Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 20 (2018), Issue 11 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

Short Paper

Review

The Generalizability of Randomized Controlled Trials of Self-Guided Internet-Based Cognitive Behavioral Therapy for Depressive Symptoms: Systematic Review and Meta-Regression Analysis

Lorenzo Lorenzo-Luaces[1], PhD; Emily Johns[1], BS; John R Keefe[2], MA

[1]Department of Psychological and Brain Sciences, Indiana University Bloomington, Bloomington, IN, United States

[2]Department of Psychology, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Lorenzo Lorenzo-Luaces, PhD
Department of Psychological and Brain Sciences
Indiana University Bloomington
1101 East 10th Street, Room 158
Bloomington, IN, 47401
United States
Phone: 1 787 222 2493
Email: lolorenz@indiana.edu

Abstract

Background: Self-guided internet-based cognitive behavioral therapies (iCBTs) for depressive symptoms may substantially increase accessibility to mental health treatment. Despite this, questions remain as to the generalizability of the research on self-guided iCBT.

Objective: We sought to describe the clinical entry criteria used in studies of self-guided iCBT, explore the criteria's effects on study outcomes, and compare the frequency of use of these criteria with their use in studies of face-to-face psychotherapy and antidepressant medications. We hypothesized that self-guided iCBT studies would use more stringent criteria that would bias the sample toward those with a less complex clinical profile, thus inflating treatment outcomes.

Methods: We updated a recently published meta-analysis by conducting a systematic literature search in PubMed, MEDLINE, PsycINFO, and EMBASE. We conducted a meta-regression analysis to test the effect of the different commonly used psychiatric entry criteria on the treatment-control differences. We also compared the frequency with which exclusion criteria were used in the self-guided iCBT studies versus studies of face-to-face psychotherapy and antidepressants from a recently published review.

Results: Our search yielded 5 additional studies, which we added to the 16 studies identified by Karyotaki and colleagues in 2017. Few self-guided iCBT studies excluded patients with severe depressive symptoms (6/21, 29%), but self-guided iCBT studies were more likely than antidepressant (14/170, 8.2%) studies to use this criterion. However, self-guided iCBT studies did not use this criterion more frequently than face-to-face psychotherapy studies (6/16, 38%). Beyond this, we found no evidence that self-guided iCBTs used more stringent entry criteria. Strong evidence suggested that they were actually less likely to use most entry criteria, especially exclusions on the basis of substance use or personality pathology. None of the entry criteria used had an effect on outcomes.

Conclusions: A conservative interpretation of our findings is that the patient population sampled in the literature on self-guided iCBT is relatively comparable with that of studies of antidepressants or face-to-face psychotherapy. Alternatively, studies of unguided cognitive behavioral therapy may sample from a more heterogeneous and representative patient population. Until evidence emerges to suggest otherwise, the patient population sampled in self-guided iCBT studies cannot be considered as less complex than the patient population from face-to-face psychotherapy or antidepressant studies.

(J Med Internet Res 2018;20(11):e10113) doi:10.2196/10113

KEYWORDS

depression; psychotherapy; CBT; internet-based therapy; pharmacotherapy; generalizability; exclusion criteria; cognitive therapy; telemedicine; drug therapy; patient selection

Introduction

Background

Depression is one of the leading causes, if not the leading cause, of disability worldwide [1,2]. A well-recognized barrier to reducing the level of disability associated with depression and other common mental disorders is the scarcity of available treatment providers [3]. Although there is a scarcity of providers, the research data support a wide variety of treatments for depression that vary in their modality (eg, self-management, psychotherapy, medications), as well as in how time and resource intensive they are for patients and providers. The combination of antidepressants and psychotherapy is widely seen as providing the best cost-benefit ratio for severe depression [4]. However, it is nearly impossible to advocate this or any other treatment approach as a first-line intervention because depression is highly heterogeneous in its prognosis [5-7]. For example, 50% to 75% of patients with first-onset major depressive episodes recover within a 3- to 6-month period, but 15% to 20% of cases are chronic [8-11]. Similarly, although 50% of patients recover and do not experience a relapse in 20- to 30-year follow-ups, approximately 35% of individuals experience a recurrent course [9,12,13]. Moreover, the presentation of depressive symptoms is highly heterogeneous. Approximately half of cases are classified as mild to moderate [14], but a proportion of cases involve psychotic symptoms [15] or suicidal risk (eg, 27%-53% [16]). As well, there is a substantial degree of comorbidity with other disorders [14].

A major advancement in the study of treatments for depression is the discovery that self-guided internet-based cognitive behavioral therapy (iCBT) can be more effective than control conditions [17]. In these interventions, patients complete interactive Web-based programs based on the principles of cognitive behavioral therapy with no therapeutic support, although sometimes technical support is available. The efficacy of these interventions can be enhanced by therapist support, although this effect may be smaller than previously thought (eg, standardized mean difference symptoms = 0.27 [18-20]). The high prevalence of depression and the ubiquity of internet access and mobile phone ownership make self-guided iCBTs hugely promising in reducing the burden of disability associated with depressive symptoms, even if they were somewhat less effective than guided iCBTs [21]. Kazdin and Blase [3] have called attention to the fact that the number of individuals with mental disorders such as depression far outnumbers the number of mental health providers available to deliver treatment. Thus, self-guided iCBT has the potential for a greater public health impact than guided iCBT or other forms of treatment that require contact with a trained professional [21]. However, a key concern in evaluating the self-guided iCBT treatment literature is the degree to which the patient population in randomized controlled trials (RCTs) is representative of the heterogeneous nature of the prognosis, severity, and comorbidity found in depressed patients. More than 20 years ago, Seligman [22] pointed out that the literature attesting to the efficacy of face-to-face psychotherapy was limited by virtue of the inclusion and exclusion criteria used, which excluded patients with subthreshold symptoms as well as comorbid conditions. Westen and Morrison [23] quantified the nature of this problem, reporting that most (68%) patients with depression were excluded from a typical psychotherapy study because of suicidality, comorbid disorders, especially substance use, and subclinical symptoms. Furthermore, these authors reported that the number of participants excluded from psychotherapy trials was related to the study-level effect size, such that studies that excluded more patients tended to find larger effect sizes for psychotherapy (see also van der Lem et al [24]). Zimmerman et al [25] made similar observations in the literature on antidepressants. More recently, Zimmerman and colleagues [26-29] reported on the inclusion and exclusion (henceforth *entry*) criteria for RCTs of antidepressant medications. They stated that most studies excluded patients on the basis of a minimum symptom threshold, suicidality, psychotic features, and substance use disorders.

Objective

We sought to review the entry criteria used in trials of self-guided iCBT. We did this by updating and coding the studies from the latest meta-analysis of self-guided iCBT, which was published by Karyotaki et al [17]. In addition to reporting the overall frequency with which specific criteria were used, we sought to explore the relationship between the specific entry criteria and the study-level outcome. Finally, to provide some context for evaluating the frequency with which self-guided iCBT studies used different exclusion criteria, we compared the frequency of use of entry criteria in self-guided iCBT trials versus the rate of inclusion criteria used in face-to-face psychotherapy trials and trials of antidepressant medications. The aim of this analysis was for face-to-face psychotherapy and antidepressant studies to serve as a kind of benchmark against which to compare the self-guided iCBT literature. A prior study suggested that face-to-face psychotherapy trials were less likely than antidepressant trials to use most of the specific entry criteria coded by the study authors [30]. Other studies similarly suggested that psychotherapy studies may be somewhat more generalizable than pharmacotherapy studies for adolescent depression [31], borderline personality disorder [32], social anxiety [33], and posttraumatic stress disorder [34], although not generalized anxiety disorder [35]. Based on prior data [23], we hypothesized that the use of more stringent entry criteria would be associated with better outcomes. However, because studies of self-guided iCBT are often fully remote and lack human support, we hypothesized that studies of self-guided iCBT would use more stringent psychiatric entry criteria than would face-to-face psychotherapy and pharmacotherapy trials.

Methods

Identification and Rating of Studies

We obtained RCTs exploring the efficacy of self-guided iCBT for depression in adults by referencing and updating the latest meta-analysis of self-guided iCBT for depression [17]. Those authors consulted a broad database of psychological treatments for depression that was constructed from a systematic literature search of free terms combining “psychotherapy” and “depression” in PubMed, EMBASE, PsycINFO, and the Cochrane Library published up to January 1, 2016. We updated

the review by applying the same search criteria (see Multimedia Appendix 1), narrowing our search for internet-based interventions. We included studies if they included adults (>18 years of age) with a diagnosis of major depressive disorder on a psychiatric semistructured interview, with elevated symptoms of depression (ie, any specific cutoff score on a depression questionnaire), or who were seeking or undergoing treatment for depression. Two of the authors (LLL, EJ) rated all articles for the presence of common psychiatric inclusion and exclusion criteria used in treatment studies [26], resolving discrepancies by consensus. Descriptive analyses summarizing the specific entry criteria of the iCBT trials are presented. We rated risk of bias according to the Cochrane Collaboration risk-of-bias assessment tool [36]. We rated the primary outcome studies. When we could not determine a specific domain category from the main outcome study, we rated protocol articles if they were available or any publicly available study registry (eg, ClinicalTrials.gov).

Meta-Regression

We calculated Hedges *g* to quantify the difference between self-guided iCBT and control conditions on symptoms of depression. Hedges *g* is derived from the difference between the average posttreatment scores on self-reported measures of depressive symptoms in the 2 groups (ie, self-guided iCBT group vs the control group), divided by the pooled standard deviation while adjusting for small-sample bias. For entry criteria that had sufficient representation in the dataset (ie, ≥3 studies used them), we conducted meta-regressions, using the R statistical computing language version R-3.5.0. (R Foundation) package metafor [37], to examine the relationship between using specific entry criteria and outcomes. First, we conducted individual meta-regressions in which we regressed the study-level effect size on each of the exclusion criteria used. Then, we regressed the number of exclusion criteria used on the study-level outcomes. Finally, we conducted a simultaneous meta-regression in which we regressed outcomes on all entry criteria.

Meta-regressions were run in a random effects meta-analytic framework, using exclusion criteria either as dummy-coded categorical variables (0/1) or as an ordinal variable (ie, the number of exclusions). In line with current best practices for calculating confidence intervals with more accurate coverage and less-inflated type I error, we used the Sidik-Jonkman random effects estimator and the Hartung-Knapp adjustment [38,39].

Benchmarking Against Antidepressant and Face-to-Face Psychotherapy

To provide a rough index of how entry criteria for guided iCBT studies compare with those of other treatments, we drew on recent reviews of the entry criteria of trials of adult depression with antidepressants [26-29] and face-to-face psychotherapy [30]. These reviews employed searches in PubMed, EMBASE, and PsycINFO, as well as individual meta-analyses and specific journals, to identify acute treatment outcomes studies for depression. The search was limited to studies in which antidepressants or face-to-face psychotherapy was compared with a control condition, so studies of multiple treatments (eg, 2 psychotherapies) were only included if a control condition was used. There are systematic differences between antidepressant and face-to-face pharmacotherapy studies regarding the type of controls employed [40]. Virtually all pharmacotherapy studies use a pill placebo, whereas the efficacy of face-to-face psychotherapy, as well as self-guided iCBT, is tested with a more diverse mix of controls, including a waiting list, treatment as usual, pill placebos, and other conditions (eg, relaxation) that are intended as a control for nonspecific effects (eg, attention). Thus, we searched for face-to-face psychotherapy trials in which a waiting list, treatment as usual, or placebo control was used. We excluded trials if they focused exclusively on comorbidities (eg, only depression and alcohol use), whether psychiatric or general medical, as they are, by definition, less inclusive. As well, we did not include trials focused on subtypes of depression, inpatients, or patients with specific symptom profiles (eg, cognitive symptoms). The application of these criteria yielded 170 studies of antidepressants and 16 studies of face-to-face psychotherapy. To compare the differences between the self-guided iCBT studies and studies of antidepressants and face-to-face psychotherapies, we used a chi-square test, or Fisher exact test when we expected any cell to have a frequency lower than 5.

Results

Study Characteristics

We identified 5 new studies (see Multimedia Appendix 1) since the publication of the review by Karyotaki et al [17] that met our inclusion criteria (see Figure 1). Thus, we coded 21 RCTs comparing self-guided iCBTs versus a control in our updated review. These studies analyzed data for 4781 participants. Like Karyotaki et al, we found risk of bias to be low across most studies (see Multimedia Appendix 1).

There were differences between the self-guided iCBT studies, the antidepressant studies, and the face-to-face psychotherapy studies in terms of sample size per study arm ($F_{2,204}$=6.45, P=.002) and country of origin (χ^2_1=48.3, P<.001), but no differences in the percentage of studies that excluded participants based on an upper age limit (χ^2_1=1.5, P=.48). Antidepressant (mean 120.69, SD 74.10) and self-guided iCBT (mean 140.61, SD 109.16) studies tended to have more participants in each study arm than face-to-face psychotherapy studies did (mean 55.12, SD 39.48). Most self-guided iCBT studies originated from single sites within European countries (14/21, 67%). By way of contrast, only 5 of the psychotherapy studies were conducted in Europe (5/16, 31%) and, instead, the face-to-face psychotherapy studies were more likely to be published in the United States (8/16, 50%). Very few antidepressants studies were published from within single European countries (15/170, 8.8%), as most were published in the United States (101/171, 59.4%).

Table 1 lists the psychiatric inclusion and exclusion criteria used in the self-guided iCBT trials.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart for self-guided internet-based cognitive behavioral therapy (iCBT) studies included in the systematic review and meta-regression analysis. CBT: cognitive behavioral therapy.

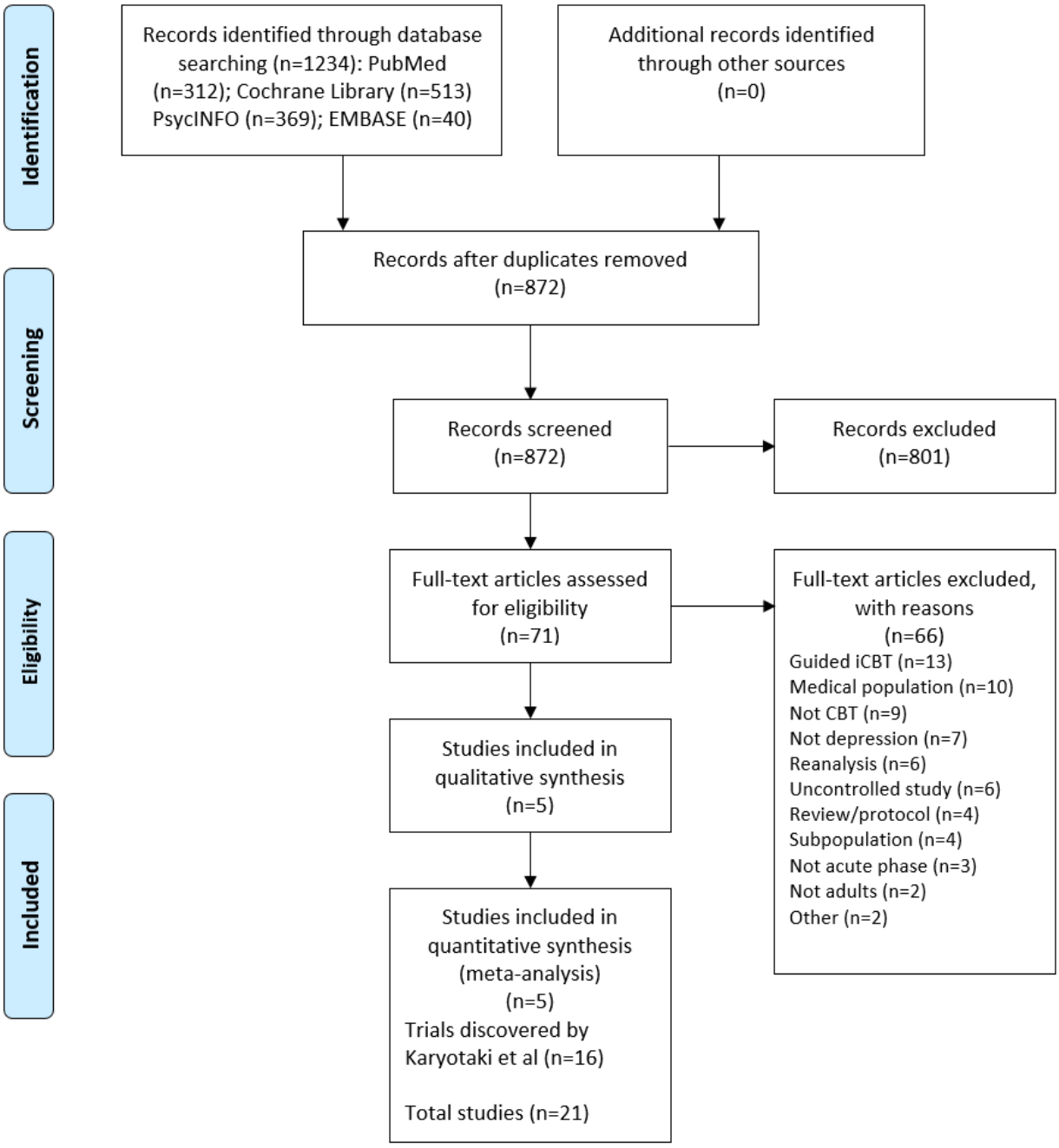


Table 1. Psychiatric exclusion criteria of randomized controlled trials of self-guided internet-based cognitive behavioral therapy for depressive symptoms (N=21).

Exclusion criteria	n (%)
Severity scale score below the cutoff (minimum score)	15 (71)
Severity scale score above the cutoff (maximum score)	6 (29)
Psychosis	12 (57)
Substance abuse or dependence	5 (24)
Significant suicidal ideation or risk	13 (62)
History of suicide attempt(s)	1 (5)
Episode length too long	1 (5)
Episode length too short	1 (5)
Any axis II disorder	0 (0)
Any axis I disorder	1 (5)
Borderline personality disorder	0 (0)
Antisocial personality disorder	0 (0)
Schizotypal personality disorder	0 (0)

At least half of the iCBT studies used the following 3 criteria: psychotic disorder or current psychotic symptoms (12/21, 57%), a minimum symptom severity on a depression scale (15/21, 71%), and significant suicidal ideation (13/21, 62%). The other criteria were used relatively infrequently. As has been noted of studies of antidepressants and face-to-face psychotherapy, there was considerable variability in the operationalization of these criteria. For example, of the 15 studies that excluded patients on the basis of a minimum score on a depression severity measure, 6 used the Patient Health Questionnaire-8 or -9 as an exclusion criterion, 4 used the Beck Depression Inventory (BDI), and 2 used the Kessler Psychological Distress Scale. The remaining 3 studies used other scales (eg, the Center for Epidemiologic Studies Depression Scale [CES-D]).

Effects of Entry Criteria on Outcomes

Only 5 of the exclusionary criteria were used with enough frequency (ie, >3 uses) that we could explore whether their use was associated with treatment outcomes: minimum symptom severity, maximum symptom severity, psychosis, substance misuse, and suicidality. Table 2 lists the results of the meta-regressions in which we regressed the study-level effect sizes in the 5 exclusionary criteria, individually as well as considered simultaneously. The updated meta-analysis found results similar to those by Karyotaki et al [17]. At the study level, self-guided iCBT was associated with more improvement in depression than were the control conditions (g=0.33, 95% CI 0.20-0.46, SE 0.06; P<.001).

We detected a high degree of between-study heterogeneity of effect sizes (I^2=76%), indicating that exclusion criteria could potentially explain between-studies differences in effects. Despite this, the meta-regressions did not find any significant effects of specific exclusionary criteria (P values >.39) or the total number of criteria used on study-level treatment outcomes (B=–0.01, 95% CI –0.09 to 0.10, SE 0.05; P=.92). The study-level exclusion criterion that appeared to have the largest effect on study-level differences was the use of a minimum of symptom severity as an exclusion (B=–0.12, 95% CI –0.41 to 0.17, SE 0.14; P=.39). A simultaneous regression considering all of the exclusion criteria in tandem likewise did not find any significant effect of the use of specific entry criteria on outcomes (P values >.28).

Comparison of Entry Criteria in Internet-Based Cognitive Behavioral Therapies, Antidepressants, and Psychotherapy

Table 3 shows the comparisons of the frequency with which different entry criteria were used in studies of self-guided iCBT versus studies of face-to-face psychotherapy versus studies of antidepressant medications. Contrary to our hypothesis, virtually all the exclusion criteria coded were used less frequently in the iCBT trials than in face-to-face psychotherapy and antidepressant trials, though not all these differences were statistically significant.

When specifically compared with studies of antidepressant medications, studies of self-guided iCBT were less likely to use almost all the exclusion criteria coded. For example, compared with studies of antidepressants (143/170, 84.1%), self-guided iCBT studies were less likely (12/21, 57%) to exclude patients on the basis of psychotic symptoms or a diagnosis (χ^2_1=7.2, P=.01). Similarly, 41.2% (70/170) of studies on antidepressants excluded patients on the basis of borderline personality disorder, although no self-guided iCBT study used this criterion (Fisher P<.001). The only exception to this general pattern of self-guided iCBT studies being more, rather than less, inclusive was on the basis of a maximum symptom severity exclusion. Only 29% (6/21) of the self-guided iCBT studies excluded patients on the basis of high symptom severity. However, this exclusion occurred more frequently in the iCBT studies than in the antidepressant studies (14/170, 8.2%; χ^2_1=6.2, P=.01). The self-guided iCBT studies were no more likely than face-to-face psychotherapy studies (6/16, 38%; χ^2_1=.05, P=.83) to use this criterion.

Table 2. Meta-regression coefficients for the relationship between individual exclusion criteria and study-level internet-based cognitive behavioral therapy controlled outcomes. No findings were significant at $P<.05$.

Exclusion criteria	Single-predictor models, *g* (95% CI)	Simultaneous-predictor model, *g* (95% CI)
Minimum symptom severity	–0.12 (–0.41 to 0.17)	–0.18 (–0.52 to 0.16)
Maximum symptom severity	–0.04 (–0.33 to 0.25)	–0.10 (–0.45 to 0.26)
Psychosis	0.07 (–0.20 to 0.34)	0.04 (–0.38 to 0.46)
Substance problems	0.02 (–0.29 to 0.33)	0.05 (–0.39 to 0.49)
Suicidality	0.09 (–0.18 to 0.36)	0.16 (–0.24 to 0.55)

Table 3. Comparison of the frequency of use of different inclusion and exclusion criteria across self-guided internet-based cognitive behavioral therapy (iCBT), face-to-face psychotherapy (F2F), and antidepressant medication (AM) trials.

Exclusion criteria	iCBT (n=21)	F2F (n=16)	AM (n=170)	*P* value		
				iCBT vs F2F vs AM	iCBT vs F2F	iCBT vs AM
Minimum symptom severity	15 (71)	13 (81)	170 (100.0)	<.001	.70	<.001
Maximum symptom severity	6 (29)	6 (38)	14 (8.2)	<.001	.83	.01
Psychosis	12 (57)	14 (88)	143 (84.1)	.02	.07	.01
Substance abuse or dependence	5 (24)	12 (75)	137 (80.6)	<.001	<.001	<.001
Suicidal risk	13 (62)	9 (56)	128 (75.3)	.14	.99	.29
Prior suicide attempt(s)	1 (5)	2 (12)	35 (20.6)	.21	.57	.13
Episode length too long	1 (5)	2 (12)	34 (20.0)	.21	.57	.13
Episode length too short	1 (5)	2 (12)	81 (47.6)	<.001	.57	<.001
Any axis II disorder	0 (0)	3 (19)	60 (35.3)	<.001	.07	<.001
Any axis I disorder	1 (5)	1 (6)	46 (27.1)	.01	>.99	.03
Borderline personality disorder	0 (0)	5 (31)	70 (41.2)	<.001	.01	<.001
Antisocial personality disorder	0 (0)	5 (31)	68 (40.0)	<.001	.01	<.001
Schizotypal personality disorder	0 (0)	5 (31)	63 (37.1)	<.001	.01	<.001

Most of the differences between self-guided iCBT and face-to-face psychotherapy studies did not meet the statistical significance threshold of $P<.05$, although the arithmetic differences were often in favor of the self-guided iCBT studies. Face-to-face therapy studies were more likely to specifically exclude participants based on the diagnoses of borderline, schizotypal, or antisocial personality disorder (each 5/16, 31%) than were the iCBT studies, which did not use this exclusion (each 0/21, 0%; Fisher $P<.001$).

By far the largest observed difference between the iCBT studies and the face-to-face and antidepressants studies was that iCBT studies infrequently excluded patients on the basis of a substance use disorder (5/21, 24%), though this exclusion was typical in antidepressant (137/170, 80.6%; Fisher $P<.001$) and face-to-face psychotherapy studies (12/16, 75%; Fisher $P=.003$).

Discussion

Principal Findings

We summarized the exclusion criteria used in self-guided iCBT studies, explored the relationship of the use of these criteria and the treatment effect size reported in the trial, and compared the frequencies with which such criteria were used in antidepressant trials or trials of face-to-face psychotherapy. Overall, self-guided iCBT studies infrequently used exclusion criteria that are the norm in studies of depression. Contrary to our hypotheses, we did not find the type and number of exclusion criteria to be related to outcomes. Also contrary to our hypothesis, self-guided iCBT trials were either equally likely or less likely to use specific psychiatric entry criteria.

Limitations

A noteworthy limitation of this review is the small number of self-guided iCBT studies. It is possible that the entry criteria used had a small effect on outcomes, but a larger number of studies is needed to detect this effect. Our review was limited to studies of self-guided iCBTs. While these interventions may have a broader public health impact than guided iCBT [21], there is evidence that they have somewhat smaller effects and lower rates of treatment completion [18-20]. Future research should study the entry criteria in studies of guided iCBT.

The fact that studies did not use a specific entry criterion (eg, substance use disorder) does not imply that the trial actually had patients representative of that feature, which is another limitation of our study. Because studies do not uniformly report on all characteristics of their patient sample (eg, substance use, duration, suicide risk), it is impossible to test for the presence of these features of the sample across all the studies. It is also

possible that studies used specific entry criteria but did not disclose them in the published report. Moreover, the psychiatric inclusion and exclusion criteria are only one aspect of the generalizability of the sample. Self-guided iCBT studies may have more selective patient samples by virtue of other criteria that they use explicitly (eg, access to the internet) or implicitly (eg, willingness to participate in an iCBT study).

While we updated the review by Karyotaki et al [17], we drew on published reviews by Lorenzo-Luaces et al [30], instead of updating this review as well. A difference between the 2 reviews is that the review by Lorenzo-Luaces and colleagues was limited to studies in which patients had a diagnosis of major depressive disorder. This difference specifically limited the number of face-to-face therapy studies eligible for coding. Our aim, however, was not to provide an exact estimate of the difference in frequency of use of exclusion criteria, but to provide a benchmarking context to evaluate the frequency with which self-guided iCBT studies excluded patients. The studies also differed by year of publications, as per the search strategy, country of origin, and size. These differences, along with the differences we identified as per our objectives, suggest that the studies are not directly or indirectly comparable (eg, as in a network meta-analysis). Our aim, however, was not to claim that the evidence base behind self-guided iCBTs is equivalent to the evidence base behind antidepressants or face-to-face psychotherapy but to provide a frame of reference against which to compare the entry criteria of self-guided iCBTs.

Implications

Only 29% of the self-guided iCBT studies excluded patients on the basis of severe depression, but it is still worth noting that self-guided iCBT studies appeared more likely than antidepressant studies to exclude participants on this basis. The perception that self-guided iCBT will not be effective for cases of more severely symptomatic depression aligns with common sense but is not supported by research data. For example, Bower et al [41] reported that the effects of self-guided internet-based therapies were more rather than less pronounced among patients high in symptom severity. In our analyses, we found no evidence that the effects of iCBT varied strongly according to exclusions by high or low symptom severity. It is possible that self-guided iCBTs are less effective for patients who have more complex presentations, with symptom severity being only one index of case complexity [5]. For example, the presence of anxiety and chronic depression duration have all been implicated in treatment outcomes in depression and may relate to lower response to iCBT [42]. Although we did not find the use of psychosis as an exclusion criterion to be related to outcomes, psychotic depression is relatively rare, and even studies allowing these patients into the trial may have had a low representation of psychotic depression. In contrast, approximately 50% of cases of depression are rated as severe or very severe. It is probable that patients self-select into treatment trials in a way that patients with severe depression avoid iCBT studies, but the data do not strongly support this conclusion. For example, Karyotaki et al [17] reported a mean score of 28 on the BDI for patients in their dataset along with a score of 26 on the CES-D, both of which are in the moderate to severe range for each measure.

Self-guided iCBT studies often excluded patients on the basis of suicidal risk. Our findings suggest that this exclusion criterion was not associated with outcomes, a finding that mirrors the results of van der Lem et al [24], as well as the fact that suicidal risk is not consistently reported as a predictor of treatment outcomes with face-to-face psychotherapy or antidepressants. A common exclusion criterion in antidepressant and face-to-face psychotherapy studies is substance use disorder, although this was missing from most self-guided iCBT studies. The findings of van der Lem et al [24] suggested that alcohol use disorder is not a strong predictor of outcomes in treatment for depression, and other studies have also not found an association between substance use disorder and outcomes in treatment for depression [43]. Given that there is no evidence to suggest negative outcomes for suicidal patients and those with substance use disorders, as well as the low rate of treatment utilization in these and other patient groups, more self-guided iCBT studies specifically recruiting from these patient groups are needed. Hoertel et al [44] have argued that using, or not using, specific criteria may influence not only the external validity and generalizability of a study but also its internal validity, by virtue of influencing outcomes. This runs counter to the prevailing logic that there is a trade-off between external and internal validity. Overall, our results suggest that the inclusion criteria commonly used in the field sacrifice external validity but provide no gains in internal validity.

Conclusions

To our knowledge, this is the first exploration of the effects of inclusion and exclusion criteria in self-guided iCBT studies, as well as the first comparison of the specific inclusion and exclusion criteria used in RCTs of iCBTs versus studies of face-to-face psychotherapy or antidepressants. Our findings can be taken to suggest that self-guided iCBT studies are more inclusive by design than studies of antidepressants or face-to-face psychotherapy. It is possible that, by using remote designs in which no individual face-to-face interviews can be conducted, self-guided iCBT trials limit their ability to exclude participants on specific features. For example, no self-guided iCBT trial excluded participants due to personality pathology broadly construed or borderline, schizotypal, or antisocial personality diagnosis, which are usually measured in psychiatric interviews, not self-reported questionnaires. By comparison, around a third of psychotherapy and around 40% of antidepressant trials used these exclusions. Given the differences in the entry criteria used, it stands to reason that, by design, studies of self-guided iCBT may be characterized by a more heterogeneous group of patients than studies of antidepressant or face-to-face psychotherapy. This level of heterogeneity in the underlying patient population increases the external validity of the research. Moreover, this variability can facilitate the discovery of process-outcome correlations [45], as well as effects of individual patient differences on outcomes [5]. This heterogeneity in the patient population may also contribute to heterogeneity in the overall effect size reported in treatment-control comparisons. Until evidence emerges to the contrary, however, it cannot be said that iCBT studies apply more stringent inclusion and exclusion criteria than studies of other treatments for depression or that the efficacy of the

treatments is inflated by using a less complex or severe patient population.

Acknowledgments

We would like to thank Eirini Karyotaki, Pim Cuijpers, and the coauthors of the meta-analysis from which we drew the studies for this work, as well as the primary authors of the studies we analyzed.

Conflicts of Interest

None declared.

Multimedia Appendix 1

List of studies, risk-of-bias assessments, and example search terms.

[PDF File (Adobe PDF File), 49KB - jmir_v20i11e10113_app1.pdf]

References

1. World Health Organization. Depression and Other Common Mental Disorders: Global Health Estimates. Geneva, Switzerland: WHO; 2017. URL: http://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf;jsessionid=D11B5ED6C69A736C263041085DAEFD55?sequence=1 [accessed 2018-10-29] [WebCite Cache ID 73X8B63Pb]
2. 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]
3. Kazdin AE, Blase SL. Rebooting psychotherapy research and practice to reduce the burden of mental illness. Perspect Psychol Sci 2011 Jan;6(1):21-37. [doi: 10.1177/1745691610393527] [Medline: 26162113]
4. Hollon SD, DeRubeis RJ, Fawcett J, Amsterdam JD, Shelton RC, Zajecka J, et al. Notice of Retraction and Replacement. Hollon et al. Effect of cognitive therapy with antidepressant medications vs antidepressants alone on the rate of recovery in major depressive disorder: a randomized clinical trial. JAMA Psychiatry. 2014;71(10):1157-1164. JAMA Psychiatry 2016 Dec 01;73(6):639-640 [FREE Full text] [doi: 10.1001/jamapsychiatry.2016.0756] [Medline: 27097060]
5. Lorenzo-Luaces L, DeRubeis RJ, van Straten A, Tiemens B. A prognostic index (PI) as a moderator of outcomes in the treatment of depression: a proof of concept combining multiple variables to inform risk-stratified stepped care models. J Affect Disord 2017 Apr 15;213:78-85. [doi: 10.1016/j.jad.2017.02.010] [Medline: 28199892]
6. Parker G. Beyond major depression. Psychol Med 2005 Apr;35(4):467-474. [Medline: 15856717]
7. Lorenzo-Luaces L. Heterogeneity in the prognosis of major depression: from the common cold to a highly debilitating and recurrent illness. Epidemiol Psychiatr Sci 2015 Dec;24(6):466-472. [doi: 10.1017/S2045796015000542] [Medline: 26081748]
8. Eaton WW, Anthony JC, Gallo J, Cai G, Tien A, Romanoski A, et al. Natural history of Diagnostic Interview Schedule/DSM-IV major depression. The Baltimore Epidemiologic Catchment Area follow-up. Arch Gen Psychiatry 1997 Nov;54(11):993-999. [Medline: 9366655]
9. Eaton WW, Shao H, Nestadt G, Lee HB, Lee BH, Bienvenu OJ, et al. Population-based study of first onset and chronicity in major depressive disorder. Arch Gen Psychiatry 2008 May;65(5):513-520 [FREE Full text] [doi: 10.1001/archpsyc.65.5.513] [Medline: 18458203]
10. Patten SB, Wang JL, Williams JVA, Lavorato DH, Khaled SM, Bulloch AGM. Predictors of the longitudinal course of major depression in a Canadian population sample. Can J Psychiatry 2010 Oct;55(10):669-676. [doi: 10.1177/070674371005501006] [Medline: 20964946]
11. Spijker J, De Graaf R, Bijl RV, Beekman ATF, Ormel J, Nolen WA. Duration of major depressive episodes in the general population: Results from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). Br J Psychiatry 2018 Jan 2;181(03):208-213. [doi: 10.1192/bjp.181.3.208]
12. Mattisson C, Bogren M, Horstmann V, Munk-Jörgensen P, Nettelbladt P. The long-term course of depressive disorders in the Lundby Study. Psychol Med 2007 Jun;37(6):883-891. [doi: 10.1017/S0033291707000074] [Medline: 17306047]
13. Steinert C, Hofmann M, Kruse J, Leichsenring F. The prospective long-term course of adult depression in general practice and the community. A systematic literature review. J Affect Disord 2014 Jan;152-154:65-75. [doi: 10.1016/j.jad.2013.10.017] [Medline: 24210624]
14. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). JAMA 2003 Jun 18;289(23):3095-3105. [doi: 10.1001/jama.289.23.3095] [Medline: 12813115]
15. Ohayon MM, Schatzberg AF. Prevalence of depressive episodes with psychotic features in the general population. Am J Psychiatry 2002 Nov;159(11):1855-1861. [doi: 10.1176/appi.ajp.159.11.1855] [Medline: 12411219]

16. Office of Applied Studies. Suicidal Thoughts, Suicide Attempts, Major Depressive Episode, and Substance Use among Adults. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2006:31.
17. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 01;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
18. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — a systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
19. Cowpertwait L, Clarke D. Effectiveness of web-based psychological interventions for depression: a meta-analysis. Int J Ment Health Addict 2013 Jan 18;11(2):247-268. [doi: 10.1007/s11469-012-9416-z]
20. Königbauer J, Letsch J, Doebler P, Ebert D, Baumeister H. 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]
21. Weisel KK, Zarski A, Berger T, Krieger T, Schaub MP, Moser CT, et al. Efficacy and cost-effectiveness of guided and unguided internet- and mobile-based indicated transdiagnostic prevention of depression and anxiety (ICare Prevent): a three-armed randomized controlled trial in four European countries. Internet Interv 2018 Apr. [doi: 10.1016/j.invent.2018.04.002]
22. Seligman ME. The effectiveness of psychotherapy. The Consumer Reports study. Am Psychol 1995 Dec;50(12):965-974. [Medline: 8561380]
23. Westen D, Morrison K. A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: an empirical examination of the status of empirically supported therapies. J Consult Clin Psychol 2001 Dec;69(6):875-899. [Medline: 11777114]
24. van der Lem R, de Wever WWH, van der Wee NJA, van Veen T, Cuijpers P, Zitman FG. The generalizability of psychotherapy efficacy trials in major depressive disorder: an analysis of the influence of patient selection in efficacy trials on symptom outcome in daily practice. BMC Psychiatry 2012 Nov 08;12:192 [FREE Full text] [doi: 10.1186/1471-244X-12-192] [Medline: 23137143]
25. Zimmerman M, Mattia JI, Posternak MA. Are subjects in pharmacological treatment trials of depression representative of patients in routine clinical practice? Am J Psychiatry 2002 Mar;159(3):469-473. [doi: 10.1176/appi.ajp.159.3.469] [Medline: 11870014]
26. Zimmerman M, Clark HL, Multach MD, Walsh E, Rosenstein LK, Gazarian D. Have treatment studies of depression become even less generalizable? A review of the inclusion and exclusion criteria used in placebo-controlled antidepressant efficacy trials published during the past 20 years. Mayo Clin Proc 2015 Sep;90(9):1180-1186. [doi: 10.1016/j.mayocp.2015.06.016] [Medline: 26276679]
27. Zimmerman M, Clark HL, Multach MD, Walsh E, Rosenstein LK, Gazarian D. Variability in the substance use disorder exclusion criterion in antidepressant efficacy trials. J Affect Disord 2016 Dec 01;198:39-42. [doi: 10.1016/j.jad.2016.03.024] [Medline: 27002283]
28. Zimmerman M, Clark HL, Multach MD, Walsh E, Rosenstein LK, Gazarian D. Symptom severity and the generalizability of antidepressant efficacy trials: changes during the past 20 years. J Clin Psychopharmacol 2016 Apr;36(2):153-156. [doi: 10.1097/JCP.0000000000000466] [Medline: 26848791]
29. Zimmerman M, Multach M, Walsh E, Rosenstein LK, Gazarian D, Clark HL. Problems in the descriptions of the psychiatric inclusion and exclusion criteria in publications of antidepressant efficacy trials: a qualitative review and recommendations for improved clarity. CNS Drugs 2016 Mar;30(3):185-191. [doi: 10.1007/s40263-016-0314-y] [Medline: 26818632]
30. Lorenzo-Luaces L, Zimmerman M, Cuijpers P. Are studies of psychotherapies for depression more or less generalizable than studies of antidepressants? J Affect Disord 2018 Dec;234:8-13. [doi: 10.1016/j.jad.2018.02.066] [Medline: 29522947]
31. Blanco C, Hoertel N, Franco S, Olfson M, He J, López S, et al. Generalizability of clinical trial results for adolescent major depressive disorder. Pediatrics 2017 Dec;140(6):e20161701. [doi: 10.1542/peds.2016-1701] [Medline: 29097612]
32. Hoertel N, López S, Wang S, González-Pinto A, Limosin F, Blanco C. Generalizability of pharmacological and psychotherapy clinical trial results for borderline personality disorder to community samples. Personal Disord 2015 Jan;6(1):81-87 [FREE Full text] [doi: 10.1037/per0000091] [Medline: 25580674]
33. Hoertel N, de Maricourt P, Katz J, Doukhan R, Lavaud P, Peyre H, et al. Are participants in pharmacological and psychotherapy treatment trials for social anxiety disorder representative of patients in real-life settings? J Clin Psychopharmacol 2014 Dec;34(6):697-703. [doi: 10.1097/JCP.0000000000000204] [Medline: 25154011]
34. Franco S, Hoertel N, McMahon K, Wang S, Rodríguez-Fernández JM, Peyre H, et al. Generalizability of pharmacologic and psychotherapy clinical trial results for posttraumatic stress disorder to community samples. J Clin Psychiatry 2016 Dec;77(8):e975-e981. [doi: 10.4088/JCP.15m10060] [Medline: 27379465]
35. Hoertel N, Le Strat Y, Blanco C, Lavaud P, Dubertret C. Generalizability of clinical trial results for generalized anxiety disorder to community samples. Depress Anxiety 2012 Jul;29(7):614-620. [doi: 10.1002/da.21937] [Medline: 22495990]
36. Higgins JP, Altman DG. Assessing risk of bias in included studies. In: Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series. Chichester, England: Wiley-Blackwell; 2008:187-241.

37. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw 2010;36(3). [doi: 10.18637/jss.v036.i03]
38. IntHout J, Ioannidis JPA, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC Med Res Methodol 2014 Feb 18;14:25 [FREE Full text] [doi: 10.1186/1471-2288-14-25] [Medline: 24548571]
39. Röver C, Knapp G, Friede T. Hartung-Knapp-Sidik-Jonkman approach and its modification for random-effects meta-analysis with few studies. BMC Med Res Methodol 2015 Nov 14;15:99 [FREE Full text] [doi: 10.1186/s12874-015-0091-1] [Medline: 26573817]
40. Huhn M, Tardy M, Spineli LM, Kissling W, Förstl H, Pitschel-Walz G, et al. Efficacy of pharmacotherapy and psychotherapy for adult psychiatric disorders: a systematic overview of meta-analyses. JAMA Psychiatry 2014 Jun;71(6):706-715. [doi: 10.1001/jamapsychiatry.2014.112] [Medline: 24789675]
41. Bower P, Kontopantelis E, Sutton A, Kendrick T, Richards DA, Gilbody S, et al. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. BMJ 2013;346:f540 [FREE Full text] [Medline: 23444423]
42. Kessler RC, van Loo H, Wardenaar KJ, Bossarte RM, Brenner LA, Ebert DD, et al. Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. Epidemiol Psychiatr Sci 2017 Feb;26(1):22-36 [FREE Full text] [doi: 10.1017/S2045796016000020] [Medline: 26810628]
43. Nunes EV, Levin FR. Treatment of depression in patients with alcohol or other drug dependence: a meta-analysis. JAMA 2004 Apr 21;291(15):1887-1896. [doi: 10.1001/jama.291.15.1887] [Medline: 15100209]
44. Hoertel N, Le Strat Y, Limosin F, Dubertret C, Gorwood P. Prevalence of subthreshold hypomania and impact on internal validity of RCTs for major depressive disorder: results from a national epidemiological sample. PLoS One 2013;8(2):e55448 [FREE Full text] [doi: 10.1371/journal.pone.0055448] [Medline: 23405152]
45. DeRubeis RJ, Gelfand LA, German RE, Fournier JC, Forand NR. Understanding processes of change: how some patients reveal more than others-and some groups of therapists less-about what matters in psychotherapy. Psychother Res 2014;24(3):419-428 [FREE Full text] [doi: 10.1080/10503307.2013.838654] [Medline: 24219275]

Abbreviations

BDI: Beck Depression Inventory
CES-D: Center for Epidemiologic Studies Depression Scale
iCBT: internet-based cognitive behavioral therapy
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 14.02.18; peer-reviewed by N Hoertel, P Bower, N Forand, T Berger; comments to author 05.03.18; revised version received 11.07.18; accepted 19.07.18; published 09.11.18.

Please cite as:
Lorenzo-Luaces L, Johns E, Keefe JR
The Generalizability of Randomized Controlled Trials of Self-Guided Internet-Based Cognitive Behavioral Therapy for Depressive Symptoms: Systematic Review and Meta-Regression Analysis
J Med Internet Res 2018;20(11):e10113
URL: http://www.jmir.org/2018/11/e10113/
doi:10.2196/10113
PMID:30413400

Review

Evidence-Based Evaluation of eHealth Interventions: Systematic Literature Review

Amia Enam[1*], MSc; Johanna Torres-Bonilla[1*], MSc; Henrik Eriksson[1], PhD

Centre for Healthcare Improvement, Technology Management and Economics, Chalmers University of Technology, Gothenburg, Sweden
[*]these authors contributed equally

Corresponding Author:
Amia Enam, MSc
Centre for Healthcare Improvement
Technology Management and Economics
Chalmers University of Technology
Vera Sandbergs Alle 8
Gothenburg, 412 96
Sweden
Phone: 46 8609449850
Email: amiaenam@gmail.com

Abstract

Background: Until now, the use of technology in health care was driven mostly by the assumptions about the benefits of electronic health (eHealth) rather than its evidence. It is noticeable that the magnitude of evidence of effectiveness and efficiency of eHealth is not proportionate to the number of interventions that are regularly conducted. Reliable evidence generated through comprehensive evaluation of eHealth interventions may accelerate the growth of eHealth for long-term successful implementation and help to experience eHealth benefits in an enhanced way.

Objective: This study aimed to understand how the evidence of effectiveness and efficiency of eHealth can be generated through evaluation. Hence, we aim to discern (1) how evaluation is conducted in distinct eHealth intervention phases, (2) the aspects of effectiveness and efficiency that are typically evaluated during eHealth interventions, and (3) how eHealth interventions are evaluated in practice.

Methods: A systematic literature review was conducted to explore the evaluation methods for eHealth interventions. Preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines were followed. We searched Google Scholar and Scopus for the published papers that addressed the evaluation of eHealth or described an eHealth intervention study. A qualitative analysis of the selected papers was conducted in several steps.

Results: We intended to see how the process of evaluation unfolds in distinct phases of an eHealth intervention. We revealed that in practice and in several conceptual papers, evaluation is performed at the end of the intervention. There are some studies that discuss the importance of conducting evaluation throughout the intervention; however, in practice, we found no case study that followed this. For our second research question, we discovered aspects of efficiency and effectiveness that are proposed to be assessed during interventions. The aspects that were recurrent in the conceptual papers include clinical, human and social, organizational, technological, cost, ethical and legal, and transferability. However, the case studies reviewed only evaluate the clinical and human and social aspects. At the end of the paper, we discussed a novel approach to look into the evaluation. Our intention was to stir up a discussion around this approach with the hope that it might be able to gather evidence in a comprehensive and credible way.

Conclusions: The importance of evidence in eHealth has not been discussed as rigorously as have the diverse evaluation approaches and evaluation frameworks. Further research directed toward evidence-based evaluation can not only improve the quality of intervention studies but also facilitate successful long-term implementation of eHealth in general. We conclude that the development of more robust and comprehensive evaluation of eHealth studies or an improved validation of evaluation methods could ease the transferability of results among similar studies. Thus, the resources can be used for supplementary research in eHealth.

(J Med Internet Res 2018;20(11):e10971) doi:10.2196/10971

KEYWORDS

evidence-based practice; program evaluation; systematic review; technology assessment

Introduction

Background

The use of electronic health (eHealth) is still driven by assumptions about the benefits of eHealth rather than its evidence [1]. With time, the trustworthiness and robustness of eHealth to facilitate safe and cost-efficient care are being questioned because of a lack of evidence [2]. This may trigger reluctance in investing and developing policies related to eHealth in organizations as well as in countries [3].

The term *eHealth* was introduced in the 1990s [4]; however, it was hardly in use until 1999 [5]. According to Eysenbach [5]:

> *e-health is an emerging field in the intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced through the Internet and related technologies.*

Eysenbach believes eHealth stands for more than internet and medicine [5]. In our study, eHealth was used as the broadest umbrella encompassing everything that comes within information and communication technology and health care, including telemedicine, mobile health, and health informatics.

Systematic evaluation can capture the evidence and criteria that evaluative judgment is based on and curtail the sources of biases [6]. The quality of an evaluation is assessed by the credibility of evidence assembled through it and using evidence in refining the policies and programs [7]. Evaluation of eHealth interventions is complex because of several reasons (eg, the need for multidisciplinary collaboration [8], context dependency [9], and differences in epistemological beliefs considering the interventions in clinical studies or including social aspects as well [10-14]). Therefore, variety exists concerning how the evaluation of eHealth interventions is performed and presented. Garnering robust evidence through evaluation becomes difficult because of these circumstances.

It is relevant to understand evidence-based medicine (EBM) while discussing the importance of evidence in eHealth interventions. A common query is how EBM can help generate evidence for eHealth interventions [15]:

> *Evidence-based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of EBM means integrating individual clinical expertise with the best available external clinical evidence from systematic research.*

As per this definition, *evidence* in EBM is conspicuously related to the clinical aspect. Although it is argued whether EBM is only about randomized controlled trials [16] or not [15], it is quite explicit that EBM usually does not contemplate anything outside clinical practices. However, eHealth interventions have more aspects to evaluate besides the clinical aspects. An extensive assessment of the aspects including sociotechnical aspects is needed through each phase of the technology's life cycle while evaluating eHealth interventions [10,11,17]. Hence, to gather evidence from eHealth intervention, the evaluation process requires a distinct approach than what is usually put forward within EBM.

Objective

Our objective was to elucidate how the evidence of effectiveness and efficiency of eHealth can be generated through evaluation. Consequently, a literature review was conducted to understand the evaluation process regarding both theories of eHealth and the practices in case studies of eHealth interventions. We decided to employ a broader perspective at the beginning of the review process to achieve our research objective. As the literature review progressed, our research objective narrowed, and the research questions were redefined several times. However, the objective was always to understand the evaluation of eHealth from a comprehensive perspective. It was pertinent to recognize the phases of eHealth interventions where evaluation occurs and the aspects of effectiveness and efficiency that are evaluated during such interventions. Our 3 research questions were as follows:

1. How is evaluation conducted in distinct eHealth intervention phases?
2. What aspects of effectiveness and efficiency are typically evaluated during eHealth interventions?
3. How have eHealth intervention case studies been evaluated?

Finally, we presented an approach to evaluate eHealth interventions by developing a model—Evidence in eHealth Evaluation. To our knowledge, this model is a novel way of looking into evaluation of eHealth interventions for comprehensive evidence. This conceptual model was based on the findings of the literature review.

Methods

Systematic Review

Identification and Screening

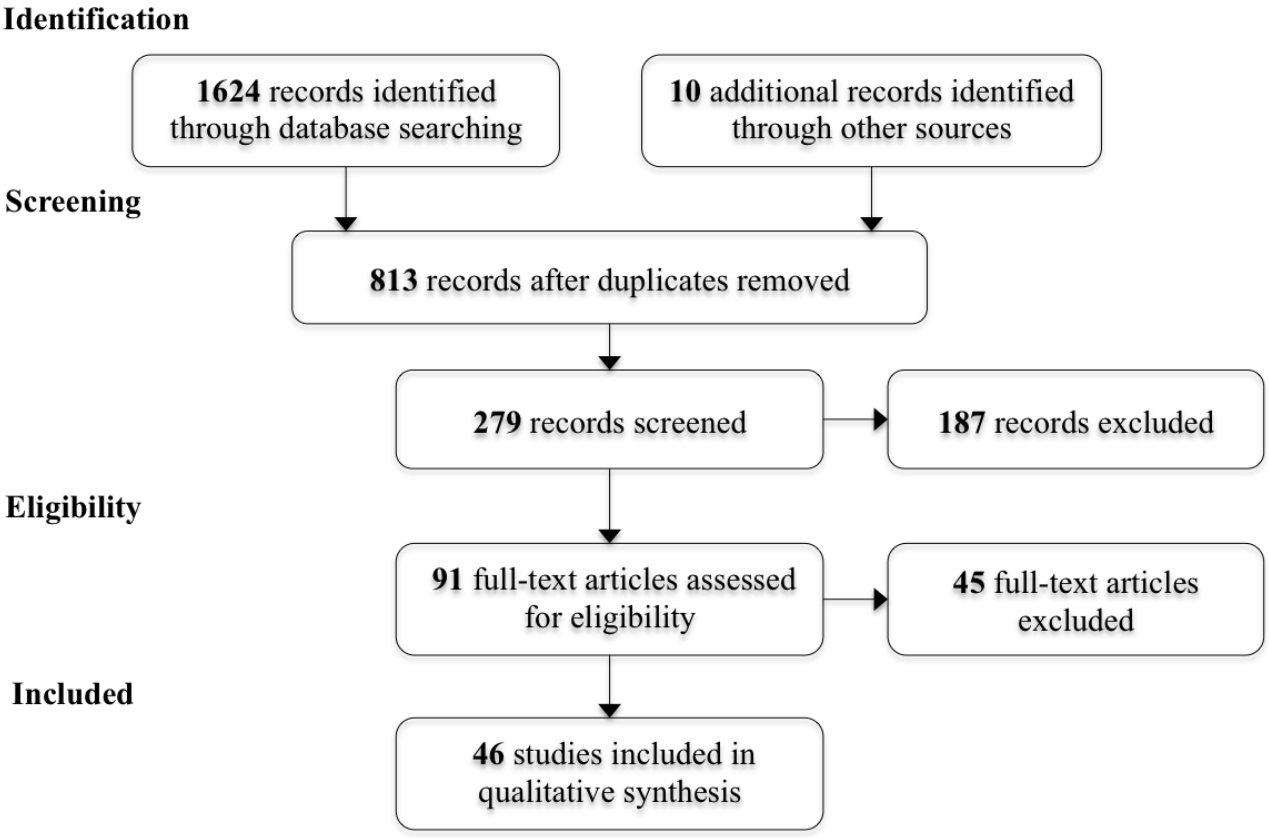
A systematic search of relevant literature was conducted following preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines [18]. Google Scholar and Scopus were used to search the following identified terms, "research methods" and "eHealth interventions," "study design" and eHealth interventions," "evaluation methods" and "eHealth interventions," "eHealth interventions" and "evaluation framework," "evidence based" and "evaluation," and "eHealth interventions." The selected set of terms is aligned to the broad scope of eHealth interventions' evaluation methods and to the aim of including and analyzing as many relevant studies as possible in the review. We included scientific papers published between 1990 and 2016. As the term eHealth evolved during the 1990s [4], we deemed it reasonable to consider the literature published on eHealth interventions since then. A total of 1624 records were found with these selected search keywords.

The screening of the papers was conducted in 3 steps. For the first 2 steps, the screening was based on the title of the manuscripts using a predefined set of exclusion and inclusion criteria. Only scientific papers were used, whereas books and patents were excluded during the search. To avoid overanalysis and repetition of the papers, the exclusion criteria for the first step were citations, literature reviews, and meta-analyses. In addition, studies addressing specific health issues designed to answer clinical research questions were excluded (ie, publications solely addressing behavior change theory, ergonomics, drugs, sedentary issues, or physical activity intervention as well as those addressing nonadult patients). The number of records was reduced to 813 after the first elimination. At this point, all the records were listed together, and duplicate records were removed. During the second step, title screening was conducted. For papers whose titles did not explicitly mention the intervention target group, the abstracts were read to decide. When in doubt, the papers were included for further scrutiny in the next step. Consequently, only those records that included either a conceptual discussion about eHealth interventions or discussion about eHealth interventions that focused on adult patients and caregivers were selected. The third step of the screening process started with 279 records; this time, all the abstracts and the methodology of the papers were read by 2 of the authors individually. Previously, we devised the inclusion parameters for this stage; specifically, the selected papers ought to (1) be aligned to the study objective, (2) have the potential to provide insight for 1 or more research question, and (3) comply with the inclusion and exclusion criteria described above. At the end of step 3, the authors discussed their observations, and 81 papers were selected for thorough reading and analysis. Besides the papers selected through systematic search, 10 records were added for further analysis. These records were found from the citation of the papers selected using the systematic literature search. All 10 papers were included in this study because of their relevance to our objective.

Eligibility and Inclusion

To extract and record useful information from the papers and to gain a general overview of evaluation in eHealth interventions, a Microsoft Excel spreadsheet was created with the criteria shown in Multimedia Appendix 1. Although most of the criteria were adopted from section 2.2.3 *Planning the topic and scope of a review* in the Cochrane review [19], criterion such as *learning point* was included by us. This section of Cochrane review was adapted to identify the potentials of the selected papers in fulfilling the research objectives. On the other hand, it seemed advantageous to record novelties of the studies concerning *learning points* for our own development. During these phases, the papers (N=91) were read meticulously, which led to the final screening. As a result, 46 papers were selected for the qualitative analysis that categorically focused on evaluation in terms of phases and aspects. The flow diagram of the papers selection process is presented in Figure 1.

Figure 1. Preferred reporting items for systematic reviews and meta-analyses flow diagram of the study selection process.



Qualitative Analysis

On the basis of the summary of studies mentioned in the previous section, the papers (N=46) were classified into 2 categories: (1) conceptual exploration of eHealth interventions (n=21) and (2) case studies of eHealth interventions (n=25). Using the summary table (Multimedia Appendix 1), the papers in the first category were divided into 2 groups: (A) evaluation of distinct eHealth intervention phases (n=10) and (B) aspects of evaluation in eHealth interventions (n=11).

Evaluation of Distinct Phases of an Electronic Health Intervention

For the papers in group A, thematic scrutiny was applied by mapping out the content of the papers and grouping the phases of intervention with similar objectives, activities, or results. The objective of the analysis was to understand whether the researchers emphasize some phases over others during evaluation and, if so, what phases are most frequently evaluated during an intervention.

Aspects of Electronic Health Intervention Supposed to Be Evaluated According to Electronic Health Literature

The studies in group B elaborated on the aspects evaluated to gather evidence of the efficiency and effectiveness of eHealth evaluations. These aspects and their key area of measurements were extracted from the papers to understand the parameters of efficiency and effectiveness that are emphasized by eHealth literature.

Evaluation Reported in Empirical Studies of Electronic Health Interventions

The studies categorized as case studies of eHealth interventions were analyzed based on several characteristics (ie, duration of the intervention, number of participants, used a framework or predefined theory for evaluation or designing the intervention, aspects assessed for evaluation, phases involved in the evaluation, data collection method, and presentation of intervention results). The purpose of the analysis was to conduct a descriptive comparison of the characteristics of evaluation performed in case studies with that of the conceptual papers.

Results

Evaluation of Distinct Phases of an Electronic Health Intervention

This subsection concentrates on how evaluation is conducted in distinct phases of an eHealth intervention. From the selected studies in group A, a spectrum of phases of an eHealth intervention was identified including *design*, *pretesting*, *pilot study*, *pragmatic trial*, *evaluation*, and *postintervention*. Table 1 provides a compilation of the phases along with the area of focus and key activities within each phase.

It can be ascertained from Table 1 that evaluation is not commonly performed in the design and pretesting phases. Although some researchers see evaluation as an ongoing process throughout the intervention [12,21,23], others believe there is value in evaluating the intervention at the end of the study period [12]. Concerning the latter, evaluation itself is one of the phases of the intervention.

Table 1. Characteristics of distinct phases of an electronic health intervention.

Phase	Area of focus	Key activities
Design phase	Conceptualization	Gather theoretical foundations and empirical evidence to detect the existing problems and identify viable solutions [20-24] and define the objectives of the to-be-developed technology [20,25]
	Contextual inquiry	Identify the end users and stakeholders to define and analyze the characteristics of the context the technology is going to be implemented on [20,21,23-25]
	Value specification	Prioritize the critical values of the technology derived from the end users and stakeholders' needs [12,22-25]
	Requirements specification	Translate the values into functional and technical requirements that frame the final design and the technology development [12,23-25]
Pretesting phase	Conduct short-term trials	Provide evidence of efficacy of the technology [12,24,26]; measure factors such as optimal intensity, timing, safety, feasibility, usability, intervention content, and logistic issues [21,22,26]; and evaluate the correspondence between technology capabilities and technology requirements [24]
Pilot study	Strategic plan	Define the preliminary plan of the pilot study (ie, objective, timeline, budget, sponsors, and team members [27,28] and identify related ethical and legal issues [20,28]
	Study design	Define the study type, duration, and participants [20,26,28] as well as data collection methods [28] and design the recruitment process to conform to statistical validity and minimize selection bias [27]
	Evaluation	Evaluate the technology and its impact simultaneously [27] and evaluate the effectiveness of the intervention [22]
Pragmatic trial phase	Execution	Administer the intervention to a larger group of participants [21,23,26] with fewer eligibility restrictions [26]
	Evaluation	Formative and summative evaluation (discussed in the evaluation phase) and internal and external evaluation (discussed in the evaluation phase)
Evaluation phase	Formative evaluation	Generate measures that provide timely feedback [12,23] and perform an evaluative iterative process, as the findings from each step are used to inform subsequent steps [21]
	Summative evaluation	Provide generalizable knowledge and benefits of the intervention [12,23]
	Internal evaluation	Perform an evaluative process intrinsic to information and communication technology implementations and conducted by the implementation team [12]
	External evaluation	Conduct the evaluation by external evaluators to provide expertise where it is needed and minimize the bias of in-house evaluators [12]
Postintervention phase		Conduct postmarketing or surveillance studies to follow up the technology once scaled up and used by a wider audience [22,26]

Aspects of Electronic Health Intervention Supposed to Be Evaluated According to Electronic Health Literature

We determined the aspects that researchers evaluated to gather the evidence of efficiency and effectiveness of an eHealth intervention. To understand the aspects of efficiency and effectiveness, we compiled the dimensions of eHealth interventions that are proposed to be measured by the studies categorized in group B (n=11). While excerpting the dimensions during the qualitative analysis, we found that they can be classified into 7 aspects: *organizational aspect*, *technological aspect*, *human and social aspect*, *clinical aspect*, *cost and economic aspect*, *ethical and legal aspect*, and *transferability aspect*. Table 2 exhibits these aspects along with their key area of measurements.

Table 2. Description of identified aspects of evaluation in electronic health interventions.

Aspects of assessment	Key areas of measurement
Organizational aspect	Organizational setting where the intervention is taking place; it can differ depending on the scale of the intervention (eg, health center, region, and country) [29]; all type of individuals or groups in the health care system that participate in the eHealth intervention, their characteristics, and expectations [30]; organizational performance and professional practice standards [30]; changes in the functions of the health care provider, skills and resource demands, and the roles of the professionals in the organization [29,31-33]; representativeness and participation rates of the health care professionals during the intervention [34]; capability of the organization to implement the intervention [30,34-38] and the extent that the technology fits the organizational strategy, operations, culture, and processes [30]; and sustainability or the degree that the technology becomes accustomed in the daily practice of an organization [29,32,34]
Technological aspect	Ensure trust [38], effectiveness, and contribution of quality of care [30,36] of the technology implemented; *system performance*: hardware and software requirements, correct functioning of the components [29,38], and system capability to meet users' needs and fit the work patterns of the health care system' professionals [30,39,40]; *usability*: broad experience of the users with the system [29,33,37,40]; *privacy and security:* safety and reliability of the technology [29], and security of the data managed in the technology [37,38]; *technical accuracy:* quality of the transfer of data [41]; *information quality*: relates to accuracy, completeness, and availability of the information produced by the system (eg, patients' records, reports, images, and prescriptions), and it depends on users' subjectivity [30,39]; *service quality:* measures the support and follow-up service delivered by the technology provider [39]; *triability:* the ability of the innovation to be tested on a small scale before the final implementation [40]; *maturity:* whether the system has been used on a sufficient number of patients to address all the technical problems [36]; and *interoperability*: communication between the technology and the pre-existing systems, the fit between the technology and the existing work practices [37]
Human and social aspect	Acceptance and usability satisfaction of the technology used in the intervention [30,31,33,36,38,39,41] where the user can be physicians, nurses and other staff, and patients, depending on the type of the participants in the intervention [41]; *system use:* volume of use, who is using, purpose of use, and motivation to use the technology [39]; *user satisfaction:* perceived usefulness, enjoyment, decision-making satisfaction, and overall satisfaction for the technology [30,39]; and psychological aspects such as satisfaction, well-being, and other psychological variables, and social aspects such as accessibility to the technology, the social relationships evolving over the transmission of care, or activities of the patients under the intervention [31]
Clinical aspect	Benefits and unanticipated negative effects of the intervention, biological outcomes including disease risk factors, behavioral outcomes of the participants, staff who deliver the intervention and the sponsors, and quality-of-life outcomes to evaluate participants' mental health and satisfaction [34] and long-term measurements of the diagnostic and clinical effectiveness [41,35,36], safety of care [33,35,36], and quality of care [33]
Cost and economic aspect	Cost analysis methods to compare the intervention with relevant alternatives in terms of costs and consequences [36]; diverse cost analysis methods can be considered (eg, cost-minimization analysis, cost-effectiveness analysis, cost-benefit analysis, cost-utility analysis, and cost-consequence analysis) [30,31,41] and are conducted from several perspectives such as societal, third-party payers, health care providers, or patient [31]; and diverse costs can be included such as investment cost, monthly user charge of equipment, costs of used communication line, education of the technology, costs of patients and their close relatives [41], wages of doctor and other staff [30,41], expenditure and revenue for the health care organization adopting the technology [36], and resource utilization and opportunity cost of the eHealth intervention [34]
Ethical and legal aspect	Ethical concerns of the app itself and its implementation including all the stakeholders' viewpoints on using the technology and the key ethical principles associated with the context in which intervention is conducted [35,36] and legal aspect identifies and analyzes the legislative documents and legal obligations that may exist in each context involved in the intervention [30,35,36]
Transferability aspect	Participation and representativeness of the intervention, percentage of persons who receive or are affected by the program, and the characteristics of participants and nonparticipants to investigate the extent that participants are representative and what population group should be a priority for future research [34] and transferability of results from studies of eHealth from one setting to another and the assessment of validity and reliability of the study [36]

Evaluation Reported in Case Studies of Electronic Health Interventions

The papers categorized as case studies of eHealth intervention show substantial variation in the approaches taken to evaluate the interventions. The use of standardized frameworks and theories for evaluating the interventions was hardly noticed in these studies. Multimedia Appendix 2 provides the result of the analysis [20,42-65].

To summarize Multimedia Appendix 2, it can be said that out of 25 case studies, 16 (64%, 16/25) evaluate clinical aspects, 12 (48%, 12/25) evaluate human and social aspects, 5 (20%, 5/25) evaluate technological aspect, and 4 (16%, 4/25) evaluate organizational aspect. The other aspects discussed by the theory-based literature (Table 2) are not evaluated in any of the case studies.

Discussion

Principal Findings

From the papers reviewed in this study, it has been revealed that numerous approaches to conceptualize and conduct eHealth intervention coexist. Several attributes of evaluation of eHealth intervention have become known through this review. There are vivid differences between how evaluation is conducted in practice (case studies) and how it is discussed in the conceptual

papers. Moreover, a wide range of variety prevails within each group. Evaluation has been depicted as both static action performed at the end of the intervention [20,24,26-28] and dynamic action dividing it further into summative and formative evaluation [12,21,23]. Depending on the evaluators, evaluation can also be classified into internal and external assessment [12]. However, all case studies conducted evaluation at the end of the intervention. Although several aspects of evaluation have been found in conceptual papers [32-39,41], the case studies mostly evaluated clinical [20,42-44,46-49,52,53,55,59-61,64,65] and human and social aspects [42,46-49,53-55,59,61,63,65].

Although analyzing standardization of eHealth evaluation was not an objective of this review, the variability found in the studies compelled us to think whether it hinders the sharing of evidence among eHealth interventions. Scarcity of evidence, in turn, could delay the growth of eHealth. It is noticeable that the need for evidence is not clearly stated in any of the papers. The evaluation of the empirical studies typically focused on the success or failure of the technology (eHealth) in that intervention. It seems that the numerous efforts taken in eHealth research are still quite disconnected, and they are thus unable to create a synergic effect on the growth of eHealth.

Evidence in Electronic Health Evaluation Model

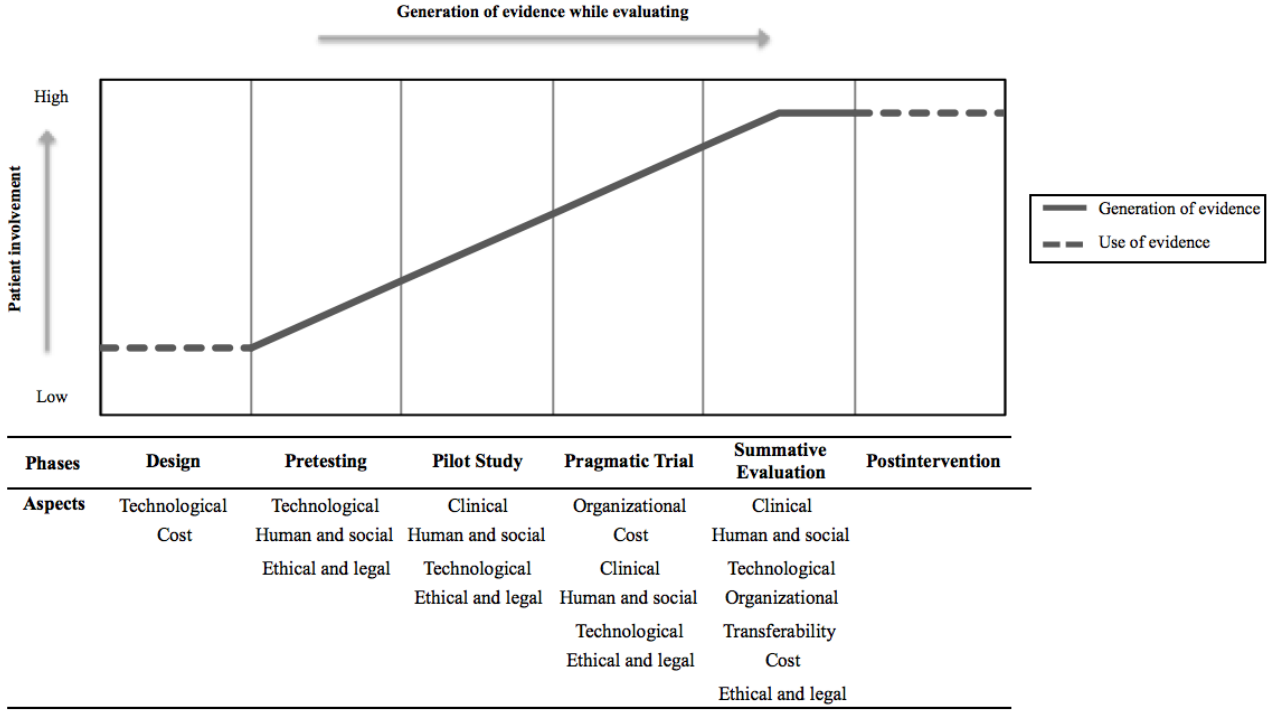
It was noticeable from the review that although some studies elaborate on the aspects of evaluation for eHealth intervention (studies from group B) and some organize the evaluation of intervention into certain phases (studies from group A), no visible interaction has been made so far between these 2 groups of works (ie, what to assess during what phase). There is a gap where a connection can be made between the distinct phases of intervention and aspects of evaluation. This led us to develop the Evidence in eHealth Evaluation model (Figure 2), which exhibits the accumulation of evidence by assessing certain aspects of evaluation in distinct intervention phases. The Evidence in eHealth Evaluation model is a novel approach to investigate the evaluation of eHealth interventions.

In this study, an eHealth intervention comprising all 6 phases (ie, *design*, *pretesting*, *pilot study*, *pragmatic trial*, *evaluation*, and *postintervention*) was conceived as a comprehensive intervention. We propose that the generation of robust evidence of effectiveness and efficiency would be plausible when the evaluation is conducted through all intervention phases. Moreover, the aspects of evaluation (ie, *organizational aspect*, *technological aspect*, *human and social aspect*, *clinical aspect*, *cost aspect*, *ethical and legal aspect*, and *transferability aspect*) would vary in each phase depending on activities of the phases. For example, when an eHealth intervention initiates with the *design* phase, the decisions are made based on the evaluation of the *technological* aspect and *cost* of technology development. The formal evaluation of the intervention begins in succeeding phases. The evaluation of *technological*, *human and social*, and *cost* aspects occurs in the *pretesting* phase. During the *pilot study* phase, the focus of evaluation shifts primarily to *clinical* aspect followed by *human and social*, *technological*, and *ethical and legal*. Depending on the evidence garnered in the pilot study, the intervention may proceed to the next phase or go back to the *design* phase. As the intervention is scaled up in the *pragmatic trial*, the evaluation is conducted to identify whether the technology-enabled care can be executed within the realistic layout of an organization. Hence, the key areas of evaluation in this phase are *organizational* and *cost* aspects along with other aspects such as *clinical*, *human and social*, *technological*, and *ethical and legal*. The last phase of gathering evidence is *summative evaluation*, where all the aspects are assessed including transferability. This comprehensive evaluation process gradually accumulates the evidence that reaches its peak in the *summative evaluation* phase and is used in the *postintervention* phase to make future decisions. The model also exhibits how the involvement of patients increases continuously from the *design* phase to the *pragmatic trial*, escalating the complexity of the evaluation process.

The inclusion of relevant information regarding other aspects besides the clinical aspect (eg, organizational aspect and cost aspect) allows creating reusable knowledge to facilitate the transfer of results to other settings [36] and to obtain useful insights for long-term implementations. It can be assumed that assessing all the aspects in a single study might conclude with a confounding result, as all the aspects are interrelated and inferior performance in an aspect can affect the performance in other aspects, which might create a misleading result. Therefore, our model proposes to extend the evaluation process throughout the 6 phases of eHealth intervention. The underlying idea is to assess specific aspects in each phase instead of evaluating all aspects in a single phase. This way of evaluating eHealth interventions can capture comprehensive evidence that is usually dynamic and complex in nature.

We acknowledge the fact that an eHealth intervention including all the phases presented in the model will become cumbersome because of high resource consumption. This conceptual model is not a prescription but just a way to show the progression of evidence in eHealth intervention in a reliable manner.

Figure 2. The evidence in electronic health (eHealth) evaluation model.



Phases	Design	Pretesting	Pilot Study	Pragmatic Trial	Summative Evaluation	Postintervention
Aspects	Technological Cost	Technological Human and social Ethical and legal	Clinical Human and social Technological Ethical and legal	Organizational Cost Clinical Human and social Technological Ethical and legal	Clinical Human and social Technological Organizational Transferability Cost Ethical and legal	

Conclusions

To date, the importance of evidence has not been discussed as rigorously as the diverse research approaches and evaluation frameworks have been discussed. In this study, the Evidence in eHealth Evaluation model was developed to exhibit how evidence can be generated by evaluating certain aspects in each intervention phase. Assessing distinct aspects during distinct phases is a novel concept discussed in this study and requires further analysis. Moreover, this study implies an inconsistency between the literary concepts and practices of eHealth intervention, which has not been noted until now.

As health interventions are context-specific, the transferability of results from eHealth studies may be difficult. Moreover, neither the conceptual nor the case studies suggested the long-term implementation of specific technology into the health care settings where it has been tested. We believe that this might be caused by a lack of or insufficiency of preliminary evidence of the effectiveness and efficiency after conducting the micro-trials or short-term tests on the effects of the technology. Consequently, it appears that lack of evidence hinders the growth of eHealth. Further research directed toward evidence-based evaluation can not only improve the quality of that intervention study but also facilitate long-term implementation of eHealth in general. We conclude that the development of more robust and comprehensive evaluation of eHealth studies or an improved validation of evaluation methods could ease the transferability of results among similar studies. Thus, the resources can be used for supplementary research in eHealth.

Limitations

This study is not devoid of limitations. We tried to include and analyze as many papers as possible; however, unknowingly and unintentionally, some papers may have been omitted. Furthermore, regarding the model, its development is in the preliminary stages; therefore, it cannot be compared with other validated frameworks.

Acknowledgments

This study was partly funded by the European Union's Horizon 2020 research and innovation programme, grant agreement number 643588.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Criteria of the summary table developed to record information from the selected papers.

[PDF File (Adobe PDF File), 29KB - jmir_v20i11e10971_app1.pdf]

Multimedia Appendix 2

Evaluation reported in empirical studies of eHealth interventions.

