyberattacks. Stock and flow variables are key tools in system dynamics to present this mechanism.

Figure 1 shows the simple core of our model, including stock and flow variables. A stock variable (eg, "cybersecurity capabilities at hospital" in Figure 1) presents accumulations such as the number of implemented programs, behaviors, and technologies, which represent a hospital's cybersecurity capabilities. A flow presents the rate at which the stock changes (see the inflow and outflow variables in Figure 1). The inflow is the capability development rate or the rate at which capabilities are added to the existing stock. The outflow is the capability erosion rate or the rate at which capabilities are removed from the existing stock.

Our interviews uncovered five major themes that described the dynamics of the cyber capability development in hospitals: uncertainty in resource availability, external pressures, end point complexity, internal stakeholder alignment, and cybercriminal activity. In this section, we expand on each theme by constructing our model and using quotes from interview data to show how our subjects described each mechanism behaving.

Uncertainty in Resource Availability

Current business trends in health care were a factor in our interview subject's minds and produced the most uncertainty with regards to how they affected cybersecurity capability development through resources available to the information security team. Two subjects stated the following:

> With all the financial pressures on hospitals right now, [cybersecurity] is not their big concern. The biggest driver is available funding.

> A larger organization can hire more security admins and come up with more purchases, or different products, or come up with more protocols, or create more tools that they need to monitor the network more closely. For an organization that either has fewer resources or chooses not to invest in that, they are more likely to look at hosted solutions.

The two main issues affecting resource availability were net revenues and talent availability. For most interviewees, net revenues were perceived to be declining, driven by flat revenues and increasing operating expenses. For organizations with declining net revenues, outsourcing IT to an organization with more expertise was an effort to increase resource availability to undertake more efforts to close cybersecurity gaps. Some of our interview subjects worked at organizations that were financially healthy enough to fund the development of purely internal solutions; however, the majority did not. Two subjects stated the following:

> We do not develop solutions, because we are too small...So we need to focus on existing solutions.

> I am increasingly looking at technology as a solution for process problems, or issues that I see in my center...I can't justify every dollar that I'm going to spend on it if it's going to give me more in return or value. But what I do know is that to make the same kind of changes in one big swoop, would require incredible patience, diplomacy, personal political capital. I can pay $20k for it and potentially other problems get solved. Or, I can pull up my sleeves and make the same changes on my own, but I would almost certainly have a Xanax dependency at the end of it.

For those who worked at organizations healthy enough to fund internal development, subjects were split as to whether self-hosting and internal development increased resource availability. On one hand, some felt that owning IT policies themselves gave them finer control over how to allocate resources in their efforts to close cybersecurity gaps. On the other hand, some felt that outsourcing security operations to a firm such as Microsoft via purchases of their cloud products simultaneously allowed them to do more with fewer resources and also tacitly allowed them to pay less attention to cybersecurity, thereby introducing an entirely new set of risks.

XSL•FO

RenderX

Figure 1. Stock and flow diagram of hospital cybersecurity capabilities.
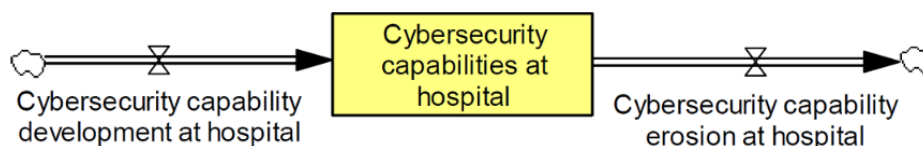
Three subjects stated the following:

> I do think there are areas where we can do a little more, because it's easy to outsource those responsibilities in a sense. So it's easy to point the finger and say we depend on these people for that.

> There's a real imperative to go to cloud-hosted services and procure those services. There's a different stack of security issues to think about if you're purchasing subscription services if you're doing it yourself.

> For an organization that either has fewer resources or chooses not to invest in [cybersecurity], they are more likely to look at hosted solutions...The push to get more things in someone else's cloud will help those organizations standardize those security practices.

For the stand-alone organizations who were not part of a larger organization or part of an urban environment, urbanization among the US population had affected their ability to hire security professionals, who mostly live in urban centers. Furthermore, because the populations of these urban centers are growing, the patient populations that those hospitals serve have grown as well. This trend had a strong impact to resource availability for hospitals outside of larger urban areas. Two subjects stated the following:

> How many trained security professionals are there in South Dakota?

> Healthcare doesn't get paid very much, so revenue doesn't go towards cybersecurity. When in banking, I would have had 25 employees at an organization of this size.

Our interviews suggested that a hospital with declining net revenues would also typically have troubles attracting talent as well. Thus, we described these two concerns in the model through the single variable "resource availability." Whether those resources were "admins, purchases, products, protocols, or tools"—as described by the interview subject above—these were the essential building blocks that allowed a hospital to make efforts to increase the capability development rate, thereby increasing the stock of capabilities at the hospital. This, in turn, would increase the cybersecurity level at the hospital, which would decrease the gaps between the actual and desired level of cybersecurity. If the gaps decreased, then so would efforts to fill out those gaps. Figure 2 presents the efforts to develop cybersecurity capabilities as driven by the variable of resource availability. Feedback loop B1 presents a balancing feedback loop that stabilizes the system (ie, filling out cybersecurity gaps) by moving it toward the desired goal.

External Pressures

In the previous section, we describe the cybersecurity level at the hospital, as influenced by the stock of cybersecurity capabilities. This cybersecurity level also drives the vulnerabilities that cybercriminals can exploit, and, if successfully exploited, the hacks and breaches which then affect hospitals. Subjects spoke of the many ways that these successful breaches translated into pressures to develop stronger capabilities.

Our interview subjects often used a successful cybercriminal exploit at another hospital to stoke higher pressure for cybersecurity capabilities by bringing the consequences of that exploit to the attention of their board or managers. They were typically speaking of the pressures imposed by the public and the media and those imposed by Health Insurance Portability and Accountability Act (HIPAA) and related regulation and, more recently, from the US Food and Drug Administration (FDA) in the arena of medical devices.

A major pressure to have stronger capabilities was the threat of a loss of public reputation. This fear is in part by design of existing regulation, which includes reporting requirements should a cyber breach occur. But the threat of a public or media backlash was real and weighed on our interview subject's minds as a "sniff test" for whether they should be doing more to develop cybersecurity capabilities. It also speaks to the importance of the media in generating awareness around cyber threats, particularly for health care organizations. Two subjects stated the following:

> Our culture wasn't this way seven years ago...It also took bad things happening sometimes. Nothing affects change like someone making a mistake.

> What Rahm [Emanuel] told us is let no emergency or crisis go unused...If you say [to the board], look Home Depot has just had this breach and this expense...Don't you think we want to avoid being in the Boston Globe in a bad way?

There were also significant pressures to have stronger cybersecurity capabilities resulting from regulation. Typically, regulation was aimed at protecting privacy, not necessarily security; nonetheless, there was some overlap. Subjects at hospitals with more resources worked with both an internal and an external audit team to assess compliance; however, all hospitals worked with an external audit team as a regulatory requirement.

Figure 2. Cybersecurity capability development with a balancing feedback loop.
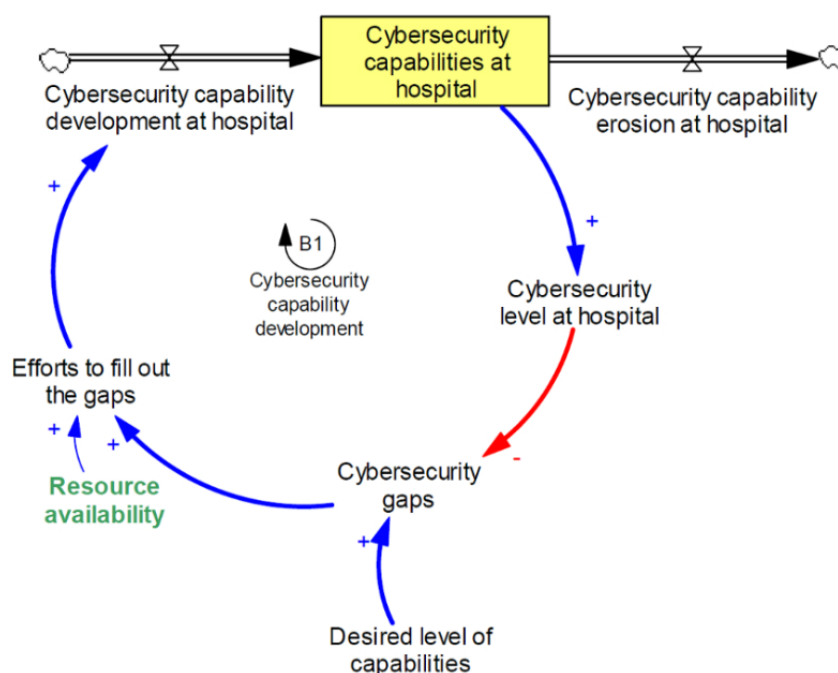
Subjects expressed a variety of views with regards to how helpful regulations were in producing good cyber hygiene, as illustrated in the following quotes:

> I think the larger part of the regulatory requirements is absent of any alignment with cybersecurity. They're largely focused purely on patient health and the patient experience of care, and...they're largely divorced from cybersecurity.

> You have to [follow regulation] because you need some common grounding and things to measure against and things to work to. It serves a good purpose.

Some felt that HIPAA created a floor of cyber capabilities that was helpful for small organizations, but not larger ones, as illustrated in the following quotes:

> I think all kind of clichés about it, which is that it's a floor, not a ceiling, and that you can be compliant but not secure. I think those are mainly right.

> So we're a small center, nobody expects me to have one FTE manager whose sole job is to walk around all day looking specifically for HIPAA violations. So that wouldn't be a reasonable standard. What is reasonable is do we train people, do I as a leader emphasize it routinely. I don't like the words minimum necessary, but it is kind of the minimum necessary needed to meet the intent [of eliminating data breaches].

> Compliance is a low bar. I guarantee that little healthcare organizations and hospitals would do nothing [without regulation]. They would have a piece of paper on a shelf called their security policy. It's needed as a backstop to get companies at least thinking about it. But being compliant does not solve the greater risk management problem.

Some felt that the pressures produced by HIPAA interacted with the target level of cybersecurity capabilities in such a way that the resultant desired level of cybersecurity capabilities encouraged hospitals to focus on the wrong things. One subject stated the following:

> Clearly, HIPAA distorted the cybersecurity programs of large organizations a little bit. I treat compliance as a separate issue from security. Let's make sure that we're plausibly compliant and let's build a program over actual security.

In particular, there was a belief that the focus on end-to-end messaging encryption was not as important for hospitals to focus on, as illustrated in the following quotes:

> Say you have an electronic medical record that sits in the same lab, but the data gets transferred between the two systems. You can say this is very safe, because in order for the bad guy to find it, they'd have to go through so many layers that they'd have to be in the hospital anyway. Some auditors say it has to be encrypted anyway, which might increase complexity, time performance, and even data availability. The more complex your systems get, the slower they run, which affects availability.

> A lot of healthcare organizations have been built around encrypted e-mail which to my mind don't have real security benefits...You want to reduce the emails people send with private info, but people don't spend a lot of time thinking about that.

Others mentioned other "best practice" security practices that are not mentioned in regulations but would help prevent data breaches, as illustrated in the following quote:

> It's impossible to have a good security without pentesting, without a very active threat hunting

program. Those are the kind of things that we really emphasize. But they're not generally contemplated in the general HIPAA regulatory regime.

And some felt that recent regulations, such as the Health Information Technology for Economic and Clinical Health Act (HITECH), were quite forward-looking and pushed their organization forward, whereas others did not, as illustrated in the following quotes:

It's really HITECH rather than HIPAA [that we focus on compliance]. That was the whole Obama-era program that resulted in audits and compliance.

I think HITECH does go above and beyond. It puts out their technical controls. And it gives you a great starting base for talking to the business owner.

HITECH is just establishing subcontractors have to abide by HIPAA. It just means there's more you have to follow. I've never heard HITECH being called out separately from HIPAA.

What is clear is that the process of external audit at least compels hospitals to adopt some cybersecurity standard (examples given were National Institute of Standards and Technology [NIST] 800-66, Control Objectives for Information and Related Technologies, and Information Technology Infrastructure Library) and try to follow it. On subject stated the following:

You need to set yourself around a common cybersecurity framework. It doesn't have to be ISO or NIST, it could be a combo. It fits the general corporate culture as it is here and now.

Interview subjects stated that external auditors varied in the degree to which they demanded rigorous compliance to that standard but that the standard gave them a helpful tool in socializing good cybersecurity practices throughout the organization. Two subjects stated the following:

It's a little confusing to be having been an auditor. You assess an organization against a set of known criteria. I see auditors doing less of that, and more taking freedoms and liberties.

We use NIST 800-66 or other NIST artifacts to decide what is the rubric for our risks and our risk mitigations. You kind of have to pick a rubric—whether it's HITRUST [the Health Information Trust Alliance]...or COBIT [Control Objectives for Information and Related Technologies] or ITIL [Information Technology Infrastructure Library]. It is not sufficient for three people to sit at a table with a bunch of beer and decide the risks. Our boards and our auditors have asked us to adopt a standard framework, whether it's Deloitte or PwC, so they can judge us against an objective framework of goodness. That's truly essential for an org.

These pressures (either internal or external), combined with the target level of cybersecurity capabilities, produce the hospital's target level of capabilities. The target level and the desired level may be different (see Figure 3 which adds the feedback loop "need for stronger capabilities"). Loop B2 is a balancing loop, with delays in how quickly vulnerabilities and cybercriminal activities ultimately affect the pressures to have stronger capabilities.

Figure 3. Balancing feedback loop of need for stronger capabilities.
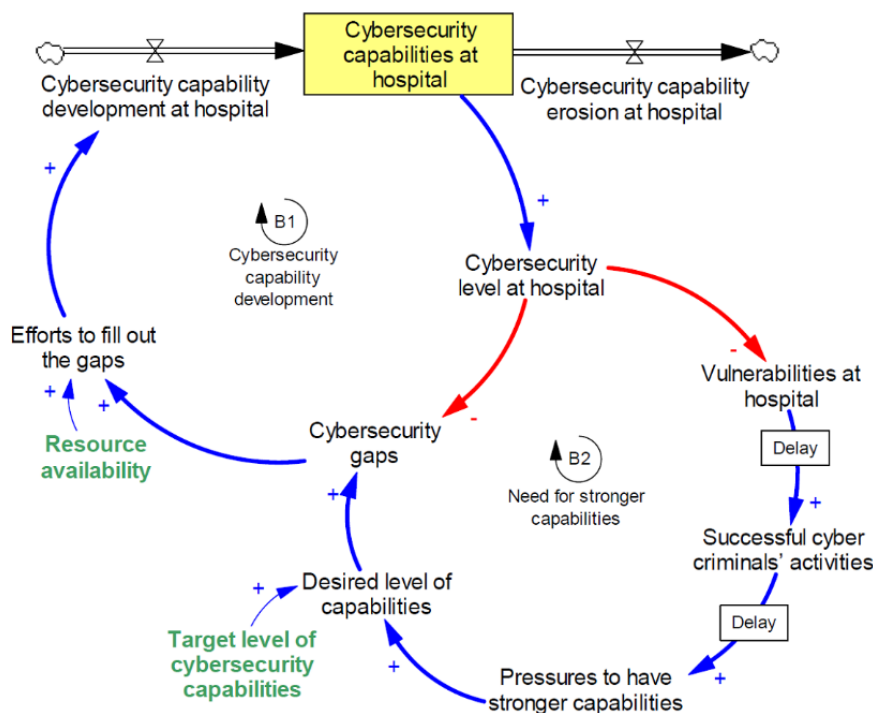
End Point Complexity

The theme explored by most interview subjects was that end point complexity made hospital cybersecurity capabilities unique. Like many organizations with a large employee base and a physical footprint, hospitals must manage the numerous devices used by both administrators, medical staff, patients, and their visitors to transact business, provide care, and pass time. One subject stated the following:

> *I have 8000 iPhones, 2000 Androids, 2000 iPads, and some Blackberries. What are the security implications of doctors and patients doing more transactions on BYOD [bring your own device] devices?*

Unlike many organizations, the bring your own device (BYOD) challenge is compounded by a bevy of instrumentation and diagnostic equipment that may also present security risks. One subject stated the following:

> *In our environment, we have about 800 families of medical devices. Most organizations have two or three dozen SCADA systems...That's an astonishingly high number. There's no counterpart to that in education or finance.*

In a competitive market, some medical device manufacturers also provide "free samples," bypassing risk assessment and management processes. Although IT and information security (IS) teams have methods of determining unauthorized connections to their networks by unexpected devices, it adds another several points of vulnerability to their organization to contend with these devices. One subject stated the following:

> *In hospitals, the interesting stuff is there's a whole underground procurement process whereby medical device vendors approach clinicians and give them lots of stuff for free that eventually makes its way on to our floors, and then a year later we get a bill for it. That's a unique quality of working in a hospital.*

Additionally, medical device manufacturers have historically not designed their products with security in mind. Interview subjects were optimistic that this might be shifting, as the FDA has waded into the regulation of the medical device market. However, they felt the process would be slow, as the FDA is slow to certify devices, creating a gap between regulation and practice that exposes patients to more risk. Three subjects stated the following:

> *We're doing infusion pump management and we're using a compromised Linux kernel. I can deliver a lethal dose and then back right out of it. We don't even have checksums in those OS's to do forensics.*

> *The cybersecurity of your pump that sits next to the patient that programmatically determines when to pump medications into an IV [intravenous], the security of that device now needs to be a primary design decision. It needs to be a motivating factor in when you create that device. So the regulation that the FDA—the guidance—feels a little behind the times. But they're catching up with the need to put out guidelines on these things.*

> *It's a highly regulated industry [by the] FDA. When you make a change to those systems, you have to go back and recertify...They end up not doing that and you end up with machines that can [be] breach[ed].*

Thus, we add end point complexity to the model. As the end point complexity increases, it increases the vulnerabilities at the hospital. Also, with the increases of end point complexity, the ability of the organization to manage the security of each end point degrades. The result is that an increase in end point complexity subsequently increases the speed at which cybersecurity capabilities erode (see Figure 4 which adds end point complexity to the system).