[PDF File (Adobe PDF File), 168KB - jmir_v20i11e10971_app2.pdf]

References

1. Rigby M, Ammenwerth E. The need for evidence in health informatics. Stud Health Technol Inform 2016;222:3-13. [Medline: 27198087]
2. Rigby M, Magrabi F, Scott P, Doupi P, Hypponen H, Ammenwerth E. Steps in moving evidence-based health informatics from theory to practice. Healthc Inform Res 2016 Oct;22(4):255-260 [FREE Full text] [doi: 10.4258/hir.2016.22.4.255] [Medline: 27895956]
3. Al-Shorbaji N. Is there and do we need evidence on eHealth interventions? IRBM 2013 Feb;34(1):24-27. [doi: 10.1016/j.irbm.2012.11.001]
4. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005;7(1):e1. [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
5. 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]
6. Mark MM, Greene JC, Shaw I, editors. The SAGE Handbook of Evaluation. London, UK: SAGE Publication Ltd; 2006.
7. Newcomer KE, Hatry HP, Wholey JS. Handbook of practical program evaluation. USA: John Wiley & Sons; 2015.
8. Pagliari C. Design and evaluation in eHealth: challenges and implications for an interdisciplinary field. J Med Internet Res 2007;9(2):e15 [FREE Full text] [doi: 10.2196/jmir.9.2.e15] [Medline: 17537718]
9. Bahati R, Guy S, Bauer M, Gwadry-Sridhar F. Where's the evidence for evidence-based knowledge in ehealth systems? In: Proceedings of the Developments in E-systems Engineering (DESE). 2010 Sep Presented at: Developments in E-systems Engineering (DESE); September, 2010; London, UK p. 29-34 URL: https://doi-org.proxy.lib.chalmers.se/10.1109/DeSE.2010.12 [doi: 10.1109/DeSE.2010.12]
10. Bates DW, Wright A. Evaluating eHealth: undertaking robust international cross-cultural eHealth research. PLoS Med 2009 Sep;6(9):e1000105 [FREE Full text] [doi: 10.1371/journal.pmed.1000105] [Medline: 19753106]
11. Catwell L, Sheikh A. Evaluating eHealth interventions: the need for continuous systemic evaluation. PLoS Med 2009 Aug;6(8):e1000126 [FREE Full text] [doi: 10.1371/journal.pmed.1000126] [Medline: 19688038]
12. Lilford RJ, Foster J, Pringle M. Evaluating eHealth: how to make evaluation more methodologically robust. PLoS Med 2009 Nov;6(11):e1000186 [FREE Full text] [doi: 10.1371/journal.pmed.1000186] [Medline: 19956674]
13. Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review--a new method of systematic review designed for complex policy interventions. J Health Serv Res Policy 2005 Jul;10 Suppl 1:21-34. [doi: 10.1258/1355819054308530] [Medline: 16053581]
14. Greenhalgh T, Russell J. Why do evaluations of eHealth programs fail? An alternative set of guiding principles. PLoS Med 2010;7(11):e1000360 [FREE Full text] [doi: 10.1371/journal.pmed.1000360] [Medline: 21072245]
15. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. 1996. Clin Orthop Relat Res 2007 Feb;455:3-5. [Medline: 17340682]
16. Geanellos R, Wilson C. Building bridges: knowledge production, publication and use. Commentary on Tonelli (2006), Integrating evidence into clinical practice: an alternative to evidence-based approaches. J Eval Clin Pract 2006 Jun;12(3):299-305. [doi: 10.1111/j.1365-2753.2006.00592.x] [Medline: 16722913]
17. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011 Jan;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
18. 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. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
19. Alderson P, Clarke M, Mulrow C, Oxman A. Introduction. In: Green S, Higgins JP, editors. Cochrane Handbook for Systematic Reviews of Interventions. 1st ed. Chichester (UK): John Wiley & Sons; 2008:6-34.
20. Harrison V, Proudfoot J, Wee PP, Parker G, Pavlovic DH, Manicavasagar V. Mobile mental health: review of the emerging field and proof of concept study. J Ment Health 2011 Dec;20(6):509-524. [doi: 10.3109/09638237.2011.608746] [Medline: 21988230]
21. Whittaker R, Merry S, Dorey E, Maddison R. A development and evaluation process for mHealth interventions: examples from New Zealand. J Health Commun 2012;17 Suppl 1:11-21. [doi: 10.1080/10810730.2011.649103] [Medline: 22548594]
22. Linn AJ, van Weert JC, Smit EG, Perry K, van Dijk L. 1+1=3? The systematic development of a theoretical and evidence-based tailored multimedia intervention to improve medication adherence. Patient Educ Couns 2013 Dec;93(3):381-388. [doi: 10.1016/j.pec.2013.03.009] [Medline: 23579041]
23. van Gemert-Pijnen JE, Maarten NN, van Limburg A, Kelders SM, Van Velsen L, Brandenburg B, et al. eHealth wiki-platform to increase the uptake and impact of eHealth technologies. In: Proceedings of 4th International Conference on eHealth,

Telemedicine, and Social Medicine eTELEMED. 2012 Presented at: 4th International Conference on eHealth, Telemedicine, and Social Medicine eTELEMED; Jan 30- Feb 4; Valencia, Spain p. 184.
24. Steele Gray C, Khan AI, Kuluski K, McKillop I, Sharpe S, Bierman AS, et al. Improving patient experience and primary care quality for patients with complex chronic disease using the electronic patient-reported outcomes tool: adopting qualitative methods into a user-centered design approach. JMIR Res Protoc 2016;5(1):e28 [FREE Full text] [doi: 10.2196/resprot.5204] [Medline: 26892952]
25. Van Velsen L, Wentzel J, Van Gemert-Pijnen JE. Designing eHealth that matters via a multidisciplinary requirements development approach. JMIR Res Protoc 2013 Jun 24;2(1):e21 [FREE Full text] [doi: 10.2196/resprot.2547] [Medline: 23796508]
26. Nguyen HQ, Cuenco D, Wolpin S, Benditt J, Carrieri-Kohlman V. Methodological considerations in evaluating eHealth interventions. Can J Nurs Res 2007 Mar;39(1):116-134. [Medline: 17450708]
27. Dansky KH, Thompson D, Sanner T. A framework for evaluating eHealth research. Eval Program Plann 2006 Nov;29(4):397-404. [doi: 10.1016/j.evalprogplan.2006.08.009] [Medline: 17950868]
28. Nykänen P, Brender J, Talmon J, de Keizer N, Rigby M, Beuscart-Zephir MC, et al. Guideline for good evaluation practice in health informatics (GEP-HI). Int J Med Inform 2011 Dec;80(12):815-827. [doi: 10.1016/j.ijmedinf.2011.08.004] [Medline: 21920809]
29. Cornford T, Doukidis G, Forster D. Experience with a structure, process and outcome framework for evaluating an information system. Omega 1994 Sep;22(5):491-504. [doi: 10.1016/0305-0483(94)90030-2]
30. Lau F, Kuziemsky C, editors. Clinical adoption framework. In: Handbook of eHealth Evaluation: An Evidence-Based Approach. Victoria: University of Victoria Press; Feb 2017:55-76.
31. Liberati A, Sheldon TA, Banta HD. EUR-ASSESS project subgroup report on methodology: methodological guidance for the conduct of health technology assessment. Int J Technol Assess Health Care 1997;13(2):186-219. [Medline: 9194352]
32. Takian A, Petrakaki D, Cornford T, Sheikh A, Barber N, National NHS Care Records Service Evaluation Team. Building a house on shifting sand: methodological considerations when evaluating the implementation and adoption of national electronic health record systems. BMC Health Serv Res 2012 Apr 30;12:105 [FREE Full text] [doi: 10.1186/1472-6963-12-105] [Medline: 22545646]
33. Cresswell KM, Sheikh A. Undertaking sociotechnical evaluations of health information technologies. Inform Prim Care 2014;21(2):78-83 [FREE Full text] [doi: 10.14236/jhi.v21i2.54] [Medline: 24841408]
34. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [Medline: 10474547]
35. Lampe K, Mäkelä M, Garrido MV, Anttila H, Autti-Rämö I, Hicks NJ, European network for Health Technology Assessment (EUnetHTA). The HTA core model: a novel method for producing and reporting health technology assessments. Int J Technol Assess Health Care 2009 Dec;25 Suppl 2:9-20. [doi: 10.1017/S0266462309990638] [Medline: 20030886]
36. Kidholm K, Ekeland AG, Jensen LK, Rasmussen J, Pedersen CD, Bowes A, et al. A model for assessment of telemedicine applications: mast. Int J Technol Assess Health Care 2012 Jan;28(1):44-51. [doi: 10.1017/S0266462311000638] [Medline: 22617736]
37. Leon N, Schneider H, Daviaud E. Applying a framework for assessing the health system challenges to scaling up mHealth in South Africa. BMC Med Inform Decis Mak 2012;12:123 [FREE Full text] [doi: 10.1186/1472-6947-12-123] [Medline: 23126370]
38. Lo O, Fan L, Buchanan W, Thuemmler C. Technical evaluation of an e-health platform. In: Proceedings of the IADIS International Conference of E-Health. 2012 Jul Presented at: IADIS International Conference of E-Health; July 17-19, 2012; Lisbon, Portugal.
39. Yusof MM, Kuljis J, Papazafeiropoulou A, Stergioulas LK. An evaluation framework for Health Information Systems: human, organization and technology-fit factors (HOT-fit). Int J Med Inform 2008 Jun;77(6):386-398. [doi: 10.1016/j.ijmedinf.2007.08.011] [Medline: 17964851]
40. Lovejoy TI, Demireva PD, Grayson JL, McNamara JR. Advancing the practice of online psychotherapy: an application of Rogers' diffusion of innovations theory. Psychotherapy (Chic) 2009 Mar;46(1):112-124. [doi: 10.1037/a0015153] [Medline: 22122574]
41. Ohinmaa A, Hailey D, Roine R. Elements for assessment of telemedicine applications. Int J Technol Assess Health Care 2001;17(2):190-202. [Medline: 11446131]
42. Aschbrenner KA, Naslund JA, Gill LE, Bartels SJ, Ben-Zeev D. A qualitative study of client-clinician text exchanges in a mobile health intervention for individuals with psychotic disorders and substance use. J Dual Diagn 2016;12(1):63-71 [FREE Full text] [doi: 10.1080/15504263.2016.1145312] [Medline: 26829356]
43. Mohr DC, Duffecy J, Ho J, Kwasny M, Cai X, Burns MN, et al. A randomized controlled trial evaluating a manualized TeleCoaching protocol for improving adherence to a web-based intervention for the treatment of depression. PLoS One 2013 Aug;8(8):e70086 [FREE Full text] [doi: 10.1371/journal.pone.0070086] [Medline: 23990896]
44. Kleiboer A, Donker T, Seekles W, van Straten SA, 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]

45. Cristancho-Lacroix V, Moulin F, Wrobel J, Batrancourt B, Plichart M, De Rotrou J, et al. A web-based program for informal caregivers of persons with Alzheimer's disease: an iterative user-centered design. JMIR Res Protoc 2014;3(3):e46 [FREE Full text] [doi: 10.2196/resprot.3607] [Medline: 25263541]
46. Klein B, Meyer D, Austin DW, Kyrios M. Anxiety online: a virtual clinic: preliminary outcomes following completion of five fully automated treatment programs for anxiety disorders and symptoms. J Med Internet Res 2011;13(4):e89 [FREE Full text] [doi: 10.2196/jmir.1918] [Medline: 22057287]
47. Pratt SI, Naslund JA, Wolfe RS, Santos M, Bartels SJ. Automated telehealth for managing psychiatric instability in people with serious mental illness. J Ment Health 2015;24(5):261-265 [FREE Full text] [doi: 10.3109/09638237.2014.928403] [Medline: 24988132]
48. Zimmer B, Moessner M, Wolf M, Minarik C, Kindermann S, Bauer S. Effectiveness of an Internet-based preparation for psychosomatic treatment: results of a controlled observational study. J Psychosom Res 2015 Nov;79(5):399-403. [doi: 10.1016/j.jpsychores.2015.09.008] [Medline: 26526315]
49. Ali L, Krevers B, Söström N, Skärsäter I. Effectiveness of web-based versus folder support interventions for young informal carers of persons with mental illness: a randomized controlled trial. Patient Educ Couns 2014 Mar;94(3):362-371. [doi: 10.1016/j.pec.2013.10.020] [Medline: 24341963]
50. Bergmo TS, Ersdal G, Rødseth E, Berntsen G. Electronic messaging to improve information exchange in primary care. In: Proceedings 5th International Conference on eHealth, Telemedicine, and Social Medicine eTELEMED. 2013 Presented at: 5th International Conference on eHealth, Telemedicine, and Social Medicine eTELEMED; Feb 24-Mar 1, 2013; Nice, France.
51. Pham Q, Khatib Y, Stansfeld S, Fox S, Green T. Feasibility and efficacy of an mhealth game for managing anxiety: “flowy” randomized controlled pilot trial and design evaluation. Games Health J 2016 Feb;5(1):50-67. [doi: 10.1089/g4h.2015.0033] [Medline: 26536488]
52. Meglic M, Furlan M, Kuzmanic M, Kozel D, Baraga D, Kuhar I, et al. Feasibility of an eHealth service to support collaborative depression care: results of a pilot study. J Med Internet Res 2010;12(5):e63 [FREE Full text] [doi: 10.2196/jmir.1510] [Medline: 21172765]
53. Ebert DD, Lehr D, Baumeister H, Boß L, Riper H, Cuijpers P, et al. GET.ON Mood Enhancer: efficacy of Internet-based guided self-help compared to psychoeducation for depression: an investigator-blinded randomised controlled trial. Trials 2014;15:39 [FREE Full text] [doi: 10.1186/1745-6215-15-39] [Medline: 24476555]
54. Skidmore ER, Butters M, Whyte E, Grattan E, Shen J, Terhorst L. Guided training relative to direct skill training for individuals with cognitive impairments after stroke: a pilot randomized trial. Arch Phys Med Rehabil 2017 Dec;98(4):673-680 [FREE Full text] [doi: 10.1016/j.apmr.2016.10.004] [Medline: 27794487]
55. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, et al. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 2011;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]
56. Robertson A, Cresswell K, Takian A, Petrakaki D, Crowe S, Cornford T, et al. Implementation and adoption of nationwide electronic health records in secondary care in England: qualitative analysis of interim results from a prospective national evaluation. BMJ 2010;341:c4564 [FREE Full text] [Medline: 20813822]
57. Bouamrane MM, 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]
58. Langrial SU, Lappalainen P. Information systems for improving mental healthix emerging themes of research. In: Proceedings of the 20th Pacific Asia Conference on Informatics Systems (PACIS). 2016 Presented at: 20th Pacific Asia Conference on Informatics Systems (PACIS); Jun 27- Jul 1, 2016; Chiayi, Taiwan.
59. Holländare F, Eriksson A, Lövgren L, Humble MB, Boersma K. Internet-based cognitive behavioral therapy for residual symptoms in bipolar disorder type II: a single-subject design pilot study. JMIR Res Protoc 2015;4(2):e44 [FREE Full text] [doi: 10.2196/resprot.3910] [Medline: 25908235]
60. Glozier N, Christensen H, Naismith S, Cockayne N, Donkin L, Neal B, et al. Internet-delivered cognitive behavioural therapy for adults with mild to moderate depression and high cardiovascular disease risks: a randomised attention-controlled trial. PLoS One 2013;8(3):e59139 [FREE Full text] [doi: 10.1371/journal.pone.0059139] [Medline: 23555624]
61. Lederman R, Wadley G, Gleeson J, Bendall S, Álvarez-Jiménez M. Moderated online social therapy: designing and evaluating technology for mental health. ACM Trans Comput Hum Interact (TOCHI) 2014 Feb 1;21(1):1-26. [doi: 10.1145/2513179]
62. Burton C, Szentagotai Tatar A, McKinstry B, Matheson C, Matu S, Moldovan R, Help4Mood Consortium. Pilot randomised controlled trial of Help4Mood, an embodied virtual agent-based system to support treatment of depression. J Telemed Telecare 2016 Sep;22(6):348-355. [doi: 10.1177/1357633X15609793] [Medline: 26453910]
63. Schaller S, Marinova-Schmidt V, Gobin J, Criegee-Rieck M, Griebel L, Engel S, et al. Tailored e-Health services for the dementia care setting: a pilot study of 'eHealthMonitor'. BMC Med Inform Decis Mak 2015 Jul 28:15-58 [FREE Full text] [doi: 10.1186/s12911-015-0182-2] [Medline: 26215731]

64. Frangou S, Sachpazidis I, Stassinakis A, Sakas G. Telemonitoring of medication adherence in patients with schizophrenia. Telemed J E Health 2005 Dec;11(6):675-683. [doi: 10.1089/tmj.2005.11.675] [Medline: 16430387]
65. de Wit J, Dozeman E, Ruwaard J, Alblas J, Riper H. Web-based support for daily functioning of people with mild intellectual disabilities or chronic psychiatric disorders: a feasibility study in routine practice. Internet Interv 2015 May;2(2):161-168. [doi: 10.1016/j.invent.2015.02.007]

Abbreviations

EBM: evidence-based medicine
eHealth: electronic health
PRISMA: Preferred reporting items for systematic reviews and meta-analyses

Edited by G Eysenbach; submitted 07.05.18; peer-reviewed by F Jahedi, R Jacobs; comments to author 07.06.18; revised version received 01.08.18; accepted 01.08.18; published 23.11.18.

Please cite as:
Enam A, Torres-Bonilla J, Eriksson H
Evidence-Based Evaluation of eHealth Interventions: Systematic Literature Review
J Med Internet Res 2018;20(11):e10971
URL: http://www.jmir.org/2018/11/e10971/
doi:10.2196/10971
PMID:30470678

Viewpoint

Rethinking Data Sharing at the Dawn of a Health Data Economy: A Viewpoint

Chunlei Tang[1,2,3*], PhD; Joseph M Plasek[1,4*], MS; David W Bates[1,2,3], MSc, MD

[1]Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Boston, MA, United States

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

[3]Clinical and Quality Analysis, Partners HealthCare System, Boston, MA, United States

[4]Department of Biomedical Informatics, School of Medicine, University of Utah, Salt Lake City, UT, United States

[*]these authors contributed equally

Corresponding Author:
Chunlei Tang, PhD
Division of General Internal Medicine and Primary Care
Brigham and Women's Hospital
1620 Tremont Street, BS-3
Boston, MA, 02120
United States
Phone: 1 (857) 600 0628
Email: ctang5@partners.org

Abstract

A health data economy has begun to form, but its rise has been tempered by the profound lack of sharing of both data and data products such as models, intermediate results, and annotated training corpora, and this severely limits the potential for triggering economic cluster effects. Economic cluster effects represent a means to elicit benefit from economies of scale from internal data innovations and are beneficial because they may mitigate challenges from external sources. Within institutions, data product sharing is needed to spark data entrepreneurship and data innovation, and cross-institutional sharing is also critical, especially for rare conditions.

(J Med Internet Res 2018;20(11):e11519) doi:10.2196/11519

KEYWORDS

economics, hospital; machine learning; models, economic; precision medicine

Data innovation and data entrepreneurship have the potential to dramatically alter the current health care landscape as health data economy is beginning to revolutionize the field [1-3]. The European Commission estimated that the value of the European Union data economy would increase to US $860 (€739) billion by 2020, up from US $331 (€285) billion in 2015 [4,5]. Data economy, wherein health care will increasingly participate, has formed, and it is lucrative and quickly growing. Sharing data is necessary to enable thriving health data economy and produce clinical advances that are not possible in the current health care environment because of siloed data resources. These data resources span from the bench to the bedside and beyond, including genetic, genomic, proteomic, clinical, imaging, patient-centered, public health, and other relevant data. Electronic health record systems enable health care organizations to share clinical data across their organization, with patients themselves through patient portals, and to a limited extent owing to a lack of interoperability, with other organizations or systems. Rethinking how we share data and data products is essential for health data economy to thrive.

Data products, such as models, intermediate results, and annotated training corpora, are the outcomes from data preparation, processing, and analysis (eg, statistical analysis, data mining, and machine or deep learning). Data products also include visualizations and dashboards created by the artistic manual work of data scientists to assist in the interpretation of the analysis in an actionable way. Data products, like data itself, are "nonrivalrous," meaning that they can be utilized by >1 data scientist at a time to create additional data products or services. For example, critical to the development of deep neural networks for image recognition tasks is the training set of >10,000 labeled images on ImageNet [6] created by manual annotation efforts that were made publicly available. Similarly, raw journal article titles can be easily searched through PubMed or MEDLINE, yet a data product from this resource that is created after standard text processing techniques (eg, tokenization and

stop-word removal) have been applied is usable for many subsequent analyses. However, similar data products at scale tend to not be available in health care, resulting in a lack of generalizability for models and concerns regarding the reproducibility of results.

Sharing data products across health care provider networks can reveal different insights into different clinical departments and may also indirectly promote the core business of health care through better revenue and profitability margins, as data products can easily be used for secondary purposes. The second benefit of data sharing is to allow data to spread beyond the current data silos, which would facilitate data entrepreneurship, data innovation, data processing, and secondary data mining.

Data products need not contain identifiable patient data that would be useful for general research purposes. Deidentified data products from clinical care must be treated with appropriate care and respect. If one had a covariance matrix and corresponding mean vector for variables, one could run regression or advanced analyses using structural equation modeling to explore latent variables that were not even postulated in the original research. The National NLP Clinical Challenges [7] provides annotated, fully deidentified corpora of clinical notes centering around particular clinical tasks, allowing researchers to start with a verified gold standard and benchmark their systems against others. As the Medical Information Mart for Intensive Care III [8] contains both structured and unstructured data and is accessible to researchers, any data products (eg, annotated clinical notes and models) built on top of this or similar resources, should they be made available, could be openly critiqued and improved upon by the community.

Learning health care systems and precision medicine are two data-driven innovations at different scales in the health care data environment, where sharing data and data products are most applicable. Learning health systems are centered on the organization where new knowledge is captured as an integral byproduct of the delivery experience [9]. For example, electronic health record data that contain rich clinical information (eg, patients' medical history, family history, surgical resection approach, and postoperative supervision) offer an opportunity to design algorithms for acute interventions, such as predicting 30-day hospital readmission or whether a patient is at risk for cardiac decompensation. Similarly, exploring care process protocols, including a combination of medications, for a specific disease could inform drug inventory management. Precision medicine represents a leading driver of the health data economy in which health care recommendations can be individually tailored on the basis of a person's genes, lifestyle, and environment [10]. Similarity-based classifiers aimed at automatically grouping patients with similar characteristics together enable improvements in assessment, diagnosis, the selection of therapeutic choice, and the prediction of prognosis. For example, abnormalities in a clinical pathway could be highlighted using trend recognition algorithms to identify a similarity cohort to allow the assessment of the complexity associated with a disease cluster. Furthermore, sharing data is critical for rare diseases, both from a learning health care system perspective to optimize the delivery of care and a precision medicine perspective to be able to effectively personalize the care plan.

We envision that economic cluster effects (ie, a geographic concentration of interconnected stakeholders and their associated institutions in a field through a nested interorganizational network of relationships) within the health data economy will emerge soon, but that the sharing of data products will be necessary to maximize their potential. Multistakeholder health data governance would be beneficial, as it would allow balancing of value for all actors (eg, clinicians, patients, and other data generators; data scientists, researchers, and other data product enhancers), which is useful in determining not only how data products should be owned but also what types of data should be shared to maximize data resource utilization toward the problems of interest to the community. The status quo is far from optimal from an economic perspective, and we collectively have poorer health [11] because of this lack of sharing and void in meaningful governance. From a technical perspective, blockchain or similar technologies can be utilized to insure the integrity of shared data and data products. Only with the wide availability and use of diverse data products will the future of learning health systems and precision medicine be truly accessible in the emerging health data economy.

References

1. Tang C. The data industry: The business and economics of information and big data. In: John Wiley & Sons. New Jersey: John Wiley & Sons, Inc; May 04, 2016:216.
2. The Economist. 2018. The world's most valuable resource is no longer oil, but data URL: https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource [WebCite Cache ID 722Mcg4ak]
3. The Economist. 2018. Data is giving rise to a new economy URL: https://www.economist.com/news/briefing/21721634-how-it-shaping-up-data-giving-rise-new-economy [WebCite Cache ID 722NA8tNd]
4. Hawksworth J, Audino H, Clarry R. PricewaterhouseCoopers. 2017. The long view: how will the global economic order change by 2050? URL: https://www.pwc.com/gx/en/world-2050/assets/pwc-the-world-in-2050-full-report-feb-2017.pdf
5. The European Commission. 2017. Building a European data economy URL: https://www.mayerbrown.com/Files/News/825879d1-6355-4235-ae2d-f2fe9144a335/Presentation/NewsAttachment/912cb16d-f749-4de1-bd5c-f57f5887b1c4/The-European-Files-Building-the-european-data-economy-Sept-2017-Issue-48.pdf [accessed 2018-10-12]

6. Department of Biomedical Informatics at Harvard Medical School. n2c2 builds on the legacy of i2b2 URL: http://dbmi.hms.harvard.edu/programs/healthcare-data-science-program/clinical-nlp-research-data-sets [accessed 2018-10-12] [WebCite Cache ID 72F1hlv2n]
7. Johnson AEW, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 24;3:160035 [FREE Full text] [doi: 10.1038/sdata.2016.35] [Medline: 27219127]
8. Lotterman E. TwinCities.com. 2018 Aug 26. Edward Lotterman: Information has value, but we often have no way to use it, August 26, 2018, archived URL: https://www.twincities.com/2018/08/26/edward-lotterman-information-has-value-but-we-often-have-no-way-to-use-it/ [WebCite Cache ID 72F2WWi2s]
9. Deng J, Dong W, Socher R, Li LJ, Li K, Li FF. Imagenet: A large-scale hierarchical image database. 2009 Presented at: Processing of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2009; US p. 807-829.
10. Olsen LA, Aisner D, McGinnis JM. The learning healthcare system: Workshop summary. In: Institute of Medicine (US) Roundtable on Evidence-Based Medicine. Washington (DC): National Academies Press; 2007.
11. Hodson R. Precision medicine. Nature 2016 Dec 08;537(7619):S49. [doi: 10.1038/537S49a] [Medline: 27602738]

Edited by G Eysenbach; submitted 18.07.18; peer-reviewed by Y Xiong, K Sward, J Goris, W Chen, Y Tan, J Sheng, Y Lin, F Chang; comments to author 29.08.18; revised version received 06.09.18; accepted 08.09.18; published 22.11.18.

Please cite as:
Tang C, Plasek JM, Bates DW
Rethinking Data Sharing at the Dawn of a Health Data Economy: A Viewpoint
J Med Internet Res 2018;20(11):e11519
URL: http://www.jmir.org/2018/11/e11519/
doi:10.2196/11519
PMID:30467103

Tutorial

Using Facebook for Large-Scale Online Randomized Clinical Trial Recruitment: Effective Advertising Strategies

Laura Akers[1], PhD; Judith S Gordon[2], PhD

[1]Oregon Research Institute, Eugene, OR, United States

[2]College of Nursing, University of Arizona, Tucson, AZ, United States

Corresponding Author:
Laura Akers, PhD
Oregon Research Institute
1776 Millrace Drive
Eugene, OR, 97403
United States
Phone: 1 541 484 2123
Fax: 1 541 484 1108
Email: lauraa@ori.org

Abstract

Targeted Facebook advertising can be an effective strategy to recruit participants for a large-scale online study. Facebook advertising is useful for reaching people in a wide geographic area, matching a specific demographic profile. It can also target people who would be unlikely to search for the information and would thus not be accessible via Google AdWords. It is especially useful when it is desirable not to raise awareness of the study in a demographic group that would be ineligible for the study. This paper describes the use of Facebook advertising to recruit and enroll 1145 women over a 15-month period for a randomized clinical trial to teach support skills to female partners of male smokeless tobacco users. This tutorial shares our study team's experiences, lessons learned, and recommendations to help researchers design Facebook advertising campaigns. Topics covered include designing the study infrastructure to optimize recruitment and enrollment tracking, creating a Facebook presence via a *fan page*, designing ads that attract potential participants while meeting Facebook's strict requirements, and planning and managing an advertising campaign that accommodates the rapid rate of diminishing returns for each ad.

(J Med Internet Res 2018;20(11):e290) doi:10.2196/jmir.9372

KEYWORDS

research subject recruitment; advertisements; social media

Introduction

In the past two decades, health interventions via electronic media (electronic health, eHealth) have become widespread. Some studies continue to recruit through traditional methods such as print and broadcast media, flyers, and word-of-mouth, whereas others have taken advantage of online methods such as social media publicity and search engine advertisements.

Facebook has become a major player in the field of digital advertising, with about US $8 billion revenue in 2015 [1]. Facebook allows advertisers to create target audiences by specifying gender, marital status, age, and geographic region as well as other personal characteristics. About 74% of adult American women and 62% of adult American men use Facebook [2]. Facebook is widely used in all age groups (78% of age 30-49 years, 65% of age 50-64 years, and 41% of age 65 years and above) and across racial and ethnic categories (67% of whites, 70% of blacks, and 73% of Hispanics) [2].

Facebook is now widely used in research. In their 2015 *American Psychologist* article, Kosinski et al [3] summarized a wide variety of ways by which social scientists are now using Facebook as a research tool, from recruiting participants (with paid advertising or snowball sampling) to tracking participants across studies and staying connected with participants over time, to collecting and using Facebook profile data, and collecting self-reports. Their detailed discussion of ethical considerations focuses on privacy, consent, and appropriate boundaries. Facebook has also been used to deliver interventions, for example, increasing physical activity among young adult cancer survivors [4] and reducing problem drinking at a university [5]. Whitaker et al [6] reviewed 35 unique studies that used Facebook to recruit participants for health, medical, or psychosocial research and found that the participants were

broadly similar to those recruited via traditional methods, although various studies reported an over-representation of women, white people, younger people, and people who were better educated or had a higher rate of income than the general public. Topolovec-Vranic and Natarajan [7] reviewed 30 studies that used both social media and at least one other method for recruitment; of the 14 studies reporting demographic information, 12 studies found that their social media sample was different from those recruited via traditional methods. Frandsen et al [8] showed that participants recruited to their smoking cessation study via social as compared with traditional media were somewhat younger (mean age: 39.3 years vs 44.9 years) but otherwise demographically similar (gender, education, and income). Those recruited via traditional media were more likely to follow the study protocol and complete the study.

To date, most of the studies reporting the use of Facebook recruitment have been surveys [9-17] or modest-sized randomized clinical trials (RCTs) [18-20] and pilot studies [21]. Facebook can also be used to recruit participants for larger-scale RCTs [22,23], but this requires careful planning and oversight to ensure the optimal use of funds allocated for recruitment. Our team relied on Facebook advertising for a 2010 pilot study, recruiting 522 women who were wives or domestic partners of smokeless tobacco users, and we used a similar strategy for a larger RCT, recruiting 1145 participants from the United States and Canada between August 2015 and November 2016. Interested women visited the study website and completed information for eligibility screening, gave informed consent, provided contact information, and completed a baseline survey entirely online. Participating women were then automatically randomized by the Web-based program to receive an intervention (access to an interactive website plus mailed booklet) or delayed treatment (receiving the intervention after completing a 7.5-month follow-up assessment). The intervention program was designed to help the women encourage their partner to consider quitting smokeless tobacco, support him if he decided to do so, and accept his decision if he was not interested in quitting [24]. All study activities were conducted online, with the exception of phone calls to complete follow-up assessments for women who did not respond to email reminders. Ethics approval for the RCT was obtained from the institutional review board at Oregon Research Institute. This study is registered at ClinicalTrials.gov (ID: NCT01885221).

The goal of this paper is to share strategies and *lessons learned* from implementing Facebook advertising to recruit large samples from a narrowly targeted population. As Facebook's advertising procedures and features change frequently, a specific *how to* article is not feasible. Instead, this paper provides general guidelines and principles that should remain valid for the foreseeable future.

Steps and Strategies

Deciding Whether Facebook Advertising Is Right for Your Study

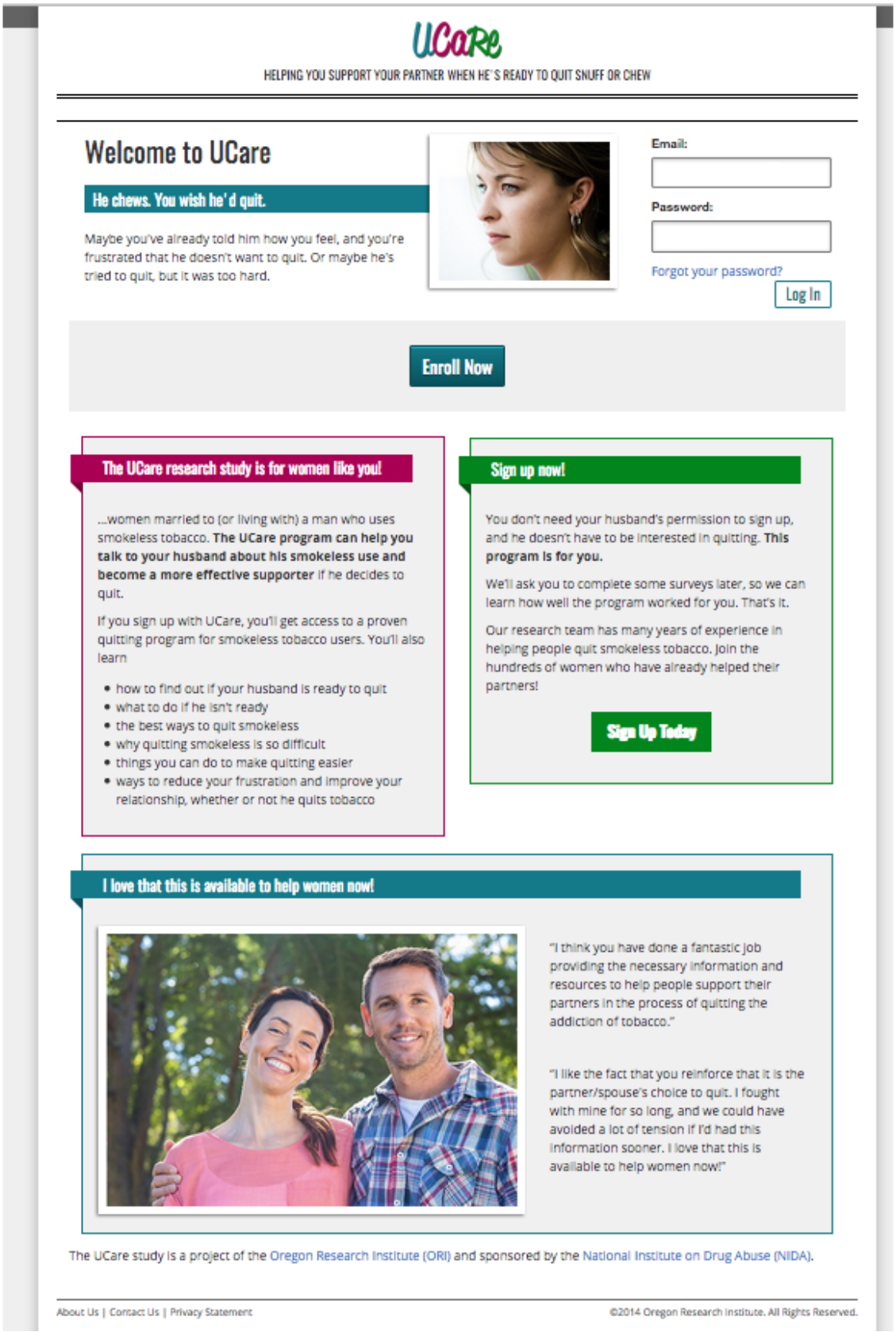
A wide variety of methods exist for recruiting participants for online RCTs, ranging from traditional print and broadcast media and in-person recruitment to innovative uses of social networks (eg, virtual snowball sampling [25]). A comparison of studies recruiting via social media and at least one other method found that for those studies reporting cost-effectiveness, about half found social media to be more cost-effective and about half did not [7]. Many companies offer online paid advertising, and many times, other services will be more cost-effective than Facebook. When recruiting participants for a study topic related to words that potential participants are likely to type into search engines, Google AdWords is an obvious option (eg, for depression prevention [26]). For studies related to specific health conditions, existing online communities may provide research studies with access to their members (although in our experience, many communities now tend to view research as exploitive and will not cooperate). Facebook advertising can be especially useful for studies targeting specific geographical or demographic groups, for conditions for which participants are not already seeking help (eg, using search engines), and for topics for which friends and family would be likely to refer them to the study (ie, by *tagging* them when they see the ad, that is, naming their friend in a comment on the ad in such a manner that the friend receives a notification). Facebook advertising is especially suitable for cases when the study team wishes to avoid exposing a particular demographic group to the advertisements—in our case, we had previously found that men interested in quitting smokeless tobacco would sometimes claim to be women (when undergoing online screening) to get access to our studies for women. To minimize this form of participant fraud, we sought to advertise in locations where only women would see the ad.

Designing Infrastructure for Enrolling and Managing Participants and for Monitoring Facebook Ads Over Time

At the intervention design stage, we recommend taking several factors into account to facilitate recruitment and enrollment in eHealth studies, which are challenged by the lack of personal contact with participants. It is imperative to have an infrastructure in place that allows easy access for potential participants and real-time monitoring of recruitment strategies linked to participant enrollment. Planning and creating this infrastructure will typically take place many months before study recruitment begins.

First, in our experience, we have found that enrollment for an online study is generally more successful if potential participants are allowed to stay within a single medium, that is, entirely online, rather than asking them to switch from being online to using a phone or mail. Changing from one medium to another may feel more burdensome as it may involve having to create a reminder to do so at a later time (eg, when the person is done with their online activities). In an earlier study, when we switched from asking participants to call a phone number they saw online to registering online directly, enrollment dramatically improved. If phone contact is an important step in enrollment for the study, we recommend that the research team initiate such contact without relying on the potential participant (eg, ask the participant to provide their phone number for the researchers to call).

Figure 1. UCare study home page.



Second, the study home page, where the potential participant will go after clicking on the ad, should be pleasing to the eye, maximally welcoming, and aimed at a lay audience. Our first attempt at a study home page was a research-oriented, text-heavy description of the study conditions and procedures. On the basis of poor response from our first few days of recruitment, we overhauled the page, rewriting the text in a friendlier style that focused on the woman's concerns about her partner and how study participation could benefit her. We added more graphics and user testimonials (see Figure 1). Our conversion of

Facebook *clicks* to enrolled participants improved immediately (eg, from 0.5 to 3 participants per day).