Internal Stakeholder Alignment

A major mechanism that our interview subjects described was of the complexity of internal stakeholder alignment—*There's no single point of decision making*. Typically, the main stakeholders described were the CIO or CISO, the IS team (if one existed), the IT team, other C-level executives, the board of directors, and finally, the medical staff. Even in a small hospital with 100 beds, employee count will easily number in the hundreds.

First, the CIO and, if relevant, the CISO, had to develop alignment on strategic IT initiatives to bring those initiatives to the strategic planning process. All of our interview subjects described some level of friction if the IT and the IS resources were separated, as illustrated in the following quote:

> *Balancing how much should come from security and from general IT is one of the perennial problems that we deal with...That friction is always at the forefront when we do operational security.*

Typically, IT advocated for the benefits of the technology, and the IS team had difficulty looking past the risks. They would come together to work out a compromise that adequately captured both of their viewpoints. Then, they would bring these perspectives to executive management during the planning process, using their organization's preferred risk or value metrics to measure the impact of new or continued investments. One subject stated the following:

> *I make the decision[s on deploying new technologies] with the help of higher management. I present [options] to the board and get their approvals and address their concerns. The metrics that we use [to evaluate] are security values.*

Figure 4. Addition of end point complexity to capability development model.
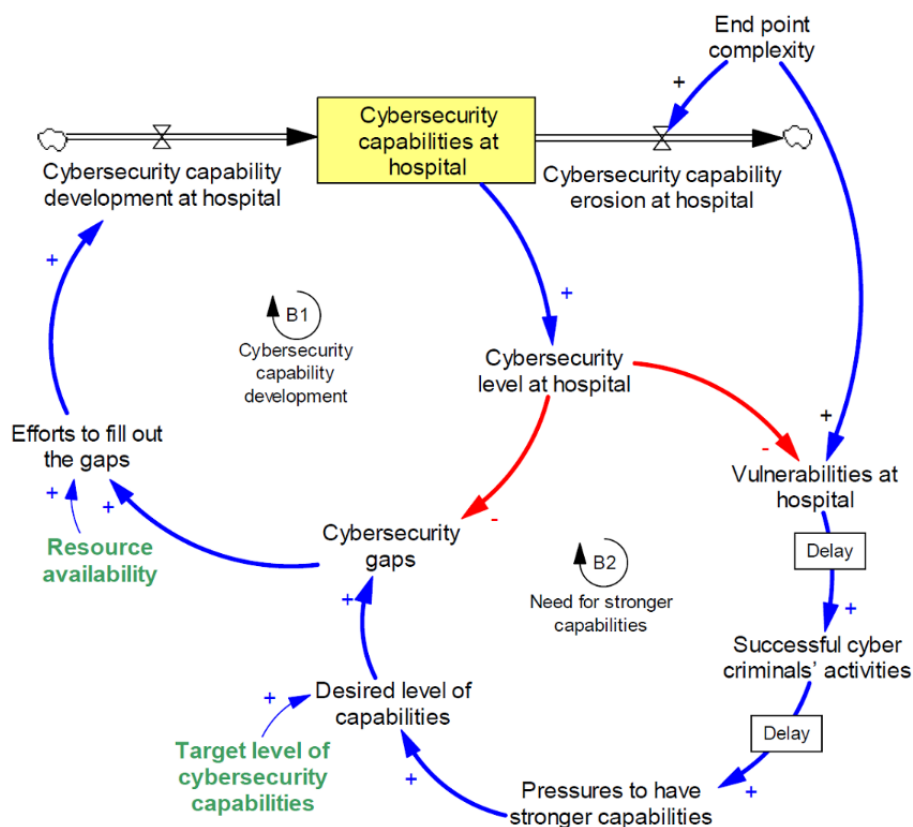
Our interview subjects who felt that top management was respective to their requests felt there were two overlapping reasons behind it: (1) there is clear evidence that underdevelopment of cybersecurity capabilities would result in a crisis for the hospital and (2) their board of directors also understood this and put pressure on the top level board to include cybersecurity as part of the hospital's core strategy, as illustrated in the following quotes:

> Boards and other fiduciary individuals have said, we need to buy down the risk. This isn't some techie thing, this is risk to your business. What happens if doctors can't do dictation? Am I going to remember everything I put down two days later? If we can't code, we can't bill, if you can't bill, you don't have revenue. If you're not for profit, you run out of resources. It's a long stretch but they're starting to get it.

> I get calls at my desk about this from board members...they are very engaged and very nervous. I have a decent level of traction despite that I'm many tiers down.

> The board contains people whose companies have had material harm done to them because of cybersecurity breaches. So when the board comes to the senior executive team, or else you'll be looking for a new job. That's probably the most significant motivator—it's some higher force than the senior management raising the issue.

It seemed that the threat of regulatory fines served to produce stakeholder alignment, to some extent, with IT, top management,

and the board of directors more easily, as illustrated in the following quote:

> It's a board level issue because of the reputational issue and fines that occur in a breach. The average cost of a breach is $300 per patient breached. When you look at the legal fees, forensics, media management, fines from Office of Civil Rights—the board says this is an existential issue. There are only a couple of ways we can fire the CEO and loss of reputation is a good way for a CEO to get fired.

For the interview subjects who did not feel that their hospital was developing cybersecurity capabilities, it was mostly because of high turnover at the C-suite level. That high turnover, in turn, led to constant shifts in strategy that became difficult to navigate as an IS specialist, leaving the organizations more reactive than proactive in developing cybersecurity capabilities. One subject stated the following:

> They understand the importance of it, but they don't understand the amount of finetuning that takes place...We do have support from the C-level. But...there's a revolving door with the C-level management, so it's hard to get someone's ear and hook on to that support.

In general, subjects felt that although the health care industry potentially lagged behind other industries, there was a growing awareness of cyber risk at both board and C-suite levels. This, in turn, gave their teams sufficient institutional power to affect the necessary changes to build cybersecurity capabilities at the administration level of the hospital. However, a hospital has

many more employees than just its administrators. Gaining stakeholder alignment with operational staff—doctors and nurses—is much more difficult. This manifested in two ways: (1) open to or at least indiscreet flaunting of IT policies and (2) an underground "procurement" process of medical devices. The latter is discussed in the theme of end point complexity. An example of skirting IT policies was provided by one interviewee:

> Say I want to use an ultrasound machine. We have regulatory requirements that require authentication to all of our IT devices. Then your password has to change every 90 days. They just want to use the ultrasound machine. It's not holding a lot of patient data, they have to memorize their passwords. Then they can't use their common username and password [because it's a different device]. They say we're putting a lot of burden on us, it's making it difficult to provide seamless patient care. So they create a shared login so that they can provide patient care.

Working against the IT and IS teams was a strong culture of "patient care first." Many of the medical professionals see cybersecurity standards as a barrier to patient data portability, which increases the paperwork that staff must prepare, increases the likelihood of error rates when patients are new or transferred, decreases the time that they can spend with patients, and, potentially, decreases the ease of collaboration among staff. Gaining buy-in from these staff was more difficult than gaining buy-in from the administration because the medical staff saw less direct impacts to their ability to perform from the consequences of cyber breaches, such as reputational loss or fines. Two interviewees stated the following:

> Doctors, that's a different story...The nature of their work—they have to get patients in and out. They're probably the least understanding.

> When Amazon is asked to open this port, or relax this firewall, the answer is NO, for no one ever. As opposed to some Nobel Laureate who wants us to relax port 1551.

Interview subjects who had gained success with medical staff had done so via three mechanisms:

(1) Direct experience with cyber crisis. One interviewee stated the following:

> Millennials and Gen Z—they've never had to deal with the old school stuff of... paper to write down information...So, when the technology turns off younger people don't know what to do...Ransomware [attacks] are wake up calls for them.

Ability to articulate patient harm. Two interviewees stated the following:

> You can solve a corporate argument about what is best for the patient. There's a consistent ability for me to push cybersecurity by focusing it on the patients.

> The one common thing is that clinicians will be more accommodating towards taking on security measures that benefit their patients.

And, designing systems such that medical staff never became aware of a way to loosen IT security policy. Two interviewees stated the following:

> My hope is that they pay very little attention to it. They shouldn't have to. Cybersecurity is my job and not theirs.

> We have a lot of ransomware attempts, but it's not a problem. We are unusual in healthcare because no one has local admin rights, so no one saves docs on a computer...and everything is fully recoverable.

Additionally, the complexity of hospitals as organizations can lead to conflicting views about what the primary mission of the hospital is, and as a result, can lead to conflicts over the purpose and acceptable impacts of security policy. The most commonly noted examples were the differences in designing security for research hospitals, teaching hospitals, and safety net hospitals. Teaching hospitals had to ensure that patient data was protected and that residents, attending physicians, and students had sufficient access to patient data to fulfill the hospital's teaching mission as well. In an example provided by a CIO working at a research hospital, a content filtering service intended to block adult content also impacted the work of a research team studying the effects of pornography on mental health:

> I personally believe that hardcore pornography has no purpose on hospital supported devices. What did I do five years ago, I put up internet content filters that prevented people from navigating to pornography. Within five minutes, the director of psychiatry calls to tell me that we have a grant to study pornography in a medical context. It's really hard in an academic medical center with a 1000 different CEOs—because every academic chair is a CEO.

Many interview subjects referred to their own work as primarily cultural rather than technical, by which they meant working with internal stakeholders to shift the perception of cybersecurity practices as a "nice-to-have," as illustrated in the following quote:

> To me, what that means is that the culture of the organization has to change. Processes are a very strong way of changing the security posture of the organization. It's not just changing the technology. It's about the vendors, the workflow you use for onboarding employees, for moving data around the organization. That's all awareness and training. It's a real cultural thing that your org has to see security through.

With regards to cybersecurity, this manifests as an increase in end point complexity. While IT attempts to manage hospital's BYOD policies and network security, they must also contend with an underground procurement process by which clinicians, who have some self-sufficiency with their own department's budget, can purchase the medical devices that they think best provide patient care. One interviewee stated the following:

> Someone has already been working with a vendor, they decide to bring the vendor on site, and decide to

work with the vendor. And then, they realize they need the technology, so they call the IT staff. So [IT does] an assessment after the fact, and works with the vendor to implement their solution. It's a backward process, it'd be nice to look at the vendor before, and go through some references, and have some sort of risk analysis and scoring, where we can say the vendor seems legitimate, and like they have the right solution in place...but that's not the current state.

As a result of all these dynamics, internal stakeholder alignment becomes a highly interconnected variable that influences many other variables and forms two reinforcing loops (R1 and R2) and two balancing loops (B3 and B4; Figure 5). As cybersecurity capabilities decrease and medical staff becomes less aware of the threats they are introducing through new devices or unsafe cyber practices, end point complexity grows. This leads to a reinforcing feedback loop that could result in an explosion of end point complexities. Additionally, if stakeholders do not align with the importance of cybersecurity, they are also more likely to undermine efforts to fill out cybersecurity gaps, reducing cybersecurity capability development, and thus decreasing capabilities. This forms another reinforcing feedback loop, captured by the attitude "We are going to get hacked

anyway?" Finally, the pressure to have stronger capabilities manifests within the variable of internal stakeholder alignment and influences how they see efforts to close cybersecurity gaps and end point complexity. This pressure forms two balancing loops, described as the pressure to maintain reputation (see Figure 5 for the addition of this variable and the corresponding feedback loops).
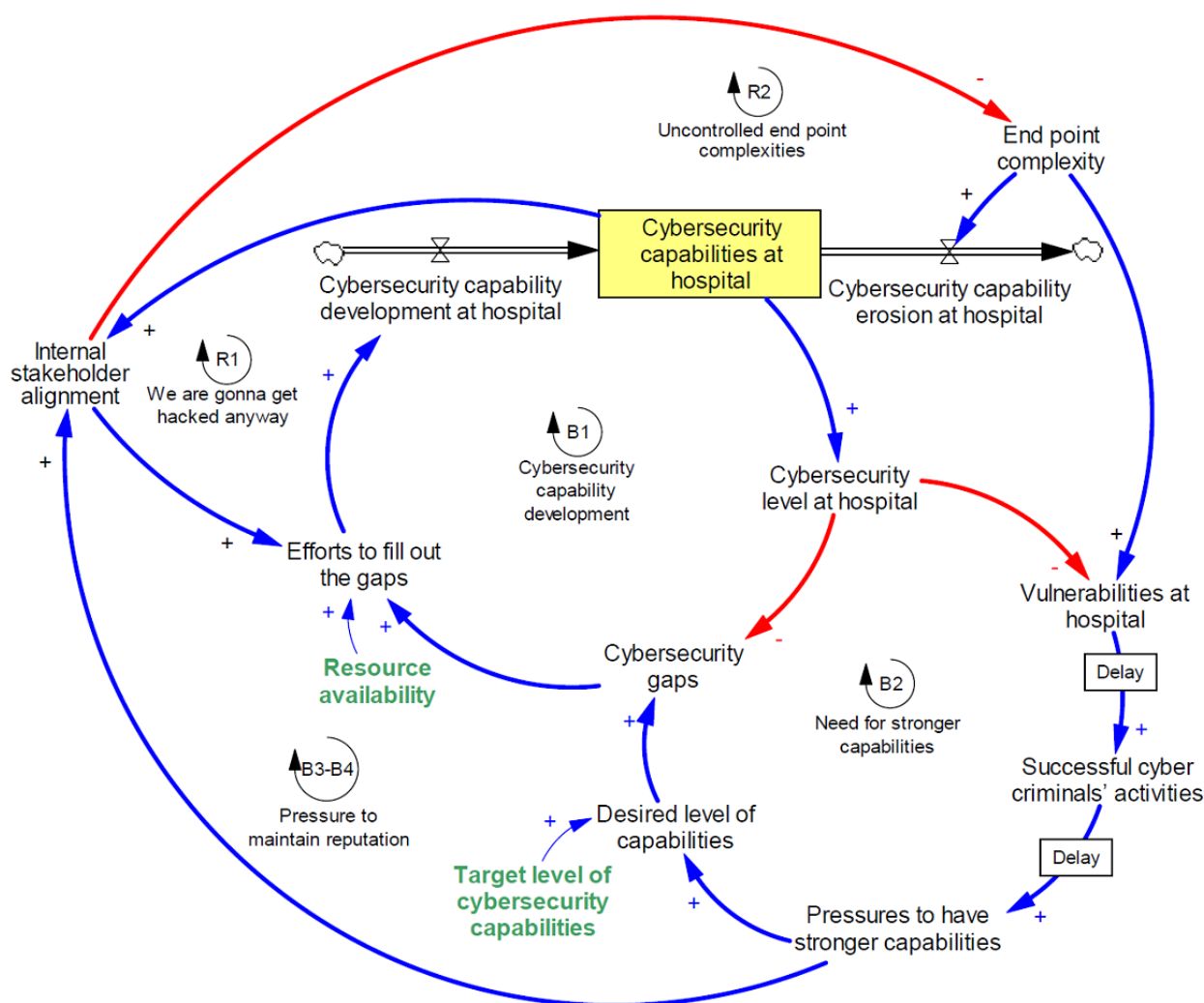
Cybercriminal Activity

Finally, the dynamics of cybercriminal activity itself is a major driver in this system. Cybercriminals have been highly active in targeting health care organizations, although subjects were split as to whether that was because of an overall increase in cyber activity, or an increase in cyber activity specifically targeted at health care. One subject stated the following:

To be clear, WannaCry was an untargeted attack, whereas Petya and BadRabbit—those appeared to be targeting, although the jury is out, no one has done attribution on that. I think the ransomware risk is real, I think organizations need to take it seriously.