Third, as part of the enrollment infrastructure, a study needs either an administrative portal (*admin page*) or other robust participant management system to make it easy to monitor enrollment and the success of recruitment strategies. For maximum usefulness, we recommend that the admin page *dashboard* displays the information needed to monitor the performance of each successive version of the ad (eg, total numbers of people completing each step of the enrollment process) [27], along with a downloadable Excel file of participant-level data for more in-depth tracking and analysis. The participant management system can send an automated email to the research staff at the same time each day with key enrollment figures such as total enrolled overall, total enrolled for each condition, and new enrolled since the previous day. In our study, we used a system custom-developed for us by our software subcontractor, but it may be possible to customize a more generic system for this purpose, such as REDCap.

Staffing the Advertising Campaign

Assign primary responsibility for the Facebook ad campaign to research staff who are detail-oriented, with good writing and graphic design skills and excellent spelling and punctuation. The research staff should have access to the study's online administrative functions and authority to use an organizational credit card. These staff will use their personal Facebook accounts to access the Facebook advertising management system (currently called "Ad Manager"). They will need to search their Facebook home page for a link to *Create Ad* or other reference to Facebook advertising; once they have begun working with Facebook ads, Facebook will likely display the link to access the Ad Manager more prominently.

Running the ad campaign will take approximately 1 hour per week, after an initial investment of 1 to 2 hours to learn how the system works and create the initial ad. Subsequent ads can generally be adapted from the first ad, which would typically take 5 to 10 min 2 or 3 times a week. Overall, the time commitment is small, but daily monitoring of the campaign will allow the staff to respond rapidly to changes in the response rate.

Establishing the Study's Facebook Presence

Before beginning the advertising campaign, it is important to create a Facebook *fan page* for the study, which will be connected to the ad. Ask friends and colleagues to *like* the page so that it will not look new to potential participants. We gave our page the name "UCare Research Study," in the category "Community," with the study home page as the "Website," and a "Short Description" summarizing how the study could potentially help the participants. Assign at least two staff members (at least one of whom should also have responsibility for the ad campaign) as managers of the page; they will receive Facebook notifications of people's interactions with both the page and the study's linked ads.

Keep the fan page active by regularly posting news and related news links to the page, also referred to as *curating* the page. We found that updating the page twice a week was a good rate, balancing the time needed to find and post news items with the desirability of keeping the page content looking fresh. Few of our links received evidence of attention (likes or shares), but their presence makes the page (and by extension, the research study) appear active. If the intervention arm involves providing information to participants or changing their behavior, take care not to reveal or overly hint at the intervention content on this page, as this could potentially contaminate the participants randomized to the control condition.

Designing the Advertising Campaign

Facebook currently uses an advertising structure of *ads* within *ad sets* within *campaigns*. The *campaign* could encompass all ads for a single study, or multiple campaigns could be created for the study. Combining all ads into 1 campaign makes it easier to monitor the costs. The *ad set* (formerly *advert set*) is where the researcher delineates the criteria of the target audience and the budget; the *ad* (formerly *advert*) is where the graphics and text for the ad are designed. Each of the 3 levels must be switched on for an ad to display. While designing your first ads, keep the overall campaign turned off until you are ready to start advertising.

Plan to create multiple *ad sets*, each targeting a meaningful subsection of your overall audience, such as by geographic region, age group, or gender. Plan to switch your ad from one ad set to another (*rotating* it) when each ad reaches *saturation* (the limit of cost-effective enrollment of participants) and new enrollments trickle to a minimum.

Design the ad campaign with a specific rotation strategy in mind (eg, geographic region or a demographic variable). We began by advertising to our target demographic throughout the United States and Canada, but after 9 days and 12,405 clicks, we had randomized only 3 participants. We then chose to rotate our ad to different geographic regions, targeting 1 region at a time and focusing on areas with high smokeless tobacco use prevalence. All our ads targeted women, aged 25 to 65 years and above, who specified their language as English and whose relationship status was married, engaged, domestic partnership, or in a relationship.

The more specifically you can target your ads, the more cost-effective they will be—if you are paying per click, you will not want to encourage clicks by people ineligible for the study. To this end, Facebook also offers more sophisticated targeting options based on *interests* (eg, *healthy eating* and *bodybuilding*); Facebook gleans this information from information that users provide about themselves, such as Facebook pages they like or Facebook groups they join [28]. In other publications about Facebook advertising, *interests* have also been referred to as *keywords* [10,12,29] and *key profile words* [14].

However, as Pedersen and Kurz [30] note, recruiting based on Facebook interests or *likes* will fail to capture many populations. They give the example of a spouse of a problem drinker who will not *like* Al-Anon groups on Facebook if they do not want their Facebook friends to know their partner is a problem drinker. In our case, there were no specific *interests* that were

suitable for our population of women who wanted their partner to quit his use of smokeless tobacco.

Furthermore, one could ask permission to post about a study in a Facebook *group*, a community of people sharing an interest, but approval must be obtained from the group moderator. For our study, there were no organized groups of women concerned about their husbands' smokeless tobacco use. Studies recruiting people with a specific medical diagnosis or shared health promotion activity may have better luck in finding Facebook groups to target. Groups may be found by typing a keyword into the main Facebook search box and then selecting *groups* to display communities fitting that criterion. Valdez et al [31] described their experiences using Facebook *groups* for recruitment; in 2 studies, the groups yielded 166 participants, which they found sufficient for qualitative but not quantitative research. Our efforts to use Facebook groups to publicize an unrelated study were unsuccessful as many groups now have firm policies against mentioning research opportunities.

Facebook offers the opportunity to pay per impression (every time they show someone the ad) or per click (every time someone clicks on the ad). The former is most suited to cases where the advertiser wants the public to be aware of some information but without taking an immediate action; a good example is a political campaign. The latter is best when you want people to take an action such as clicking on the ad to visit your study home page and consider enrolling. We chose to pay per click. The payment amount for Facebook outcomes such as clicks is determined by a highly complex auction process. We chose the default, *automatic bidding*, which (according to the Ad Manager) will "let Facebook set the bid that helps you get the most link clicks at the best price."

Decide how much to budget per day for the ad. We budgeted for US $250 per day, based on our previous experience in Facebook advertising with this population, which had cost US $70 per participant. We anticipated that the US $250 per day budget would yield 3 to 4 participants per day and allow us to complete our recruitment in 10 to 11 months. In the summer of 2016, Facebook decided not to observe the daily budgets strictly but to let them average out over time. Assign a credit card to pay for the ads automatically, and be sure that the credit limit is at least twice the amount budgeted per month so the ads can continue while the previous month's bill is being processed and paid. If it takes longer to pay the bill at your institution, then adjust your credit limit accordingly. After the daily budget is entered into the Ad Manager, it will display the "estimated daily results reach"—the number of people likely to see the ad on a given day. Our numbers were typically 15,000 to 40,000 people, which yielded about 5 participants per day the first time we ran the ad in each location and later yielded 1 to 3 participants per day. When determining their daily advertising budget, researchers should take into account both their study's desired recruitment period and the risk of sinking costs into an ineffective campaign. It is important not to spend more per day than necessary and yet to spend enough so that effective ads can be distinguished from ineffective ads. The Advertising Costs section below discusses the cost issue more generally.

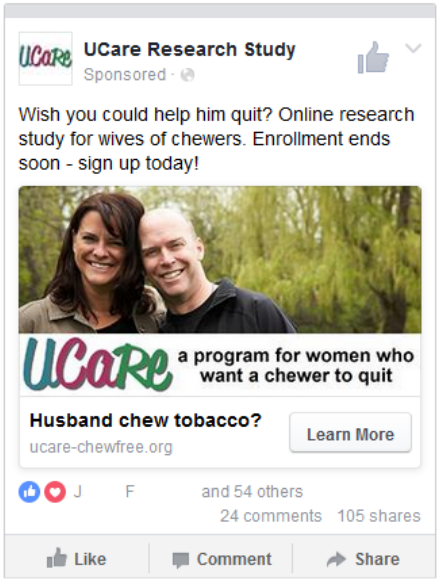
Designing the Ad

Facebook ads have strict design specifications that generally involve a graphic image and several categories of text, which are displayed in varying ways depending on where the ad will appear, for example, in the right margin, in the newsfeed, and on Instagram (Instagram, owned by Facebook, may be better for reaching adolescents than Facebook itself [32].). Ad specifications change frequently. When we began advertising for our pilot study in 2010, Facebook ads could be text-only, with an optional image. Later, the image became a requirement and was generally small and square. When we began recruiting for our new study in August 2015, we found that the image was again required and the dimensions must be 1200x628 pixels (a horizontal orientation). In early 2013, Facebook began regulating the amount of text that could appear within the image, and for some time, it had a limit of 20% [33]. As of late 2016, there was no official percentage limit on text within graphics, but any such text (including product logos) was highly discouraged.

For our study, the graphic theme was happy couples, and because such images convey no information about quitting smokeless tobacco, we did include text in our image to help catch the eye of potential participants (our UCare logo and the phrase "a program for women who want a chewer to quit"). We used stock images for our graphics, showing a smiling younger-middle-aged couple, which seemed sufficiently generic to appeal to both younger and older couples, all of whom are in our target audience. The mean age of women in our final sample was 43.2 years (SD 9.5, range 19-78). We also went with white couples for the ad because the vast majority of smokeless users are white.

For the text part of the ad, Facebook currently allows a brief title and a brief block of descriptive text. The text should be succinct and read like an advertisement, without implying that you are trying to sell something. For maximum credibility, have the text mention that recruitment is for a research study conducted by a hospital, university, or nonprofit organization [34]. An ideal ad would encourage those eligible for the study to click the ad and those who are not eligible (or do not have eligible friends) to ignore it. Note, however, that Facebook prohibits advertisement that "asserts or implies personal attributes including disability or medical condition (including physical or personal health)." They prefer that ads say that help for a condition is available than attempt to engage the reader more directly, for example, "Depression counseling available" rather than "Depression getting you down? Get help now." [35].

Figure 2. Sample Facebook ad.



For our ad title, we used "Husband chew tobacco?" throughout the study (which may now violate the policy just described), and we varied the main block of text to see what would work best. We also changed the text depending on the stage of the study. The first ads mentioned that the study was "new," and toward the end of recruitment, we used "enrollment ends soon" as a method to motivate potential participants to take quick action. For ads in between, we found that "Everything he wishes you knew about quitting smokeless. Online research study – sign up now!" was especially effective. We tested ads that stated that participating in the study was "free," but those ads had a low response rate. We carefully avoided mentioning the financial incentives we provide for completing follow-up surveys, either in the ad or on the study's main marketing page (the home page), to avoid attracting fraudulent participants [36,37]; other studies have been able to mention financial incentives by using an insider knowledge check to screen for fraud (eg, when recruiting veterans, checking to see that rank and pay grade at discharge matched [19,23]). See Figure 2 for an example of our ad. Research teams may do well to conduct formative work to get feedback from target participants on sample ads, including both text and graphics, before the recruitment campaign.

As part of creating the ad, it is important to link the ad both to the study URL (the place where participants are directed when they click on the ad) and to the Facebook fan page.

The "create a similar ad" or "duplicate advert" function can be useful for a campaign that will be rotating among different audiences. Sometimes, the function allowed us to set up the same ad in a different ad set with minimal input, and at other times, we were required to enter all the details again. All the graphics uploaded for use in an ad campaign are saved in the campaign's "image library" and can be accessed again easily.

Each ad will require Facebook approval, which typically takes less than an hour, although we did find that some would occasionally take up to 12 hours. The display often said the ad was approved immediately, but it was not actually approved until we received a notification later. Facebook has specific criteria for approving or denying ads to prevent illegal marketing of certain products or services. For researchers studying addiction, this can result in erroneous denial of ads. Once we had an ad denied for mentioning tobacco—the reviewer (whether human or automated) had assumed that we had fallen afoul of the ban on tobacco advertising. We were able to file an appeal, following instructions we received in the email from Facebook telling us about the denial, but processing the appeal took several days.

Monitoring Facebook Users' Interaction With the Ad

There are 2 aspects of responses to the ad that the researcher can attend to: Facebook activity related to the ad and participants enrolling in the study. The Facebook notification feature will allow the researchers to see easily throughout the day when people are interacting with the ad with *likes*, *shares*, and *comments*, which typically include people *tagging* their friends to see the ad. The staff members who have been assigned as managers for the Facebook fan page and who are creating ads for the study will receive these notifications, which will give them clues regarding the popularity of the study and the times of days the ad is being shown.

In the comments on the ads, many women *tagged* specific friends, which gave us an implied endorsement from the tagger when the tagged woman saw the ad. Our study was unique and may have met a need that many potential participants had been expressing to their friends (eg, "I wish he'd quit!"); advertising on Facebook allowed us to capitalize on this. Many women also tagged their husbands to let them know about their interest in the study, and the husbands often responded. Many men encouraged their wives to sign up for the study or at least assured them that they would try to quit smokeless tobacco.

Tracking Enrollment

If the study admin page allows for the researcher to see at a glance how many participants have enrolled, then monitoring enrollment will take a trivial effort; however, for greatest accuracy in tracking, this monitoring should be done at the same time daily. As mentioned earlier, an automated email with key

enrollment figures can be sent to the research staff to facilitate tracking. We manually tracked the enrollment from each ad, recording the ad targeting and text, the dates, the number enrolling between those dates, and the number enrolling per day. We were able to enroll about 5.6 participants per day during the first 9 weeks of the study (the first pass through all geographic areas targeted for recruitment); the rest of the study averaged 1.9 participants per day.

Strategies for Sustaining a Long-Term Advertising Campaign

Our primary campaign strategy was to rotate the ad among distinct geographic regions, focusing on 1 region at a time so that we could easily see how well each ad was doing (by monitoring the locations of enrollees on our admin page). We would typically give an ad at least 2 days to have an impact, and once people had begun enrolling in response to an ad, we would wait for responses to taper off and then move the ad once it seemed fairly dormant for 2 days. Facebook notifications of user interaction with the ad let us judge whether interest had truly dwindled or whether people were too busy to complete the enrollment on a given day.

We recruited throughout the United States and Canada. As noted above, we began by rotating through broad regions with a high prevalence of smokeless tobacco use. This phase of our campaign lasted about 2 months. Next, we began switching up the graphic and text and rotated through our broad regions again for 4 months. Then, we targeted very specific areas where our ad had done well for about 3 months, making sure to have approximately 15,000 to 40,000 in Facebook's projected *reach* for the ad. To target specific areas, we used a map to identify the named communities in the region, which we entered into Facebook as a list. Facebook allows communities to be designated with a user-specified surrounding radius, allowing suburbs and more rural communities to be included or excluded. For some states with high numbers of smokeless tobacco users but low overall smokeless tobacco prevalence, we excluded the urban areas from the recruitment zone by specifying the names of the cities we wanted to exclude. (We did so anticipating that urban people would use up the budget if not excluded; see the discussion of ineligible or uninterested people clicking the ad in the Saturation: The Problem of Diminishing Returns section below). Researchers can use their own expertise to identify groups they would expect to respond to the ad and tailor their campaigns accordingly.

The time of year is important to take into account. September has historically been a good recruitment time for us, but for this study, September coincided with both the study launch and the "ends soon" ad campaign; therefore, we cannot conclude that the time of year was itself a key factor. January and early spring, like September, are also traditional *new beginnings* times and could potentially be good as well for an intervention focusing on changing health behaviors. Our recruitment was very low from mid-November to early January, even though we tried a boost of advertising in many areas simultaneously during early January to try to capitalize on New Year's resolutions. The appeal to resolutions may be relevant for changing some behaviors such as quitting tobacco but did not appear to engage supporters. Researchers may want to consider the times of year that their prospective participants would be most responsive to their message.

By the time we had finished advertising in the very specific regions described above, we had accumulated around 1500 *likes* on our Facebook fan page, which allowed us to try 2 additional Facebook targeting options. The first of these was an advertisement that targeted Facebook friends of people who had liked the page, not limited by geographic region. This ad received a fairly high response rate: 32 participants in 13 days. Next, we used the "Lookalike" function in which Facebook used an internal algorithm to identify 1,000,000 people who they considered similar to the people who had liked our page. We then targeted members of this Lookalike sample, starting with an ad to all Lookalike sample members, then again going region by region (still specifying women aged 25 to 65 years and above who were married, etc, as described above). This strategy revitalized the campaign. To conclude the campaign, we returned to the most successful ads that included our "ends soon" message, rotating between them until we had reached our recruitment goal.

The "friends of people who like the page" and "Lookalike" options will be most useful when recruiting for a large study with a target audience that is demographically homogeneous (eg, wives of smokeless tobacco users). This type of advertising will be less useful for smaller studies and studies enrolling participants with health conditions that are relatively infrequent and randomly distributed. For example, a breast cancer patient is unlikely to have many friends who are also breast cancer patients, and people who are demographically similar to her will not be appreciably more likely to have breast cancer than people who are demographically different.

Saturation: The Problem of Diminishing Returns

Perhaps the most important thing to know about Facebook advertising is this: although enrollment from each ad will taper off within a few days of its launch, people will continue to click on the ad (and use up your budget) every day. If you are paying per click rather than per impression (view), you are paying for these clicks. Whether these extra *clickers* are people who for whatever reason click on anything or bots who scour the internet to follow any available link [38-40], little can be done about the waste, other than to note that the ad will quickly see diminishing returns (reach saturation), and thus, it should be shown to a fresh target audience every few days.

It may be that fewer than 1% of the people who click on an ad will enroll in a study, especially if the study is a full-scale RCT. In our RCT, we enrolled 1145 people from 371,472 clicks, a rate of 0.3%. When we first began advertising on Facebook, the ratio of enrolled participants to clicks was much higher but still not more than about 1%. Others have had similar experiences: Ramo et al [29] converted 5875 clicks into 79 participants in an RCT for young adult smokers (1.3%), and Adam et al [18] converted 1001 clicks into 45 participants for a pregnancy-related RCT (4.5%). In our case, enrollment required first completing an eligibility screening, followed by informed consent, registration with personal contact information, and finally completion of a baseline survey that took perhaps 20

min to complete. This series of hurdles to enrollment may have discouraged many potential participants, but it also suggests that those who do complete the process may be more motivated to be actively involved in the intervention, which we consider a good trade-off. Simpler studies (eg, those requiring completion of a single survey) can get higher rates for converting Facebook ad clicks to participants (with some offering financial incentives and others not doing so). For example, Ramo and Prochaska [16] converted 14,808 clicks into 1548 completed surveys on tobacco and marijuana use (10.5%); Tan [17] converted 280 clicks into 59 completed surveys on math students' learning preferences (21.1%). Conversely, Kapp et al [13] received 280 clicks on a Facebook ad to recruit women to complete a health survey on mammography, but no surveys were completed. A human papillomavirus study requiring an online questionnaire and a self-collected penile swab converted 41,811 clicks into 535 study completers (1.3%) [41].

Table 1 displays the recruitment results for our study, showing the numbers of people "reached," clicking the link, beginning the enrollment process, eligible, consenting, and randomized. The *reach* and click data may be found in the Facebook Ad Manager, and the totals for beginning the process, eligible, consenting, and randomized were shown on the online administrative participant management system we had designed for this purpose. As the table indicates, Facebook displayed the ad to about 6.6 million women meeting our criteria, and 5.63% (371,472/6,600,839) of them clicked on the ad to see our study home page. A very small fraction of those then began the eligibility screening, and 30.63% (1554/5074) of those who began the eligibility screening completed it and were eligible to participate in the study. Of the eligible people, 96.99% (1416/1460) completed the informed consent, and 80.86% (1145/1416) of consenters provided their personal contact information, completed the lengthy baseline survey, and were randomized. Overall, 22.57% (1145/5074) of those who began the eligibility screening were randomized; 78.42% (1145/1460) of those who were eligible were randomized.

Examining Ad Effectiveness

We analyzed the registration patterns associated with each of the 91 ads we used during the study. Almost 30% of participants were enrolled from the first 7 regional ads in approximately 2 months. The most successful ad in terms of total enrolled participants was the first ad targeting the Midwest, which yielded 95 participants from 13 days of advertising. After this period, from mid-October to late December, 10 ads yielded only 113 participants. The New Year's simultaneous advertising plan (counted as a single ad for analyses) yielded 43 participants but was very expensive. Regional advertising was then hit-and-miss for several months, with some ads doing acceptably but others poorly. The "friends of friends" ad yielded 32 participants, and the Lookalike ads were moderately successful, with many yielding more than 15 participants but some far fewer.

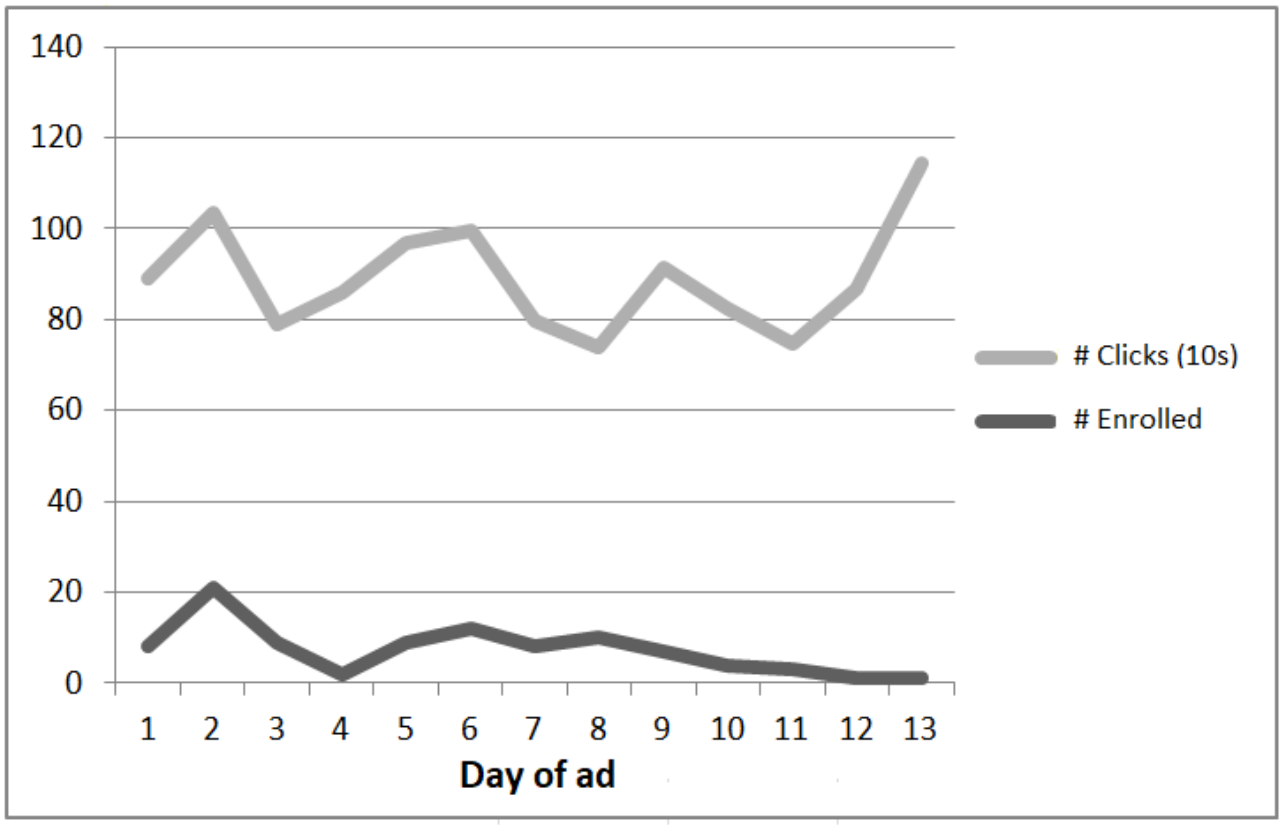
Figure 3 displays the recruitment pattern from the Midwest ad yielding 95 participants. The figure illustrates 2 key points: enrollment gradually tapers off within a few days of launching even a successful ad like this one (the saturation effect), whereas ad clicks remain high throughout (700 to 1135 clicks per day). For this ad, enrollment dropped sharply on a holiday (day 4, Labor Day) but resumed thereafter; we discontinued the ad after day 13.

The vast majority of the participants (94.2%) lived in the target area for an ad and enrolled while the ad was running or within 2 days of its conclusion. Of the others, 2.2% could plausibly have been stragglers from an earlier ad (enrolling within several weeks of an ad beginning in their region and assigned an identification number based on that ad period) and were attributed to that ad. The other 3.6% lived outside the current ad's target area (sometimes nearby but sometimes on the other side of the country), and we expect that most of these people were tagged or otherwise notified of the study by a friend seeing the ad; we attributed their enrollments to the current or most recent ad. Enrollment was consistent across the days of the week, from a low of 13.0% of the total sample enrolling on a Thursday to a high of 15.7% enrolling on a Sunday. Ads were placed on each day of the week, ranging from 6 ads that began on Sundays to 20 ads that began on Fridays.

Table 1. From "Reach" to randomization: UCare study recruitment data.

Enrollment step	Number	Proportion continuing from prior step
"People Reached"	6,600,839	Total viewing Facebook ad
Link (ad) clicks	371,472	5.63% of "reached"
Began online eligibility screening	5074	1.37% of clicks
Completed online eligibility screening	1554	30.63% of began screening
Eligible	1460	93.95% of completed screening
Consenting	1416	96.99% of eligible
Randomized	1145	80.86% of consenting

Figure 3. Enrollment versus ad clicks (tens of clicks) for 1 ad by day of ad.



Smokeless tobacco prevalence data are available by state. We designed our study admin page to provide a downloadable Excel file with participant registration data, including their mailing addresses. These data allowed us to calculate per capita response rates by state to see which states were most welcoming (very simply, state adult population x smokeless tobacco user rate=estimated smokeless tobacco users in that state, divided by women enrolled from that state=state enrollment rate). For example, 33 women enrolling from Idaho gave us our best rate, at 47 per 100,000 smokeless tobacco users; 11 women enrolling from Texas gave us our worst rate, at 1 per 100,000 smokeless tobacco users. (As our previous smokeless tobacco cessation study [42] enrolling chewers directly had been very popular in Texas, we made extra efforts to improve recruitment of women there, but with no success until we were able to run a "lookalike" ad, as described above.) By tracking this type of information, researchers can target geographical regions that may be more responsive than others or take extra steps to enroll participants from audience subsections that are lagging behind expectation.

If a particular ad does poorly, it may reflect a lack of interest among the targeted group. This may be temporary (eg, severe local weather conditions). It may also reflect a decision by Facebook on how much to promote the ad, for example, if they deem this ad has "too much" text (although it may be identical to an ad recently used successfully). The same ad could be tried again in a few weeks, or a new (and if desired, essentially identical) ad could be created to see if the results improve.

Advertising Costs

This advertising campaign resulted in a cost of US $112.48 per randomized participant (US $0.35 per click). In our pilot 2010 Facebook recruitment campaign with the same study population, we spent just under US $70 per participant. As the cost per click for that campaign is no longer available, we are unable to discern whether the rising cost was due to lower rates of converting clicks to participants or higher advertising prices. In their 2016 review of Facebook recruitment for health and psychosocial research, Thornton et al [34] reported that 110 studies had experienced costs per participant ranging from US $1.36 to US $110 (mean=US $17.48, SD US $23.06). Whitaker et al's review of 35 studies found a median value of US $14.41 per health study participant recruited using Facebook [6]. The magnitude of the difference between these costs and our own may be attributable in part to the fact that most of these studies were not RCTs, and the burden of enrollment in terms of time, effort, and commitment would thus be far smaller. These studies could also be targeting populations that are easier to reach than those targeted in our study. Our cost of US $112 per participant is in line with the costs we incurred for using traditional advertising media in 2004 to 2005 (US $92 per participant for free publicity in newspaper, radio, and TV news and US $115 per participant for newspaper display ads); in that study, Google AdWords was far more cost-effective (US $7 per participant). None of these advertising methods would have allowed us to exclude men from seeing the study and attempting to enroll, an outcome we wanted to avoid.

The costs for recruiting our population may not be representative of other groups. Pedersen et al recruited 793 young veteran drinkers to a *very brief* intervention in only 7 days for less than US $5 per participant [23]. For this study, 4.4% of clicks were converted to participants, in comparison with our rate of 0.3%.

Ethical Considerations

There are 2 important ethical concerns to using Facebook advertising: Facebook's potential use of the data and the risks from interacting with social networks. When a user clicks on an ad, and especially when they *like* an ad or a campaign's Facebook page, information about these choices is stored by the user's browser cookies [19], and Facebook's algorithms may conclude that the user has an interest in the topic and begin showing the user related links. This may make the user uncomfortable, and they may conclude that the study is violating their privacy. One way to partially mitigate this risk would be to include text on the study's fan page cautioning the user that *liking* the page or clicking on the embedded link to the study's own homepage may be used by Facebook marketing. Visitors could be provided with a nonhypertext URL to the study, with the suggestion that they access the study by copying and pasting the URL to a new browser tab. This option may be weighed against the risk to the study of reducing its page *likes*, which is an important element of a long-term advertising campaign, as described above.

The risks from interacting with social networks are largely related to privacy, especially for stigmatizing health conditions. When someone sees an ad and *tags* a friend who they think may be interested, others can see this tagging and will assume there is some association between the tagged person and the condition. Unless Facebook provides a way to eliminate tagging (eg, by disabling comments on the ad), there is no realistic way to avoid this risk other than not advertising for this population. One study recruiting for a domestic violence intervention took 2 measures to protect potential participants' safety: placing their ads at the side of the screen rather than in the user's newsfeed and asking women's organizations to post the ad for them such that women would already be following those pages to receive the information about the study in their newsfeed. They also included accompanying text for all ads, "Please open the link in a new browser window" and "Share only if safe to do so" [43].

Detailed ethical guidelines for protecting users' privacy during social media recruitment have been developed by Bender et al [44], using Privacy by Design principles.

Key Points for Using Facebook Advertising

Following are our key recommendations for using Facebook advertising:

- First, decide whether Facebook is really an optimal advertising medium for the study.
- For a large-scale RCT, a substantial advertising budget may be necessary.
- Before launching recruitment, design an infrastructure for enrolling and managing participants, including a simple method to monitor the success of each ad daily.
- Before beginning a Facebook advertising campaign, create a fan page on Facebook for the study to connect with the ad; plan to keep the page active with regular posts.
- Be aware that Facebook changes its ad requirements often. Consider getting feedback from targeted participants on the text and graphics of ads before using them.
- Plan for each Facebook ad to reach saturation quickly. Design the ad campaign with a specific rotation strategy in mind (eg, geographic region or a demographic variable), and create an easy way to track enrollment by this variable to help monitor each ad's effectiveness.
- The "friends of people who like the page" and "Lookalike" options will be most useful when recruiting for a large study with a target audience that is demographically homogeneous but less so for a target audience whose members are heterogeneous, for example, those diagnosed with a particular low-incidence disease.
- Monitor the advertising campaign regularly to avoid overcommitting the advertising budget and be flexible and ready to change strategies if the responses are not as expected.
- Above all, be aware that a great many people will click on Facebook ads without the intention of enrolling in the study, and an ample budget will be necessary.

Conclusions

Our experiences have shown that it is possible to recruit a moderately large RCT sample via Facebook. This recruitment method provides considerable flexibility to monitor and modify the advertising tactics based on feedback. In our case, we were advertising for participating in a study that would be very appealing to potential participants if they became aware of it and that was easy for people to promote to their Facebook friends who they knew may also be interested. This method of recruitment may not work for all types of projects or participant populations. Furthermore, Facebook and other social media are in a constant state of flux, and the methods described in this paper may not work the same way in the future.

Acknowledgments

Funding for this research was provided by the National Institute on Drug Abuse (R01-DA033422) and the National Cancer Institute (R21-CA131461). This manuscript has benefitted from feedback helpfully provided by Christi Patten, Devon Noonan, Ed Lichtenstein, Judith Gordon, and Judy Andrews. Erin Ross contributed useful information. Katie Clawson provided assistance with manuscript preparation.

Conflicts of Interest

None declared.

References

1. Ingram M. Fortune. How Google and Facebook have taken over the digital ad industry URL: http://fortune.com/2017/01/04/google-facebook-ad-industry/ [accessed 2018-05-07] [WebCite Cache ID 6zFRabQAn]
2. Smith A, Anderson M. Pew Internet & American Life Project. 2018. Social media use in 2018 URL: http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/ [accessed 2018-05-03] [WebCite Cache ID 6z9KRPRej]
3. Kosinski M, Matz SC, Gosling SD, Popov V, Stillwell D. Facebook as a research tool for the social sciences: opportunities, challenges, ethical considerations, and practical guidelines. Am Psychol 2015 Sep;70(6):543-556. [doi: 10.1037/a0039210] [Medline: 26348336]
4. Valle CG, Tate DF, Mayer DK, Allicock M, Cai J. A randomized trial of a Facebook-based physical activity intervention for young adult cancer survivors. J Cancer Surviv 2013 Sep;7(3):355-368 [FREE Full text] [doi: 10.1007/s11764-013-0279-5] [Medline: 23532799]
5. Ridout B, Campbell A. Using Facebook to deliver a social norm intervention to reduce problem drinking at university. Drug Alcohol Rev 2014 Nov;33(6):667-673. [doi: 10.1111/dar.12141] [Medline: 24689339]
6. 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]
7. Topolovec-Vranic J, Natarajan K. The use of social media in recruitment for medical research studies: a scoping review. J Med Internet Res 2016 Dec 07;18(11):e286 [FREE Full text] [doi: 10.2196/jmir.5698] [Medline: 27821383]
8. Frandsen M, Thow M, Ferguson SG. The effectiveness of social media (Facebook) compared with more traditional advertising methods for recruiting eligible participants to health research studies: a randomized, controlled clinical trial. JMIR Res Protoc 2016 Aug 10;5(3):e161 [FREE Full text] [doi: 10.2196/resprot.5747] [Medline: 27511829]
9. Amon KL, Campbell AJ, Hawke C, Steinbeck K. Facebook as a recruitment tool for adolescent health research: a systematic review. Acad Pediatr 2014;14(5):439-447.e4. [doi: 10.1016/j.acap.2014.05.049] [Medline: 25169155]
10. Carter-Harris L, Bartlett ER, Warrick A, Rawl S. Beyond traditional newspaper advertisement: leveraging Facebook-targeted advertisement to recruit long-term smokers for research. J Med Internet Res 2016 Jun 15;18(6):e117 [FREE Full text] [doi: 10.2196/jmir.5502] [Medline: 27306780]
11. Chu JL, Snider CE. Use of a social networking web site for recruiting Canadian youth for medical research. J Adolesc Health 2013 Jun;52(6):792-794. [doi: 10.1016/j.jadohealth.2012.12.002] [Medline: 23352727]
12. Gilligan C, Kypri K, Bourke J. Social networking versus Facebook advertising to recruit survey respondents: a quasi-experimental study. JMIR Res Protoc 2014 Sep;3(3):e48 [FREE Full text] [doi: 10.2196/resprot.3317] [Medline: 25230740]
13. Kapp JM, Peters C, Oliver DP. Research recruitment using Facebook advertising: big potential, big challenges. J Cancer Educ 2013 Mar;28(1):134-137. [doi: 10.1007/s13187-012-0443-z] [Medline: 23292877]
14. Lohse B, Wamboldt P. Purposive Facebook recruitment endows cost-effective nutrition education program evaluation. JMIR Res Protoc 2013 Aug;2(2):e27 [FREE Full text] [doi: 10.2196/resprot.2713] [Medline: 23948573]
15. Nguyen P, Gold J, Pedrana A, Chang S, Howard S, Ilic O, et al. Sexual health promotion on social networking sites: a process evaluation of The FaceSpace Project. J Adolesc Health 2013 Jul;53(1):98-104. [doi: 10.1016/j.jadohealth.2013.02.007] [Medline: 23583509]
16. Ramo DE, Prochaska JJ. Broad reach and targeted recruitment using Facebook for an online survey of young adult substance use. J Med Internet Res 2012 Feb 23;14(1):e28 [FREE Full text] [doi: 10.2196/jmir.1878] [Medline: 22360969]
17. Tan H. 2010 Contemporary Approaches to Research in Mathematics, Science, Health and Environmental Education Symposium. Melbourne: Deakin University; 2010. URL: https://blogs.deakin.edu.au/steme/past_events/2010-contemporary-approaches-to-research-in-mathematics-science-health-and-environmental-education-symposium/ [WebCite Cache ID 6zFSMY9t9]
18. Adam LM, Manca DP, Bell RC. Can Facebook be used for research? Experiences using Facebook to recruit pregnant women for a randomized controlled trial. J Med Internet Res 2016 Sep 21;18(9):e250 [FREE Full text] [doi: 10.2196/jmir.6404] [Medline: 27655184]
19. Pedersen ER, Osilla KC, Helmuth ED, Tolpadi A, Gore K. Reaching concerned partners of heavy drinking service members and veterans through Facebook. Mil Behav Health 2017 Apr 13;5(3):265-273. [doi: 10.1080/21635781.2017.1316804]
20. Prescott TL, Phillips IG, DuBois LZ, Bull SS, Mustanski B, Ybarra ML. Reaching adolescent gay, bisexual, and queer men online: development and refinement of a national recruitment strategy. J Med Internet Res 2016 Aug 04;18(8):e200 [FREE Full text] [doi: 10.2196/jmir.5602] [Medline: 27492781]
21. Jones R, Lacroix LJ, Porcher E. Facebook advertising to recruit young, urban women into an HIV prevention clinical trial. AIDS Behav 2017 Nov;21(11):3141-3153. [doi: 10.1007/s10461-017-1797-3] [Medline: 28528463]
22. Brief DJ, Rubin A, Keane TM, Enggasser JL, Roy M, Helmuth E, et al. Web intervention for OEF/OIF veterans with problem drinking and PTSD symptoms: a randomized clinical trial. J Consult Clin Psychol 2013 Oct;81(5):890-900. [doi: 10.1037/a0033697] [Medline: 23875821]
23. Pedersen ER, Naranjo D, Marshall GN. Recruitment and retention of young adult veteran drinkers using Facebook. PLoS One 2017 Mar;12(3):e0172972 [FREE Full text] [doi: 10.1371/journal.pone.0172972] [Medline: 28249027]
24. Akers L, Andrews JA, Gordon JS. A multimedia support skills intervention for female partners of male smokeless tobacco users: use and perceived acceptability. JMIR Formativ Res 2018 May 28;2(1):e10. [doi: 10.2196/formative.9948]

25. Baltar F, Brunet I. Social research 2.0: virtual snowball sampling method using Facebook. Internet Res 2012 Jan 27;22(1):57-74. [doi: 10.1108/10662241211199960]
26. Morgan AJ, Jorm AF, Mackinnon AJ. Internet-based recruitment to a depression prevention intervention: lessons from the Mood Memos study. J Med Internet Res 2013 Feb;15(2):e31 [FREE Full text] [doi: 10.2196/jmir.2262] [Medline: 23403043]
27. Mattingly WA, Kelley RR, Wiemken TL, Chariker JH, Peyrani P, Guinn BE, et al. Real-Time enrollment dashboard for multisite clinical trials. Contemp Clin Trials Commun 2015 Oct 30;1:17-21 [FREE Full text] [doi: 10.1016/j.conctc.2015.09.001] [Medline: 26878068]
28. Facebook. Facebook, Advertising Targeting Options URL: https://www.facebook.com/business/products/ads/ad-targeting [accessed 2018-10-15] [WebCite Cache ID 73CV3DTfs]
29. Ramo DE, Rodriguez TM, Chavez K, Sommer MJ, Prochaska JJ. Facebook recruitment of young adult smokers for a cessation trial: methods, metrics, and lessons learned. Internet Interv 2014 Apr;1(2):58-64 [FREE Full text] [doi: 10.1016/j.invent.2014.05.001] [Medline: 25045624]
30. Pedersen ER, Kurz J. Using Facebook for health-related research study recruitment and program delivery. Curr Opin Psychol 2016 Jun;9:38-43. [doi: 10.1016/j.copsyc.2015.09.011] [Medline: 26726313]
31. Valdez RS, Guterbock TM, Thompson MJ, Reilly JD, Menefee HK, Bennici MS, et al. Beyond traditional advertisements: leveraging Facebook's social structures for research recruitment. J Med Internet Res 2014 Oct;16(10):e243 [FREE Full text] [doi: 10.2196/jmir.3786] [Medline: 25348050]
32. Lomas N. TechCrunch. Teens favoring Snapchat and Instagram over Facebook, says eMarketer URL: https://techcrunch.com/2017/08/22/teens-favoring-snapchat-and-instagram-over-facebook-says-emarketer/ [accessed 2017-11-09] [WebCite Cache ID 6upsbun8W]
33. Darwell B. Adweek. Facebook updates cover photo and News Feed ad policy, limits text to 20% of image URL: http://www.adweek.com/digital/facebook-updates-cover-photo-and-news-feed-ad-policy-limits-text-to-20-of-image/ [accessed 2017-11-09] [WebCite Cache ID 6upseCkIl]
34. Thornton L, Batterham PJ, Fassnacht DB, Kay-Lambkin F, Calear AL, Hunt S. Recruiting for health, medical or psychosocial research using Facebook: systematic review. Internet Interv 2016 May;4:72-81 [FREE Full text] [doi: 10.1016/j.invent.2016.02.001] [Medline: 30135792]
35. Facebook. Advertising Policies: Personal Attributes URL: https://www.facebook.com/policies/ads/prohibited_content/personal_attributes [WebCite Cache ID 73CVAxx3h]
36. Kramer J, Rubin A, Coster W, Helmuth E, Hermos J, Rosenbloom D, et al. Strategies to address participant misrepresentation for eligibility in Web-based research. Int J Methods Psychiatr Res 2014 Mar;23(1):120-129 [FREE Full text] [doi: 10.1002/mpr.1415] [Medline: 24431134]
37. Teitcher JE, Bockting WO, Bauermeister JA, Hoefer CJ, Miner MH, Klitzman RL. Detecting, preventing, and responding to "fraudsters" in internet research ethics and tradeoffs. J Law Med Ethics 2015;43(1):2200. [doi: 10.1111/jlme.12200] [Medline: 25846043]
38. Clark DB. The New Republic. 2015. The bot bubble: How click farms have inflated social media currency URL: https://newrepublic.com/article/121551/bot-bubble-click-farms-have-inflated-social-media-currency [accessed 2018-10-16] [WebCite Cache ID 73E58C4gN]
39. Elgin B, Riley M, Kocieniewski D, Brustein J. Bloomberg Businessweek. How much of your audience is fake? URL: https://www.bloomberg.com/features/2015-click-fraud/ [WebCite Cache ID 6upshUleT]
40. LaFrance A. The Atlantic. The Internet is mostly bots URL: https://www.theatlantic.com/technology/archive/2017/01/bots-bots-bots/515043/ [WebCite Cache ID 6upsjTwoC]
41. Das R, Machalek DA, Molesworth EG, Garland SM. Using Facebook to recruit young Australian men into a cross-sectional Human Papillomavirus study. J Med Internet Res 2017 Nov 17;19(11):e389 [FREE Full text] [doi: 10.2196/jmir.8739] [Medline: 29150417]
42. Severson HH, Gordon JS, Danaher BG, Akers L. ChewFree.com: evaluation of a web-based cessation program for smokeless tobacco users. Nicotine Tob Res 2008 Feb;10(2):381-391. [doi: 10.1080/14622200701824984] [Medline: 18236303]
43. Tarzia L, Valpied J, Koziol-McLain J, Glass N, Hegarty K. Methodological and ethical challenges in a web-based randomized controlled trial of a domestic violence intervention. J Med Internet Res 2017 Mar 28;19(3):e94 [FREE Full text] [doi: 10.2196/jmir.7039] [Medline: 28351830]
44. Bender JL, Cyr AB, Arbuckle L, Ferris LE. Ethics and privacy implications of using the internet and social media to recruit participants for health research: a privacy-by-design framework for online recruitment. J Med Internet Res 2017 Apr 06;19(4):e104 [FREE Full text] [doi: 10.2196/jmir.7029] [Medline: 28385682]

Abbreviations

eHealth: electronic health
RCT: randomized clinical trial

Edited by G Eysenbach; submitted 08.11.17; peer-reviewed by L Carter-Harris, M Kotter, E Pedersen; comments to author 15.03.18; revised version received 27.07.18; accepted 14.08.18; published 08.11.18.