For those who did feel health care was specifically targeted, subjects described three reasons for why there has been increased cybercriminal activity in health care.

Figure 5. Introduction of internal stakeholder alignment variable.

If the cybercriminal's goal is financial, the value of medical data is relatively high compared with other types of data, as illustrated in the following quotes:

> If I remember my data correctly, a hospital record costs 20x more than the $10 that you get for someone's social security number. So hackers aren't looking for credit cards and bank accounts anymore, they're looking for medical records which contain a lot more information.

> The number, amount, and variety are probably only comparable to a bank that deals with all those transaction volumes. And that this information is required for us to do the most basic activities that we are engaged in.

If the cybercriminal's goal is ideological or terrorism-related, disrupting the feeling of safety that health care might otherwise inspire is an attractive target to induce fear, as illustrated in the following quote:

> I think they do it because it's scary. Because...when you hit the place the space cares for you when you're sick, it's scary...It might have a stronger impact. It makes it to the news faster...If a private organization gets shut down for the day, no one might even know. But if you have to turn away patients from your emergency room because you can't get your IT up, that's scary.

And, the health care industry lags other industries in cyber resilience, making them an easy target irrespective of any other qualities, as illustrated in the following quote:

> I would aim at a hospital. Healthcare is typically five years behind the power curve. Why not go for soft targets?

In any case, it is clear that health care organizations have been an attractive target recently. Even with an increase in cybersecurity capabilities, the first two reasons for their attractiveness to criminals will remain in place. This overall increasing trend in cybercriminal activities can be incorporated in our model, next to successful cybercriminal activity.

Not specifically included in our model but present nonetheless with our subjects was the feeling, whether substantiated or not, that the motivations of hackers had shifted from being "kids in pajamas" to more malicious organizations: either organized crime or nation-state backed activities. Two subjects stated the following:

> The nature of attacks is increasingly sophisticated. It used to be my biggest threat was MIT students. Today, it's state-sponsored attacks, terrorism, and organized crime. It's more threats than ever before of a more serious nature.

> It's either economic espionage or geopolitical espionage. Corporate espionage. Basically hacker organizations. It's gone away from script kiddies, and now it's nationally sponsored hackers that are actually getting paid to do this.

Figure 6. Impact of intravulnerabilities on intervulnerabilities and attractiveness of hospitals system to cybercriminals.
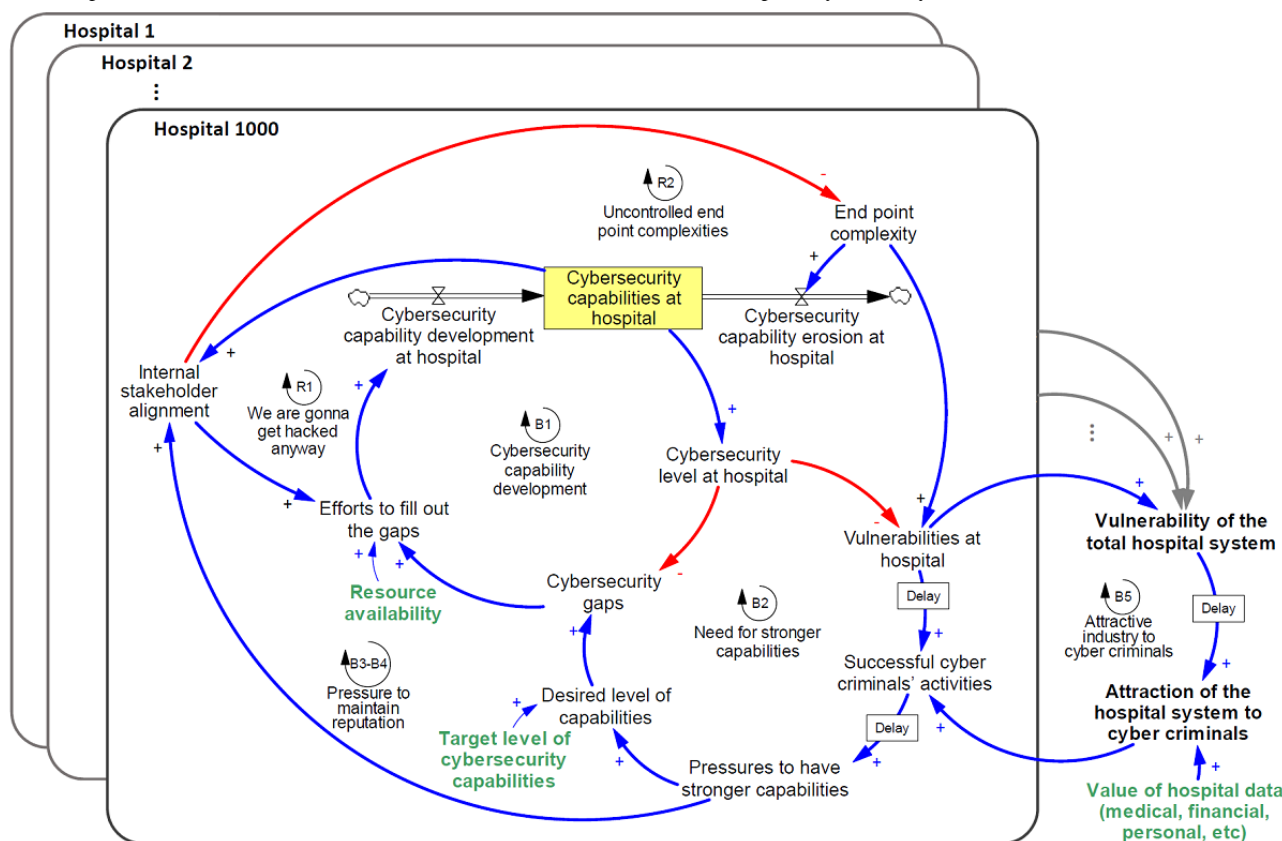
A Large "System" of Hospital Systems

The mechanisms discussed above manifest themselves not just over a single hospital but over the entire hospital system. The cyber vulnerability of a country's hospital infrastructure is the result of not just one hospital but rather many hospitals. To represent this in our model, we include 1000 hypothetical hospitals, each with different levels of resource availability and target level of cybersecurity capabilities, and show how the vulnerability of the hospital system, combined with the attractiveness of hospital data would become attractive to cybercriminals (see Figure 6 for this final addition to the modeled system).

Discussion

Overview

In the previous section, we used interview data to develop our model. Here, we use simulation analysis to illustrate how the mechanisms in the model might influence each other and can distinguish more resilient hospitals from less resilient ones. We use this to derive potential levers for IT and information security professionals in hospitals to improve cyber resiliency and to identify questions for future potential research.

To study the impact of one parameter in our simulation analyses, we change only that parameter of interest and keep the rest of the model parameters at a hypothetical baseline. The baseline is hypothesized based on resource availability=0.2, initial end point complexity=0.8, initial stakeholder alignment=0.2, and desired target level for capabilities=0.5. These parameters are fractions (ie, changing between zero and one representing lowest and highest possible level, respectively). We analyze the effects of a variable at three levels: low (set to 0.1), medium (0.5), and high (0.9). Furthermore, successful cybercriminals' activities (ie, "vulnerabilities at hospital" × "attraction of the hospital system to cybercriminals") is assumed to be zero at the beginning of the simulation (Time=0). All simulations are conducted for 60 months.

Sensitivity of Internal Stakeholder Alignment to Pressures to Improve Capabilities

Given that there are several causal links between internal stakeholder alignment and the eventual pressure to have stronger capabilities (see Figure 5), we wanted to show how sensitive pressures to improve capabilities is to internal stakeholder alignment. We looked at the behavior of pressure to improve capabilities in a single hospital over time in three scenarios: low, medium, and high; as discussed above (Figure 7).

Figure 7. Effects of stakeholder alignment on pressures to improve capabilities over time (a); trends of successful cybercriminals' activity given the variability in internal stakeholder alignment (b); the variability in end point complexity (c); and the variability in resource availability (d). All y-axes are fractions, changing between zero and one representing lowest and highest possible level, respectively.
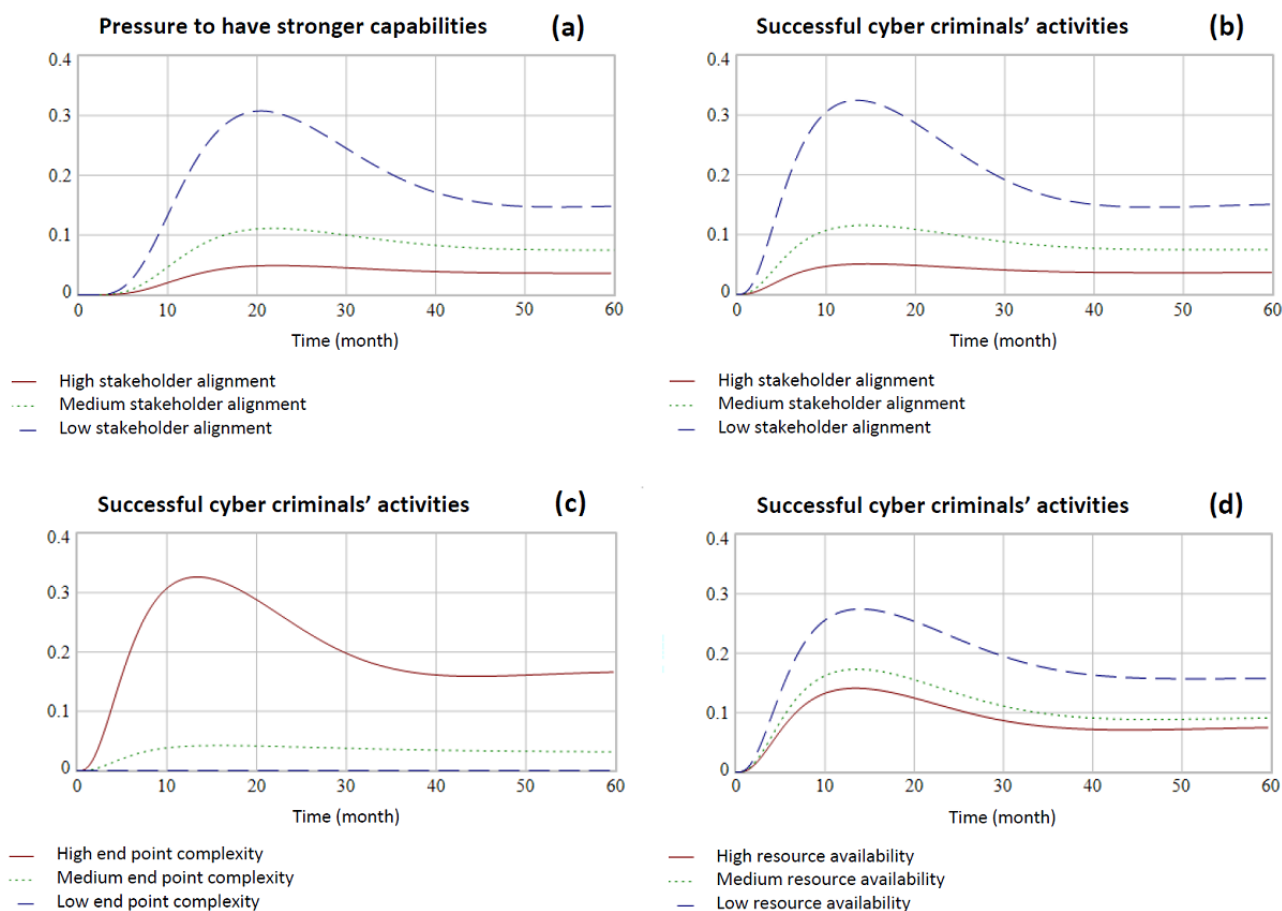
Figure 7 presents that what might seem to be an initially counterintuitive behavior: Medium to high stakeholder alignment results in low pressure to have stronger capabilities. However, consider that hospitals with medium to high stakeholder alignment likely already have a higher target and desired level of cybersecurity capabilities, the result being that they are less likely to become the victim of a cyberattack. Hospitals with a low stakeholder alignment, however, would be more likely to become the victim of a cyber incident, thus creating pressure to have stronger capabilities. The result of a low stakeholder alignment environment, therefore, would be a high pressure one.

This emphasizes the importance of a CIO's or CISO's job in producing stakeholder alignment across the hospital organization.

Impact of Internal Stakeholder Alignment on Cyberattacks

We also wanted to investigate the direct impact that internal stakeholder alignment has on cyberattacks ("successful cybercriminals' activity" in the model). We looked at the behavior of successful cybercriminal activity in a single hospital over time in three variations of stakeholder alignment (low, medium, and high; as discussed above; Figure 7).

All three scenarios eventually reach an equilibrium in which the likelihood of a cybercriminal activity is positive. However, both medium and high stakeholder alignment reach this equilibrium without experiencing a period of time during which the likelihood of cybercriminal activity is greatly heightened.

In practical terms, this variability speaks to the importance of the board's role in cybersecurity governance. If, through governance, the board can create strong stakeholder alignment on the importance of cybersecurity to the organization, this will help minimize the likelihood of cyberattacks.

Impact of Variability in End Point Complexity on Cyberattacks

Most of our interview subjects suggested that end point complexity was the characteristic that most strongly defined the hospital environment. As a result, we wanted to review to what extent it influenced cyberattacks ("successful cybercriminals' activity" in the model; see Figure 7). High end point complexity (set to 0.9) had the strongest impact on successful cybercriminal activity relative to moderate (set to 0.5) and low (set to 0.1) end point complexity. With low end point complexity, however, successful cybercriminal activity dropped almost to 0. Although the high-innovative nature of medical environments is associated with the introduction of new technologies that is usually increasing the end point complexity, our results show that minimizing and managing end point complexity across a hospital is an important lever to decrease successful cybercriminal activity. Therefore, cybersecurity professionals should seek for effective interventions to control end point complexity that do not hurt innovation.

Impact of Variability in Resource Availability on Cyberattacks

Interview subjects were split as to how variability in resource availability affected the hospital system. Some interview subjects felt that a reason that the WannaCry attacks were so harmful to the UK's National Health Service was that the interconnected nature of their systems raised the resource needs for the maintenance of that large system, and thus lowered the resource availability for cybersecurity initiatives. Others felt that even one hospital among many with fewer available resources for cybersecurity was a threat to the entire infrastructure of health care. Anecdotally, subjects who worked at small outposts of consolidated health care organizations felt protected and safe by their parent organization's IT and security teams. One subject stated the following:

> I think healthcare because of its relatively decentralized nature is particularly vulnerable...There's consolidation going on in healthcare, there are literally thousands of different organizations across the country.

We investigated the impact of variability in resource availability on cybercriminal activity over time (see Figure 7). Again, we see that higher resource availability decreases the likelihood of successful attack. Interestingly, however, the spread between outcomes for cybercriminal activity is not as wide as it is for variability in internal stakeholder alignment or end point complexity. Additionally, all three scenarios have a ramp-up period before settling into equilibrium. This suggests that given the levers available to them, it is a better use of energy to pursue increasing internal stakeholder alignment and decreasing end point complexity than blindly pursuing higher resource availability.

Relative Importance of Target Level of Cybersecurity Capabilities and Resource Availability

Most of our interview subjects acknowledged that they operated in a context in which they did not have sufficient resources or were not in full control of the resources at their team's disposal. This might be for a variety of reasons, including a lack of priority during the hospital's budgeting process or declining hospital revenues. We use our model to show how the interaction between resource availability and target level of cybersecurity capabilities impacts successful cyber incidents.