Please cite as:
Akers L, Gordon JS
Using Facebook for Large-Scale Online Randomized Clinical Trial Recruitment: Effective Advertising Strategies
J Med Internet Res 2018;20(11):e290
URL: http://www.jmir.org/2018/11/e290/
doi:10.2196/jmir.9372
PMID:30409765

Review

Internet-Delivered Early Interventions for Individuals Exposed to Traumatic Events: Systematic Review

Naomi Ennis[1*], MA; Iris Sijercic[1*], BA (Hons); Candice M Monson[1*], PhD

Department of Psychology, Ryerson University, Toronto, ON, Canada

[*]all authors contributed equally

Corresponding Author:
Candice M Monson, PhD
Department of Psychology
Ryerson University
350 Victoria Street
Toronto, ON, M5B 2K3
Canada
Phone: 1 4169795000 ext 6209
Email: candice.monson@psych.ryerson.ca

Abstract

Background: Over 75% of individuals are exposed to a traumatic event, and a substantial minority goes on to experience mental health problems that can be chronic and pernicious in their lifetime. Early interventions show promise for preventing trauma following psychopathology; however, a face-to-face intervention can be costly, and there are many barriers to accessing this format of care.

Objective: The aim of this study was to systematically review studies of internet-delivered early interventions for trauma-exposed individuals.

Methods: A literature search was conducted in PsycINFO and PubMed for papers published between 1991 and 2017. Papers were included if the following criteria were met: (1) an internet-based intervention was described and applied to individuals exposed to a traumatic event; (2) the authors stated that the intervention was intended to be applied early following trauma exposure or as a preventive intervention; and (3) data on mental health symptoms at pre-and postintervention were described (regardless of whether these were primary outcomes). Methodological quality of included studies was assessed using the Downs and Black checklist.

Results: The interventions in the 7 studies identified were categorized as *selected* (ie, delivered to an entire sample after trauma regardless of psychopathology symptoms) or *indicated* (ie, delivered to those endorsing some level of posttraumatic distress). Selected interventions did not produce significant symptom improvement compared with treatment-as-usual or no intervention control groups. However, indicated interventions yielded significant improvements over other active control conditions on mental health outcomes.

Conclusions: Consistent with the notion that many experience natural recovery following trauma, results imply that indicated early internet-delivered interventions hold the most promise in future prevention efforts. More studies that use rigorous methods and clearly defined outcomes are needed to evaluate the efficacy of early internet-delivered interventions. Moreover, basic research on risk and resilience factors following trauma exposure is necessary to inform indicated internet-delivered interventions.

(J Med Internet Res 2018;20(11):e280) doi:10.2196/jmir.9795

KEYWORDS

psychological trauma; secondary prevention; trauma and stressor-related disorders; internet

Introduction

Background

Approximately 75% of individuals are exposed to a traumatic stressor in their lifetime that involves exposure to actual or threatened death, serious injury, or sexual violence [1]. The types of traumatic exposures that are most commonly experienced include sexual assault, witnessing another person getting killed or badly injured, sudden unexpected death, and life-threatening motor vehicle accidents [1]. Following trauma, the majority of individuals will experience subclinical symptoms

of distress that abate over time without intervention, considered *natural recovery* [2-4]. Although natural recovery is the most common trajectory following trauma exposure, a subset of trauma-exposed individuals experience significant distress and impairment that require intervention to facilitate recovery to healthy levels of functioning [1]. These individuals may be diagnosed with posttraumatic stress disorder (PTSD), major depressive disorder, generalized anxiety disorder, panic attacks, and health-risk behavior such as substance abuse [5-7]. These disorders can be chronic and pernicious but may be preventable if interventions are delivered early following trauma exposure [8]. The purpose of this paper was to systematically review early interventions delivered through the internet for individuals exposed to trauma.

Although natural recovery is expected for most trauma-exposed individuals, trauma experts recommend that mental health professionals should not wait to provide care until problems are chronic and purport the value of early preventive interventions [9]. The Institute of Medicine defines prevention as efforts to reduce the incidence of a disorder, as opposed to reducing the prevalence [10]. Within this definition, preventive interventions have been organized into 3 categories: universal, selected, and indicated [11]. Universal interventions are provided to all members of a population regardless of risk for developing a disorder, for example, interventions applied to an entire population before a traumatic event regardless of trauma exposure. Selected interventions are intended for those who exhibit risk factors for the disorder but show no signs or symptoms of the disorder, such as individuals exposed to a traumatic event who may or may not be experiencing symptoms of the disorder. Indicated interventions are provided to only those who have subthreshold symptoms of the disorder or a subclinical diagnosis, for example, those who screen positive as experiencing symptoms of distress following trauma [10].

Universal intervention delivered before trauma exposure has been argued to be infeasible and too costly. However, as compared with other mental health disorders that have a prodromal phase (eg, schizophrenia) or a waxing and waning course (eg, depression), disorders that have an onset subsequent to trauma exposure (eg, PTSD and acute stress disorder) have a clear onset, providing a unique window for selected or indicated prevention. Following a traumatic event, individuals may present to emergency rooms or be seen by health care providers, and these circumstances are opportune for the provision of intervention [12]. In light of these considerations, this paper focuses on selected and indicated interventions delivered to individuals already exposed to trauma.

Reviews of early interventions for trauma-exposed individuals demonstrate that efficacy varies across modalities. The literature consistently contraindicates psychological debriefing as an intervention following trauma [9,11,13]. Interventions based on cognitive and behavioral principles have been found to be valuable in the prevention of posttraumatic mental health problems [9,11,14]. Despite their potential value, Feldner et al [11] highlight that face-to-face preventive interventions can be intensive, time-consuming, and costly. Moreover, many individuals who experience distress following trauma do not receive intervention [15] due to perceived stigma, difficulty scheduling appointments, and lack of access to care (eg, due to living in a remote location, lack of transportation, lack of financial resources [6]). Given these considerations, the internet may be a valuable platform to deliver early intervention. The internet provides a medium to deliver interventions with a wide reach as websites can be accessed remotely at a low cost, and users can maintain anonymity. Such interventions may be particularly valuable in reducing the economic and psychological burden of natural disasters, terrorist attacks, or war because internet-delivered interventions can be delivered on a large scale and accessed by the entire affected communities.

Objective

Research suggests that internet-based interventions are feasible [16] and efficacious in reducing PTSD and other psychopathologies (eg, anxiety disorders) that may follow from trauma [17,18]. However, to date no research has systematically reviewed the literature on early interventions for trauma-exposed individuals delivered via the internet. Given the potential value of such interventions for targeting a large number of people following trauma in a cost-effective way, an understanding of the available interventions and their efficacy in preventing or ameliorating posttraumatic distress is important. This study systematically reviewed internet-delivered interventions intended to be delivered acutely following trauma exposure, and the empirical data on mental health symptom change following these interventions.

Methods

Literature Search

A search was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement [19]. The literature search was conducted in PsycINFO and PubMed, using the search terms “Trauma” OR “Posttraumatic Stress” OR “Posttraumatic Stress Disorder” or “Recent Trauma” OR “PTSD” and “Early Intervention” OR “Preventive Intervention” and “Online” OR “Web-based” OR “Internet-based therapy” OR “Internet-delivered.” Searches were limited to papers published after 1991 when the internet became available in North America. Reference lists of reviews and meta-analyses on early interventions [8,14] and internet-delivered interventions for PTSD [17] were also reviewed.

Selection Criteria

Two reviewers (NE and IS) screened identified abstracts and titles to identify full-text studies. Due to the nascency of internet-delivered interventions, and to review all available early posttrauma interventions, inclusion criteria were broad and not limited to controlled trials. Papers were included if the following criteria were met: (1) an internet-based intervention was described and applied to individuals exposed to a traumatic event; (2) the authors stated that the intervention was intended to be applied early following trauma exposure or as a preventive intervention; and (3) data on mental health symptoms at pre- and postintervention were described (regardless of whether these were primary outcomes). Internet-based interventions were defined as interventions delivered online via a computer

or mobile phone platform. The terms internet-based and Web-based were used synonymously to describe such interventions. Papers that described stepped care interventions whereby 1 aspect of a broader intervention involved an internet-based intervention were included [20].

Papers were excluded if (1) the intervention was delivered via the telephone or videoconferencing [21]; (2) an intervention was described and implemented, but empirical data on participants' mental health symptoms pre- and postintervention were not reported (eg, protocols for randomized controlled trials [22] and studies that only describe feasibility data [23,24]); and (3) the authors did not explicitly state that the intervention was preventive or intended to be delivered early following trauma. Level of agreement between the 2 reviewers (NE and IS) was 100%.

Methodological Quality of Included Studies

A total of 2 raters (NE, IS) evaluated the methodological quality of each included empirical study (ie, only studies examining empirical evidence for the intervention) using the Downs and Black Checklist [25]. This checklist was selected because it evaluates the quality of both randomized and nonrandomized trials, given that both were included in this review. The checklist assesses items under the following subscales: reporting, external validity, internal validity (bias and confounding), and power. A modified version of the power item was used [26]. With this modification, a study can achieve a total possible score of 28. The test-retest reliability (r=.88), inter-rater reliability (r=.75), and internal consistency (Kuder-Richardson formula 20=.89) of the checklist are good [25]. Higher scores indicate greater quality. In this study, inter-rater reliability ranged from 85% to 100% on each article reviewed. Raters discussed each discrepancy and achieved consensus for all discrepantly rated items.

Data Extraction

After screening for relevance, full papers were examined. Data on the intervention, sampling, recruitment, methodology, and design were extracted from all included studies. Data on mental health outcomes at pre- and postintervention as well as longer-term follow-up, if available, were extracted (ie, feasibility data were not reviewed). Any measures of mental health symptoms (eg, depression, anxiety, and worry) were reviewed, given the range of psychological responses that individuals might experience after trauma. Results from intent-to-treat analyses and completer analyses were extracted. Interventions were coded as "selected" or "indicated." Selected interventions were defined as interventions that were delivered to the entire sample, regardless of whether or not they endorsed mental health symptoms. Interventions were coded as indicated when delivered only to participants endorsing cut-off criteria of mental health symptoms.

Results

Database searches yielded a total of 2346 articles. Review of reference lists of relevant articles yielded an additional 15 papers. Abstracts and titles were screened for inclusion. A total of 7 articles were included in the review based on selection criteria (see Figure 1). The most common reason for exclusion was that the intervention was not delivered via the internet or not an early or preventive intervention (eg, intended for chronic PTSD).

Interventions

Descriptions of interventions are provided in Multimedia Appendix 1. Each paper described a different intervention. With regard to age of target population, 4 interventions were designed for adults [20,27-29], 1 for adults and adolescents [30], and 2 for children [31,32]. In terms of the type of trauma that interventions were designed to address, 4 interventions were intended for survivors of physical injury or medical events [20,27,31,32], 2 for individuals exposed to natural disasters [28,30], and 1 for veterans following combat [29]. No interventions specifically targeted survivors of interpersonal trauma such as sexual assault.

Each of the 7 interventions described were based on cognitive and behavioral principles. Moreover, 3 interventions were completely self-guided and consisted only of self-help psychoeducational materials [28,30,31]. Of the interventions, 1 intervention [32] was formatted as an interactive game with a storyline whereby children chose characters exposed to different types of trauma. The goal of the game was for users to help people in a town whose emotions had been *zapped*.

One intervention [20] was a stepped care approach in which patients were given laptops at their hospital bedside with access to an online community forum website, and they also met clinicians who consulted them about their intervention preferences. In the intervention described by Van Voorhees et al [29], patients received instant messages from clinicians and peers trained as counselors to encourage continued use of the website. The authors described this approach as motivational interviewing instant messages. The intervention described in Mouthaan et al [27] also had a peer support forum.

Summary of Research Findings

The included studies are described in Table 1 and Multimedia Appendix 2. The quality of the studies according to the modified Downs and Black Checklist [26] ranged from 18 to 24, considered fair to good quality, with a median score of 21 (see Multimedia Appendix 1). Across all studies, assessors were not blind to patient intervention conditions. In only 1 study [29], adverse events that may have been important during the trial were reported, and only 2 studies [27,31] were adequately powered to detect a clinically important effect. Only 1 study [27] assessed symptom outcomes using gold standard clinician-administered measures and blinded assessment. There were 15 different assessment measures used across the studies because studies included multiple outcome measures. There were 3 assessment measures that were used across more than 1 study, and in each case, these measures were used in 2 studies only (see Table 1).

Across studies, participants were recruited from hospital emergency departments, intensive care units or surgical wards [20,27,31,32], random digit-dial methods in disaster-affected areas [30], online advertisements [29], and outpatient clinics [28]. There was variability with respect to the time the

participants were recruited following trauma as well as the length of time that participants were followed after the intervention (Multimedia Appendix 2).

With regard to outcomes, in controlled studies of selected interventions (ie, all trauma-exposed participants received the intervention regardless of mental health symptoms), the interventions were not better than control conditions in reducing mental health symptoms. Means of outcome measures for intervention and control groups are reported in Multimedia Appendix 3. There was 1 exception to this finding. Ruggiero et al [30] found marginally statistically significant decreases in PTSD and depression symptoms in the group that received the *Bounce Back Now* intervention as compared with the assessment-only control group at 12-months postbaseline.

Figure 1. Flow diagram of literature search. PTSD: posttraumatic stress disorder.

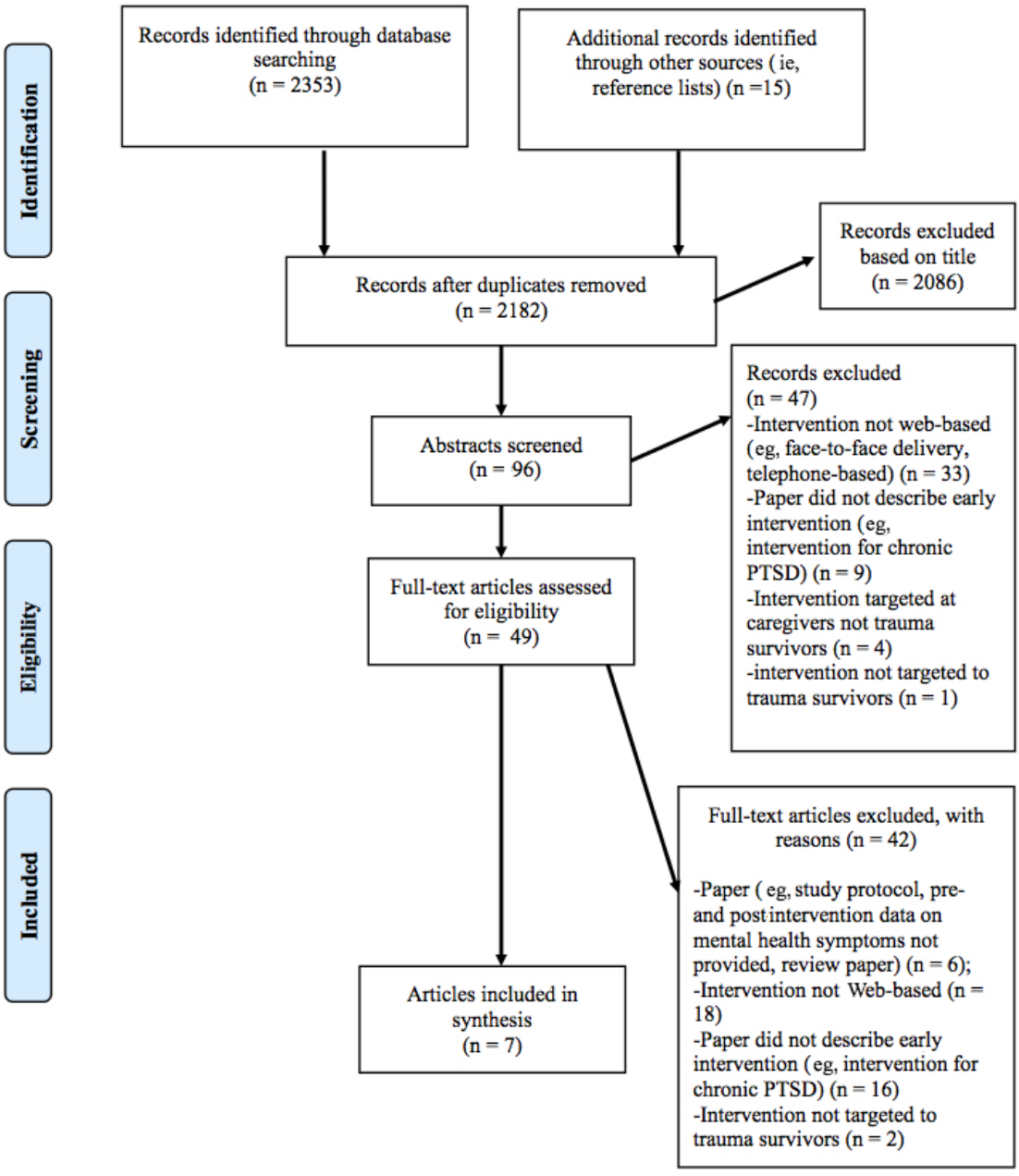


Table 1. Empirical data on internet-delivered early interventions.

Study (year)	Outcome measures	Intent-to-treat results	Completer results
Cox et al (2010) [31]	TSCC-A[a]; IES-R[b]	No advantage of intervention on any outcome at any assessment	Advantage of intervention on anxiety only, at both follow-up assessments
Kassam-Adams et al (2016) [32]	CPSS[c]; PedsQL[d]	No advantage of intervention on any outcome at any assessment	NR[e]
Mouthaan et al (2013) [27]	CAPS[f]; IES-R; HADS[g]	No advantage of intervention on any outcome at any assessment	Results similar to intention-to-treat (ITT) analyses for all outcome measures (statistics not reported in paper)
Ruggiero et al (2015) [30]	National Survey of Adolescents PTSD[h], depression, substance use modules	No advantage of intervention on any outcome at any assessment	Results similar to ITT analyses for all outcome measures (statistics not reported in paper)
Steinmatz et al (2012) [28]	PSS[i]; CSE[j]; MPSS[k]; CES-D[l]; PSWQ[m]	Advantage of intervention on worry at postassessment.	NR
Van Voorhees et al (2012) [29]	CES-D 10; PCL-M[n]	Advantage of intervention on depression at 4 and 12 weeks and on PTSD at 4, 8, and 12 weeks	Advantage of intervention on depression at 4 and 12 weeks and PTSD at 4, 8, and 12 weeks
Zatzick et al (2015) [20]	PCL-C[o]; PHQ-9[p]	Advantage of intervention on PTSD at 6 months	NR

[a]TSCC-A: Trauma-Symptom Checklist for Children-A.

[b]IES-R: Impact of Event Scale-Revised.

[c]CPSS: The Child PTSD Symptom Scale.

[d]PedsQL: Pediatric Quality of Life Inventory.

[e]NR: not reported.

[f]CAPS: Clinician-Administered PTSD Scale.

[g]HADS: Hospital Anxiety and Depression Rating Scale.

[h]PTSD: posttraumatic stress disorder.

[i]PSS: Perceived Stress Scale.

[j]CSE: Coping Self-Efficacy Scale for Trauma.

[k]MPSS: Modified PTSD Symptoms Scale.

[l]CES-D: Center for Epidemiologic Studies Depression Scale.

[m]PSWQ: Penn State Worry Questionnaire.

[n]PCL-M: PTSD-Checklist Military version.

[o]PCL-C: PTSD Checklist Civilian version.

[p]PHQ-9: Patient Health Questionnaire-9.

The Bounce Back Now intervention is a combination of selected and indicated prevention wherein all participants had experienced trauma and were enrolled regardless of symptom presentation. However, within the intervention, modules were indicated based on participants' symptomatology (eg, participants with a clinical level of symptoms of depression were invited and encouraged to use the depression module).

In the 3 studies in which interventions were indicated [20,28,29], significant reductions on some mental health symptoms were found compared with control conditions. Steinmatz et al [28] found that the group receiving the *My Disaster Recovery* intervention endorsed significantly greater reduction in worry over time as compared with the groups receiving online information only and the intervention-as-usual group. No significant differences among intervention conditions were found for symptoms of depression, PTSD, perceived stress, or coping self-efficacy. Van Voorhees et al's [29] study did not have a control group, and they found significant decreases in depression and PTSD over time. Zatzick et al [20] found significant reductions in PTSD symptoms among participants in the intervention group from baseline to 6-month follow-up as compared with their usual care control. They did not find clinically or statistically significant differences between groups in depression scores over time. Although the intervention was delivered to all those who were exposed to trauma, regardless of symptom presentation, Kassam-Adams et al [32] conducted exploratory analyses separating those participants considered *at-risk* for posttraumatic stress (as indicated by baseline scores of 15 or higher on the Child PTSD Symptom Scale; [33]) from those *not-at-risk*. They found medium to large between-intervention-group effect sizes from baseline to 6 weeks (d=−0.84) and for baseline to 12 weeks (d=−0.68) for those at-risk, in favor of the intervention group. Small effect sizes were found for those not-at-risk between baseline to 6 weeks (d=−0.15) and for baseline to 12 weeks (d=−0.24).

Discussion

Principal Findings

The aim of this paper was to systematically review the literature on studies of internet-delivered early interventions for trauma-exposed individuals. Although previous reviews have identified numerous internet-delivered interventions for PTSD [17] and other chronic psychological disorders following trauma [16], this review identified only 7 studies that evaluated *early* internet-delivered interventions for trauma-exposed individuals. The lack of research on internet-based interventions following traumatization is interesting, given the potential low cost and wide-reaching impact of such prevention efforts.

Overview of Included Studies

The quality of the included studies ranged from fair to good. Studies were generally not adequately powered to detect differences. In addition, most studies did not employ gold standard clinician-administered assessments and lacked long-term follow-up data. The objectives of studies varied as mental health symptoms were not the primary outcome of interest in several studies [28,32], and the same assessment measures were not used across studies. Mental health outcomes of interest also varied across studies (eg, PTSD, worry, and depression). The heterogeneity in outcomes assessed and general poor quality of assessment measures limit the conclusions that can be drawn about the effects of such interventions.

Despite the authors of these included studies describing the interventions as preventive or early interventions, the time since trauma varied widely across studies. For example, Van Voorhees et al's [29] intervention was designed to be delivered acutely after return from deployment but was not delivered and tested until up to 5 years after deployment. The heterogeneity across studies in terms of when these interventions were provided is problematic. Interventions in these studies may not actually be *early* interventions, given the time elapsed since trauma, and therefore, findings may not represent the value of early interventions or be generalizable to individuals recently exposed to trauma. In addition, given that the trajectory of change in posttraumatic distress is most robust in the first 3 months following trauma exposure [3], it is recommended that early posttrauma intervention be tested within 3 to 6 months. Before conclusions can be drawn about the results of these studies, rigorous research is needed, and interventions should be delivered in the window of time after trauma within which the intervention was designed to be employed.

Findings and Interpretation

Although limited conclusions can be drawn about the effects of the interventions because of the varied study quality and potential problems with timing of delivery in the included studies, a pattern emerged whereby interventions that were indicated (as compared with selected) tended to yield more promising results. In these studies, greater symptom improvement was found in intervention groups compared with controls. Interestingly, Kassam-Adams et al [32] provided the intervention to all trauma-exposed children and found no effect compared with the wait-list control. However, exploratory analyses showed larger effect sizes of the intervention for children at risk of PTSD compared with children not at risk. In studies of selected interventions, intervention groups faired similarly to control groups over time. This finding is in line with the notion that natural recovery is expected following trauma, and it suggests that the interventions may be most advantageous when delivered specifically to individuals demonstrating early symptoms of psychopathology. Delivering interventions when they are not indicated may not be cost effective since recovery without intervention is expected for the majority of those exposed to trauma.

Although data from this review and theory support that indicated interventions may be superior to selected early interventions, the potential for sampling bias across these types of interventions should be considered. For example, in over half of the included studies, the proportion of those who were approached to participate compared with the number who agreed was not reported, and no information was provided on how participating individuals might have differed from those who declined to participate. In indicated interventions, participants present with mental health symptoms. Individuals who self-select for a study and agree to participate may have more awareness of their symptoms and potentially more motivation for intervention, and this may contribute to better outcomes than studies of selected interventions. It is possible that stigma-related biases might render trauma-exposed individuals more likely to agree to participate in selected interventions where all individuals are given the intervention, regardless of whether they demonstrate natural recovery. Sampling bias must be considered when comparing outcomes across selected and indicated interventions, and reported in future studies.

Underscoring the importance of indicated over selected intervention, Feldner et al [11] hold that, given the high frequency of trauma exposure in the population, interventions should be delivered to those most at risk. Data on moderators of the interventions can contribute to better understanding of who will benefit the most from these interventions. However, none of the included studies examined moderators of intervention outcomes. In addition, potential mechanisms of interventions were not examined in any of the included studies, and thus, conclusions about the specific potency of specific strategies in these programs are limited.

Although the current data point to the potential benefits of indicated interventions, finding out whether users find them acceptable and sustainable and how far-reaching the interventions were is arguably as important as it is to ascertain whether the interventions are effective (ie, what communities they could reach). Data provided on accessibility, sustainability, and reach differed across studies when reported, with the majority not demonstrating far reach, given that the design was a pilot study. For example, Kassam-Adams et al [32] found that 35 of the 36 participants (participants were recruited from a hospital setting) accessed the website and roughly half completed the intervention. In contrast, in Ruggiero et al's study [30] in which 2000 families were approached to participate in the intervention (via random digit dialing), approximately half of those approached accessed the website, and 37.5% completed one module. Just over a quarter of those who completed a

module, completed the entire intervention. The differences in reach between studies and findings related to sustained use warrant further exploration.

Limitations and Future Directions

Findings highlight several areas for future investigation. More research is needed before conclusions can be drawn about the efficacy and cost-benefit analysis of selected versus indicated prevention. The most common methodological weaknesses across the studies included variability in timing of the intervention (ie, interventions posited to be early interventions were not delivered acutely following trauma), lack of adequate power to detect significant differences, and potential sampling bias. To address design weaknesses in the current literature, rigorous and adequately powered research with clearly defined objectives should be conducted on existing interventions that utilize gold standard, blinded-clinician assessment, control groups, and follow-up data. Blinded-clinician assessment reduces potential bias, and clearly defined objectives may increase internal validity of the study.

Interventions should be delivered and tested in the intended acute phase after trauma. Accuracy in the timing of intervention delivery will ensure validity and generalizability of results. Due to the potential value of indicated over selected prevention, more research on posttrauma risk and resilience factors is necessary to determine for whom and how to target interventions. Participants should be recruited from diverse settings to determine the generalizability and efficacy of such interventions across survivors of different types of trauma. Moreover, to address potential sampling bias, recruiting participants from settings that draw trauma survivors for reasons other than mental health purposes (eg, motor vehicle accident reporting centers, emergency departments, or family physician offices) could be employed. By recruiting individuals who do not necessarily present for mental health posttrauma care, samples may include individuals who do not self-select based on predetermined symptoms of trauma. Future research should also identify barriers to accessing internet-delivered interventions.

In addition to lack of rigor across studies, the literature is limited in that none of the studies evaluated early internet-delivered intervention in survivors of interpersonal trauma (eg, sexual assault) specifically. Interpersonal trauma may be especially important to target, given its prevalence [34] and that most cases of posttrauma pathology stems from interpersonal trauma [35]. Online interventions aimed at victims of interpersonal trauma may be valuable in these populations and increase intervention seeking, because victims may be at risk for experiencing shame and stigma after trauma that could hinder them from seeking face-to-face mental health care [36].

Despite calls for interpersonal-based early posttrauma interventions [11] and findings that lack of posttraumatic social support is a potent risk factor for psychopathology [37,38], none of the interventions reviewed were interpersonally based (although 3 included peer support [20,27,29]). As none of the included studies examined mediators of intervention outcomes, little is known about whether peer support groups offered unique benefits compared with other aspects of the interventions. No studies examined the use of early interventions for individuals with poor posttraumatic social support, despite consistent findings that these individuals are at greater risk of posttraumatic pathology [37,38]. In studies where interventions were delivered to only those at-risk, risk was not defined in terms of social support (ie, always as symptom elevation). Feldner et al [11] hold that preventive interventions that mobilize social support may be best suited to naturalistic settings (eg, schools, religious communities), but the internet may serve as a valuable platform through which connections between socially isolated trauma survivors can be fostered. Researchers should continue to investigate interpersonal risk factors for posttraumatic psychopathology to develop and target such interventions.

In addition, although there was an approximately equal distribution of interventions included in this review that were targeted at adults, compared with children and adolescents, only 7 studies were reviewed. Researchers should continue to develop and study early posttraumatic interventions targeted at different age groups, given the prevalence of trauma across the life span [1]. There were also differences in terms of the complexity of the interventions delivered across the included studies, as some interventions [20,29,30] employed hybrid approaches (ie, stepped care, combination of a support group and clinician messages, self-help for parents in addition to an online intervention). Hybrid approaches may require more resources and could be more expensive [39]. However, there is a potential that such interventions may yield better results in alleviating psychopathology [39]. Researchers should continue to explore hybrid interventions compared with single treatments to determine the most parsimonious, cost-effective way to effectively prevent posttrauma psychopathology.

Conclusions

To conclude, data suggest the potential efficacy of *indicated* early internet-delivered interventions in reducing mental health symptoms among trauma-exposed individuals experiencing elevated mental health symptoms. However, more high-quality, adequately powered studies are necessary before concrete conclusions can be drawn about the efficacy of such interventions.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Description of interventions in included studies.

[DOCX File, 16KB - jmir_v20i10e280_app1.pdf]

Multimedia Appendix 2

Information on internet-delivered early interventions.

[PDF File (Adobe PDF File), 127KB - jmir_v20i11e280_app2.pdf]

Multimedia Appendix 3

Mean scores on outcome measures at pre- and postintervention.