In Figure 8, we see that at low levels of resource availability on the x-axis (eg, <0.5), even a high target level of cybersecurity capabilities (on the y-axis) does not offer protection from cyber incidents. At high levels of resource availability though, the bigger driver in minimizing successful cyber incidents is the target level of cybersecurity capabilities.

In practice, a hospital that does not have sufficient resources will struggle to develop cybersecurity capabilities and meet a target level of cybersecurity capabilities. They will almost certainly be the victim of a cyberattack, and following the attack, will likely increase resources for cybersecurity (ie, a reactive mode). In our interviews, many of the interviewees felt that their hospital had been at this point a few years ago.
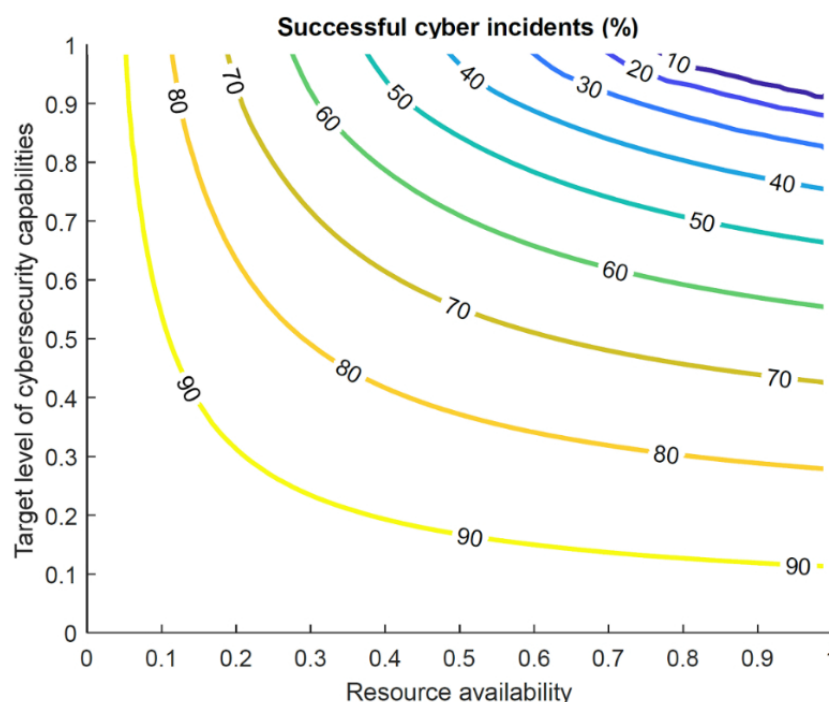
Figure 8. Rate of successful cyber incidents based on target level of cybersecurity capabilities and stakeholder alignment and resource availability.

Figure 8 shows that, at high levels of resource availability (eg, >0.5), even at the same target level for cybersecurity capabilities, the likelihood of a successful attack only slightly decreases when more resources are available. However, a CIO or CISO could more significantly decrease the likelihood of a successful attack by simultaneously raising the target level of cybersecurity capabilities for their organization—eg, by designing a prevention plan. Most CIOs or CISOs did attempt to raise their target security levels, typically by publicizing the impacts of cyberattacks on hospital operations at many different levels (board members, C-suite, and staff).

In Multimedia Appendix 1, we provide further analyses on the effects of heterogeneity in resource availability and end point complexity on cybercriminal activities. We also discuss limitations in this study and present suggestions for future research.

Conclusions

We used interview data to study the complexities of cybersecurity capability development at hospitals. On the basis of the interview data, we developed a system dynamics model and used it to understand how individual hospitals could improve cybersecurity capabilities most effectively. Developing the model demonstrated the existence of three primary levers to improve cybersecurity capabilities available to CIO or CISOs:

Reducing end point complexity: the end point complexity of the hospitals environment is rich with exploitation opportunities for cybercriminals. The tension between decreasing the complexity of this environment and providing excellent patient care is a challenging trade-off. If, however, CIOs and CISOs can decrease the end point complexity of their hospitals, it will have a dramatic impact on decreasing the likelihood of cyberattack. Some of the ways, among many others, that our interviewees achieved the outcome of reducing end point complexity were

- Moving to cloud-hosted services when resource availability was a constraint
- Using technology to detect unauthorized devices on networks
- Maintaining firewalled networks for patients, staff, and medical devices
- Stricter policies on technology procurement

It should also be noted that the end points and what they are connected to are both critical; hence, in addition to the focus on end points, the base architecture needs to be optimized to control the complexity.

Improving internal stakeholder alignment: improving internal stakeholder alignment also reduces the likelihood of cyberattacks. We showed that low internal stakeholder alignment decreases the effectiveness of capability development and increases the erosion of capabilities (by not maintaining them). Our experience shows that soft variables such as stakeholder alignments are often forgotten in cybersecurity management.

Resource availability: finally, while we showed that variability in resource availability did not have the strongest impact on successful cybercriminal activities, we also showed a moderate level of resources is required to have any success in fending off attacks at all. Securing more resources is required to achieve the lowest likelihood of cyberattack, but without internal stakeholder alignment, capabilities are not built and maintained effectively. Furthermore, in the absence of sufficient resources for cybersecurity, setting a high target level of cybersecurity capabilities (beyond those required by policies and regulations) can relatively offset the lack of resources.

Additionally, we used the model to understand what the impact of variability in resource availability within the US health care

system has on cybercriminal activity. Our analysis suggests that efforts to homogenize resource availability across hospitals reduce the likelihood of cybercriminal attacks. This effort could be achieved in a few ways. There have already been some efforts to centralize and unify health record data. Using policy as a way to set target levels of cybersecurity capabilities around this health record data could raise the required "floor." As of today, policy does not specifically address data security, but rather data privacy. Another way would be to work within a single system that assigns resources and set policies to control end point complexity across different hospitals. Although the United States is unlikely to move to the most extreme application of this system (eg, a single player health care), market consolidations have already merged some single hospitals together into a larger system.

Our interview data presents some of the main challenges of cybersecurity capability development at hospitals. Our model also provides an explanatory platform to analyze the complexities development of cybersecurity capabilities in hospitals. For instance, cybersecurity experts believe that

resource utilization correlates strongly with infrastructure age: with the increasing arrival of security patches to a hospital IT department, the number of patches increases with the age of systems. These patches need to be tested for their impacts on internal systems, which is a losing endless loop of resource burden. This mechanism can be explained by the general feedback loop B1 in the model, where with the aging systems at a hospital, the cybersecurity level decreases, which in turn requires resources to build capabilities to fill out the cybersecurity gaps.

The potential consequences of cybersecurity risks promoted the Congress to establish the health care industry cybersecurity task force (see [26] for more information), and our study helps complement the work of the task force. It also opens up additional questions for future research, most notably the quantification of the variables introduced in our model. Using this systemic perspective, however, researchers and practitioners can seek to activate or minimize reinforcing processes as their health organization seeks to develop cybersecurity capabilities, thereby improving their resiliency to cyberattacks.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary information including four sections: (1) interview data summary, (2) effects of heterogeneity in resource availability on cyber-criminal activities, (3) effects of heterogeneity in end point complexity on cyber-criminal activities, and (4) limitations and suggestions for future research.

[PDF File (Adobe PDF File), 731KB - jmir_v20i5e10059_app1.pdf]

References

1. Perakslis ED. Cybersecurity in health care. N Engl J Med 2014 Jul 31;371(5):395-397. [doi: 10.1056/NEJMp1404358] [Medline: 25075831]
2. Claunch D, McMillan M. Determining the right level for your IT security investment. Healthc Financ Manage 2013 May;67(5):100-103. [Medline: 23678697]
3. Ponemon Institute. Sixth annual benchmark study on privacy & security of healthcare data URL: https://www.ponemon.org/local/upload/file/Sixth%20Annual%20Patient%20Privacy%20%26%20Data%20Security%20Report%20FINAL%206.pdf [WebCite Cache ID 6yYqScUxi]
4. Kruse C, Frederick B, Jacobson T, Monticone D. Cybersecurity in healthcare: a systematic review of modern threats and trends. Technol Health Care 2017;25(1):1-10. [doi: 10.3233/THC-161263] [Medline: 27689562]
5. Smet M. Cost characteristics of hospitals. Soc Sci Med 2002 Sep;55(6):895-906. [Medline: 12220092]
6. Riazul Islam SM, Daehan K, Humaun Kabir M, Hossain M, Kyung-Sup K. The internet of things for health care: a comprehensive survey. IEEE Access 2015;3:678-708. [doi: 10.1109/ACCESS.2015.2437951]
7. Brookman J. Protecting privacy in an era of weakening regulation. Harv Law Policy Rev 2015;9:355-374.
8. Institute of Medicine. Crossing the quality chasm: a new health system for the 21st century. The National Academies Press 2001:1-8. [doi: 10.17226/10027]
9. Menachemi N, Burke D, Diana M, Brooks R. Characteristics of hospitals that outsource information system functions. J Healthc Inf Manag 2005;19(1):63-69. [Medline: 15682678]

10. Menachemi N, Burke D, Diana M, Brooks R. Characteristics of hospitals that outsource information system functions. J Healthc Inf Manag 2005;19(1):63-69. [Medline: 15682678]

11. Angst CE, Block E, D'Arcy J, Kelley K. When do IT security investments matter? accounting for the influence of institutional factors in the context of healthcare data breaches. MISQ 2017 Mar 3;41(3):893-916. [doi: 10.25300/MISQ/2017/41.3.10]

12. -. 37% of hospitals perform cybersecurity incident response exercises annually. Trustee 2015 Sep;68(8):44, 1. [Medline: 26591236]

13. Park W, Seo S, Son S, Lee M, Kim S, Choi E, et al. Analysis of information security management systems at 5 domestic hospitals with more than 500 beds. Healthc Inform Res 2010 Jun;16(2):89-99 [FREE Full text] [doi: 10.4258/hir.2010.16.2.89] [Medline: 21818429]

14. Namoğlu N, Ulgen Y. Network security vulnerabilities and personal privacy issues in Healthcare Information Systems: a case study in a private hospital in Turkey. Stud Health Technol Inform 2013;190:126-128. [Medline: 23823398]

15. Zarei J, Sadoughi F. Information security risk management for computerized health information systems in hospitals: a case study of Iran. Risk Manag Healthc Policy 2016;9:75-85 [FREE Full text] [doi: 10.2147/RMHP.S99908] [Medline: 27313481]

16. Laabes EP, Nyango DD, Ayedima MM, Ladep NG. Physician use of updated anti-virus software in a tertiary Nigerian hospital. Niger J Med 2010;19(3):289-294. [Medline: 20845633]

17. Humaidi N, Balakrishnan V. Indirect effect of management support on users' compliance behaviour towards information security policies. Health Inf Manag 2018 Jan;47(1):17-27. [doi: 10.1177/1833358317700255] [Medline: 28537207]

18. Sittig DF, Singh H. A socio-technical approach to preventing, mitigating, and recovering from ransomware attacks. Appl Clin Inform 2016;7(2):624-632 [FREE Full text] [doi: 10.4338/ACI-2016-04-SOA-0064] [Medline: 27437066]

19. Harrell M, Bradley M. Data collection methods: semi-structured interviews and focus groups. RAND Corporation 2009:1-146.

20. Patton M. Qualitative research and evaluation methods. London: SAGE; 2002.

21. Luna-Reyes LF, Andersen DL. Collecting and analyzing qualitative data for system dynamics: methods and models. Syst Dyn Rev 2004;19(4):271-296. [doi: 10.1002/sdr.280]

22. Jalali M, Siegel M, Madnick S. Arxiv. 2017. Decision making and biases in cybersecurity capability development: evidence from a simulation game experiment URL: https://arxiv.org/abs/1707.01031 [accessed 2018-05-07] [WebCite Cache ID 6zEox3wUw]

23. Jalali M, Kaiser J, Siegel M, Madnick S. SSRN. 2017. The internet of things (IoT) promises new benefits—and risks: a systematic analysis of adoption dynamics of IoT products URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3022111 [accessed 2018-05-07] [WebCite Cache ID 6zEpECXlS]

24. Jalali M, Rahmandad H, Bullock S, Ammerman A. Dynamics of implementation and maintenance of organizational health interventions. IJERPH 2017 Aug 15;14(8):917. [doi: 10.3390/ijerph14080917]

25. Jalali MS, Ashouri A, Herrera-Restrepo O, Zhang H. Information diffusion through social networks: the case of an online petition. Expert Syst Appl 2016 Feb;44:187-197. [doi: 10.1016/j.eswa.2015.09.014]

26. Jarrett MP. Cybersecurity—a serious patient care concern. JAMA 2017 Oct 10;318(14):1319. [doi: 10.1001/jama.2017.11986]

Abbreviations

BYOD: bring your own device
CIO: chief information officer
CISO: chief information security officer
COBIT: Control Objectives for Information and Related Technologies
FDA: Food and Drug Administration
HIPAA: Health Insurance Portability and Accountability Act
HITECH: Health Information Technology for Economic and Clinical Health Act
HITRUST: Health Information Trust Alliance
IS: information security
IT: information technology
ITIL: Information Technology Infrastructure Library

XSL•FO

RenderX

Original Paper

A Mental Health Storytelling Intervention Using Transmedia to Engage Latinas: Grounded Theory Analysis of Participants' Perceptions of the Story's Main Character

MarySue V Heilemann[1], RN, PhD, FAAN; Adrienne Martinez[1], RN, MSN; Patricia D Soderlund[1], RN, PhD

School of Nursing, University of California, Los Angeles, Los Angeles, CA, United States

Corresponding Author:
MarySue V Heilemann, RN, PhD, FAAN
School of Nursing
University of California, Los Angeles
Factor Building, Room 5252
Box 956919
Los Angeles, CA, 90095 6919
United States
Phone: 1 310 206 4735
Fax: 1 310 206 3241
Email: mheilema@sonnet.ucla.edu

Abstract

Background: Transmedia storytelling was used to attract English-speaking Latina women with elevated symptoms of depression and anxiety to engage in an intervention that included videos and a webpage with links to symptom management resources. However, a main character for the storyline who was considered dynamic, compelling, and relatable by the target group was needed.

Objective: We conducted interviews with 28 English-speaking Latinas (target group) with elevated symptoms of depression or anxiety who participated in an Internet-accessible transmedia storytelling intervention. The objective of this study was to examine participants' perceptions of the lead character of the story. Development of this character was informed by deidentified data from previous studies with members of the target group. Critique of the character from a panel of therapists informed editing, as did input from women of the target group.

Methods: All interviews were conducted via telephone, audio-recorded, and transcribed. Data analysis was guided by grounded theory methodology.

Results: Participants embraced the main character, Catalina, related to her as a person with an emotional life and a temporal reality, reported that they learned from her and wanted more episodes that featured her and her life. Grounded theory analysis led to the development of one category (She "just felt so real": relating to Catalina as a real person with a past, present, and future) with 4 properties. Properties included (1) relating emotionally to Catalina's vulnerability, (2) recognizing shared experiences, (3) needing to support others while simultaneously lacking self-support, and (4) using Catalina as a springboard for imagining alternative futures. Participants found Catalina's efforts to pursue mental health treatment to be meaningful and led them to compare themselves to her and consider how they might pursue treatment themselves.

Conclusions: When creating a story-based mental health intervention to be delivered through an app, regardless of type, careful development of the main character is valuable. Theoretical guidance, previous deidentified data from the target group, critique from key stakeholders and members of the target group, and preliminary testing are likely to enhance the main character's relatability and appropriateness, which can increase sustained engagement.