[PDF File (Adobe PDF File), 129KB - jmir_v20i11e280_app3.pdf]

References

1. Van Ameringen M, Mancini C, Patterson B, Boyle MH. Post-traumatic stress disorder in Canada. CNS Neurosci Ther 2008;14(3):171-181. [doi: 10.1111/j.1755-5949.2008.00049.x] [Medline: 18801110]
2. Ehlers A, Mayou RA, Bryant B. Psychological predictors of chronic posttraumatic stress disorder after motor vehicle accidents. J Abnorm Psychol 1998;107(3):508-519. [doi: 10.1037/0021-843X.107.3.508]
3. Riggs DS, Rothbaum BO, Foa EB. A prospective examination of symptoms of posttraumatic stress disorder in victims of nonsexual assault. J Interpers Violence 2016 Jul 02;10(2):201-214. [doi: 10.1177/0886260595010002005]
4. Rothbaum BO, Foa EB, Riggs DS, Murdock T, Walsh W. A prospective examination of post-traumatic stress disorder in rape victims. J Trauma Stress 1992 Jul;5(3):455-475. [doi: 10.1007/BF00977239]
5. Galea S, Ahern J, Resnick H, Kilpatrick D, Bucuvalas M, Gold J, et al. Psychological sequelae of the September 11 terrorist attacks in New York City. N Engl J Med 2002 Mar 28;346(13):982-987. [doi: 10.1056/NEJMsa013404] [Medline: 11919308]
6. Hoge CW, Castro CA, Messer SC, McGurk D, Cotting DI, Koffman RL. Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. N Engl J Med 2004 Jul 01;351(1):13-22. [doi: 10.1056/NEJMoa040603] [Medline: 15229303]
7. Kilpatrick DG, Ruggiero KJ, Acierno R, Saunders BE, Resnick HS, Best CL. Violence and risk of PTSD, major depression, substance abuse/dependence, and comorbidity: results from the National Survey of Adolescents. J Consult Clin Psychol 2003 Aug;71(4):692-700 [FREE Full text]
8. Roberts NP, Kitchiner NJ, Kenardy J, Bisson JI. Systematic review and meta-analysis of multiple-session early interventions following traumatic events. Am J Psychiatry 2009 Mar;166(3):293-301. [doi: 10.1176/appi.ajp.2008.08040590] [Medline: 19188285]
9. Litz BT, Gray MJ, Bryant RA, Adler AB. Early intervention for trauma: current status and future directions. Clin Psychol Sci Prac 2002;9:112-134. [doi: 10.1093/clipsy.9.2.112]
10. Mrazek PG, Haggerty RJ. Reducing Risk for Mental Disorders: Frontiers for Preventive Intervention Research. Washington, DC: National Academy Press; 1994.
11. Feldner MT, Monson CM, Friedman MJ. A critical analysis of approaches to targeted PTSD prevention: current status and theoretically derived future directions. Behav Modif 2007 Jan;31(1):80-116. [doi: 10.1177/0145445506295057] [Medline: 17179532]
12. Qi W, Gevonden M, Shalev A. Prevention of post-traumatic stress disorder after trauma: current evidence and future directions. Curr Psychiatry Rep 2016 Feb;18(2):20 [FREE Full text] [doi: 10.1007/s11920-015-0655-0] [Medline: 26800995]
13. Agorastos A, Marmar CR, Otte C. Immediate and early behavioral interventions for the prevention of acute and posttraumatic stress disorder. Curr Opin Psychiatry 2011 Nov;24(6):526-532. [doi: 10.1097/YCO.0b013e32834cdde2] [Medline: 21941180]
14. Kliem S, Kröger C. Prevention of chronic PTSD with early cognitive behavioral therapy. A meta-analysis using mixed-effects modeling. Behav Res Ther 2013 Nov;51(11):753-761. [doi: 10.1016/j.brat.2013.08.005] [Medline: 24077120]
15. Elhai JD, Terry N, Frueh CB. Health service use predictors among trauma survivors: a critical review. Psychol Serv 2005;2(1):3-19 [FREE Full text]
16. Amstadter AB, Broman-Fulks J, Zinzow H, Ruggiero KJ, Cercone J. Internet-based interventions for traumatic stress-related mental health problems: a review and suggestion for future research. Clin Psychol Rev 2009 Jul;29(5):410-420 [FREE Full text] [doi: 10.1016/j.cpr.2009.04.001] [Medline: 19403215]
17. Sijbrandij M, Kunovski I, Cuijpers P. Effectiveness of internet-delivered cognitive behavioral therapy for posttraumatic stress disorder: a systematic review and meta-analysis. Depress Anxiety 2016 Dec;33(9):783-791. [doi: 10.1002/da.22533] [Medline: 27322710]
18. Sloan DM, Gallagher MW, Feinstein BA, Lee DJ, Pruneau GM. Efficacy of telehealth treatments for posttraumatic stress-related symptoms: a meta-analysis. Cogn Behav Ther 2011 Jun;40(2):111-125. [doi: 10.1080/16506073.2010.550058] [Medline: 21547778]
19. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med 2009;6:1-6. [Medline: 19622511]

20. Zatzick D, O'Connor SS, Russo J, Wang J, Bush N, Love J, et al. Technology-enhanced stepped collaborative care targeting posttraumatic stress disorder and comorbidity after injury: a randomized controlled trial. J Trauma Stress 2015 Oct;28(5):391-400 [FREE Full text] [doi: 10.1002/jts.22041] [Medline: 26467327]
21. Todder D, Matar M, Kaplan Z. Acute-phase trauma intervention using a videoconference link circumvents compromised access to expert trauma care. Telemed J E Health 2007 Feb;13(1):65-67. [doi: 10.1089/tmj.2006.0039] [Medline: 17309357]
22. Freedman SA, Dayan E, Kimelman YB, Weissman H, Eitan R. Early intervention for preventing posttraumatic stress disorder: an Internet-based virtual reality treatment. Eur J Psychotraumatol 2015;6:25608 [FREE Full text] [doi: 10.3402/ejpt.v6.25608] [Medline: 25843345]
23. Price M, Gros DF, McCauley JL, Gros KS, Ruggiero KJ. Nonuse and dropout attrition for a web-based mental health intervention delivered in a post-disaster context. Psychiatry 2012 Jan;75(3):267-284 [FREE Full text] [doi: 10.1521/psyc.2012.75.3.267] [Medline: 22913502]
24. Ruggiero KJ, Resnick HS, Acierno R, Coffey SF, Carpenter MJ, Ruscio AM, et al. Internet-based intervention for mental health and substance use problems in disaster-affected populations: a pilot feasibility study. Behav Ther 2006 Jun;37(2):190-205. [doi: 10.1016/j.beth.2005.12.001] [Medline: 16942971]
25. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998 Jun 01;52(6):377-384. [doi: 10.1136/jech.52.6.377] [Medline: 9764259]
26. Trac MH, McArthur E, Jandoc R, Dixon SN, Nash DM, Hackam DG, et al. Macrolide antibiotics and the risk of ventricular arrhythmia in older adults. Can Med Assoc J 2016 Apr 19;188(7):E120-E129 [FREE Full text] [doi: 10.1503/cmaj.150901] [Medline: 26903359]
27. Mouthaan J, Sijbrandij M, de Vries GJ, Reitsma JB, van de Schoot R, Goslings JC, et al. Internet-based early intervention to prevent posttraumatic stress disorder in injury patients: randomized controlled trial. J Med Internet Res 2013 Aug 13;15(8):e165 [FREE Full text] [doi: 10.2196/jmir.2460] [Medline: 23942480]
28. Steinmetz SE, Benight CC, Bishop SL, James LE. My Disaster Recovery: a pilot randomized controlled trial of an Internet intervention. Anxiety Stress Coping 2012 Sep;25(5):593-600 [FREE Full text] [doi: 10.1080/10615806.2011.604869] [Medline: 21834731]
29. Van Voorhees BW, Gollan J, Fogel J. Pilot study of internet-based early intervention for combat-related mental distress. J Rehabil Res Dev 2012;49(8):1175-1190 [FREE Full text] [Medline: 23341310]
30. Ruggiero KJ, Price M, Adams Z, Stauffacher K, McCauley J, Danielson CK, et al. Web intervention for adolescents affected by disaster: population-based randomized controlled trial. J Am Acad Child Adolesc Psychiatry 2015 Sep;54(9):709-717 [FREE Full text] [doi: 10.1016/j.jaac.2015.07.001] [Medline: 26299292]
31. Cox CM, Kenardy JA, Hendrikz JK. A randomized controlled trial of a web-based early intervention for children and their parents following unintentional injury. J Pediatr Psychol 2010 Jul;35(6):581-592. [doi: 10.1093/jpepsy/jsp095] [Medline: 19906829]
32. Kassam-Adams N, Marsac ML, Kohser KL, Kenardy J, March S, Winston FK. Pilot randomized controlled trial of a novel web-based intervention to prevent posttraumatic stress in children following medical events. J Pediatr Psychol 2016 Jun;41(1):138-148 [FREE Full text] [doi: 10.1093/jpepsy/jsv057] [Medline: 26089554]
33. Foa E, Johnson K, Feeny N, Treadwell K. The child PTSD symptom scale: a preliminary examination of its psychometric properties. J Clin Child Psychol 2001;30(3):376-384. [doi: 10.1207/S15374424JCCP3003_9}]
34. Centers for Disease Control Prevention. CDC. 2014. Preventing intimate partner violence URL: https://www.cdc.gov/violenceprevention/pdf/IPV-Factsheet.pdf [WebCite Cache ID 6wFYYdmQs]
35. Luthra R, Abramovitz R, Greenberg R, Schoor A, Newcorn J, Schmeidler J, et al. Relationship between type of trauma exposure and posttraumatic stress disorder among urban children and adolescents. J Interpers Violence 2009 Nov;24(11):1919-1927. [doi: 10.1177/0886260508325494] [Medline: 18945918]
36. Ullman SE, Brecklin LR. Sexual assault history, PTSD, and mental health service seeking in a national sample of women. J Community Psychol 2002 Apr 05;30(3):261-279. [doi: 10.1002/jcop.10008]
37. Brewin CR, Andrews B, Valentine JD. Meta-analysis of risk factors for posttraumatic stress disorder in trauma-exposed adults. J Consult Clin Psychol 2000;68(5):748-766. [doi: 10.1037/0022-006X.68.5.748]
38. Ozer EJ, Best SR, Lipsey TL, Weiss DS. Predictors of posttraumatic stress disorder and symptoms in adults: a meta-analysis. Psychol Bull 2008;129(1):52-73. [doi: 10.1037/1942-9681.S.1.3]
39. Ho FY, Yeung WF, Ng TH, Chan CS. The efficacy and cost-effectiveness of stepped care prevention and treatment for depressive and/or anxiety disorders: a systematic review and meta-analysis. Sci Rep 2016 Dec 05;6:29281 [FREE Full text] [doi: 10.1038/srep29281] [Medline: 27377429]

Abbreviations

CAPS: Clinician-Administered PTSD Scale
CES-D: Center for Epidemiologic Studies Depression Scale
CPSS: The Child PTSD Symptom Scale

CSE: Coping Self-Efficacy Scale for Trauma
HADS: Hospital Anxiety and Depression Rating Scale
IES-R: Impact of Event Scale-Revised
ITT: intention-to-treat
MPSS: Modified PTSD Symptoms Scale
NR: not reported
PCL-C: PTSD Checklist Civilian version
PCL-M: PTSD-Checklist Military version
PedsQL: Pediatric Quality of Life Inventory
PHQ-9: Patient Health Questionnaire-9
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
PSS: Perceived Stress Scale
PSWQ: Penn State Worry Questionnaire
PTSD: posttraumatic stress disorder
TSCC-A: Trauma-Symptom Checklist for Children-A

Edited by G Eysenbach; submitted 02.02.18; peer-reviewed by K Ruggiero, V White, K Goniewicz; comments to author 04.04.18; revised version received 12.07.18; accepted 17.07.18; published 14.11.18.

Please cite as:
Ennis N, Sijercic I, Monson CM
Internet-Delivered Early Interventions for Individuals Exposed to Traumatic Events: Systematic Review
J Med Internet Res 2018;20(11):e280
URL: https://www.jmir.org/2018/11/e280/
doi:10.2196/jmir.9795
PMID:30429113

Original Paper

Evaluation of a Web-Based Intervention for Multiple Health Behavior Changes in Patients With Coronary Heart Disease in Home-Based Rehabilitation: Pilot Randomized Controlled Trial

Yan Ping Duan[1], PhD; Wei Liang[1], MEd; Lan Guo[2], PhD; Julian Wienert[3], PhD; Gang Yan Si[4], PhD; Sonia Lippke[3], PhD

[1]Department of Sport and Physical Education, Faculty of Social Sciences, Hong Kong Baptist University, Hong Kong, China (Hong Kong)

[2]Cardiac Rehabilitation Centre, Guangdong General Hospital, Guangzhou, China

[3]Department of Psychology & Methods, Jacobs University Bremen, Bremen, Germany

[4]Sport Psychology Centre, Hong Kong Sports Institute, Hong Kong, China (Hong Kong)

Corresponding Author:
Yan Ping Duan, PhD
Department of Sport and Physical Education
Faculty of Social Sciences
Hong Kong Baptist University
12/F, Shek Mun Campus
8 On Muk Street, Shek Mun, Shatin
Hong Kong,
China (Hong Kong)
Phone: 852 3411 3080
Fax: 852 3411 3357
Email: duanyp@hkbu.edu.hk

Abstract

Background: Web-based and theory-based interventions for multiple health behaviors appears to be a promising approach with respect to the adoption and maintenance of a healthy lifestyle in cardiac patients who have been discharged from the hospital. Until now, no randomized controlled trials have tested this assumption among Chinese rehabilitation patients with coronary heart disease using a Web-based intervention.

Objective: The study aim was to evaluate the effect of an 8-week Web-based intervention in terms of physical activity (PA), fruit and vegetable consumption (FVC), lifestyle changes, social-cognitive outcomes, and health outcomes compared with a waiting control group in Chinese cardiac patients. The intervention content was theory-based on the health action process approach. Self-reported data were evaluated, including PA, FVC, healthy lifestyle (the synthesis of PA and FVC), internal resources (combination of intention, self-efficacy, and planning), and an external resource (social support) of PA and FVC behaviors, as well as perceived health outcomes (body mass index, quality of life, and depression).

Methods: In a randomized controlled trial, 136 outpatients with coronary heart disease from the cardiac rehabilitation center of a hospital in China were recruited. After randomization and exclusion of unsuitable participants, 114 patients were assigned to 1 of the 2 groups: (1) the intervention group: first 4 weeks on PA and subsequent 4 weeks on FVC and (2) the waiting control group. A total of 2 Web-based assessments were conducted, including 1 at the beginning of the intervention (T1, N=114), and 1 at the end of the 8-week intervention (T2, N=83). The enrollment and follow-up took place from December 2015 to May 2016.

Results: The Web-based intervention outperformed the control condition for PA, FVC, internal resources of PA and FVC, and an external resource of FVC, with an eta-squared effect size ranging from 0.06 to 0.43. Furthermore, the intervention effect was seen in the improvement of quality of life ($F_{1,79}$=16.36, P<.001, η^2=.17). When predicting a healthy lifestyle at follow-up, baseline lifestyle (odds ratio, OR 145.60, 95% CI 11.24-1886; P<.001) and the intervention (OR 21.32, 95% CI 2.40-189.20; P=.006) were found to be significant predictors. Internal resources for FVC mediated the effect of the intervention on the adoption of a healthy lifestyle (R^2_{adj}=.29; P=.001), indicating that if the intervention increased the internal resource of behavior, the adoption of a healthy lifestyle was more likely.

Conclusions: Patients' psychological resources such as motivation, self-efficacy, planning, and social support as well as lifestyle can be improved by a Web-based intervention that focuses on both PA and FVC. Such an intervention enriches extended rehabilitation approaches for cardiac patients to be active and remain healthy in daily life after hospital discharge.

Trial Registration: ClinicalTrials.gov NCT01909349; https://clinicaltrials.gov/ct2/show/NCT01909349 (Archived by WebCite at http://www.webcitation.org/6pHV1A0G1)

(J Med Internet Res 2018;20(11):e12052) doi:10.2196/12052

KEYWORDS

eHealth; physical activity; diet; cardiac rehabilitation; health resources

Introduction

Background

Cardiovascular disease (CVD) is the leading cause of death worldwide, with coronary heart disease (CHD) being the major component [1]. In China, CVD was the number one cause of death in 2017, and CHD was the dominant component responsible for 51.6% of all CVD deaths [2]. Cardiac rehabilitation is widely recognized as an effective intervention to avoid further progression and relapse of CVD [3-5]. CHD patients receive the guidance provided on healthy lifestyle changes regarding physical activity (PA) and a healthy diet during rehabilitation in the hospital [6,7]. However, several studies have revealed that it is often difficult for patients to integrate and transfer these recommendations and learning outcomes into their daily life after discharge from the hospital [8,9]. Thus, they need internal and external resources, which can be supported by an extended rehabilitation aftercare when they are at home [10].

With the prevalence of internet use, Web-based interventions are widely developed and applied in the field of cardiac rehabilitation [11-13]. The effectiveness of Web-based interventions on PA and fruit and vegetable consumption (FVC) has been proven to facilitate healthy lifestyle changes after discharge among cardiac patients and people with cardiovascular risk profiles in the Western Hemisphere [14-16]. However, in the Eastern Hemisphere, and especially in China, a large body of Web-based rehabilitation interventions only focus on knowledge, education, and learning. Very few integrate individualized and comprehensive interventions that include educational, cognitive, and psychological elements [17,18].

This study applied the Health Action Process Approach (HAPA) [8] as the theoretical backdrop, which suggests that there are 2 distinctive phases during the health behavior change process. The first phase is the *motivation phase*, which plays an important role before a goal is set. Subsequently, individuals enter the *volition phase*, which plays an important role in planning and maintaining health behavior changes. This distinction allows interventions to be tailored based on variables that correspond to either the motivation or the volition phase. In particular, motivation must be built up before individuals can actually change unhealthy behaviors (eg, physical inactivity). Therefore, individuals with an unhealthy lifestyle and no intention to change it might benefit most from interventions that will increase risk perception, self-efficacy, and positive outcome expectancies, to support the formation of goal intentions that focus on healthy lifestyle changes [19]. The purpose is to lead toward a specific intention (eg, "I intend to eat five portions of fruit and vegetables today"). Once the intention is formed and the goal is set, individuals enter the volition phase.

In the volition phase, people benefit most from planning interventions, which can help cross the so-called intention behavior gap [14]. In this phase, individuals learn how to make specific plans (eg, when, where, and how to eat fruit and vegetables), determine priorities, and translate their action plans into behavior. Such self-regulatory planning skills are necessary to maintain progress once people initiate a healthy behavior. Positive behavior change will be guided by self-efficacy, as self-efficacy regulates how much effort is invested in goal achievement, and how much persistence is maintained if obstacles and setbacks occur. In addition, promoting behavior-specific social support from people's social context is equally important in preventing relapse [8]. In general, cardiac rehabilitation patients need internal (forming and improving intention, self-efficacy, and planning) and external (forming behavior-specific social support) resources to support them to adopt and maintain a healthy lifestyle after their discharge from the hospital.

Until now, how internal and external resources interplay with lifestyle change in cardiac patients (eg, following recommendations for PA and FVC) has not been fully addressed. Interventions that integrate both internal and external resources may enhance social cognitions, which in turn can lead to the adoption of a healthy lifestyle. Thus, mediation analysis might disclose the underlying mechanisms of such an intervention. All of this has been researched in the Western Hemisphere [15,20] but not in the Eastern Hemisphere, and therefore, the purpose of this study is to fill this gap.

Aim and Hypotheses

The main research aim of this study was to examine the efficacy of an 8-week Web-based intervention compared with a waiting control group (WCG) to improve PA and FVC in Chinese rehabilitation patients with CHD after hospital discharge. Effects on single health behaviors, healthy lifestyle (the synthesis of PA and FVC), social-cognitive indicators for PA and FVC (internal and external resources), and perceived health outcomes (body mass index [BMI], quality of life, and depression) were tested (see hypotheses 1 to 3). The second aim of this study was to examine whether changes in internal resources (combination of intention, self-efficacy, and planning) and an external resource (social support) mediate the effect of the intervention on changes in lifestyle (hypothesis 4).

The main intervention effects were hypothesized in terms of (1) self-reported increase in PA and FVC behavior (single behavior indicators; hypothesis 1a) and adoption of a healthy lifestyle (hypothesis 1b); (2) more improvements in indicators of internal resources (combination of intention, self-efficacy, and planning) and an indicator of external resources (social support; hypothesis 2); (3) improvement in health outcomes (BMI, quality of life, and depression level; hypothesis 3); and (4) the expectation that those patients who had increased internal resources (combination of intention, self-efficacy, and planning) and an external resource (social support) because of the intervention would be more likely to adopt a healthy lifestyle (mediation effect; hypothesis 4).

Methods

Study Design, Procedure, and Participants

This study was a pilot randomized controlled trial involving 1 intervention group (IG) and 1 WCG. There were 2 assessments: a pretest after randomization and before intervention start or the waiting time (T1, n=114) and a posttest at 8 weeks after the pretest (T2, n=83). The WCG was allowed to access the 8-week Web-based intervention after T2, once the IG had finished the intervention. All patients were informed about the purpose of the study with an informed consent form. The study received ethical approval by the Committee for the Use of Human & Animal Subjects in Teaching & Research of Hong Kong Baptist University (FRG1/12-13/064) and was registered with ClinicalTrials.gov (NCT01909349; Multimedia Appendix 1 [12]).

Enrollment and follow-up took place from December 2015 to May 2016. Outpatients with CHD were recruited face-to-face by the physician of the research team, with the assistance of a nurse in the cardiac rehabilitation center of a hospital in south China. The inclusion criteria were as follows: aged between 18 and 75 years, no restriction of physical mobility under the cardiac function at entry, no restriction of other relevant diseases such as diabetes or fruit allergies, sufficient reading and writing skills in Chinese, internet access via a computer at home, and mobile access.

Figure 1 shows the flowchart of patients from enrollment in the study to allocation to the 2 groups (IG and WCG) and posttest after 8 weeks. A total of 136 patients with CHD were recruited. Of these, 16.2% (22/126) were excluded after random allocation due to restrictions in terms of PA or FVC (n=11), no internet access via a computer at home (n=4), or because they declined to participate (n=7). Subsequently, 83.8% (114/136) eligible patients fulfilled the online registration and provided personal information within 1 week during the period of hospital rehabilitation. In total, 60 patients (52.6%) in the IG and 54 patients (47.4%) in the WCG were included. During the last week before they were discharged from the hospital, patients in both groups were invited to complete the first online questionnaire (pretest; T1). Patients in the IG were encouraged to access a Web-based health program via their home computer once a week during the following 8 weeks, whereas patients in the WCG did not receive any treatment. Upon completion of the 8-week intervention, patients in both groups were invited to fill in the second online questionnaire (posttest; T2). The final longitudinal sample consisted of 83 (72.8%) patients, including 44 (53%) in the IG, and 39 (47%) in the WCG (Figure 1).

All website links for the questionnaire surveys at T1 and T2 as well as for the weekly intervention program were delivered via email. To boost the engagement of patients, short message service (SMS) text messages were sent as reminders. Furthermore, the nurse contacted participants via phone calls once per week before each intervention session and at the 2 measurement points to remind the patients. Moreover, patients were offered a 60 renminbi (RMB; US $9) telephone recharge card as an incentive in exchange for their participation in the study and completion of the 2 data collection waves.

Intervention

The 8-week Web-based intervention started with PA in the first 4 weeks, followed by the FVC in the later 4 weeks. This sequential design falls into the previous study results, suggesting PA might act as gateway behavior [21,22], and PA modules are the most favored modules in tailored electronic health (eHealth) lifestyle promotion [23].

The intervention content was designed based on the HAPA theory. Accordingly, social-cognitive variables of PA and FVC were targeted in the intervention (Multimedia Appendix 2). In particular, the targeted variables were as follows: week 1 and week 5—risk perception, outcome expectancies, and goal setting; week 2 and week 6—development of action plans; week 3 and week 7—revision and adjustment of previous action plans and development of coping plans; and week 4 and week 8—revision and adjustment of previous coping plans and development of behavior-specific social support. Self-efficacy was a fixed intervention variable that was included from week 1 to week 8. Multimedia Appendix 2 indicates intervention variables and content for each week.

In addition, selected behavior change techniques were addressed in the intervention to facilitate the implementation and maintenance of behavior [24]. For example, 2 types of feedback were provided: patients received individualized feedback on their self-reported behavior performance 4 weeks ago, 3 weeks ago, 2 weeks ago, and 1 week ago. Criterion-based feedback was also presented (eg, accumulated at least 150 min with moderate intensity of PA per week and 5 portions of FVC per day). The examples of the feedback can be found in a separate publication [25]. Moreover, examples of role models were provided throughout the intervention to support patients to set goals, develop plans, and increase their self-efficacy.

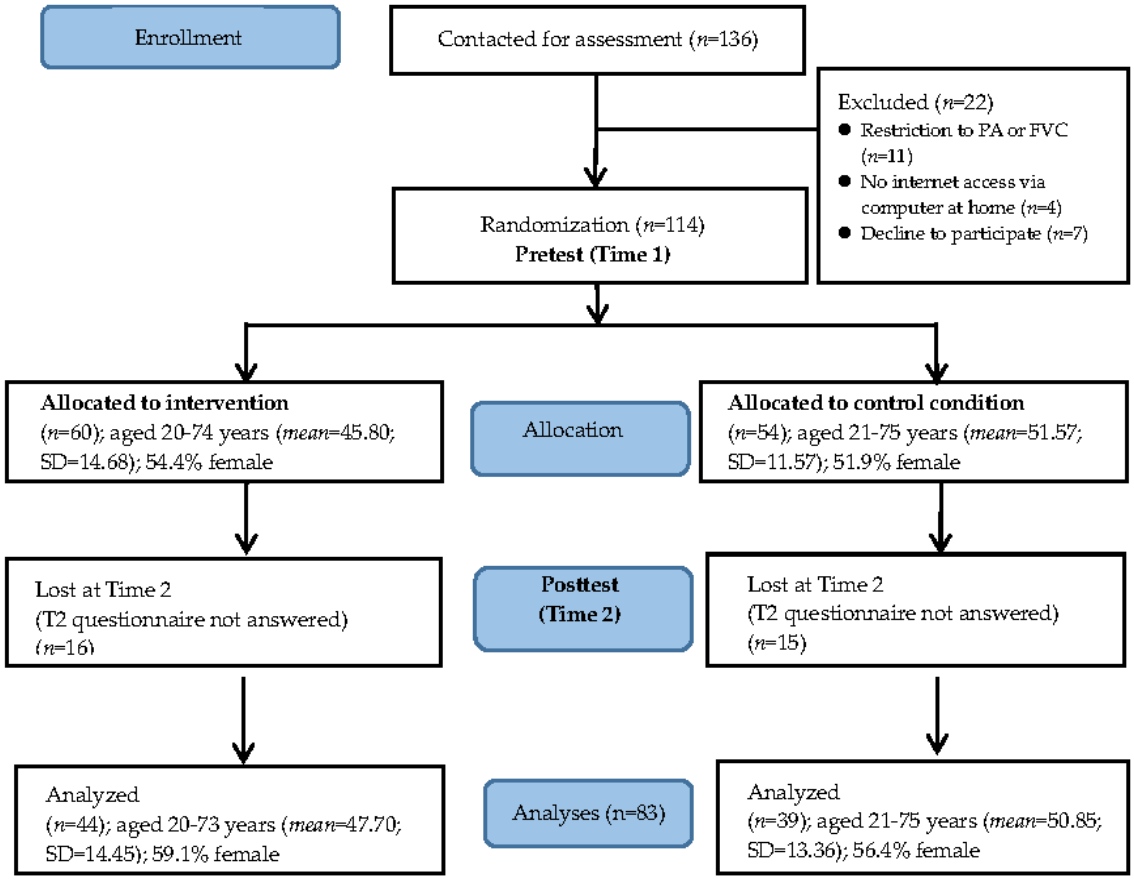
Measures

All variables were self-reported online at pretest (T1) and posttest after the 8-week intervention period (T2).

Demographic Information

Sociodemographic information such as gender, age, level of education (low-level: primary school; high-level: high school, and university or college), and relationship status (single or in a relationship) was assessed. In addition, self-reported body height (cm) and weight (kg) were collected at T1 and T2.

Figure 1. Flowchart of participant progress throughout the study phases. FVC: fruit and vegetable consumption; PA: physical activity.



Health Behavior (Single Behavior Indicators)

The level of PA was assessed using an adaption of the Chinese short version of the International Physical Activity Questionnaire [26]. Patients indicated how often per week and how long each time they performed vigorous PAs (eg, fast bicycling and intensive swimming) and moderate physical activities (eg, carrying light loads and bicycling at a regular pace) during the last week. The total PA score (in min/week) for each patient was obtained by summing up all questions.

FVC during the past 7 days was assessed with 4 items, including *raw vegetables*, *fruits*, *fruit or vegetable juice*, and *cooked or steamed vegetables* [27]. For each item, patients were asked to count the number of portions or glasses of liquid, fruit, and vegetables they consumed on average per day (11 options such as 0, 0.5, 1, 1.5, 2, 2.5..., 5, or above). The total consumed portion was the sum of each item.

Combined Healthy Lifestyle Indicator

To combine the 2 behaviors, PA and FVC behaviors were categorized according to whether or not the patients met the health recommendations [20]. After exploring the usefulness of different criteria, the thresholds of 150 min of PA with moderate intensity per week and 5 portions of fruit and vegetables were chosen. Both criteria were validated and have been shown to be effective in improving health [28,29]. At T1, 60% of patients (50/83 eligible patients participating in T1 and T2) did not perform 150 or more min of PA with moderate intensity per week. Moreover, 71% (59/83) of patients did not eat 5 or more portions of fruit and vegetables per day. When both behaviors were combined, 84% (70/83) participants met only 1 or none of these 2 behavioral criteria and were categorized as having an unhealthy lifestyle at T1. Overall, 15% (13/83) of the patients met both behavior recommendations and were categorized as having a healthy lifestyle.

At T2, 50% of patients (42/83 eligible patients participating in T1 and T2) did not perform 150 or more min of PA with moderate intensity per week. In addition, 60% (50/83) of patients did not eat 5 or more portions of fruit and vegetables per day. When both behaviors were combined, 73% (61/83) of participants met only 1 or none of these 2 behavioral criteria and were categorized as having an unhealthy lifestyle at T1. In total, 26% (22/83) of patients met both behavior recommendations and were categorized as having a healthy lifestyle.

Social-Cognitive Indicators as Internal Resources of Behavior Change

Intention

Intention concerning PA was measured with the stem "On 5 days a week for 30 min, I have the intention to perform..." followed by 3 items such as "...strenuous sports activities," "...moderate PA," and "...mild PA, like walking" (Cronbach alpha=.40) [30]. Intention concerning FVC was assessed by the stem "I intend to..." followed by 3 items such as "...eat at least 5 servings of fruits and vegetables each day," "...eat fruit and vegetables at every meal," and "drink at least one fruit or vegetable juice every day" (Cronbach alpha=.51) [30]. The answers were indicated on a 4-point scale ranging from 1 "not true" to 4 "exactly true."

Self-Efficacy

Self-efficacy for PA was assessed with 5 items such as "I am certain that I can permanently be physically active for at least 5 days a week for 30 minutes each day" (Cronbach alpha=.89) [31]. Self-efficacy for FVC was assessed by 5 items such as "I am certain that I can eat 5 portions of fruits and vegetables a day, even if it is sometimes difficult" (Cronbach alpha=.95) [32]. The answers were given on a 5-point scale ranging from 1 "don't agree at all" to 5 "totally agree."

Planning

The planning indicator was categorized into *action planning* and *coping planning*. Action planning was assessed by the stem "For the next month I have carefully planned..." followed by 3 items for PA such as "...which PA I will pursue" (Cronbach alpha=.94) or followed by 3 items for FVC such as "...what I will eat" (Cronbach alpha=.91) [32,33]. Coping plans were measured by the stem "For the next month I have carefully planned..." followed by 3 items for PA such as "...What I can do in difficult situations to stick to my intentions" (Cronbach alpha=.90) or followed by 3 items for FVC such as "...How I can eat healthy, even if something happened" (Cronbach alpha=.92) [32,33]. The answers were given on a 5-point scale ranging from 1 "strongly disagree" to 5 "strongly agree."

Social-Cognitive Indicator as External Resources of Behavior Change

Social Support

This indicator was measured with 3 items for PA (Cronbach alpha=.72) and with 3 items for FVC (Cronbach alpha=.87) such as "My partner helps me/my family helps me/my friends and acquaintances help me to stay physically active" or "My partner helps me/my family helps me/my friends and acquaintance help me to eat healthy." Answers were given on a 4-point scale ranging from 1 "totally disagree" to 4 "totally agree" [34].

Health Outcomes

Body Mass Index

BMI was calculated using the formula "BMI=weight (kg)/height (m^2)." Body weight and body height were reported by the study participants.

Quality of Life

This indicator was measured using the Hong Kong version of the World Health Organization's Quality of Life questionnaire (Brief version) [35]. Respondents were first asked about their general quality of life by the item "How would you rate your quality of life?" Study participants were assessed with a 5-point scale ranging from 1 "very poor" to 5 "very good" and by 7 items on physical health subdomains. Such subdomains were "How satisfied are you with your ability to perform your daily living activities?" Response categories used a 5-point scale ranging from 1 "very dissatisfied" to 7 "very satisfied" (Cronbach alpha=.87).

Depression

Level of depression was measured using the Chinese translated Center for Epidemiological Studies-Depression (CES-D) scale. Respondents were asked with the stem "In the past week how often I feel..." followed by 20 items such as "...I was bothered by things that usually don't bother me." Respondents were asked to indicate the frequency of symptoms on a 4-point scale (0=less than a day, 1=1 to 2 days, 2=3 to 4 days, 3=5 to 7 days). Positively formulated items were reversed. The total score consisted of the sum of all 20 items and ranged from 0 to 60 (Cronbach alpha=.88). A CES-D score ≥16 indicated the likelihood of a clinically significant depression [36].

All questionnaires above stem from validated and well-tested measurement tools in the Chinese version used in previous studies [25,26,35,36]. In addition, pretests were conducted to ensure the usability and technical functionality of the electronic versions of the questionnaires before the main study.

Data Analysis

All data analyses were performed using IBM SPSS 24.0. The analyses of dropout and comparison of participant characteristics at T1 were conducted with independent sample *t* tests and chi-square tests. The 5% level (2-tailed) was used as the statistical significance cutoff point.

Intervention effects on PA and FVC behaviors (hypothesis 1a) were tested with repeated-measures analysis of variance (ANOVA). Hypothesis 1b on the combined healthy lifestyle indicator was tested using chi-square and logistic regression analysis (determining odds ratio, OR). In addition, intervention effects on social-cognitive variables (indicators of internal resources and external resources; hypothesis 2) and health outcomes (BMI, quality of life, and depression level; hypothesis 3) were evaluated by conducting a series of repeated-measures analyses of covariance (ANCOVA) with baseline behavior as the covariate. The 2 factors for both ANOVAs and ANCOVAs were time (T1 and T2) as the within-participants factor and treatment (IG vs WCG) as the between-participants factor. Hypothesis 4 on the multiple mediation models was performed using an SPSS macro [37]. Before the mediation analysis, the diagnostics of multicollinearity were performed. The criteria of serious multicollinearity problem include high correlation (r>.85), low tolerance (≤0.01), high variance inflation factor (VIF >10), low eigenvalue (approach zero), and large condition index (>30) among mediators [38]. Residualized change scores were used, and 95% CIs of the standardized effects of the

intervention were estimated by applying the bootstrap approach (5000 bootstrap samples).

Results were reported based on those cardiac patients who participated in both measurement points. As less than 5% of items were missing and belonged to a randomized distribution, the expectation-maximization method was adopted to impute the missing data within each measurement point in time [39].

Results

Sample Characteristics and Randomization Check

There were 114 cardiac patients who participated in data collection at T1 (IG=60, WCG=54), and 83 of them completed data collection at T2 (IG=44, WCG=39). The dropout rate of participants was 27.2% (31/114) from T1 to T2. For participants in the IG, 44 (T2) out of 60 (T1) patients (73%) adhered to the whole 8-week intervention program. The 83 final patient sample included more females (57%, 48/83) than males, aged from 20 to 75 years (mean=49.18 years, SD=13.96). The majority of patients were in a relationship (94%, 78/83). Most patients had a high education level (89%, 74/83). The average BMI of patients was 23.65 (SD 2.89, range 16.71-31.25).

Participants at T1 and T2 did not significantly differ from dropouts at T2 regarding the gender (χ^2_1=2.3, *P*=.13), education level (χ^2_1=1.6, *P*=.20), relationship status (χ^2_1=0.3, *P*=.56), age (t_{112}=−0.77, *P*=.44), amount of PA at T1 (t_{109}=−1.78, *P*=.08), and FVC at T1 (t_{108}=−0.88, *P*=.38).

Results also indicated that there were no significant differences across the 2 groups at T1 with respect to single behaviors of PA and FVC, internal resources for PA and FVC (combination of intention, self-efficacy, and planning), an external resource for PA and FVC (social support), age, or health outcomes (BMI, quality of life, and depression; all *P*=.11 to .86). In addition, there were no significant differences between the 2 groups on gender (χ^2_1=0.4, *P*=.54), education level (χ^2_1=0.8, *P*=.49), relationship status (χ^2_1=0.02, *P*=.90), and combined health lifestyle indicator (combination of both behaviors; χ^2_1=0.5, *P*=.50).

Evaluation of Time and Treatment on Single-Behavior Indicators as well as Social-Cognitive Indicators of Behavior Change

Table 1 presents the evaluation result of the time, treatment, and time x treatment effect. Out of 6 interaction effects, 5 were statistically significant, with an effect size of η^2ranging from 0.06 to 0.43. The mean values of PA, FVC, internal resources of PA, internal resources of FVC, and social support of FVC between 2 groups over time are presented in Figure 2.

Although there was no significant interaction effect on social support for PA (*P*=.08), the mean value of social support for PA (see Figure 2, panel C) indicated that there was a descriptive difference between the 2 groups, which was clearly in favor of the IG.

Evaluation of Time and Electronic Health Treatment on Combined Healthy Lifestyle Indicators

As the purpose of the intervention was not only to change single behaviors but also to improve patients' lifestyle including both PA and FVC behaviors, change in the combined health indicator was evaluated. According to their PA and FVC behaviors, patients were categorized into groups based on whether or not they met the recommended criteria. The frequencies and percentage of a healthy and an unhealthy lifestyle at T2 depending on T1 lifestyle and treatment are presented in Table 2.

Table 1. Intervention efficacy evaluated in terms of changes over time assessed in a 2-factorial repeated measures analysis of variance.

Variables	Time			Treatment			Baseline behavior			Interaction time x treatment		
	F test (*df*)	η^2	*P* value	*F* test (*df*)	η^2	*P* value	*F* test (*df*)	η^2	*P* value	*F* test (*df*)	η^2	*P* value
Physical activity (minutes/week)												
Behavior[a]	2.88 (1,81)	.03	.09	1.33 (1,81)	.02	.25	N/A[b]	N/A	N/A	9.25 (1,81)	.10	.003
Internal resources[c]	14.13(1,80)	.15	<.001	11.17 (1,80)	.12	.001	13.66 (1,80)	.15	<.001	13.0 (1,80)	.14	.001
External resource (social support)	3.26 (1,80)	.04	.08	2.51 (1,80)	.03	.12	0.19 (1,80)	<.01	.66	3.08 (1,80)	.04	.08
Fruit and vegetable consumption (portions/day)												
Behavior	23.69 (1,80)	.23	<.001	7.27	.08	.009	N/A	N/A	N/A	59.01 (1,80)	.43	<.001
Internal resources	11.06 (1,79)	.12	.001	9.21 (1,79)	.10	.003	10.21 (1,79)	.11	.002	24.96 (1,79)	.24	<.001
External resource (social support)	1.59 (1,79)	.02	.21	1.05 (1,79)	.01	.30	0.46 (1,79)	<.01	.31	4.61 (1,79)	.06	.04

[a]Behavior indicators: moderate-vigorous physical activity or fruit and vegetable consumption.

[b]N/A: not applicable.

[c]Internal resources: intention, self-efficacy, and planning.

Figure 2. Mean values for intervention group (solid line) and waiting control group (dotted line) at T1 and T2. (A) Performed physical activity (PA) (minutes/week). (B) Internal resources for PA (combination of intention, self-efficacy, and planning). (C) External resources for PA (social-support). (D) Fruit and vegetable consumption (FVC) behaviour (portions/day). (E) Internal resources for FVC (combination of intention, self-efficacy, and planning). (F) External resources for FVC (social-support).

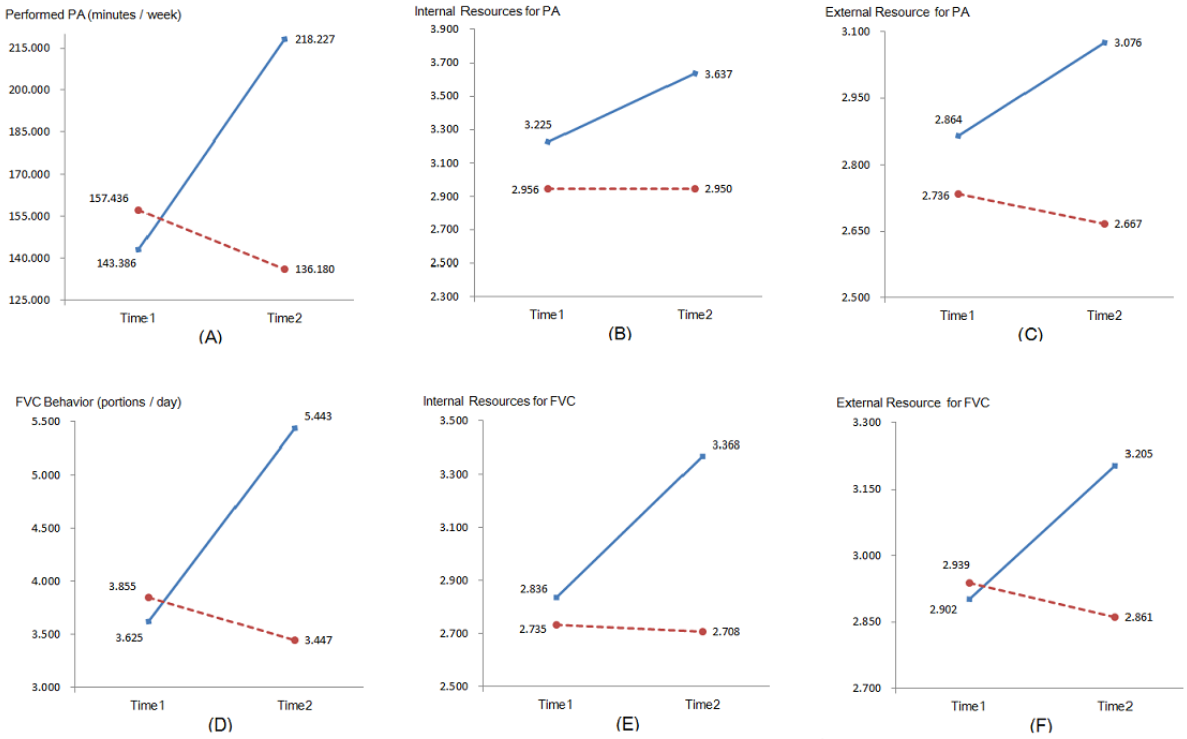


Table 2. Performance of a healthy and an unhealthy lifestyle at T2 based on lifestyle and treatment at T1.

Lifestyle at T1 and treatment	Lifestyle T2[a], n (%)		Total, N
	Unhealthy T2	Healthy T2	
Unhealthy T1			
Intervention group	26 (72)	10 (28)	36
Control group	33 (97)	1 (3)	34
Healthy T1			
Intervention group	0 (0)	8 (100)	8
Control group	2 (40)	3 (60)	5
Unhealthy and Healthy T1			
Intervention group	26 (59)	18 (41)	44
Control group	35 (90)	4 (10)	39
Total	61 (74)	22 (27)	83

[a]Lifestyle T1 or T2=1 indicates meeting both behavior recommendations (conducting ≥150 min of moderate-vigorous physical activity per week and adopting ≥5 portions of fruit and vegetables per day); lifestyle T1 or T2=0 indicates not meeting both behavior recommendations (conducting ≥150 min of physical activity per week and adopting ≥5 portions of fruit and vegetables per day).

Descriptively, among those with an unhealthy lifestyle at T1, the IG outperformed the WCG in performing healthy lifestyle behavior (ie, 24.9% more). For those with a healthy lifestyle at T1, 100% (8/8) of patients in the IG still maintained their healthy lifestyle throughout treatment, compared with only 60% of patients in WCG. If all patients were considered together, it was found that 40% (18/44) of patients in the IG adopted a healthy lifestyle through treatment compared with only 10% of (4/39) patients in the WCG. However, due to the small sample size, the differences could not be statistically tested.

To further explore which variables predict healthy lifestyle behavior at T2, logistic regression analyses were employed with 3 models (Table 3). First, gender and age (all at T1) were used as predictors for the adoption of a healthy lifestyle at T2. Neither

of these 2 variables, however, was a significant predictor (P=.30 to .47). Lifestyle at T1 was then added as a predictor in model 2, which exhibited a significant correlation to a healthy lifestyle at T2: patients who had a healthy lifestyle at T1 were about 47.51 times more likely to also adopt or maintain a healthy lifestyle at T2 (Table 3). Furthermore, the treatment was included as a predictor in model 3, which was also a significant predictor for a healthy lifestyle at T2 ($P \leq .006$; Table 3). In other words, those patients receiving the intervention were about 21.32 times more likely to practice or maintain a healthy lifestyle in comparison with patients in the control group. With model 3, lifestyle at T1 and treatment could attribute to 37% of the variance of a healthy lifestyle at T2.

Evaluation of Time and Electronic Health Treatment on Health Outcomes

Analyses of the time and treatment effect on health outcomes in terms of BMI, quality of life, and depression level were performed. A significant difference was yielded for the interaction factor (time x treatment) on quality of life ($F_{1,79}$=16.36, P<.001, η^2=.17). Regarding the BMI and depression level, however, the interaction of time and treatment were not significant, respectively ($F_{1,78}$=3.35, P=.07, η^2=.04; $F_{1,80}$=0.001, P=.98, η^2<.001). The mean values for BMI, quality of life, and depression at T1 and T2 are displayed in Figure 3.

Testing Mechanism of How the Electronic Health Treatment Facilitated a Healthy Lifestyle

Finally, a multiple mediator analysis tested whether the effects of the intervention on lifestyle change could be explained by changes in internal and external resources (Figure 4). The multicollinearity diagnostics indicated that there was no serious multicollinearity problem among mediators (r=0.26-0.55, tolerance=0.48-0.60, VIF=1.66-2.07, eigenvalue=0.25-2.43, and condition index=1.00-3.11). Residualized change scores obtained by regression T2 scores on T1 scores were chosen for the proposed mediators (Figure 4). Group assignment predicted changes in all social-cognitive variables, in particular, in internal resources for PA (beta=.44, SE .81; P<.001), external resource for PA (beta=.25, SE .94; P=.02), internal resources for FVC (beta=.57, SE .68; P<.001, shown in bold in Figure 4), and external resource for FVC (beta=.28, SE .92; P=.009). Lifestyle change, as operationalized by meeting the recommendation toward PA and FVC, was predicted by changes in internal resources for FVC (beta=.54, SE .13; P<.001, shown in bold in Figure 4) and by no other variables. After controlling for changes in these predictor variables, the relation between group assignment and behavior change was no longer significant (beta=.26, SE .23; P=.27; without controlling: beta=.74, SE .21; P<.001), which indicates that internal resources for FVC was a full mediator of the intervention effectiveness. The multiple mediator model accounted for 33% of the variance (R^2_{adj}=.29; P=.001) in lifestyle.

Table 3. Predicting lifestyle at T2.

Variable	Model 1		Model 2		Model 3	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Constant	0.71	N/A[a]	0.05	N/A	0.002	N/A
Gender	1.46 (0.53-4.04)	.47	2.60 (0.65-10.29)	.18	2.89 (0.65-12.79)	.16
Age	0.98 (0.95-1.02)	.30	1.01 (0.97-1.06)	.56	1.03 (0.98-1.08)	.30
Lifestyle in T1[b]	N/A		47.51 (6.97-324.02)	<.001	145.60 (11.24-1886)	<.001
Treatment[c]	N/A		N/A		21.32 (2.40-189.20)	.006
R^2	.02	<.05	.27	<.001	.39	<.001
$\triangle R^2$	N/A	N/A	.25	<.001	.12	<.001

[a]N/A: not applicable.

[b]Lifestyle T1 or T2=1 indicates meeting both behavior recommendations (conducting ≥150 min of physical activity/week and adopting ≥5 portions of fruit and vegetables/day); lifestyle T1 or T2=0 indicates not meeting both behavior recommendations (conducting ≥150 min of moderate-vigorous physical activity/week and adopting ≥5 portions of fruit and vegetables/day).

[c]Treatment: 0 indicates waiting control condition; 1 indicates intervention condition.

Figure 3. Mean values of health outcomes (Body mass index, quality of life, and depression) for intervention group (solid line) and waiting control group (dotted line) at T1 and T2. (A) Body mass index. (B) Quality of life. (C) Depression.

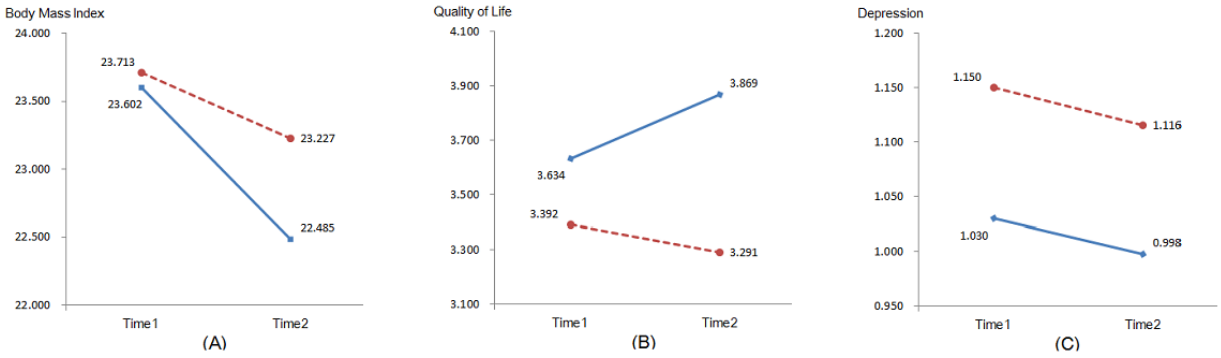
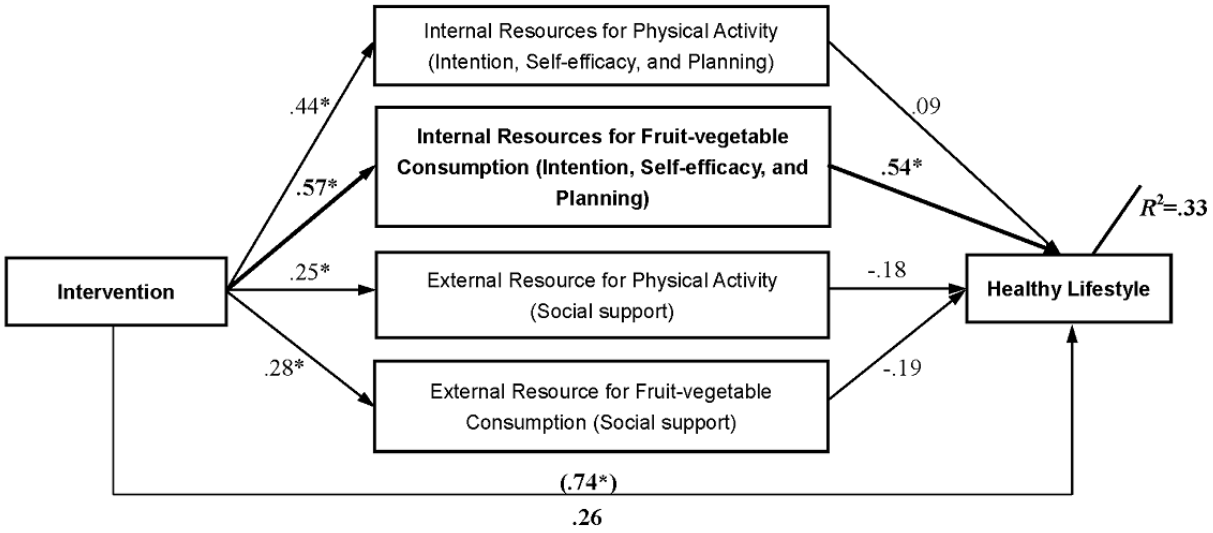


Figure 4. Mediation of the effect of the intervention on lifestyle changes by internal and external resources. Significant changes are indicted by an asterisk.



Discussion

Principal Findings

The aim of this study was to gain insights into an 8-week Web-based health promotion intervention for Chinese cardiac rehabilitation patients after discharge from hospital. The majority of the research hypotheses were supported.

The main expected intervention effects on single behavior and combined lifestyle were identified. Compared with patients in the control group, patients in the IG reported significantly less decrease of PA and FVC over time (hypothesis 1a). The findings on multiple behavior change are in line with previous studies, which showed effects on both behaviors for patients with metabolic syndrome [40,41].

Furthermore, testing the effect of the intervention on lifestyle indicated that more patients in the IG than in the control group improved their lifestyle by meeting the recommendations for PA and FVC, with a 21 times higher likelihood to practice or maintain a healthy lifestyle in comparison with patients in the control group, after controlling for baseline behavior (hypothesis 1b). In addition, patients who already had a healthy lifestyle at baseline were 47 times more likely to engage in a healthy lifestyle than patients in the control group, after controlling for gender and age, which were both insignificant. The effect of baseline behavior matches previous findings. For instance, this was found in another study targeting employees at the workplace [20] and in rehabilitation settings in North America and Europe [15,33,42]. The findings of this study performed in China support the merits of a Web-based extended program for cardiac rehabilitation patients to improve their lifestyle and thus comparably improve their health, for the Western and the

Eastern Hemisphere. Taking the different findings together, hypothesis 1a and hypothesis 1b were confirmed.

The hypothesized intervention effects on internal (combination of intention, self-efficacy, and planning) and external (social support) resources of behaviors were also identified—3 out of 4 social-cognitive test variables were found to be significant. The Web-based intervention increased the internal resources in the PA domain; it was even able to improve nutrition-related internal and external resources compared with the control group. The majority of hypothesis 2 was confirmed.

When testing the hypothesis on the intervention effects on health outcome, we found patients in the IG were more successful in improving their quality of life, which is coherent with the latest literature using a similar intervention program among university students [25]. However, the significant time x treatment interaction was not detected for BMI, although differential intervention effects surfaced on a descriptive level. In a systematic review on interactive computer-based interventions for weight loss or weight maintenance in overweight or obese people, it was indicated that treatment duration of such interventions significantly leading to weight loss is 6 months [42]. Likewise, the reduction of BMI through Web-based interventions in cardiac patients also needs long-term treatment duration in the future. In addition, we did not find the expected intervention effect on depression. The possible cause might be that the high percentage of cardiac rehabilitation patients (68%, 56/82) who reported their depression score at the start of the study was equal or higher than 16, which reflects the critical mental status of many patients. To reduce depression levels in cardiac patients, long-term and combined provision of guidance, information, stress management, and relaxation skills incorporated into the intervention would need to be developed in the future. This suggestion is also supported by another intervention study in rehabilitation patients with myocardial infarction [43]. Overall, hypothesis 3 was confirmed with one-third of the tested variables.

The hypothesized changes in internal and external resources of behavior change (combination of intention, self-efficacy, and planning; social support) were found in the majority of the tested variables in the mediation analysis. In addition, based on the sequential intervention mode in this study (first 4 weeks for PA, followed by latter 4 weeks for FVC), the gateway effect of the PA mechanism on nutrition was confirmed with the increasing close relationship from PA to nutrition on internal ($R^2$from .44 to .57) and external ($R^2$from .25 to .28) resources, which was also found before [23,44]. However, internal resource for FVC was the only facilitator of the intervention effectiveness. In particular, in the mediation model, it was found that patients in the IG who gained more internal resources on nutrition (eg, increased intention, higher self-efficacy, and more specific planning; Figure 2, panel E) were more likely to adopt or maintain a healthy lifestyle (Figure 4). Thus, it is important to empower the motivation to change, as well as the self-regulation ability of cardiac rehabilitation patients regarding nutrition behavior. This can boost patients to eat healthier and to maintain or adopt PA. Referring back to hypothesis 4, the data support the assumption of a mediator. However, only the internal resource of FVC was revealed to be a mediator but no internal resources of PA or external resource of PA and FVC. Thus, hypothesis 4 was only partially supported.

Limitations

This study is subject to some limitations. First, as it was a small study, it provides only a limited sample size and power. Thus, the statistical analysis of chi-square could not be computed due to the small cell distributions. Second, all variables used in the study were self-report measures. Therefore, the inclusion of objective measures such as accelerometers, wearable cameras, or height and weight measurements is desirable in future studies. Third, physical health outcomes in the study only included BMI, which is too simple to reflect the comprehensive health situation of cardiac patients. More clinical health indicators such as blood pressure, total cholesterol, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, triglycerides, peak VO_2, VO_2@AT, and HR@AT should be added in the future. Fourth, only short-term effects were investigated. Long-term effects should be addressed in depth in the future with such an eHealth intervention. Fifth, patients received monetary incentives in exchange for their participation in the study, which might lead to a bias when evaluating the treatment effects. Moreover, the study could only include patients with access to a computer and the internet; therefore, the results cannot be generalized to patients without such access. Finally, although previously tested interventions in the Western Hemisphere [15,20] were tested with this study in the Eastern Hemisphere, no direct comparative study was conducted. It can only be concluded in this study that the intervention was as effective in Eastern Hemisphere as in Western Hemisphere. Future study should be warranted to test the same intervention in a comparative study to tease out differences in a systematic way.

Conclusions

To conclude, the study demonstrated the potential of a Web-based multiple health behavior intervention among Chinese rehabilitation patients with CHD. The majority of study hypotheses were supported. Such an intervention enriches the extended rehabilitation approaches for cardiac patients to be active and remain healthy in daily life when they are discharged from the hospital. The mechanism of exploration indicates that internal psychological resource (eg, high intention and good self-regulation capability) is a central variable and should especially be considered in cardiac patient health promotion. In addition, gateway or transfer effects, from 1 behavior to another, should be addressed in depth to explore synergetic effects.

Acknowledgments

This research was supported by a junior research group grant from the Wilhelm-Stiftung für Rehabilitationsforschung in Germany and Faculty Research Grant from Hong Kong Baptist University in Hong Kong (FRG1/12-13/064). All authors would like to thank Ms Guilan Wu, the chief nurse of Cardiac Rehabilitation Centre of Guangdong General Hospital, for her active assistance and engagement in recruiting cardiac patients during the research process.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT‐EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 440KB - jmir_v20i11e12052_app1.pdf]

Multimedia Appendix 2

Intervention variables and content.

[PNG File, 247KB - jmir_v20i11e12052_app2.png]

References

1. World Health Organization. World health statistics 2018: monitoring health for the SDGs, sustainable development goals URL: http://www.who.int/gho/publications/world_health_statistics/2018/en/ [accessed 2018-10-28] [WebCite Cache ID 73V5mmK5b]
2. Chen WW, Gao RL, Liu LS, Zhu ML, Wang W, Wang YJ, et al. China Cardiovascular Disease Report 2017 (Summary) [In Chinese]. Chin Circ J 2018 Jan;33(1):1-8 [FREE Full text] [doi: 10.3969/j.issn.1000-3614.2018.01.001]
3. Taylor RS, Brown A, Ebrahim S, Jolliffe J, Noorani H, Rees K, et al. Exercise-based rehabilitation for patients with coronary heart disease: systematic review and meta-analysis of randomized controlled trials. Am J Med 2004 May 15;116(10):682-692. [doi: 10.1016/j.amjmed.2004.01.009] [Medline: 15121495]
4. Williams MA, Ades PA, Hamm LF, Keteyian SJ, LaFontaine TP, Roitman JL, et al. Clinical evidence for a health benefit from cardiac rehabilitation: an update. Am Heart J 2006 Nov;152(5):835-841. [doi: 10.1016/j.ahj.2006.05.015] [Medline: 17070142]
5. Wenger NK. Current status of cardiac rehabilitation. J Am Coll Cardiol 2008 Apr 29;51(17):1619-1631 [FREE Full text] [doi: 10.1016/j.jacc.2008.01.030] [Medline: 18436113]
6. Russell KL, Bray SR. Self-determined motivation predicts independent, home-based exercise following cardiac rehabilitation. Rehabil Psychol 2009 May;54(2):150-156. [doi: 10.1037/a0015595] [Medline: 19469604]
7. Lachat C, Otchere S, Roberfroid D, Abdulai A, Seret FM, Milesevic J, et al. Diet and physical activity for the prevention of noncommunicable diseases in low- and middle-income countries: a systematic policy review. PLoS Med 2013;10(6):e1001465 [FREE Full text] [doi: 10.1371/journal.pmed.1001465] [Medline: 23776415]
8. Schwarzer R. Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. Appl Psychol 2008 Jan;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
9. Fleig L, Lippke S, Pomp S, Schwarzer R. Intervention effects of exercise self-regulation on physical exercise and eating fruits and vegetables: a longitudinal study in orthopedic and cardiac rehabilitation. Prev Med 2011 Sep;53(3):182-187. [doi: 10.1016/j.ypmed.2011.06.019] [Medline: 21784096]
10. Kordy H, Theis F, Wolf M. [Modern information and communication technology in medical rehabilitation. Enhanced sustainability through Internet-delivered aftercare]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2011 Apr;54(4):458-464. [doi: 10.1007/s00103-011-1248-7] [Medline: 21465402]
11. Liu SQ. Constructionanalysis of application effect of a new network platform in the high quality nursing system for CHD patients after PCI. J Qi Lu Nurs 2016 May;22(6):28-30 [FREE Full text] [doi: 10.3969/j.issn.1006-7256.2016.06.012]
12. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
13. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]
14. Reinwand DA, Crutzen R, Storm V, Wienert J, Kuhlmann T, de Vries H, et al. Generating and predicting high quality action plans to facilitate physical activity and fruit and vegetable consumption: results from an experimental arm of a

randomised controlled trial. BMC Public Health 2016 Apr 12;16:317 [FREE Full text] [doi: 10.1186/s12889-016-2975-3] [Medline: 27066779]
15. Storm V, Dörenkämper J, Reinwand DA, Wienert J, De Vries H, Lippke S. Effectiveness of a web-based computer-tailored multiple-lifestyle intervention for people interested in reducing their cardiovascular risk: a randomized controlled trial. J Med Internet Res 2016 Apr 11;18(4):e78 [FREE Full text] [doi: 10.2196/jmir.5147] [Medline: 27068880]
16. Widmer RJ, Allison TG, Keane B, Dallas A, Lerman LO, Lerman A. Using an online, personalized program reduces cardiovascular risk factor profiles in a motivated, adherent population of participants. Am Heart J 2014 Jan;167(1):93-100. [doi: 10.1016/j.ahj.2013.09.019] [Medline: 24332147]
17. Li L, Shao XP. Research progress of cardiac rehabilitation in patients with coronary heart disease. Nurs Integr Tradit Chin West Med 2017;3(7):32-35 [FREE Full text] [doi: 10.11997/nitcwm.201707011]
18. Wu Y, Chen MX. Research progress on web-based nursing intervention. Chin Nurs Res 2016 May;30(5):1552-1553 [FREE Full text] [doi: 10.3969/j.issn.1009-6493.2016.013.005]
19. Schwarzer R, Lippke S, Luszczynska A. Mechanisms of health behavior change in persons with chronic illness or disability: the Health Action Process Approach (HAPA). Rehabil Psychol 2011 Aug;56(3):161-170. [doi: 10.1037/a0024509] [Medline: 21767036]
20. Lippke S, Fleig L, Wiedemann AU, Schwarzer R. A computerized lifestyle application to promote multiple health behaviors at the workplace: testing its behavioral and psychological effects. J Med Internet Res 2015;17(10):e225 [FREE Full text] [doi: 10.2196/jmir.4486] [Medline: 26429115]
21. Tucker M, Reicks M. Exercise as a gateway behavior for healthful eating among older adults: an exploratory study. J Nutr Educ Behav 2002;34(Suppl 1):S14-S19. [doi: 10.1016/S1499-4046(06)60306-0] [Medline: 12047824]
22. Carraça EV, Silva MN, Coutinho SR, Vieira PN, Minderico CS, Sardinha LB, et al. The association between physical activity and eating self-regulation in overweight and obese women. Obes Facts 2013;6(6):493-506 [FREE Full text] [doi: 10.1159/000356449] [Medline: 24217426]
23. Schulz DN, Kremers SPJ, De Vries H. Tailored eHealth lifestyle promotion: which behavioral modules do users prefer? J Health Commun 2015;20(6):663-672. [doi: 10.1080/10810730.2015.1012243] [Medline: 25928161]
24. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
25. Duan YP, Wienert J, Hu C, Si GY, Lippke S. Web-based intervention for physical activity and fruit and vegetable intake among Chinese university students: a randomized controlled trial. J Med Internet Res 2017 Apr 10;19(4):e106 [FREE Full text] [doi: 10.2196/jmir.7152] [Medline: 28396306]
26. Macfarlane DJ, Lee CC, Ho EY, Chan KL, Chan DT. Reliability and validity of the Chinese version of IPAQ (short, last 7 days). J Sci Med Sport 2007 Feb;10(1):45-51. [doi: 10.1016/j.jsams.2006.05.003] [Medline: 16807105]
27. Rafferty AP, Anderson JV, McGee HB, Miller CE. A healthy diet indicator: quantifying compliance with the dietary guidelines using the BRFSS. Prev Med 2002 Jul;35(1):9-15. [doi: 10.1006/pmed.2002.1056] [Medline: 12079436]
28. LeFevre ML, US Preventive Services Task Force. Behavioral counseling to promote a healthful diet and physical activity for cardiovascular disease prevention in adults with cardiovascular risk factors: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med 2014 Oct 21;161(8):587-593. [doi: 10.7326/M14-1796] [Medline: 25155419]
29. Aune D, Giovannucci E, Boffetta P, Fadnes LT, Keum N, Norat T, et al. Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality-a systematic review and dose-response meta-analysis of prospective studies. Int J Epidemiol 2017 Dec 01;46(3):1029-1056 [FREE Full text] [doi: 10.1093/ije/dyw319] [Medline: 28338764]
30. Lippke S, Ziegelmann JP, Schwarzer R, Velicer WF. Validity of stage assessment in the adoption and maintenance of physical activity and fruit and vegetable consumption. Health Psychol 2009 Mar;28(2):183-193 [FREE Full text] [doi: 10.1037/a0012983] [Medline: 19290710]
31. Luszczynska A, Sutton S. Physical activity after cardiac rehabilitation: evidence that different types of self-efficacy are important in maintainers and relapsers. Rehabil Psychol 2006;51(4):314-321. [doi: 10.1037/0090-5550.51.4.314]
32. Luszczynska A, Tryburcy M, Schwarzer R. Improving fruit and vegetable consumption: a self-efficacy intervention compared with a combined self-efficacy and planning intervention. Health Educ Res 2007 Oct;22(5):630-638. [doi: 10.1093/her/cyl133] [Medline: 17060349]
33. Lippke S, Schwarzer R, Ziegelmann JP, Scholz U, Schüz B. Testing stage-specific effects of a stage-matched intervention: a randomized controlled trial targeting physical exercise and its predictors. Health Educ Behav 2010 Aug;37(4):533-546. [doi: 10.1177/1090198109359386] [Medline: 20547760]
34. Jackson J, Lippke S, Gray C. Stage-specific prediction of physical activity in orthopaedic patients after rehabilitation treatment. Int J Sport Psychol 2011;42(6):586-609 [FREE Full text]
35. Yao G, Wu C. Similarities and differences among the Taiwan, China, and Hong-Kong versions of the WHOQOL Questionnaire. Soc Indic Res 2008 Sep 30;91(1):79-98. [doi: 10.1007/s11205-008-9326-4]
36. Rankin SH, Galbraith ME, Johnson S. Reliability and validity data for a Chinese translation of the Center for Epidemiological Studies-Depression. Psychol Rep 1993 Dec;73(3 Pt 2):1291-1298. [doi: 10.2466/pr0.1993.73.3f.1291] [Medline: 8115582]

37. Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav Res Methods 2008 Aug;40(3):879-891. [doi: 10.3758/BRM.40.3.879] [Medline: 18697684]
38. Schroeder MA. Diagnosing and dealing with multicollinearity. West J Nurs Res 1990 Apr;12(2):175-184. [doi: 10.1177/019394599001200204] [Medline: 2321373]
39. Schafer JL, Olsen MK. Multiple imputation for multivariate missing-data problems: a data analyst's perspective. Multivariate Behav Res 1998 Oct 01;33(4):545-571. [doi: 10.1207/s15327906mbr3304_5] [Medline: 26753828]
40. Avram C, Iurciuc M, Craciun L, Avram A, Iurciuc S, Oancea C, et al. Dietary and physical activity counseling in high-risk asymptomatic patients with metabolic syndrome?A primary care intervention. J Food Agric Environ 2011 Jan 1;9(3&4):16-19 [FREE Full text]
41. Gomez-Huelgas R, Jansen-Chaparro S, Baca-Osorio AJ, Mancera-Romero J, Tinahones FJ, Bernal-López MR. Effects of a long-term lifestyle intervention program with Mediterranean diet and exercise for the management of patients with metabolic syndrome in a primary care setting. Eur J Intern Med 2015 Jun;26(5):317-323. [doi: 10.1016/j.ejim.2015.04.007] [Medline: 25907985]
42. Wieland LS, Falzon L, Sciamanna CN, Trudeau KJ, Brodney S, Schwartz JE, et al. Interactive computer-based interventions for weight loss or weight maintenance in overweight or obese people. Cochrane Database Syst Rev 2012;8:CD007675 [FREE Full text] [doi: 10.1002/14651858.CD007675.pub2] [Medline: 22895964]
43. Wang W, Chair SY, Thompson DR, Twinn SF. Effects of home-based rehabilitation on health-related quality of life and psychological status in Chinese patients recovering from acute myocardial infarction. Heart Lung 2012;41(1):15-25. [doi: 10.1016/j.hrtlng.2011.05.005] [Medline: 21974926]
44. Fleig L, Kerschreiter R, Schwarzer R, Pomp S, Lippke S. 'Sticking to a healthy diet is easier for me when I exercise regularly': cognitive transfer between physical exercise and healthy nutrition. Psychol Health 2014 Jul;29(12):1361-1372. [doi: 10.1080/08870446.2014.930146] [Medline: 24894668]
45. Schulz DN, Kremers SP, De Vries H. Tailored eHealth lifestyle promotion: which behavioral modules do users prefer? J Health Commun 2015;20(6):663-672. [doi: 10.1080/10810730.2015.1012243] [Medline: 25928161]

Abbreviations

ANOVA: analysis of variance
ANCOVA: analyses of covariance
BMI: body mass index
CHD: coronary heart disease
CVD: cardiovascular disease
eHealth: electronic health
FVC: fruit and vegetable consumption
HAPA: Health Action Process Approach
IG: intervention group
PA: physical activity
OR: odds ratio
VIF: variance inflation factor
WCG: waiting control group

Edited by G Eysenbach; submitted 28.08.18; peer-reviewed by R Schwarzer, R Wurst; comments to author 08.10.18; revised version received 20.10.18; accepted 20.10.18; published 19.11.18.

Please cite as:
Duan YP, Liang W, Guo L, Wienert J, Si GY, Lippke S
Evaluation of a Web-Based Intervention for Multiple Health Behavior Changes in Patients With Coronary Heart Disease in Home-Based Rehabilitation: Pilot Randomized Controlled Trial
J Med Internet Res 2018;20(11):e12052
URL: http://www.jmir.org/2018/11/e12052/
doi:10.2196/12052
PMID:30455167

Original Paper

A Web-Based Telemanagement System for Improving Disease Activity and Quality of Life in Patients With Complex Inflammatory Bowel Disease: Pilot Randomized Controlled Trial

Javier Del Hoyo[1*], MD; Pilar Nos[1,2,3], MD, PhD; Raquel Faubel[4,5], PhD; Diana Muñoz[1], RN; David Domínguez[6], CS; Guillermo Bastida[1,2,3], MD, PhD; Bernardo Valdivieso[5,7,8], MD, PhD; Marisa Correcher[9], PhD; Mariam Aguas[1,2,3*], MD, PhD

[1]Gastroenterology Department, La Fe University and Polytechnic Hospital, Valencia, Spain

[2]CIBEREHD (Networked Biomedical Research Center for Hepatic and Digestive Diseases), Valencia, Spain

[3]Health Research Institute La Fe, Valencia, Spain

[4]Physiotherapy Department, University of Valencia, Valencia, Spain

[5]Joint Research Unit in Biomedical Engineering (eRPSS: IIS La Fe-UPV), Valencia, Spain

[6]Connected Health Services, Valencia, Spain

[7]Home Care and Telemedicine Department, La Fe University and Polytechnic Hospital, Valencia, Spain

[8]Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC), Madrid, Spain

[9]Systems Department, La Fe University and Polytechnic Hospital, Valencia, Spain

[*]these authors contributed equally

Corresponding Author:
Mariam Aguas, MD, PhD
Gastroenterology Department
La Fe University and Polytechnic Hospital
Av Fernando Abril Martorell 106
Valencia, 46026
Spain
Phone: 34 600859948
Fax: 34 9612462
Email: m.aguas.peris@gmail.com

Abstract

Background: The reported efficacy of telemedicine in patients with inflammatory bowel disease (IBD) is inconsistent among studies, and data for complex IBD are lacking.

Objective: We aimed to evaluate the impact of remote monitoring using a Web system—*Telemonitorización de la Enfermedad de Crohn y Colitis Ulcerosa* or Telemonitoring of Crohn's Disease and Ulcerative Colitis (TECCU)—as compared to standard care and telephone care on health outcomes and health care in patients with complex IBD.

Methods: We performed a 3-arm randomized controlled trial. Adult patients with IBD who received immunosuppressants and biological agents were recruited from the IBD Unit of a tertiary university hospital. The patients were randomized into groups to receive remote monitoring (G_TECCU), nurse-assisted telephone care (G_NT), or standard care with in-person visits (G_control). All patients completed the study visits at baseline and at 12 and 24 weeks in addition to each type of intervention. The primary outcome was the percentage of patients in remission at 24 weeks. Secondary health outcomes were quality of life, medication adherence, adverse effects, satisfaction, and social activities. Data on the number of outpatient visits and telephone calls, emergency visits, hospitalizations, IBD-related surgeries, and corticosteroid courses were also collected.

Results: A total of 63 patients were selected (21 patients in each group). During the study, 90.5% (19/21) of patients in G_control, 95.2% (20/21) in G_NT, and 85.7% (18/21) in G_TECCU were compliant to the intervention. After 24 weeks, the percentage of patients in remission was higher in G_TECCU (17/21, 81%) than in G_NT (14/21, 66.7%) and G_control (15/21, 71.4%). A higher improvement in disease activity was observed in G_TECCU than in G_control in terms of the Harvey-Bradshaw/Mayo (odds ratio=0.12, 95% CI=0.003-2.162, *P*=.19) and Harvey-Bradshaw/Walmsley (odds ratio=0.11, 95% CI=0.004-1.55, *P*=.13) indexes. Improvement in disease activity was associated with a larger reduction in fecal calprotectin values in G_TECCU compared

to G_control (estimated intervention effect: odds ratio=–0.90; 95% CI=–1.96 to 0.16, *P*=.11). All completers adhered to treatment in G_TECCU. In addition, the quality of life, social activities, and satisfaction improved in all 3 groups. Although the number of outpatient visits and telephone calls was lower in G_TECCU than in G_NT and G_control, the safety profile was similar in all 3 groups.

Conclusions: This pilot clinical trial suggests that the TECCU Web-based system is a safe strategy for improving health outcomes in patients with complex IBD and reducing the use of health care resources.

Trial Registration: ClinicalTrials.gov NCT02943538; https://clinicaltrials.gov/ct2/show/NCT02943538 (Archived by WebCite at http://www.webcitation.org/746CRRtDN).

(J Med Internet Res 2018;20(11):e11602) doi:10.2196/11602

KEYWORDS

Crohn disease; e-health; inflammatory bowel disease; information and communication technology; telemedicine; ulcerative colitis

Introduction

Inflammatory bowel disease (IBD), comprising ulcerative colitis (UC) and Crohn disease (CD), is a chronic relapsing disorder characterized by inflammation of the gastrointestinal tract. Its natural progression includes flares and periods of remission, and IBD requires continuous and personalized follow-up to ensure long-term remission. Patients with IBD use health care resources significantly more often than patients with other conditions [1]. Importantly, 30%-45% of adult patients with IBD are nonadherent to treatment [2], which increases the probability of relapse by 4-fold and consequently increases the health care costs [3]. The high percentage of nonadherence is associated with behavioral and psychological factors and the physician-patient relationship [2]. In addition, IBD is related to high levels of school absenteeism and work disability [4], interference with social activities, and impairment of health-related quality of life (HRQoL) [5]. Therefore, IBD has a significant medical, social, and financial impact.

To address these difficulties, telemedicine uses information and communication technologies (ICTs) to provide health care services remotely. Despite the heterogeneous nature and progress of IBD, ICTs allow tailored follow-up with better communication between physicians and patients [6-8] and provide educational resources that promote patient empowerment [6] and optimize treatment [9,10]. ICTs were initially used to support the management of other chronic diseases such as congestive heart failure [11], diabetes mellitus [12], and chronic obstructive pulmonary disease [13] and showed excellent acceptance by patients and improvement in HRQoL. Owing to these positive results, telemedicine systems have been evaluated in patients with IBD [14,15].

A ground-breaking evaluation of telemedicine in patients with IBD by Cross et al. [16], who designed the Home Automated Telemanagement system, was acceptable, increased patients' knowledge about their disease, and facilitated greater self-control of IBD symptoms [17]. However, in a randomized controlled trial of 47 patients with UC in remission or with mild UC activity, no significant differences were noted in disease activity, HRQoL, or medication adherence between patients who followed the telemedicine system and those who received usual care at 12 months [18]. These results were partially explained by a higher discontinuation rate in the Home Automated Telemanagement group due to the platform design. To avoid similar problems, the authors subsequently designed a mobile phone-based Web program to determine the impact of telemedicine on health outcomes [19]. Web programs and mHealth interventions are increasingly used to provide information and health care at a distance. Unlike other telehealth systems, Web-based apps do not require home installation and are less expensive. Web-based systems adapted to IBD are safe and feasible for adults and adolescents [20-22], improve medication adherence, and reduce the duration of relapses [6]. A recent multicenter, randomized controlled trial including more than 900 patients with different subtypes of IBD showed that compared with standard care, empowerment through these systems reduces outpatient visits and hospital admissions, with potential cost savings [8]. Web-based systems are currently the most useful platforms, and their application allows development of distance care projects aimed at providing access to health care in remote areas [23,24]. These models promote collaboration and knowledge exchange between specialists, potentially reducing variability in clinical practice, and could modify the future structure of health systems if they prove to be cost-effective. However, the effect of telemedicine systems on disease outcomes is inconsistent and varies across studies according to the population and health care system in which they are applied [6,8,18,20,22]. Moreover, no specific studies have thus far focused on the impact of telemedicine in patients with complex IBD.