(J Med Internet Res 2018;20(5):e10028) doi:10.2196/10028

KEYWORDS

depression; anxiety; transmedia storytelling; Internet; cell phone; mental health; eHealth; mood disorders; smartphone

XSL•FO
RenderX

Introduction

Background

Innovative ideas that harness the use of digital storytelling to engage users in mental health interventions involving smartphones, computers, tablets, or other devices merit exploration. Technological apps aimed at early detection, resource use, or Web-based mental health treatment that feature a human character [1,2], embodied conversational agent [3], virtual agent or coach [4,5], avatar, or fantasy character [6] can benefit from attention as to how characters affect user engagement. From Mexico [7] to the United States, photonovelas [8,9], webnovelas [10], and telenovelas to engage Latinos have been used in story-based research on a variety of physical health–related topics [11-16]. The expanding accessibility of the Internet and the growing popularity of smartphones worldwide support the use of story-based media to attract and reach previously unreached populations in terms of mental health or well-being. Many psychiatric outpatients own smartphones; Torous et al [17] found that 200 of the 320 psychiatric outpatients in their study owned a smartphone, and most (70.6%, 141/200) were willing to use their smartphones to run apps to monitor their mental health. Because smartphone use is higher for US Latino adults (aged 18-49 years) than it is for adults of other ethnicities [18], there is great promise for mental health engagement using smartphones with this population. Many US Latinos obtain important health information online. According to an American Trends Panel Survey, Latino adults used smartphones more often than non-Latino whites in 2014 to access online health information; that is, nearly three-fourths (72.94%, 755.5/1035) of Latino adults reported using their smartphones in the last year to access information about a health condition compared with 58.00% (1244.68/2146) of non-Latino whites [19]. Usage of smartphones is also high for individuals living with financial difficulties in the United States; for example, those with a household income of less than US $30,000 per year are more reliant on smartphones to access the Internet (13.00%, 195/1500) compared with individuals earning US $75,000 per year (1.00%, 15/1500) [19].

Recreational use of the Internet is high among US Latinos who speak English (94.0%, 771.7/821) and those who speak Spanish (86.0%, 583.9/679) [20]. Digital and electronic media are important potential avenues for reaching Latinos who may be struggling with untreated mental health issues because of their widespread use. For example, 51.00% (765/1500) of Latino adults sampled by the Pew Research Center reported they played video games on a cell phone or other device [21]. Many video games depict specific characters, but critics complain that few desirable Latino characters are featured in games [22]. When Latina characters are portrayed, they are often depicted stereotypically as maids with heavy accents in provocative clothing [23]. Although electronic games can provide effective, appealing mental health interventions, they must be designed with attention to characteristics of the clients they target, including culture and gender [24]. Attention to problematic depictions of leading characters in electronic games and other media technology apps is crucial to engagement. Historically,

entertainment-education media strategies in radio and television broadcast series have involved the use of characters whom audiences could identify with, which led to high viewer ratings and robust social discourse. This has been linked to viewers' engagement in health promotion activities and positive lifestyle choices, which were influenced by the behavior modeled by characters of the show [25].

Theoretical Guidance for Development of Characters in Story-Based Media

Theorists have explained that a key element of story-based approaches is how much an audience can relate to the main character. Albert Bandura's social cognitive theory (SCT) holds that characters who are perceived as interesting or relatable [26,15] are key to engagement with media. A viewer who feels a sense of connection with the main character will have a deeper engagement, which may be enduring over time, leading the viewer to come back for more. For example, in their research with an interactive video drama series aimed at helping men quit smoking, Bottoroff et al [1] found that men who related more strongly to the main character experienced higher levels of support from the series.

SCT [27] purports that communication involves psychosocial mechanisms that influence the thoughts, feelings, and actions of individuals. When communication takes the form of stories dramatized in media, viewers learn vicariously by observing characters as they deal with expected rewards, social consequences, and the influence of values. The story and characters inform viewers in terms of their thoughts and judgments. Due to the power of the story, SCT holds that direct engagement with story-based media can lead to behavior change by informing, motivating, or guiding viewers [26]. Viewers can be further influenced to change their behavior through personal discussions with other people who found the story worth talking about [28]. SCT can be used to guide the development of characters that are featured in any app to enhance the engagement of users. However, creation of a compelling and relatable main character requires thoughtful strategic development with input from the target group of users.

Transmedia: Storytelling Across Multiple Digital Platforms

Transmedia [29] is a dynamic, Internet-accessible media approach that features compelling characters and uses dramatic storytelling that extends across various digital platforms, including computers, tablets, and smartphones. When designed to engage, entertain, and educate, transmedia productions can be leveraged to attract and sustain the attention of audiences for the purposes of health [30]. Transmedia storytelling attracts users through compelling content and offers users decision-making power to choose when and how they will engage which parts of a story and for how long. This encourages the participant to be dynamically engaged rather than passively entertained. Participants might choose to expand their engagement via story extension videos, bonus videos (featuring a particular character speaking outside of the storyline about related topics), or by using other interactive components of the transmedia package. The participant decides which links to click and which content to explore and does so in a discreet way if

they are using their own personal device. The confidentiality of this feature is particularly attractive for smartphone users [31] who might be reluctant, afraid, or unwilling to share anything about their mental health symptoms, their distress, or their interest in mental health treatment resources.

A successful transmedia production to enhance sexual health among teens that has been very popular with Latinos is the Emmy-nominated webisode series, *East Los High*, created by Wise Entertainment, Inc. [32]. Having just completed its fourth season on Hulu, this English-language series featured an all-Latino cast of teen characters attending a fictitious high school in a fictitious community with plots involving romance, humor, sex, friendship, betrayal, success, and loss. The story extended beyond episodes to the show's companion webpage which had, during its first season alone, 123,728 unique visitors with 870,684 page views from all 50 US states and 163 countries [32]. Although not a prime focus of the show, a trio of story extender videos in the third season explored how one teen character sought support from a therapist as she dealt with past experiences of sexual abuse.

Objectives

Although the specific focus of how Latina adults deal with mental health symptoms has not been central to any transmedia production to date, US Latino audiences have a high interest in story-based media. According to Nielsen's Total Audience Report [33], 96% of US Latinos watched television in the second quarter of 2017, 92% engaged apps on a smartphone (including Web-based), and 78% accessed videos on a smartphone. Latinos continue to be over-represented in movie audiences in the US and Canada according to the Motion Picture Association of America [34]; although Latinos are only 18.00% (62,366,197.1/346,478,873) of the population, they accounted for 23.00% (79,690,140.8/346,478,873) of frequent moviegoers in 2016. However, characters depicted in story-based media such as streaming or network television shows or films include relatively few Latinos characters; in 2016, 5.80% (605.7/10,444) of characters with speaking roles were Latinos, and of them, only 37.9% (229.7/605.7) were female characters [35].

Due to the growing population of English-speaking Latinos in the United States and the expansive reach of transmedia via the Internet, our research team created a dramatic story-based transmedia intervention featuring a Latina main character with the aim of attracting and engaging untreated symptomatic Latinas. Informed by SCT, the goal of our overall mixed methods study was to engage viewers to increase their awareness of mental health symptoms, enhance early detection of symptoms, provide an interactive experience to help women contemplate their own situation, inform them of available resources for symptom management, and ultimately increase their behavioral intentions or actions taken to get help. The purpose of this qualitative analysis was to explore, describe, and interpret participants' perceptions of the main character
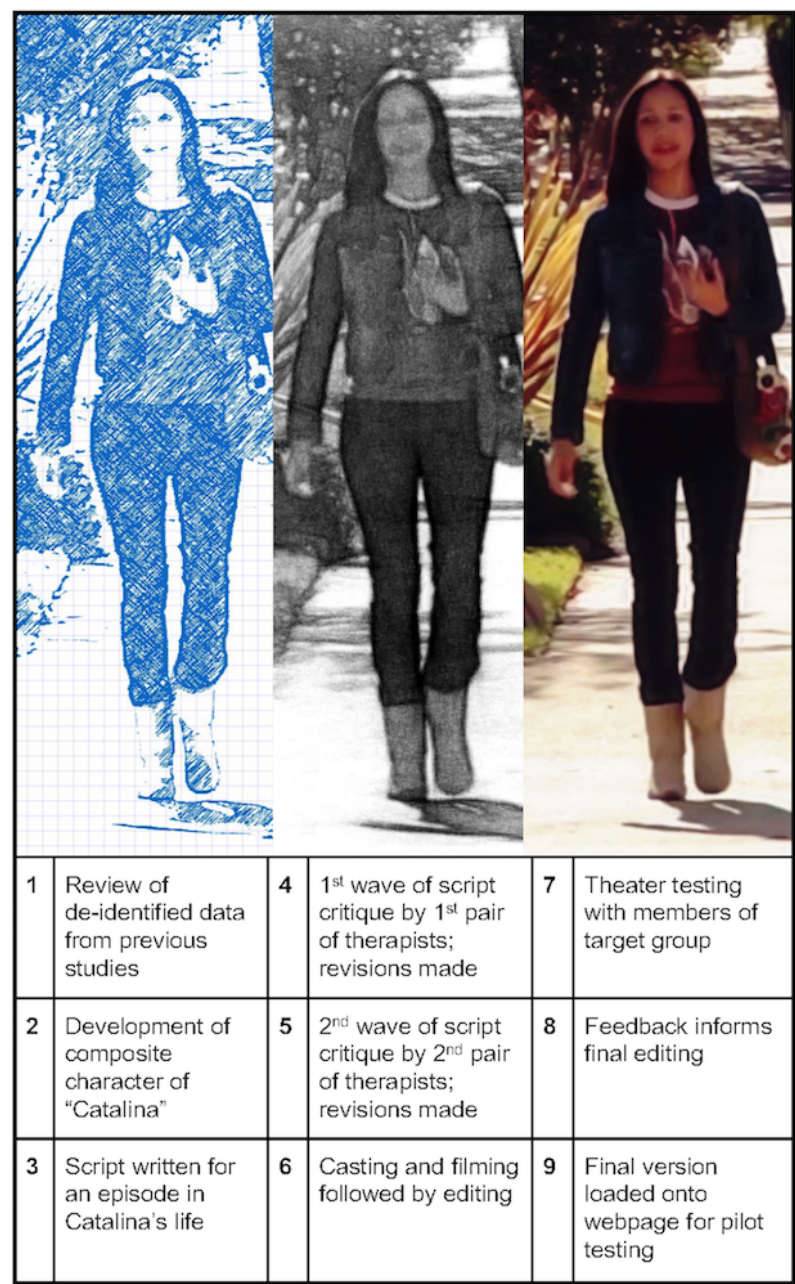
(Catalina) and to discuss our findings in relation to implications for future development of characters for various technological apps aimed at enhancing mental health or well-being.

Prior Work

Mindful of SCT [26-28], we invested in a rigorous process for creating our transmedia including the main character, Catalina, which balanced believability with interest, so participants would find it compelling, acceptable, and relevant to their own lives. Under the guidance of the University of California, Los Angeles Institutional Review Board (IRB), a composite sketch for Catalina was developed by the researcher informed by deidentified data from previous community-based, qualitative, or mixed methods studies with English-speaking Latinas of Mexican or Central American descent struggling with depression. From the array of attributes compiled from previous data, the researcher proposed a sketch of the main character, Catalina, including her backstory, life circumstances, concerns, motives, and goals. Catalina was proposed as a 28-year-old Latina who was a single mother of a 4-year-old daughter, living with her parents. The sketch and a basic storyline were presented to a Latino scriptwriter with extensive Hollywood experience who also served as our director and cinematographer; he created the settings, the script, and the trajectory of the drama. A community advisory board, comprising 4 therapists who had significant experience with Latinas in outpatient mental health settings (2 of whom identified as Latinas) read and critiqued the script in 2 waves, paying special attention to the main character. After each wave, the script was revised, and nuanced tailoring was done of the main character to enhance content validity in terms of believability, accuracy, and sociocultural appropriateness. Auditions were held in Hollywood, and Latino actors were hired; filming was done over a 5-day period. After the video scenes were filmed, editing was done. A composer provided a soundtrack. Subsequently, a total of 19 Latina adults from community settings participated in theater testing; feedback was collected via 1:1 interviews and focus groups. This feedback, including input about the portrayal of Catalina, informed the next phase of editing and the final version of the media plus the title, "Catalina: Confronting My Emotions." Finally, a 3-min introduction video was created. For this, the researcher created the script, one of the character actors did the voice over, and an editor wove together clips from the story to fit the script (see Figure 1).

The story of Catalina is told through events of her everyday life, but a tension develops as the drama unfolds. Over the course of 3 videos, Catalina is portrayed in her own home, on the street in her neighborhood, in a social situation at a party at the home of her best friend, and finally, in a city scene as she leaves a clinic and walks to a bus stop. In each scene, she is either interacting with others or dealing with her thoughts and feelings alone. Through this, Catalina becomes aware that she is experiencing mental health symptoms and grapples with what she should do about it.

Figure 1. Development of Catalina and her story.

1	Review of de-identified data from previous studies	4	1st wave of script critique by 1st pair of therapists; revisions made	7	Theater testing with members of target group
2	Development of composite character of "Catalina"	5	2nd wave of script critique by 2nd pair of therapists; revisions made	8	Feedback informs final editing
3	Script written for an episode in Catalina's life	6	Casting and filming followed by editing	9	Final version loaded onto webpage for pilot testing

The researcher also collaborated with a computer programmer to create a user-friendly project webpage, as was described elsewhere [2]. The goal was to create an interface that had ease of navigation [31], the same aesthetic quality of the videos, plus an attractive, organized display of links on a black background with photos and clear, simple captions so features could be accessed in a sequence. This included links to 3 different videos through which the story of Catalina unfolded (14-, 4-, and 4-min long). The videos were designed to attract the attention of users and enhance their motivation to click the link to watch the next video when the previous one concluded. The 3 videos completed 1 arc in Catalina's story (1 episode), so there was some resolution at the end. However, to continue with the transmedia intervention, the story-based videos were followed by another set of video logs and an interactive feature involving a character who portrayed Catalina's nurse therapist, whom we named

Veronica Sanchez, RN described elsewhere [2]. The final link on the website led to a blog written from Veronica's point of view that provided links and contact information for local and online mental health resources including hotlines. An interview was held with participants after engagement with the transmedia. At that time, an additional access code was provided to participants so that they could share the videos, blog, and links with friends and family members of their choice. Results showed that women rated their ability to relate to Catalina at 6.95 on a 10-point scale; approximately 6 weeks after first engaging with Catalina in the media, 75% (21/28) of participants reported they were still thinking about the story-based media, and 79% (22/28) were still talking about it with others [2].

Our in-depth analysis of participants' perceptions of Catalina was catalyzed, in part, by requests made by participants during

postmedia engagement interviews for additional episodes that featured her and her life beyond what was available in the intervention. Some participants expressed hope that there would be future episodes to further elucidate the role played by Catalina's mother; others wanted to know how Catalina's romantic life would unfold, what would happen if she pursued higher education, or how her engagement in therapy turned out. Overall, many participants wanted more opportunities to engage future episodes that depicted what might happen to Catalina in the future. In making a direct request for more from the research team, one woman summed up how the story-based videos featuring Catalina made an impact, provided an opportunity for reflection, and motivated her. She said:

> ...I think it would be great as the study goes if you guys could put more videos out there, more pamphlets, more information on how she [Catalina] actually got out of her situation, because for girls like me that are truly impacted by the videos and it's a reflection of how you're feeling, it's going to be a motivation to see [her] move on with her life. I don't know if you guys are going to have seminars. I don't know what the next step [is] with your studies, but it will be great to...have a video to see that she moved on.

In response to such direct requests from participants themselves, we designed this grounded theory (GT) analysis and opened ourselves to gain insight into what made Catalina so compelling to users.

Methods

Study Design

Our mixed methods design led to quantitative findings on the feasibility, acceptability, and limited efficacy of the transmedia intervention that have been reported elsewhere [2]. This qualitative analysis was guided by GT methodology to explore, analyze, describe, and interpret interview data on participants' perceptions of the main character of the transmedia storyline.

Recruitment and Sampling

After securing IRB approval, purposeful sampling was used to recruit a sample of 28 English-speaking Latina adults with elevated symptoms of depression or anxiety located in a Southern California metropolitan area, described elsewhere [2]. Latinas were eligible if they could read and speak English, were within the age range of 21 to 55 years, had access to the Internet, and met criteria for moderate to severe symptoms of depression (score of 10 or more on the Patient Health Questionnaire-9 [36]) or anxiety (score of 10 or more on the general anxiety disorder-7 [37]). After eligible women provided online consent, completed an online survey (described elsewhere [2]), engaged the media, and were interviewed, they received a US $60.00 gift card via US mail, text message, or email.

Data Collection

Within 72 hours of viewing the media, individual 1:1 interviews were conducted with participants over the telephone by the third author who is a psychiatric mental health nurse practitioner. Each interview was audiotaped with the participants' permission

and lasted an average of 45 min (range: 29-75 min). Since there was no attrition in this study, all women who participated in the transmedia intervention were interviewed (N=28). By design, each participant was interviewed only once. An interview guide was developed during the planning stages of the study to address a variety of topics; however, participant-initiated divergence from any topic was honored to encourage free sharing about any aspect of their experience with the transmedia or any related topic they found relevant during the interview. After the first participant was interviewed, subsequent interviews with other participants were influenced by earlier interviews, so some questions were added to enhance clarity as the interviewing process progressed. However, the focus of this analysis is limited to women's perceptions of the main character, Catalina, including what participants thought and felt about Catalina, what they found memorable about her, as well as how and why they did or did not relate to her. After verbatim transcription, all interview transcripts were checked for accuracy and loaded onto Atlas.ti (Qualitative Data Analysis by Scientific Software Development GMbH) [38] to help with data management.

Analysis Plan

Using the methodology of GT [39,40], initial coding was done using gerunds to focus on the action of the participant. To begin, each line of 10 transcripts was carefully considered and coded. The codes that occurred most frequently or were of particular significance were identified and aggregated into what Charmaz termed focused codes [40]. The focused codes were then used to guide analysis with the rest of the 18 interviews. For this analysis, all coded quotations that invoked women's perceptions of Catalina were identified, scrutinized, and discussed at length among the 3 researchers. The constant comparison of data with data across all 28 interviews was useful in sorting and analyzing quotations. Simultaneously, coding was done and analytic memos were written so each researcher could ask deeper questions of the data and check for patterns and connections across cases. Reflexive memos helped identify biases that might have been at play because of assumptions of the researchers. All 3 researchers had experience in mental health research and clinical care with Latinas; the second author self-identifies as a Latina. To enhance sensitivity in relation to the processes of coding, dialog among the researchers facilitated recognition of situational and contextual factors for each woman. The sharing of coding processes and memos among the researchers was facilitated by Atlas.ti [38]. Charts were created to organize data and allowed us to more effectively compare codes from all 28 women to develop and test properties within the category. Interactive sessions involving all 3 researchers provided opportunities for debate and discussion of our analytic products (the category and its properties). Moving deeper into analysis, diagrams and drawings were created to clarify dynamics within data.

Results

Sample

A detailed report of sampling, demographics, and media viewing habits of the final sample was reported elsewhere [2]. A total of 28 eligible Latinas between the ages of 21 and 48 years

provided informed consent, accessed the transmedia website, and completed the entire study.

Qualitative Analysis

One Overarching Category Depicted the Sample's Perception of Catalina as a Real Person

Knowing that participants expressed a desire for more episodes of Catalina's life, we moved forward to analyze participants' perceptions of this character, Catalina, and why they wanted a more prolonged engagement with her and her story. Analyses revealed that participants invoked the character of Catalina as a real person who had a past, present, and future. This led us to use temporality as a heuristic device which allowed a deeper analysis of the data and generated more questions, which is congruent with GT methodology [39]. We explored why participants related to the character of Catalina, how Catalina's timeline influenced their perceptions of her, and what utility their temporal view of Catalina served as they examined their own life histories, timelines, and circumstances.

Our use of the heuristic device of temporality led to the development of the category: She "just felt so real": Relating to Catalina as a real person with a past, present, and future. We

found that while participants were reflecting on a scene about Catalina's life, it was not uncommon for them to share that it sparked a past memory in a way that brought greater awareness or meaningful insight into their present reality. For some, these memories expanded into thoughts about an imagined future for themselves (see Figure 2).

The women seemed to move easily forward and backward on Catalina's timeline as they talked about their own. Although the specific details were different for each woman, participants mixed comments about Catalina with comments about themselves as they shared thoughts about their own histories.

Four Properties of the Category Revealed How Women Related to Catalina

We identified 4 properties of this category, which showed how women related to Catalina. These included (1) relating emotionally to Catalina's vulnerability, (2) recognizing shared experiences, (3) needing to support others while lacking self-support, and (4) using Catalina as a springboard for imagining alternatives. Each of these properties hinged on the reciprocal assumptions made by the women that, just as they had a past, a present, and an imagined future, so too did Catalina (see Figure 3).

Figure 2. Diagram showcasing one example of how a participant would compare an event on Catalina's timeline with her own past, present, or future.
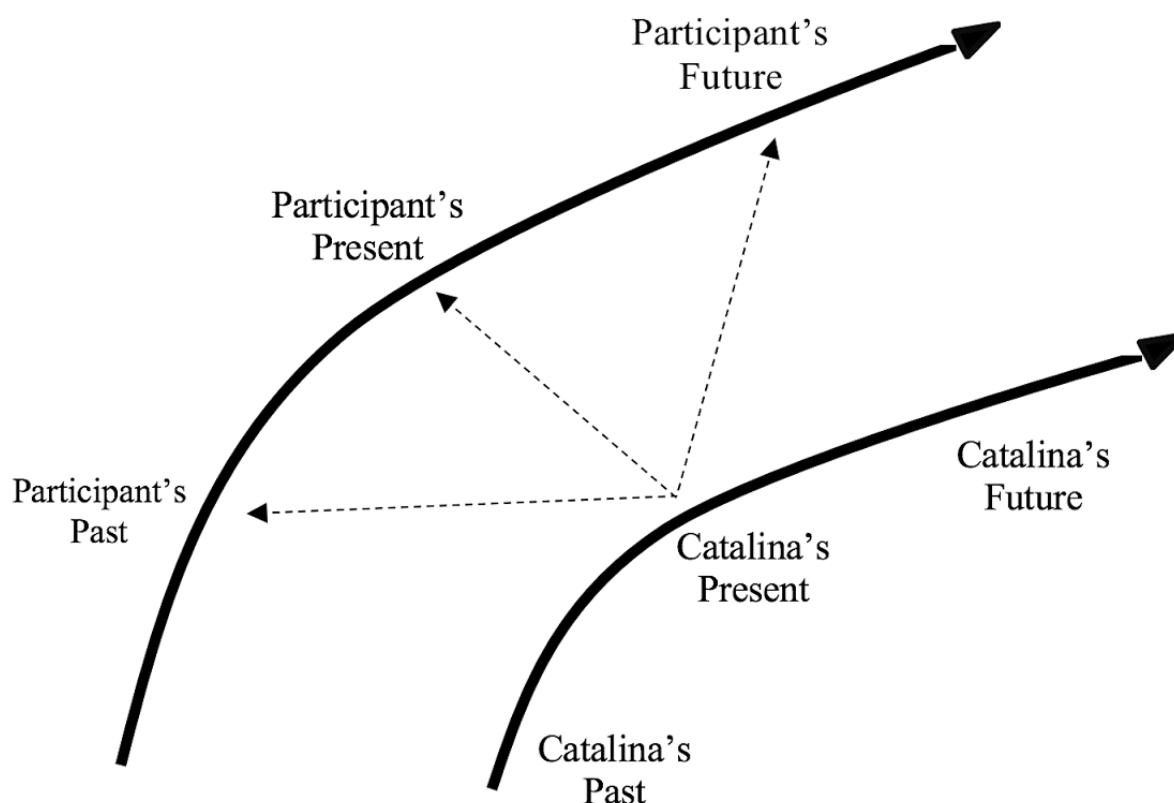
Figure 3. The overarching category (She "just felt so real": Relating to Catalina as a real person with a past, present, and future) and its four properties.
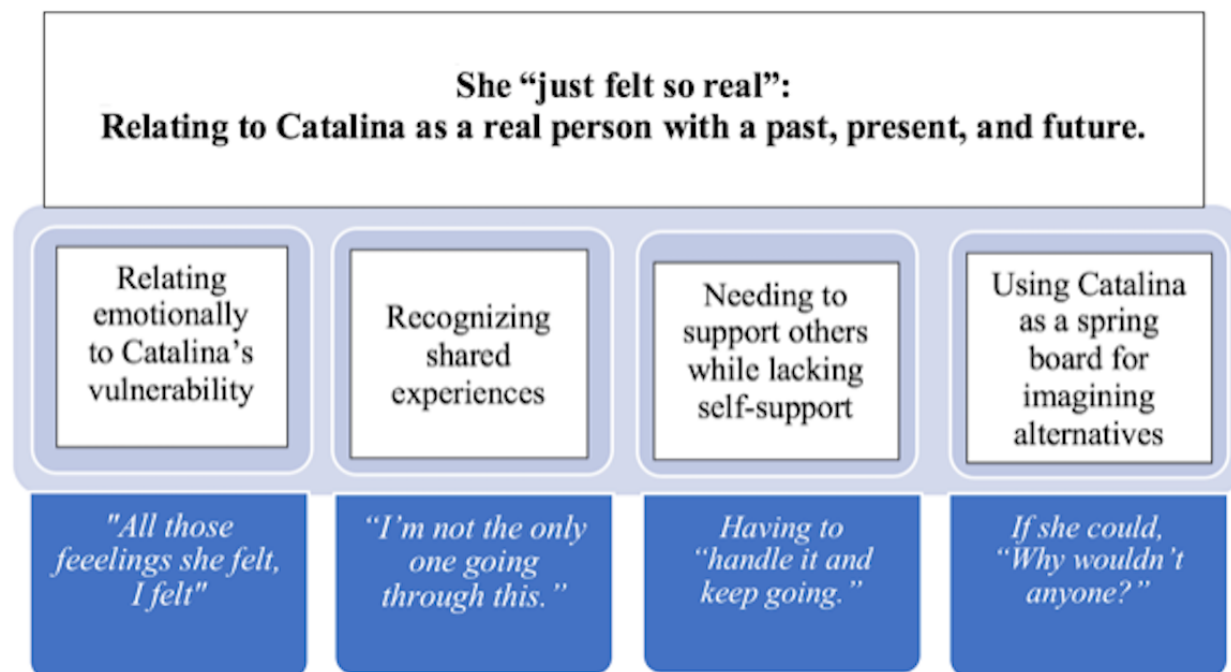
Relating Emotionally to Catalina's Vulnerability: "All Those Feelings That She Felt, I Felt"

When participants started talking about their perceptions of Catalina, their comments easily shifted to a focus on themselves and their own feelings. The women said they related to Catalina and used self-referential phrases, including, "That's me!" or "I remember that feeling." Women described how, watching Catalina, they found themselves able to reflect on their emotions although it made them feel vulnerable. One woman admitted:

> I know what it feels to be depressed, confused, lost, when you ask yourself, self-doubt[ing], who you really are.

Another connected to Catalina and a feeling of being lost. She explained that she related to a feeling of having "no sense of direction." Others related to Catalina's good intentions as someone who was just "trying to survive."

Several participants had similar emotional reactions to scenes in the drama, but this was not always the case. The same scene did not always stir the exact same emotions in every woman. Nonetheless, the emotional reactions of women connected them to the character. Reflecting on the drama, women seemed to juxtapose Catalina's emotional reality on their own in the past or the present. One woman stated:

> I related to the frustration that [Catalina] felt, to the loneliness, to overwhelmed feelings, a feeling where it's many things at once. So, overall, a lot of the emotions and feelings that she was feeling in the video, it's pretty much how I feel right now.

This woman, like many others, used comparisons of herself and Catalina to further articulate details of her current emotional state. Another woman said:

> I do feel at times, I feel like I'm going to like, blow up. And, that's one thing that she did mention about how sometimes she feels like her head is about to explode, and like, I feel the same way sometimes too. Like, sometimes I feel like it's too much to handle.

Some women related to how Catalina expressed herself emotionally, such as when she cried in one particular scene. After interpreting Catalina's feelings in that scene, one woman said:

> It's almost like if you sit there and have a pity party for yourself, and that's not the solution. I mean, I understand what the issue is, I just, the frustration of not knowing how to get out of that.

Speaking about that same scene, another woman extended Catalina's feelings of vulnerability and helplessness at that moment by expressing her own, exclaiming:

> ...it's almost as if there's nobody out there that can actually help you.

Other emotions spurred by Catalina were linked to feeling "down," "worthless," or lacking "confidence." Women related to what they interpreted as Catalina's feeling of "low self-esteem."

Recognizing Shared Experiences: "I'm Not the Only One Going Through This"

In addition to feelings, women related to the situational predicaments of Catalina's life. Specific aspects of Catalina's struggles at various time points throughout her story resonated with participants' struggles. This enhanced their ability to connect. Women said they realized that if Catalina was experiencing this, then they were not alone. One woman said:

> I'm not the only one going through this.

Another woman summed it up saying she related not only to the demographic and emotional situation of Catalina but also the way she engaged in negative self-talk:

> I related to a lot of the parts—like the child, the going out part, the being sad part about a guy—what else?...And then, she was just putting herself down.

Women recalled particular instances when they had grappled with specific problems and how their struggle felt similar to Catalina's struggle. For example, one woman related to Catalina's disappointment with a romantic partner saying:

> It just made me really sad, because I know how that feels when you get ditched by somebody. I just felt bad for her, even while she was just like a character in the movie, [she] just felt so real. Like, I can relate to her.

Women spoke about how their past impacted their present situations. For some, this included memories of a past event or situation that was troubling them now. Some women compared themselves and pointed out that they had made better decisions than Catalina did, for example, in romantic relationships. Others felt their experiences were worse than Catalina's. For example, thinking about the story, one participant identified with Catalina's decision to accept a ride home from a man whom she just met at a party. She shared a memory when something similar but "worse" had happened to her at a young age, a traumatic incident from her past. However, rather than discouraging her from continuing to engage the media, the scene encouraged a deeper connection with Catalina. She found the scene to be "really, really scary," but it led her to relate and then confide how this negative past experience "made" her "extra cautious" from that point on in her life, shaping the way she looked at the dangers of social encounters.

Various women related to Catalina's demographic situation as a single, working mother who was going to school. One woman emphasized how, like Catalina, she was "still" living with her parents. She likened her experience to Catalina's saying:

> I feel like I'm stuck as well.

Other participants described a connection with Catalina despite demographic differences. One woman who self-identified as a lesbian said she identified with Catalina although, unlike her, the character was involved romantically with a man and had a child. She said:

> It would be nice to feel like I would identify with her if she were gay or anything, but that's okay. I can still identify or relate [to her] on another note, in another, you know, whatever [an]other aspect of her life...I did feel as though I've been in her shoes before. I've never been a mother and I'm not a heterosexual woman. Actually, I identify as lesbian [but] I do have a working class family and I did reside at home with my family before moving in with my partner.

Although married, another woman who said that she had a supportive husband related to Catalina's sense of being alone. She said:

> [He] just doesn't understand me, like how Catalina would want people to understand her.

Many women found a connection because Catalina seemed to be asking herself deeper questions about which direction she should take in life. One woman related to Catalina's experience of not knowing what to do, saying:

> It doesn't have to be that I'm a single parent or whatever. Just whatever life throws at you, at times, we don't really know how to deal with it.

Needing to Support Others While Lacking Self-Support: Having to "Handle It and Keep Going"

Participants related to Catalina as a mother, daughter, or a partner who needed to support others even if she did not have much support for herself. Several women felt close to a breaking point which they also saw in Catalina. Women recounted how "stressed" and "overwhelmed" they felt in their present situation, how they lacked sufficient support to carry the load, and how it seemed that there was no end in sight for their struggle. To them, their troubles seemed to be ongoing into the future without a viable resolution. Women disclosed how they felt responsible in their role as a parent and tended to ignore their own needs by deciding to "handle it and keep going." Recalling a scene when Catalina was talking about her role as a mom, one woman said:

> [Catalina] saw her baby kind of lying there as she slept, and she's kind of like freaking out inside. I can relate to that. Even though I'm the adult, I don't know the answer to everything, but I have to keep it together because someone else depends on me.

Women related to how diligently Catalina "struggled" to hold things together for herself and her family. One woman said:

> She's just trying to succeed and do better for her kid.

Another woman confided how she felt compelled to work weekends so she could "pay all the bills" so she related to how "hard" it was for Catalina to be a single mother; she too lacked time to spend with her children or to "do stuff for yourself."

Women explained that, just like Catalina, the pressure to stay solvent financially brought stress that eroded confidence in themselves and their parenting abilities. One woman who lived with her spouse disclosed that she felt:

> ...overwhelmed with money situations and trying to survive and knowing if I'm a good Mom.

Another contrasted herself with Catalina and explained her belief that she had to make sacrifices when she became a parent, for better or worse, which eliminated options and reduced her freedom. She explained that as a single mother, she had not gone out "for fun" since her daughter was born 3 years ago, but she reasoned:

> ...that's just the things that you give up, in my opinion, [for] kids.

After watching Catalina interact with her family in the story, some women admitted how their efforts to support their family had negatively impacted their lives. They acknowledged that, like Catalina, they wanted to have more support. They found

themselves wishing they "could talk to someone" about their problems or that they could meet a romantic partner who would be supportive. But, many women felt that there was no way to actually meet their needs. They expected themselves to be able to handle all the responsibilities of motherhood on their own. A participant who was a single, working mother and student reluctantly admitted just how difficult her day-to-day life had become and how she was trying her best to hold herself together. She felt disillusionment in her struggle to persevere in her situation but still tried to be optimistic. Echoing the words that Catalina spoke in the story, she said:

> *I mean, everything's going good, but I do feel at times, like I'm going to like, blow up.*

Just how long participants would be able to carry on without sufficient support was left as an open question for many women.

Imagining Alternatives Through Catalina: If She Could, "Why Wouldn't Anyone?"

Participants related to Catalina and used comparison to imagine a future situation for themselves. When they perceived Catalina as navigating an experience they had not yet encountered, they contemplated this as a future possibility in their own lives. This led them to reflect on their past and, in tandem, to project into their future. Positive aspects of Catalina's story motivated some women, whereas others were catalyzed by negative aspects. For example, scenes that depicted Catalina before she got help for her emotions led some women to realize how much they did not want to live like that in the future. One woman said:

> *Seeing how the video shows that since her baby was an infant, I think she was just a newborn when [Catalina] split with her partner, and me seeing that years had gone by [in a subsequent scene], because the little girl was bigger and [Catalina's] still trying to get out of her situation and she's still depending on that father to provide for her financially, it made me see that this is not how I want to be. I don't want years to go by in the same situation. I don't want that. Yeah, I don't want that.*

Another woman used the video instrumentally. She thought about what she saw happening to Catalina, and this "pushed" her toward taking action. She said:

> *I was planning on getting [help] anyways, but once I saw the video it pushed me to not wait on it. I used the video to kind of help me on like okay, you can't wait on therapy because if you do, it can prolong. And as time goes on, and like Catalina seemed so much down, I think it's not healthy emotionally, physically, and mentally. So, the video did give me a little bit of a push to do it now instead of waiting and see where I can go, you know?*

Other women focused on what they perceived as positive in that Catalina acted on her own behalf to get help. Women talked about the scene in the story when Catalina was coming out of a neighborhood clinic where she had engaged in therapy sessions with Veronica, another Latina character who was a nurse therapist. They compared this situation with their own and reflected on choices they could have made in the past or might

make in the future to get help for themselves. Several women perceived Catalina to have greater self-awareness and more positive feelings at this point in the story. They saw her as having more self-esteem, as carrying herself with more confidence. They proceeded to imagine how, like Catalina, they could adjust their own paths moving forward. One woman said she learned helpful lessons from Catalina:

> *Catalina was teaching me how to start getting help and how to go get help and understand myself. Then after, it seems like, after the first session she was feeling much better about herself, and she enjoyed it [therapy].*

Through the scenes of Catalina's life, whether focusing on the negative or the positive, participants started envisioning a future strategy for themselves. One woman focused on Catalina's courage to get help and surmised:

> *Just seeing that, like, you know, the kind of life that she had, if she was able to find time and make it possible then, like, why wouldn't anyone?*

Others imagined seeking help but recognized personal fears about seeing a therapist. They related but looked past the fear to see the bigger picture. One woman said:

> *Like, towards the end of the video, you can tell her personality was changed. She seemed happy. And she was scared to see a therapist, like myself, and that was actually a positive outcome.*

Seeing Catalina as a Latina who overcame barriers to get help, one participant articulated her perspective about cultural expectations in her community, feelings of shame about needing mental health care, and the pressure to hide emotional issues. She said:

> *It's very difficult, [in] the Hispanic culture, to get help, because Hispanic women keep it to themselves for a very long time. And they feel ashamed. But on the video, it makes you just want to go get help because of the situation Catalina was having. Because, you know, sometimes you say to yourself, "No, this cannot happen to me. No, this will never happen to me." But, in reality, it is happening to you.*

Discussion

Principal Findings

Having received requests from Latinas in our sample for more videos featuring the main character of our story-based transmedia intervention, we did a GT analysis of participants' perceptions of the main character (Catalina) and found that they related to her socially and emotionally. Participants found Catalina's life and story to be familiar and realistic to the contexts and situations of their own lives and those of their friends or family. As discussed by Bandura [26], Catalina was relatable, compelling, and likable as a Latina who dealt with her situation by getting help. Our participants related to Catalina's emotional experiences and variously defended her, critiqued her, or supported her actions as reasonable in her situation. Although we did not ask how entertained women were

by the story, none of the women dropped out of the 6-week study after watching the introduction video, and all women chose to click to watch all 3 story-based videos of Catalina's story [2]. This implies that women were engaged in the transmedia intervention based on Calvo and Peters' [41] definition of engagement as a "combination of motivated commitment and sustained attention." During interviews, women spoke openly about Catalina, often relating to her emotional vulnerability and the ways she supported others even when she did not have the support she needed. An indication of the impact Catalina had on participants is that most (75%, 21/28) of the sample were still thinking and talking (79%, 22/28) about Catalina's story and the transmedia 6 weeks after first engaging with it [2].

The sample's engagement with the intervention seems to have been enhanced not only by their ability to relate to Catalina but by a dynamic of comparison that the production spawned. The women actively compared themselves with Catalina through a process that involved reflection and imagination, and they accomplished this with the help of temporality. Because they perceived Catalina to be like a real person with a past, present, and future, they used temporality in a fluid way to traverse her timeline and then reflect on or imagine themselves at various points in their own lives. Because of Catalina, they transported themselves to different time points in their own lives, remembering some of their own life events, reviewing their emotions, and comparing their own experiences with hers. Participants tracked the events of their lives and made sense of them while using the Catalina character as a benchmark. This allowed them to move beyond her story to focus on their own. They engaged in a dynamic reconsideration of the past, a focus on the present, or a sizing up of future possibilities.

In addition to relating to and comparing themselves to Catalina, participants seemed to take the next step described by Bandura [26] and learned from her. It is likely that various aspects of the character contributed to this phenomenon such as who she was from their point of view, how she acted and reacted in situations, and what she shared about herself and her feelings throughout the story. Just as Botoroff et al [1] found with their smoking cessation–oriented video drama series, it was the challenges of the main character that participants seemed to find the most compelling. The portrayal of both successes and failures in Catalina's life may have deepened women's ability to relate to her and to learn vicariously as they imagined alternatives for their own lives, as proposed with SCT [27,26]. The person women perceived Catalina to be in the past, present, or future and how she reacted at different time points, all seemed to serve as grist for learning about themselves as they compared and contrasted her life with their own. Through multiple connections and comparison points, participants could consider possible outcomes they desired, hoped for, feared, or wished to avoid.

By watching Catalina, some women voiced a newfound acceptance of their own experiences as normal. Some identified with Catalina's feelings of being overwhelmed to the point of exasperation, and others related to her disappointment in not being successful at navigating a romantic relationship. Ultimately, the women in our study found Catalina to be relatable as a person who, just like them, sometimes failed to make good decisions, and at other times took courageous steps.

When focusing on the future, Catalina served as a catalyst for women to think about possibilities for seeking help to enhance their emotional health. Engagement with Catalina and her story supported their ability to consider alternate decisional pathways they had not used before. They were able to contemplate certain actions they could take in the future by considering actions they saw Catalina take in the videos. They wondered if these could be a possible direction for them. For many, using Catalina as a springboard added a dynamic of hope or inspiration.

In the present, women across our sample identified specifically with how Catalina reasoned and responded to life experiences and how she conveyed emotions. By embracing her as an actual person, Catalina's ongoing timeline seemed to become a vehicle for a deeper connection to an inner life they recognized and valued in the present moment.

With a focus on Catalina's past decisions and predicaments, some women simultaneously projected themselves into the role and were able to identify patterns, make connections, or learn from past events in their own histories. This gave some insight they previously had not appreciated and brought a different way of thinking about their own life course and past circumstances. Observing Catalina's life history and past behaviors had the effect of allowing the women to contextualize the factors that set up their own situations and recall how they had reacted emotionally as a result. This identification with Catalina's past allowed some of the women to be more accepting of their past feelings and emotional reactions.

Some women saw Catalina as stronger and more insightful after she sought therapy, as depicted in the last story scene. Some expressed positive thoughts about this ending scene, reported feeling inspired, and linked this to a pivot in their willingness to pursue the option of therapy for themselves. This highlights the importance of what Calvo and Peters identified as the peak-end rule [41]—that is, of all aspects of their experience with our transmedia intervention, participants may have remembered most vividly both the peak of their experience with Catalina and also their feelings at the end of the media experience. Thus, as app designers, we were wise to pay careful attention to how we presented Catalina within the most exciting or gripping parts of story and also how we showcased her in the last scene of the story when she was talking about her feelings after having gone to therapy, a scene that many women found memorable. It is possible that participants' ability to relate to Catalina and desire for her storyline to continue beyond the intervention's bounds, was in part due to the women's perceptions at these key moments and what they ultimately remembered about her.

Overall, the character of Catalina seemed not only to represent who the women were but potentially who they might become. Connection to the character via the intervention seemed to represent a viable and important mechanism for them to continue to think and talk about her and potentially to learn more about themselves and their options.

Implications

A rigorous and layered understanding of how and why women related to Catalina provides guidance for researchers planning to create characters for story-based technology apps and interventions. The time we devoted to a solid methodological approach to developing Catalina was crucial for maximizing the power of the character; this included creating a composite based on previously collected, deidentified data from Latinas from the target group, allowing multiple phases of script critique from key stakeholders such as community-based therapists who were very familiar with the target group, theater testing with members of the target group to identify Latinas' reactions to her, and editing before pilot testing. As Singhal emphasized, relatable characters are more entertaining and can serve as valuable role models by positively influencing viewers' lifestyle choices and health promotion activities [25].

Commitment to the ethical protection of study participants who provide data to be used for character development is of high importance when creating media for interventions. The goal is to create characters, settings, storylines, plots, and scripts that participants will find relatable, culturally appropriate, and compelling but not offensive or demeaning. Thus, it is best for the research team to include members of the target group and their advocates, so their input is inherent in the development of the narrative. Active consultation with the IRB will help maximize appropriateness of characters and storylines.

Future research can build on this work. A randomized controlled trial is needed to test the effectiveness of our story-based videos compared with videos that are psychoeducational only, without characters. Another valuable study would be to share the transmedia program featuring the character of Catalina with women of other ethnic groups and investigate how relatable the character was to them and if it had an impact on confidence, help-seeking intentions, or action taken. In addition, the transmedia could be shared with men who are struggling with mental health symptoms themselves to determine the impact on help seeking or the impact on their actions to help other women in their lives who are struggling with untreated depression or anxiety symptoms.

Limitations

This transmedia intervention involved multiple videos but was limited to one story arc. Also, access to the transmedia webpage was tracked only for 6 weeks with this feasibility study. Since we used purposive sampling with a flier that described the study as involving the use of a smartphone, tablet, or computer, self-selection of participants may have occurred. Women who were already interested in and felt competent with technology may have been especially attracted to the study. Their past experience could have made their use of the technology more comfortable and may have positively impacted their experience with the transmedia overall or the characters in particular.

Conclusions

Stories with a compelling, desirable, relatable, socially, and culturally appropriate main character offers a powerful strategy for sustaining user adherence and engagement in a variety of technology apps that aim to enhance the mental health and well-being of various populations. Theory-based designs, data-informed characters, and culturally appropriate stories that are compelling to the target group are possible if collaborations include not only technology and mental health experts but also stakeholders and members of the target group. More creative experimentation holds abundant promise for deepening and extending user engagement in technology apps to enhance mental health.

Conflicts of Interest

None declared.

References

1. Bottorff JL, Sarbit G, Oliffe JL, Kelly MT, Lohan M, Stolp S, et al. 'If I Were Nick": men's responses to an interactive video drama series to support smoking cessation. J Med Internet Res 2015 Aug 10;17(8):e190 [FREE Full text] [doi: 10.2196/jmir.4491] [Medline: 26265410]

2. Heilemann MV, Soderlund PD, Kehoe P, Brecht ML. A transmedia storytelling intervention with interactive elements to benefit Latinas' mental health: feasibility, acceptability, and efficacy. JMIR Ment Health 2017 Oct 19;4(4):e47 [FREE Full text] [doi: 10.2196/mental.8571] [Medline: 29051135]

3. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. J Med Internet Res 2017 May 09;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]

4. Tielman M. A Virtual Agent for Post-Traumatic Stress Disorder Treatment. Delft, Netherlands: Delft University of Technology; 2018:434-437.

5. Tielman M, Neerincx MA, Brinkman WP. Generating situation-based motivational feedback in a PTSD E-health system. 2017 Presented at: 17th International Conference on Intelligent Virtual Agents; August 30, 2017; Stockholm, Sweden URL: https://repository.tudelft.nl/view/tno/uuid:1f4e7b87-e57e-46f6-8786-26a306bf3c11/ [doi: 10.1007/978-3-319-67401-8_54]

6.   Morie JF, Antonisse J, Bouchard S, Chance E. Virtual worlds as a healing modality for returning soldiers and veterans. Stud Health Technol Inform 2009;144:273-276. [Medline: 19592780]

7.   Villasenor M, Galindo C, Sabido M. Handbook for Reinforcing Social Values Through Day-Time TV Serials. Mexico City, MX: Mexican Institute of Communications Studies; 1982.

8.   Cabassa LJ, Molina GB, Baron M. Depression fotonovela: development of a depression literacy tool for Latinos with limited English proficiency. Health Promot Pract 2012 Nov;13(6):747-754 [FREE Full text] [doi: 10.1177/1524839910367578] [Medline: 21051325]

9.   Hernandez MY, Organista KC. Qualitative exploration of an effective depression literacy fotonovela with at risk Latina immigrants. Am J Community Psychol 2015 Sep;56(1-2):79-88. [doi: 10.1007/s10464-015-9729-8] [Medline: 25987298]

10. Andrade EL, Evans WD, Edberg MC, Cleary SD, Villalba R, Batista IC. Victor and Erika Webnovela: an innovative generation @ audience engagement strategy for prevention. J Health Commun 2015;20(12):1465-1472 [FREE Full text] [doi: 10.1080/10810730.2015.1018648] [Medline: 26252644]

11. Lalonde B, Rabinowitz P, Shefsky ML, Washienko K. La Esperanza del Valle: alcohol prevention novelas for Hispanic youth and their families. Health Educ Behav 1997 Oct;24(5):587-602 [FREE Full text] [Medline: 9307895]

12. Grigsby TJ, Unger JB, Molina GB, Baron M. Evaluation of an audio-visual novela to improve beliefs, attitudes and knowledge toward dementia: a mixed-methods approach. Clin Gerontol 2017;40(2):130-138. [doi: 10.1080/07317115.2016.1211210] [Medline: 28452671]

13. Larkey LK, McClain D, Roe DJ, Hector RD, Lopez AM, Sillanpaa B, et al. Randomized controlled trial of storytelling compared to a personal risk tool intervention on colorectal cancer screening in low-income patients. Am J Health Promot 2015;30(2):e59-e70. [doi: 10.4278/ajhp.131111-QUAN-572] [Medline: 25615708]

14. Crist JD. Cafecitos and telenovelas: culturally competent interventions to facilitate Mexican American families' decisions to use home care services. Geriatr Nurs 2005;26(4):229-232. [Medline: 16173135]

15. Wilkin HA, Valente TW, Murphy S, Cody MJ, Huang G, Beck V. Does entertainment-education work with Latinos in the United States? Identification and the effects of a telenovela breast cancer storyline. J Health Commun 2007;12(5):455-469. [doi: 10.1080/10810730701438690] [Medline: 17710596]

16. Vaughn S. Stroke and heart disease prevention education via telenovela: a focus group's evaluation. Rehabil Nurs 2012;37(5):215-219. [doi: 10.1002/rnj.53] [Medline: 22949274]

17. Torous J, Chan SR, Yee-Marie Tan S, Behrens J, Mathew I, Conrad EJ, et al. Patient smartphone ownership and interest in mobile apps to monitor symptoms of mental health conditions: a survey in four geographically distinct psychiatric clinics. JMIR Ment Health 2014;1(1):e5 [FREE Full text] [doi: 10.2196/mental.4004] [Medline: 26543905]

18. Nielsen. 2016 Jul 18. Television is still top brass, but viewing differences vary with age URL: http://www.nielsen.com/us/en/insights/news/2016/television-is-still-top-brass-but-viewing-differences-vary-with-age.html [accessed 2017-07-22] [WebCite Cache ID 6sB529Ali]

19. Smith A. Pew Research Center. 2015 Apr 01. U.S. Smartphone use in 2015 URL: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/ [WebCite Cache ID 6tTfMCGdy]

20. Brown A, Lopez G, Lopez MH. Pew Research Center. 2016 Jul 20. Internet use among Hispanics URL: http://www.pewhispanic.org/2016/07/20/1-internet-use-among-hispanics/ [WebCite Cache ID 6sB5lir09]

21. Duggan M. Pew Research Center. 2015 Dec 15. Gaming and gamers URL: http://www.pewinternet.org/2015/12/15/gaming-and-gamers/ [WebCite Cache ID 6x5qBpE22]

22. Lavandier AM. Elmundotech. 2016 May 31. ¿Dónde están?: Latino characters and video games URL: http://www.elmundotech.com/2016/05/31/latino-characters-and-video-games/ [WebCite Cache ID 6x5rSo0N8]

23. Aldama FL. Getting Your Mind/Body On: Latinos in Video Games. New York: Palgrave Macmillan; 2013.

24. Horne-Moyer HL, Moyer BH, Messer DC, Messer ES. The use of electronic games in therapy: a review with clinical implications. Curr Psychiatry Rep 2014 Dec;16(12):520 [FREE Full text] [doi: 10.1007/s11920-014-0520-6] [Medline: 25312026]

25. Singhal A, Rogers EM. The entertainment-education strategy in communication campaigns. In: Rice RE, Atkin CK, editors. Public Communication Campaigns, 3rd Edition. Thousand Oaks, CA: Sage Publications; 2000.

26. Bandura A. Social Cognitive Theory of mass communication. In: Bryant J, Zillman D, editors. Media Effects: Advances in Theory and Research. Hillsdale, NJ: Lawrence Erlbaum; 2002:122-153.

27. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice-Hall; 1986.

28. Bandura A. Social cognitive theory for personal and social change by enabling media. In: Signal A, Cody MJ, Rogers EM, editors. Entertainment-Education and Social Change: History, Research, and Practice. Mahwah, NJ: Lawrence Erlbaum; 2004:75-96.

29. Jenkins H. Convergence Culture: Where Old and New Media Collide. New York: NYU Press; 2006.

30. Weinreich NK. Transmedia storytelling for health: an integrated behavior change model. 2013 Presented at: National Conference on Health Communication, Marketing and Media; CDC; 2013; Atlanta, GA URL: https://www.researchgate.net/publication/267886669_Transmedia_Storytelling_for_Health_An_Integrated_Behavior_Change_Model [WebCite Cache ID 6ymqk3LFP]

XSL•FO
RenderX

31. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]

32. Wang H, Singhal A. East Los High: transmedia edutainment to promote the sexual and reproductive health of young Latina/o Americans. Am J Public Health 2016 Jun;106(6):1002-1010. [doi: 10.2105/AJPH.2016.303072] [Medline: 27077336]

33. Nielsen. 2017. The Nielsen total audience series URL: http://www.nielsen.com/content/dam/corporate/us/en/ reports-downloads/2017-reports/total-audience-report-q2-2017.pdf [accessed 2018-02-10] [WebCite Cache ID 6x7OAn5AF]

34. Motion Picture Association of America. 2016. Theatrical market statistics 2016 URL: https://www.scribd.com/document/ 342731953/2016-Theatrical-Market-Statistics-Report-MPAA [WebCite Cache ID 6x6zHQM89]

35. Smith SL, Choueiti M, Pieper K. USC Annenberg. 2016. Inclusion or invisibility? Comprehensive Annenberg report on diversity in entertainment URL: https://annenberg.usc.edu/sites/default/files/2017/04/07/ MDSCI_CARD_Report_FINAL_Exec_Summary.pdf [WebCite Cache ID 6x6zaeH7E]

36. 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] [Medline: 11556941]

37. Spitzer RL, Kroenke K, Williams JB, 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]

38. Atlas.ti. 2017. The No.1 software for qualitative data analysis URL: http://atlasti.com [WebCite Cache ID 6ta4j74Vt]

39. Corbin J, Strauss A. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. 4th ed. Thousand Oaks, CA: Sage Publications; 2015.

40. Charmaz K. Constructing Grounded Theory: A Practical Guide Through Qualitative Research. 2nd ed. Thousand Oaks, CA: Sage Publications; 2014.

41. Calvo RA. In: Peters D, editor. Positive Computing: Technology for Wellbeing and Human Potential. Cambridge, MA: The MIT Press; 2014:143.

Abbreviations

GT: grounded theory
IRB: institutional review board
SCT: social cognitive theory

Letter to the Editor

Comment on "An Online Intervention Comparing a Very Low-Carbohydrate Ketogenic Diet and Lifestyle Recommendations Versus a Plate Method Diet in Overweight Individuals With Type 2 Diabetes: A Randomized Controlled Trial"

Andrew Nathan Reynolds[1], BSc, MSc, PhD

Department of Human Nutrition, University of Otago, Dunedin, New Zealand

Corresponding Author:
Andrew Nathan Reynolds, BSc, MSc, PhD
Department of Human Nutrition
University of Otago
PO Box 56
Dunedin, 9010
New Zealand
Phone: 64 34795690
Email: andrew.reynolds@otago.ac.nz

This letter is regarding the recent publication of "An Online Intervention Comparing a Very Low-Carbohydrate Ketogenic Diet and Lifestyle Recommendations Versus a Plate Method Diet in Overweight Individuals With Type 2 Diabetes: A Randomized Controlled Trial" [1]. The authors have been innovative in their use of an online platform, and present findings that are both interesting and relevant. The effect size of the intervention is considerable, highlighting the potential of lifestyle interventions delivered in this manner. However, the interpretation of the trial results requires further discussion.

The authors attribute the improvements in body weight and glycaemic control to a very low-carbohydrate ketogenic diet and lifestyle recommendations. To single out the very low-carbohydrate ketogenic diet recommendation above the other lifestyle recommendations made in the intervention is unsupported. The paper states that the self-reported dietary assessment method "should not be considered validated." Furthermore, data to demonstrate dietary change due to the intervention, such as the measured urinary acetoacetate, has not been provided.

The observed effect of the intervention is at least as likely due to the other lifestyle recommendations, or recommendations acting in concert. Advice to increase physical activity [2] and

sleep [3] as well as the provision of psychological support [4], such as mindfulness training and goal setting, were included in the intervention and have been shown in meta analyses to independently improve the primary outcomes of the current paper. By contrast, a recent meta analysis of trials with type 2 participants randomised to lower or higher carbohydrate diets showed no difference in body mass index and blood glucose control (HbA_{1c}) at 12 months, although there was a transient effect on HbA_{1c} at 3 and 6 months [5]. To further demonstrate the potentially unwarranted focus on the very low-carbohydrate ketogenic diet, the intervention of the current paper also included training on: awareness of hunger, fullness, craving, taste satisfaction and triggers for overeating, scheduling, noticing and savouring positive events, developing self-compassion, practicing positive reappraisal, gratitude, acts of kindness, and being aware of personal strengths. In contrast, the control group only received dietary advice. Given these disparities and the lack of supporting data it is unfeasible of the authors to attribute a change due to their recommendation to follow a very low-carbohydrate ketogenic diet.

Differences in outcomes between the intervention and control groups may further be attributed to the dissimilar level of participant support. Participants in the intervention group were

contacted once a week for 16 weeks, and then once a fortnight for a further 16 weeks (24 times). The control group were contacted once a week for four weeks, and then once every four weeks (11 times). This disparity in the level of support and engagement is a probable driver for the participant drop out rate in the control (6 of 13) and intervention (1 of 12).

That the sample size is grossly inadequate is evident by differences in the baseline attributes of participants who, although randomised, were on average 18.8kg heavier and 5.2 years younger in the intervention group than those in the control group. An estimate for an adequate sample size would provide greater context to the results of this pilot study with the author's previous work in this area providing them with the necessary data to do so. Finally, that change in body weight is the sole measure of body fatness is a further concern, with additional markers not confounded by baseline body weight required.

The findings of this study are undoubtedly interesting. The authors should be acknowledged for their work highlighting the potential of online interventions to deliver multiple lifestyle recommendations. The findings also suggest at the level of support required to enable behavioural change towards healthier behaviours. What the findings do not support is the promotion of very-low carbohydrate ketogenic diets.

Conflicts of Interest

None declared.

References

1. Saslow LR, Mason AE, Kim S, Goldman V, Ploutz-Snyder R, Bayandorian H, et al. An Online Intervention Comparing a Very Low-Carbohydrate Ketogenic Diet and Lifestyle Recommendations Versus a Plate Method Diet in Overweight Individuals With Type 2 Diabetes: A Randomized Controlled Trial. J Med Internet Res 2017 Feb 13;19(2):e36 [FREE Full text] [doi: 10.2196/jmir.5806] [Medline: 28193599]
2. Umpierre D, Ribeiro PAB, Kramer CK, Leitão CB, Zucatti ATN, Azevedo MJ, et al. Physical activity advice only or structured exercise training and association with HbA1c levels in type 2 diabetes: a systematic review and meta-analysis. JAMA 2011 May 04;305(17):1790-1799. [doi: 10.1001/jama.2011.576] [Medline: 21540423]
3. Cappuccio FP, Taggart FM, Kandala NB, Currie A, Peile E, Stranges S, et al. Meta-analysis of short sleep duration and obesity in children and adults. Sleep 2008 May;31(5):619-626 [FREE Full text] [Medline: 18517032]
4. Ismail K, Winkley K, Rabe-Hesketh S. Systematic review and meta-analysis of randomised controlled trials of psychological interventions to improve glycaemic control in patients with type 2 diabetes. Lancet 2004 May 15;363(9421):1589-1597. [doi: 10.1016/S0140-6736(04)16202-8] [Medline: 15145632]
5. Snorgaard O, Poulsen GM, Andersen HK, Astrup A. Systematic review and meta-analysis of dietary carbohydrate restriction in patients with type 2 diabetes. BMJ Open Diabetes Res Care 2017;5(1):e000354 [FREE Full text] [doi: 10.1136/bmjdrc-2016-000354] [Medline: 28316796]

XSL•FO

RenderX

Letter to the Editor

Authors' Reply: Comment on "An Online Intervention Comparing a Very Low-Carbohydrate Ketogenic Diet and Lifestyle Recommendations Versus a Plate Method Diet in Overweight Individuals With Type 2 Diabetes: A Randomized Controlled Trial"

Laura Saslow[1], PhD; Ashley E Mason[2], PhD; Sarah Kim[2], MD; Veronica Goldman[2]; Robert Ploutz-Snyder[3], PhD; Hovig Bayandorian[4]; Jennifer Daubenmier[5], PhD; Frederick M Hecht[2], MD; Judith T Moskowitz[6], PhD

[1]School of Nursing, Department of Health Behavior and Biological Sciences, University of Michigan, Ann Arbor, MI, United States

[2]University of California, San Francisco, San Francisco, CA, United States

[3]University of Michigan, Ann Arbor, MI, United States

[4]University of California Berkeley, Berkeley, CA, United States

[5]Institute of Holistic Health, Department of Health Education, San Francisco State University, San Francisco, CA, United States

[6]Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

Corresponding Author:
Laura Saslow, PhD
School of Nursing
Department of Health Behavior and Biological Sciences
University of Michigan
Room 2178, 400NIB
Ann Arbor, MI,
United States
Phone: 1 (734) 764 7836
Email: laura.saslow@gmail.com

This letter is in response to the letter from Dr Andrew Reynolds [1] about our publication, "An Online Intervention Comparing a Very Low-Carbohydrate Ketogenic Diet and Lifestyle Recommendations Versus a Plate Method Diet in Overweight Individuals With Type 2 Diabetes: A Randomized Controlled Trial" [2].

As Dr Reynolds notes, our study had several differences between the two groups including different diets, psychological tools, sleep education, and number of lessons. Given the combination of intervention components, we did not attribute all of the differences in outcomes to the nutritional composition of the diet, but to the diet "and lifestyle recommendations." It was our goal, if this preliminary trial showed promise, to follow up with factorial screening experiments to better assess each component of our multicomponent very low-carbohydrate diet intervention.

We are currently testing our very low-carbohydrate diet intervention in this way, and we hope to vary the dietary component in future research. We agree that a randomized trial that varies only diet is an ideal way to ascertain to what extent the nutritional composition of the diet, specifically, contributes to the outcomes.

However, important information can be learned about the overall effects of diet and lifestyle interventions without such screening experiments and intervention optimization. For example, the landmark Diabetes Prevention Program trial, with more than 3000 patients with prediabetes, compared a multicomponent low-fat, calorie-reduced diet and lifestyle intervention to metformin or medicine placebo groups [3]. The intact, multicomponent program is now used nationally, but to our knowledge, no research has carefully varied the dietary

component to examine if the low-fat dietary recommendations are optimal. Despite this, this multicomponent program is supported by the Centers for Disease Control and Prevention and will be soon reimbursed by Medicare. Even so, we believe that the field needs to do more screening experiments and intervention optimization.

We acknowledge the baseline differences on some of our outcome measures. This can happen even with classic randomized controlled trials. Nevertheless, our statistical evaluations emphasize interaction effects and, as such, they emphasize relative changes and fully incorporate baseline information. Therefore, we disagree that the interaction effects are due primarily to these baseline differences. With respect to a post-hoc power analysis, a power analysis reveals the *likelihood* of observing a significant effect. We disagree that once a study has been completed there is any value to a post-hoc power analysis. After study completion, we know for certain whether effects are significant ("power" = 100% for significant effects, and 0% for non-significant effects), and these results utilize observed variability rather than estimates of variability that would have informed a pre-study analysis.

We agree that meta-analyses are ideal for judging the effects of different diets, however, we have a different interpretation of the meta-analysis to which he refers [4]. This meta-analysis of carbohydrate-restricted trials in type 2 diabetes itself notes that the "effect on glycemic control increased with the reported degree of carbohydrate restriction," and that "recent trials suggest that LCD [low-carbohydrate diets] may be superior to HCD [high-carbohydrate diets] with respect to glucose level and postprandial excursions, but only as long as the subject adheres to the diet." This suggests that when individuals can maintain adherence to long-term carbohydrate restriction they experience improved glycemic control. Moreover, only two of the seven studies in this meta-analysis followed participants beyond 12 months. We view this meta-analysis as support for the short-term benefits of low-carbohydrate diets in the treatment of type 2 diabetes, and that it highlights a need to conduct longer trials of low-carbohydrate diets that focus on long-term dietary adherence.

We appreciate the opportunity to respond to Dr Reynolds' letter.

Conflicts of Interest

Frederick Hecht is on the Scientific Advisory Board for Virta Health. No other author declares any conflict of interest.

References

1. Reynolds AN. Comment on "An Online Intervention Comparing a Very Low-Carbohydrate Ketogenic Diet and Lifestyle Recommendations Versus a Plate Method Diet in Overweight Individuals With Type 2 Diabetes: A Randomized Controlled Trial". J Med Internet Res 2018;10(5):e180 [FREE Full text] [doi: 10.2196/jmir.7672]
2. Saslow RL, Mason EA, Kim S, Goldman V, Ploutz-Snyder R, Bayandorian H, et al. An Online Intervention Comparing a Very Low-Carbohydrate Ketogenic Diet and Lifestyle Recommendations Versus a Plate Method Diet in Overweight Individuals With Type 2 Diabetes: A Randomized Controlled Trial. J Med Internet Res 2017 Feb 13;19(2):e36 [FREE Full text] [doi: 10.2196/jmir.5806] [Medline: 28193599]
3. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 2002 Feb 07;346(6):393-403 [FREE Full text] [doi: 10.1056/NEJMoa012512] [Medline: 11832527]
4. Snorgaard O, Poulsen GM, Andersen HK, Astrup A. Systematic review and meta-analysis of dietary carbohydrate restriction in patients with type 2 diabetes. BMJ Open Diabetes Res Care 2017;5(1):e000354 [FREE Full text] [doi: 10.1136/bmjdrc-2016-000354] [Medline: 28316796]

XSL·FO

RenderX

XSL•FO

RenderX