Our research group designed a Web-based telemanagement system known as *Telemonitorización de la Enfermedad de Crohn y Colitis Ulcerosa* or Telemonitoring of Crohn's Disease and Ulcerative Colitis (TECCU) for remote monitoring of patients with moderate-to-severe IBD activity or who initiate treatment with corticosteroids, immunosuppressants, and biological agents. We performed a randomized controlled trial to assess the safety of care provided through this system and its impact on disease activity, HRQoL, medication adherence, satisfaction, work productivity, social activities, and use of health care resources and compared these outcomes with those of nurse-assisted telephone care and standard face-to-face visits.

Methods

Study Design

A 3-arm, parallel-group, randomized controlled trial was performed to compare the impact of the Web-based telemanagement system TECCU, nurse-assisted telephone care, and standard face-to-face visits on health outcomes and outpatient visits in patients with complex IBD. Neither the patients nor the researchers were masked to the intervention, but the results were analyzed by an independent statistician who was blinded to group identification.

Ethical Considerations

The study protocol was reviewed and approved by the local independent ethics committee of La Fe University and Polytechnic Hospital, Valencia, Spain; the regional independent ethics committee (Comité Ético Autonómico de Estudios Clínicos de Medicamentos y Productos Sanitarios de la Comunitat Valenciana); and the Spanish Agency of Medicines and Medical Devices (Agencia Española de Medicamentos y Productos Sanitarios). According to the physicians involved in the study, the risks did not outweigh the potential benefits, and each participant provided informed consent without coercion before inclusion in the study. The trial is registered at ClinicalTrials.gov with the identifier NCT02943538.

Patient Selection

Patients were recruited from the IBD Unit of a tertiary university hospital. All participants were diagnosed with IBD at least 6 months previously and according to the internationally accepted criteria [25,26].

Inclusion and Exclusion Criteria

The inclusion criteria were age ≥ 18 years; initiation of therapy with corticosteroids, immunosuppressants, and biological agents due to disease activity; and provision of written informed consent to participate in the study. The exclusion criteria were inability to speak and read Spanish; inability to manage a mobile phone or tablet or the Internet or not having a telephone line; participation in other clinical trials during the inclusion period; uncontrolled medical or psychiatric disease; presence of ileorectal or ileal pouch-anal anastomosis; receipt of definitive ileostomy; perianal disease; and pregnancy.

Recruitment

This project was funded in 2012, and the platform was installed and configured in 2013-2014. Enrollment was started in October 2014 and ended in June 2016. The follow-up ended in December 2016. Patients were included consecutively from the Outpatient Clinic of the IBD Unit or the Gastroenterology Department if they were admitted for a flareup of IBD. During the visit, the inclusion and exclusion criteria were verified, and the patient was informed about the study through the patient-information sheet. If the patient agreed to participate, he or she signed the informed consent document.

Randomization

Eligible patients were randomized to 1 of the 3 groups to receive remote monitoring (G_TECCU), nurse-assisted telephone care (G_NT), or standard care with in-person visits (G_control) in a 1:1:1 ratio by using a block randomization method through a Web-based tool [27] in order to generate a random-allocation sequence and ensure allocation concealment. Once a number was assigned, it could not be reassigned, and members of the research team who were in contact with patients did not have access to the randomization tables or lists.

Study Outcomes

Participants completed study visits at the baseline, 12 weeks, and 24 weeks, in addition to routine visits scheduled for their clinical care. The variables measured at baseline were sociodemographic variables, disease profile and activity, HRQoL, adverse events, medication adherence, and patient satisfaction.

Primary Outcome

The primary objective of the study was to determine the percentage of patients in clinical remission at 24 weeks. Remission was evaluated using the modified Harvey-Bradshaw index (HBI) for patients with CD [28]. For patients with UC, we used the Simple Clinical Colitis Activity Index (SCCAI, also known as the Walmsley index) [29] for remote checkups together with the partial Mayo score for face-to-face visits [30].

Patients with CD and an HBI < 5 were considered to be in clinical remission, whereas patients with scores of 5-7, 8-16, or >16 were considered to have mild, moderate, or severe activity, respectively [28]. For remote checkups in patients with UC, clinical remission was defined as a Walmsley score ≤ 2, whereas mild-to-moderate and severe activities were defined as scores of 3-5 and >5, respectively [31]. In the face-to-face visits, clinical remission was defined as a partial Mayo score ≤ 2 and no individual Mayo subscore > 1; scores of 2-5, 6-8, and 9 were defined as mild, moderate, and severe disease activity, respectively [30].

Laboratory parameters were measured at baseline and each subsequent visit according to individual patient's schedule and included a complete blood analysis with nutritional profile and C-reactive protein level (mg/L). Fecal calprotectin (FC, µg/g) was assessed at baseline, 12 weeks, and 24 weeks after inclusion. Changes in the medication were made on the basis of the results obtained, in association with clinical activity indexes, and were prescribed according to specific intervention plans adapted to the severity of each alert.

Secondary Outcomes

The HRQoL of patients at inclusion and at week 24 was evaluated using the specific questionnaire Inflammatory Bowel Disease Questionnaire 9 (IBDQ-9) and the generic questionnaire EuroQol-5D (EQ-5D). IBDQ-9 is a validated questionnaire comprising 9 items distributed in 4 dimensions: bowel symptoms, systemic symptoms, emotional function, and social function. It correlated very well with the Spanish version of the 36-item IBDQ [32]. Each item is scored on a 7-point Likert scale, with an overall score ranging from 7 (lowest QoL) to 63 (highest QoL) points, and calculated over 100 points. EQ-5D is a generic questionnaire that has been used for patients with various chronic diseases such as IBD and was previously

validated in Spain [33]. This instrument provides a global value for HRQoL, with 5 questions related to 5 dimensions: mobility, selfcare, usual activities, pain or discomfort, and anxiety or depression. Additionally, a visual analog scale (VAS) was used. Each question was scored with 1, 2, or 3 points, depending on whether impairment in the dimension assessed was nonexistent, moderate, or extreme, respectively; subsequently, the results were converted for each dimension according to the specific coefficients calculated for the Spanish population [34]. The maximum score of 1 represents the best health status, a score of 0 represents a health state considered equivalent to death, and negative index scores represent health states considered worse than death. The VAS rates overall health between 0 (poorest imaginable health) and 100 (best imaginable health).

We assessed the impact of disease on work productivity and activities of daily living using the Work Productivity and Activity Impairment (WPAI) questionnaire [35], which patients completed at baseline and week 24. The questionnaire comprises 6 questions on the effect of the disease on work and activities of daily living during the previous 7 days. A higher score on the questionnaire indicates a more pronounced effect on work and daily activities. The Spanish version has been validated and is reproducible in patients with CD.

Medication adherence was evaluated using the Morisky-Green index [36], which has been used to evaluate adherence in patients with IBD in clinical trials [18]. We considered adherence to be adequate when the patient answered all questions correctly and inadequate if any answer was associated with nonadherence.

We recorded the number of outpatient visits and telephone consultations, as registered in the NOMHADCHRONIC app (version V2.RC6; Connected Health Services SL, Valencia, Spain) for patients in G_TECCU and in the Orion Clinic information system (as part of daily clinical practice at our hospital) for all 3 groups during the study. Patients who adhered to >80% of checkups in the study protocol were considered to be compliant.

The safety of each intervention was assessed by measuring the number of visits to the emergency department, hospitalizations, IBD-related surgeries, corticosteroid courses, and adverse effects to medication. Finally, patient satisfaction with the care received was evaluated at 24 weeks by using an adapted version of the Client Satisfaction Questionnaire, which comprises 6 questions (measured on a scale of 0-10) on the quality, usefulness, and viability of the system of care applied for each case [37].

Setting and Interventions

This pilot trial was developed at the tertiary referral center La Fe University and Polytechnic Hospital, Valencia, Spain. The hospital serves more than 1500 patients with IBD, has 2 specialist IBD nurses, and provides an email and telephone consultation structure for IBD patients to contact the hospital.

The trial had 3 arms and compared remote monitoring through a Web-based telemanagement system, nursing care by telephone, and usual care provided in our IBD Unit (Outpatient Clinic). All patients completed the study visits at baseline and 12 and 24 weeks, in addition to routine visits to the IBD clinic, telephone consultations, or Web telemonitoring based on group assignment at randomization. Disease activity, HRQoL, adverse effects, adherence, and use of health care resources were measured at baseline and during the 24-week follow-up.

To adequately control disease activity and adverse effects of immunosuppressants, patients treated with these drugs alone or in combination with biological agents were monitored every 1-2 weeks during the first month, every 2-4 weeks between months 1 to 3, and every 4 weeks from month 3 until the end of follow-up. Patients treated with biological agents alone were monitored every 2-4 weeks during the entire follow-up period. Patients from all 3 arms who took the same type of drug underwent these follow-up schedules; the schedules only differed in terms of the monitoring: patients in G_TECCU were monitored via the NOMHADCHRONIC app, those in G_NT were monitored via a telephone line managed by nursing staff from the IBD Unit, and those in G_control were monitored by the usual face-to-face visits combined with telephone calls made by physicians according to standard clinical practice. Moreover, additional clinical visits were performed, if required, during patient evolution in any of the 3 arms.

Telemonitoring Apps

In G_TECCU, follow-up and monitoring were performed telematically using the integrated platform for management of chronically ill patients (NOMHADCHRONIC app), which was designed to meet the specific needs of this group. The NOMHADCHRONIC platform is an innovative technological system that was designed to boost the rollout of services for the management of chronically ill patients. The telemanagement care system was developed in collaboration with patients, the La Fe telemedicine Unit, and *Tecnologías para la Salud y el Bienestar* SA. The protocol of the Web-based telemanagement system (TECCU) has been described elsewhere [27]. TECCU is a secure webpage with an HTTP app for mobile phones, tablets, and computers. Patients connected to the platform via the Internet using a computer or an app on a mobile phone or tablet had to self-complete questionnaires. In addition, they received advice, reminders, educational material about their disease, and information on prevention. This information was received by the case managers and filtered using an intelligent prioritization system with generation of alerts and push notifications according to an integrated intervention protocol.

Patients monitored via the TECCU used NOMHADhome [38]. The resources of the NOMHADhome platform were also available on the NOMHAD mobile app, which patients could download onto their mobile phones. After inclusion, these patients automatically received an email with a personal code that allowed them to access the platform. Each patient profile contained the following information: contact information, active IBD medications, testing schedule (blood and stool tests), log of disease activity, medication use, body weight, vital signs, HRQoL, alerts and action plans, progress of inflammatory activity and vital signs in the form of graphs, electronic messaging to the study nurse coordinator and health care provider, and educational tips.

Patients in G_TECCU had to answer several simple questions related to their IBD symptoms and possible adverse effects since

the last evaluation via text messaging. In the main menu of the platform, patients accessed the questionnaires by clicking on specific icons designed for this purpose and could answer the multiple choice and yes or no questions needed to complete these items (Figures 1 and 2). We established individualized alerts and action plans based on the answers to questions about the activity index, adverse effects, and blood biochemistry results. A scale of values was assigned for each alert depending on the severity (green, yellow, orange, and red zone).

After receiving an alert, the specialized medical staff, in collaboration with the nurses, used the general recommendations of the action plans to guide medication adjustments in biological therapy, immunomodulators, corticosteroids, and salicilates. These treatment changes were performed with the support of the platform messaging system in combination with telephone calls or in-person visits when it was necessary to train patients for the administration of medications such as subcutaneous biological agents. Once the disease was in remission again (green zone), the patient had to continue with the initially programmed follow-up. All patients had access to tools that informed them about their disease. In G_TECCU, information was available on the NOMHADhome platform itself; in G_control and G_NT, patients received written documents with the same information.

Figure 1. Homepage of the NOMHADhome platform for access to questionnaires (patient version).

Figure 2. Home page of the NOMHADmobile platform for access to vital signs (patient version).

Usual Care Provided in the Inflammatory Bowel Disease Unit

The G_control patients received the normal care provided in the IBD Unit (Outpatient Clinic) for patients with moderately to highly complex IBD, based on national and European clinical guidelines [25,26,39]. Treatment was adjusted according to the evolution of disease activity and medication adherence, which was measured using specific indexes and biological markers used to report the study outcomes during office visits or telephone calls. In addition, to measure the time in remission, clinical activity of patients was self-recorded in a diary on paper at home weekly during the first 12 weeks and every 2 weeks subsequently, until the end of the follow-up.

This care was complemented by ad hoc hospital care in case of flareups or if the patient's health deteriorated for any reason. Ad hoc intensive care was maintained until the patient's condition stabilized, at which point he or she returned to follow-up based on standard care in the Unit. For comparability among the 3 groups, we provided information on all available on-paper educational materials about IBD for remotely monitored patients as well as information on prevention and written action plans in case of flareups. Questionnaires on HRQoL, satisfaction, and work productivity were recorded at baseline and at the end of the study.

Nursing Care by Telephone

The G_NT patients were asked about their health through telephone calls by the nursing staff in the IBD Unit. We performed telephone assessment periodically by using structured interviews to evaluate health status, and clinical activity was self-recorded at home, as described for the G_control patients. The interventions depended on the results of the interview and changes in the medication or follow-up schedule established by nurses with the support of medical staff, according to the alerts and action plans designed in the intervention protocol [27]. Furthermore, we provided these patients with all educational elements made available to the other 2 groups.

Notification of Adverse Effects

TECCU is a minimum-risk system. The characteristics of the interventions in the present study are designed such that patients do not experience direct adverse effects. The interventions were proposed to control and achieve long remission periods in patients with IBD and were not expected to cause lesions or damage to the patients' health, since we did not test a new experimental drug. We evaluated a Web-based telemonitoring system designed to improve communication between patients and health care providers, with a proactive remote follow-up method, allowing changes in the patients' treatment and follow-up schedule when flareups are detected or when the flareups increase in severity as per the alert system. In addition, we incorporated educational elements in G_TECCU and the other 2 groups to make them more comparable.

With respect to the platform, a risk-minimization study was performed to ensure accomplishment of ethical norms and that the appropriate ethics committees approved the study. We also recorded all adverse events from the time the patient provided consent to participate in the study until 28 days after study completion.

Power Calculation

The most-efficient means of determining differences between the 3 groups was by contrasting differences in activity indexes for CD and UC. Considering the differences in the scales, the analysis was stratified by performing a comparison of patients with UC and another comparison of patients with CD. The sample size was calculated by estimating the size required to detect a difference of 3 points in the HBI; considering an SD of 4, a power of 80%, and an alpha significance level of .05, a total of 30 patients with CD (10 per arm) was required. In addition, for a 2-point difference in the Mayo index, an SD of 2.5, a power of 80%, and an alpha significance level of .05, a total of 30 patients with UC (10 per arm) was required. Therefore, the overall sample size was 60 patients (20 per arm). We also stratified patients globally (CD and UC) by comparing

patients in remission or with inflammatory activity, irrespective of disease severity (mild, moderate, or severe).

Statistical Analysis

First, we described the characteristics of patients in the test and control groups by applying appropriate estimators (means, medians, or proportions) according to the type of variables and evaluated possible differences between the groups in the main and secondary outcomes by using regression models to assess differences in these measures. We evaluated the effects of the different treatment groups on the probability of remission using mixed-effects logistic regression models including an interaction between week and treatment groups. Individuals were included in the model as a random intercept to correct for the nonindependent data. Differences in slopes for weeks 12 and 24 were evaluated by assessing the interaction between week and treatment groups. *P* values were estimated using the Satterthwaite approximation for degrees of freedom. We used mixed-effects ordinal logistic regression models to evaluate progress according to the IBDQ-9 and EQ-5D scores in the 3 study groups between weeks 0 and 24. However, owing to convergence problems in assessing medication adherence (because of the reduced sample size and the ordinal character of the Morisky-Green test), Bayesian techniques were applied in the ordinal logistic regression model. Differences in the percentage of missed work hours, work impairment, and social impairment were analyzed using mixed-effects linear regression and mixed-effects beta regression models (% work hours missed). Profile likelihood 95% confidence intervals were calculated for all estimations, and $P<.05$ was considered statistically significant. A likelihood-based analysis for all patients included was performed using R (version 3.5.1) with the lme4 (version 1.1-17), lmerTest (version 3.0-1), brms (version 2.2.0), clickR (version 0.4.04), ordinal (version 2018.4-19), and glmmADMB (0.8.3.3) packages [40].

Results

Study Sample

A total of 68 patients with complex IBD were invited to participate in this study between October 2014 and June 2016, of which 3 (4.4%) declined to participate owing to inaccessibility to the Internet at home and 2 (2.9%) did not meet the inclusion criteria. The remaining 63 eligible patients provided informed consent and were randomly assigned to the 3 groups (21 patients in each group, Figure 3). During the study period, all patients except 1 patient in G_TECCU (95.2%) continued to use the Web-based telemanagement system and 2 other patients did not respond to >80% of checkups. The remaining 18 patients (85.7%) in G_TECCU showed good adherence to the study protocol as compared to 19 patients (90.5%) in G_control and 20 patients (95.2%) in G_NT. We did not observe any differences between the completers and patients who did not adhere to the study protocol. The baseline characteristics of the 3 study groups are shown in Table 1.

Disease Activity

According to the HBI and the partial Mayo scores, 47.6% (10/21), 38.1% (8/21), and 57.1% (12/21) of patients were in clinical remission at baseline in G_TECCU, G_NT, and G_control, respectively. The baseline percentage of remission according to the Walmsley score for UC was the same for each group and improved progressively at 12 and 24 weeks in all 3 groups. This percentage was higher in G_TECCU, even after considering dropouts, with 81% of patients (17/21) inactive after 24 weeks compared with 66.7% (14/21) in G_NT and 71.4% (15/21) in G_control (using both Mayo and Walmsley scores for evaluation of UC); therefore, the percentage of patients in remission increased by 33.4% in G_TECCU, 28.6% in G_NT, and 14.3% in G_control (Figure 4). In the intention-to-treat analysis, a higher probability of remission was observed in G_TECCU than in G_control at 12 weeks (odds ratio (OR)=0.12, 95% CI=0.003-2.103, *P*=.18) and 24 weeks (OR=0.12, 95% CI=0.003-2.162, *P*=.19) using the HBI-Mayo scores. With the HBI-Walmsley scores, no differences were observed between G_TECCU and G_control at 12 weeks (OR=0.92, 95% CI=0.08-9.71, *P*=.94), but the probability of remission was higher in G_TECCU at 24 weeks (OR=0.11, 95% CI=0.004-1.55, *P*=.13). The median time in remission was 17.9 weeks (interquartile range, 12-24 weeks) in G_TECCU compared to 17.3 weeks (interquartile range, 10-24 weeks) in G_NT and 14.3 weeks (interquartile range, 8-24 weeks) in G_control.

Disease activity was evaluated on the basis of FC levels throughout the study. At 24 weeks, the median FC level for clinical activity improved progressively from a baseline value of 490 µg/g to 137 µg/g in G_TECCU and from 526 µg/g to 115.5 µg/g in G_NT; however, this reduction was smaller in G_control, from 330 µg/g to 230 µg/g. The improvement in FC levels was larger in G_TECCU than in G_control, with an estimated intervention effect of –0.76 (95% CI=–1.85 to 0.336, *P*=.18) in the reduction of FC values at 12 weeks and –0.90 (95% CI=–1.96 to 0.16, *P*=.11) at 24 weeks. Similarly, a larger improvement was noted in G_NT than in G_control at the same time points, with an estimated intervention effect of –0.61 (95% CI=–1.62 to 0.39, *P*=.25) at 12 weeks and –0.91 (95% CI=–1.96 to 0.15, *P*=.10) at 24 weeks (Figure 5).

Figure 3. Flowchart of study participants. G_CONTROL: group receiving standard care with in-person visits; G_NT: group receiving nurse-assisted telephone care; G_TECCU: group receiving remote monitoring; IBD: inflammatory bowel disease.

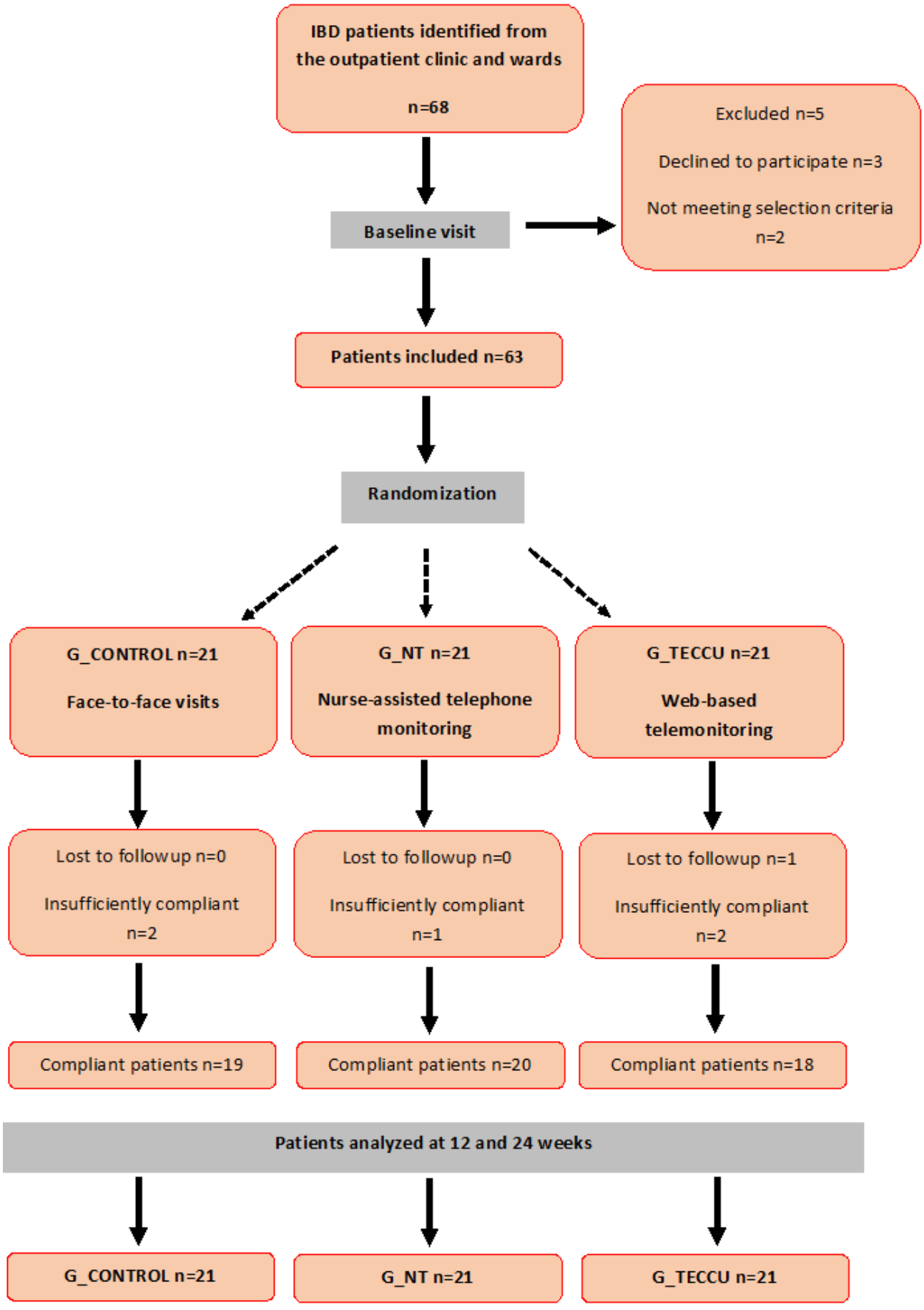


Table 1. Baseline characteristics of patients.

Characteristics	Control group (n=21)	Telephone group (n=21)	TECCU[a] group (n=21)
Median age, years (range)	39.31 (22-61)	40.91 (24-60)	41.32 (19-66)
Sex, n (%)			
Men	12 (57.1%)	12 (57.1%)	9 (42.9%)
Women	9 (42.9%)	9 (42.9%)	12 (57.1%)
Education, n (%)			
Primary	4/21 (19%)	4/21 (19%)	5/21 (23.8%)
Secondary	9/21 (42.9%)	6/21 (28.6%)	6/21 (28.6%)
University	8/21 (38.1%)	11/21 (52.4%)	10/21 (47.6%)
Disease profile			
Crohn disease, n (%)	14/21 (66.7%)	13/21 (61.9%)	13/21 (61.9%)
Ulcerative colitis, n (%)	7/21 (33.3%)	8/21 (38.1%)	8/21 (38.1%)
Median time since diagnosis, months (range)	123.32 (6-427)	108.27 (7-452)	146.72 (7-424)
Treatment (%)			
Immunomodulators	10/21 (47.6%)	10/21 (47.6%)	9/21 (42.9%)
Biological monotherapy	4/21 (19%)	4/21 (19%)	4/21 (19%)
Combination therapy	6/21 (28.6%)	5/21 (23.8%)	6/21 (28.6%)
Corticosteroids	1/21 (4.8%)	2/21 (9.5%)	2/21 (9.5%)
Clinical Remission, n (%)	12/21 (57.1%)	8/21 (38.1%)	10/21 (47.6%)
Crohn disease	10/14 (71.4%)	6/13 (46.2%)	9/13 (69.2%)
Ulcerative colitis	2/7 (28.6%)	2/8 (25%)	1/8 (12.5%)
Median calprotectin level, µg/g (interquartile range)	330 (103-617)	526 (115-1724)	490 (23-2016)
Quality of life			
Median IBDQ-9[b] score (interquartile range)	38.50 (33.25-46.75)	37.50 (28.75-46.25)	42.00 (33.75-47.50)
Median EQ-5D[c] score (interquartile range)	0.816 (0.754-0.914)	0.825 (0.710-0.914)	0.825 (0.576-0.914)
Median VAS[d], % (interquartile range)	60.5% (50%-85%)	62.5% (50%-80%)	60% (40%-90%)
Medication adherence, n (%)	14/21 (66.7%)	7/21 (33,3%)	12/21 (57,1%)
WPAI[e]			
Not working, n (%)	8/21 (38.1%)	7/21 (33.3%)	5/21 (23.8%)
Percentage of work hours missed, median (interquartile range)	27.5% (0%-52%)	40% (15%-62.5%)	32.5% (7.5%-57.5%)
Work impairment score, median (interquartile range)	7 (2.75-10)	7 (3-10)	10 (2.25-10)
Social impairment score, median (interquartile range)	3.5 (1-5.75)	3.5 (2-7)	6 (2.75-8)
Satisfaction score, median (interquartile range)	49.5 (42.5-53.75)	53 (50-59)	52 (47.5-55)

[a]TECCU: Telemonitoring of Crohn's Disease and Ulcerative Colitis.

[b]IBDQ-9: Inflammatory Bowel Disease Questionnaire 9.

[c]EQ-5D: EuroQol-5D.

[d]VAS: visual analog scale.

[e]eWPAI: Work Productivity and Activity Impairment.

Figure 4. Evolution of disease activity over the study period in the 3 groups. G_CONTROL: group receiving standard care with in-person visits; G_NT: group receiving nurse-assisted telephone care; G_TECCU: group receiving remote monitoring; HBI: Harvey-Bradshaw index.

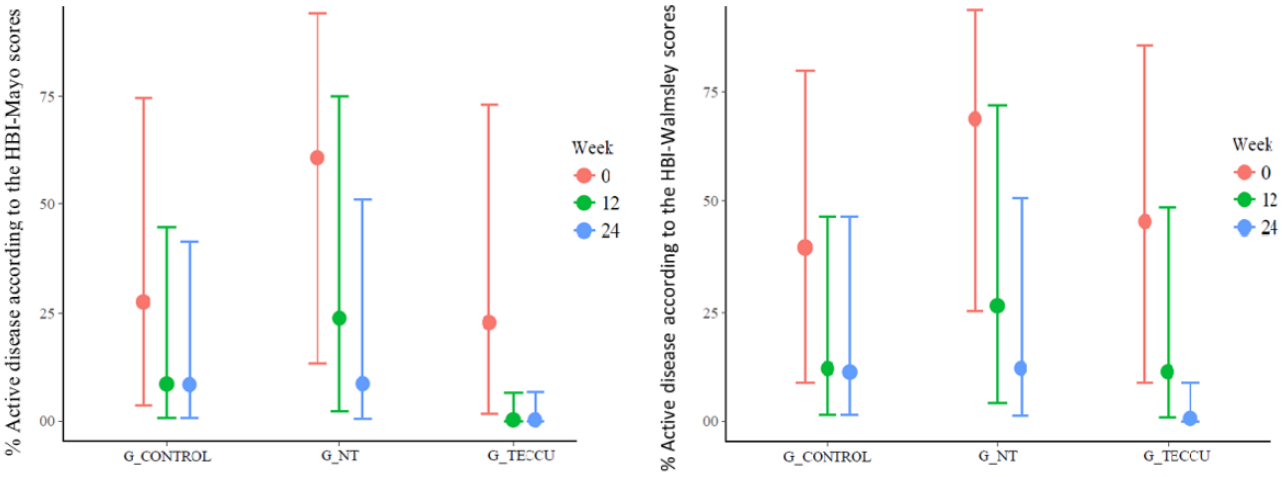
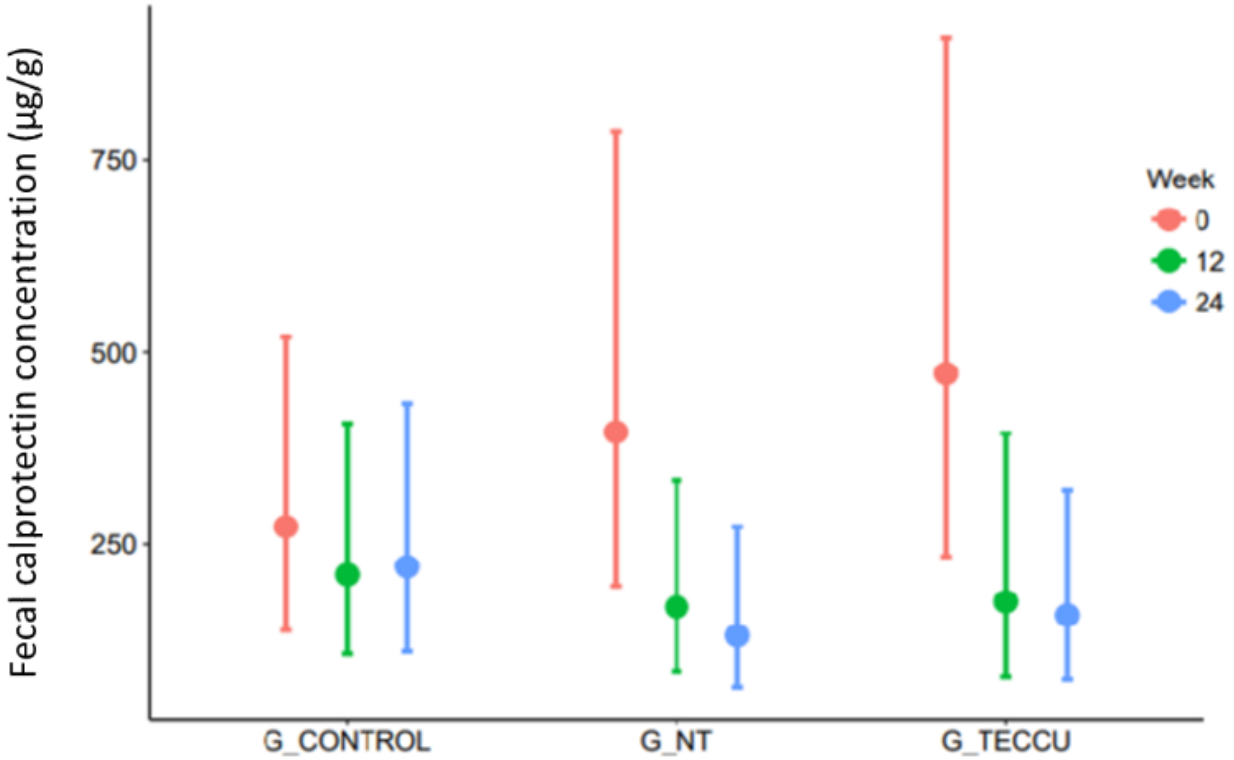


Figure 5. Evolution of the median fecal calprotectin levels over the study period in the 3 groups. G_CONTROL: group receiving standard care with in-person visits; G_NT: group receiving nurse-assisted telephone care; G_TECCU: group receiving remote monitoring.



Health-Related Quality of Life

HRQoL scores were similar between all groups at baseline, and the HRQoL of patients from all 3 groups improved after 24 weeks. The median IBDQ-9 scores increased from 38.5 to 53 in G_control, from 37.5 to 53 in G_NT, and from 42 to 52.5 in G_TECCU (overall intervention effect on the IBDQ-9 score: OR=8.42, 95% CI=3.98-17.81, *P*<.001). The median EQ-5D score improved from 0.816 to 1.00 in G_control and from 0.825 to 1.00 in G_NT and G_TECCU (overall intervention effect on the EQ-5D score: OR=1.99, 95% CI=1.09-3.63, *P*<.001). However, the improvement in HRQoL was not significantly different among groups with regard to the IBDQ-9 score (G_TECCU vs G_control: OR=1.25, 95% CI=0.49-3.15, *P*=.64; G_NT vs G_control: OR=0.79, 95% CI=0.32-1.98, *P*=.62) and the EQ-5D score (G_TECCU vs G_control: OR=1.42, 95% CI=0.49-4.13, *P*=.52; G_NT vs G_control: OR=1.68, 95% CI=0.59-4.81, *P*=.33). In addition, a significant improvement in the HRQoL of patients from the 3 arms was evident in the EQ-5D-VAS scores, with an increase in the median score from 60.5% to 85% in G_control, from 62.5% to 70% in G_NT, and from 60% to 80% in G_TECCU (overall intervention effect on the score increase: OR=3.53, 95% CI=1.79-6.95, *P*<.001); however, the improvement was not significantly different between the groups at 24 weeks (G_TECCU vs G_control: OR=0.63, 95% CI=0.26-1.57, *P*=.32; G_NT vs G_control:

OR=0.69, 95% CI=0.28-1.71, *P*=.43; Figure 6). The HRQoL improved in 4 patients in G_TECCU who were followed up via mobile phone, from a median baseline IBDQ-9 score of 31.5 to 47.75 and a median baseline EQ-5D score of 0.71 to 0.94, with an improvement in the median EQ-5D-VAS score from 50% to 70% at study completion.

Work Productivity and Social Activities

With regard to work productivity and activity impairment, we compared the median percentage of work hours missed owing to disease on the basis of the answers to question 2 of the WPAI questionnaire ("During the past seven days, how many hours did you miss from work because of problems associated with your ulcerative colitis/Crohn's disease?"). We found wide variability in the percentage of hours missed in the 3 arms, with no significant reduction in any group (OR=0.99, 95% CI=0.43-2.29, *P*=.93) and no differences among the groups at 24 weeks (G_NT vs G_control: OR=0.64, 95% CI=0.21-1.99, *P*=.44; G_TECCU vs G_control: OR=0.97, 95% CI=0.31-3.08, *P*=.96). Moreover, the median scores of impairment in work productivity over 12 weeks showed a larger reduction in G_TECCU than in G_control (OR=0.14, 95% CI=0.01-1.46, *P*=.10; G_NT vs G_control: OR=0.13, 95% CI=0.01-1.13, *P*=.06), but this improvement was not maintained at 24 weeks (G_TECCU vs G_control: OR=0.32, 95% CI=0.03-2.94, *P*=.31; G_AT vs G_control: OR=0.21, 95% CI=0.03-1.67, *P*=.14). With regard to social impairment, we found a significant improvement in daily activities and a significant reduction in the median score of each group at 12 weeks (OR=0.16, 95% CI=0.05-0.49, *P*=.002) and 24 weeks (OR=0.26, 95% CI=0.09-0.77, *P*=.02). This reduction in social impairment was larger in G_TECCU than in G_control at 24 weeks (OR=0.27, 95% CI=0.05-1.49, *P*=.13).

Medication Adherence

Medication adherence was poorer in G_NT than in the other 2 groups at baseline: 33.3% (7/21) of patients adhered to medication in G_NT; 66.7% (14/21), in G_control; and 57.1% (12/21), in G_TECCU. Despite these differences, medication adherence improved significantly in the 3 arms: 81% (17/21) in G_control, 71.4% (15/21) in G_NT, and 85.7% (18/21) in G_TECCU at 24 weeks (overall intervention effect on Morisky-Green score reduction: OR=0.051, 95% CI=0.001-0.769). In addition, all completers adhered to treatment in G_TECCU (Morisky-Green score=0), and the reduction in the Morisky-Green score was significantly more pronounced in G_TECCU than in G_control (OR=0.0001, 95% CI=1.02e^{-10} to 0.517; Figure 7).

Safety Intervention

The safety of the 3 interventions was similar, and no differences were noted among the groups for visits to the emergency department (0 in G_TECCU, 2 in the G_NT, and 1 in the G_control), IBD-related surgeries (1 in each group), hospitalizations (2 in G_TECCU, 2 in G_NT, and 1 in G_control), and corticosteroid courses (4 in G_TECCU, 4 in G_NT, and 3 in G_control). Adverse effects of medication were reported in 8 patients (38.1%) in G_TECCU, 7 patients (33.3%) in G_NT, and 9 patients (42.9%) in G_control after 24 weeks (*P*=.92 for G_TECCU vs G_control). No patients died during the study, and adverse effects related to the follow-up intervention were not reported.

Use of Health Care Resources

The total number of outpatient visits to the gastroenterologist or nurse after 24 weeks was lower in G_TECCU (72 visits; 25% of total) and G_NT (85 visits; 29.5% of total) than in G_control (131 visits; 45.5% of total). Similarly, the number of telephone calls was lower in G_TECCU (12 calls; 6.8% of total) than in G_NT (118 calls; 66.7% of total) and G_control (47 calls; 26.5% of total).

Figure 6. Evolution of the quality of life over the study period in the 3 groups. G_CONTROL: group receiving standard care with in-person visits; G_NT: group receiving nurse-assisted telephone care; G_TECCU: group receiving remote monitoring; IBDQ-9: Inflammatory Bowel Disease Questionnaire 9.

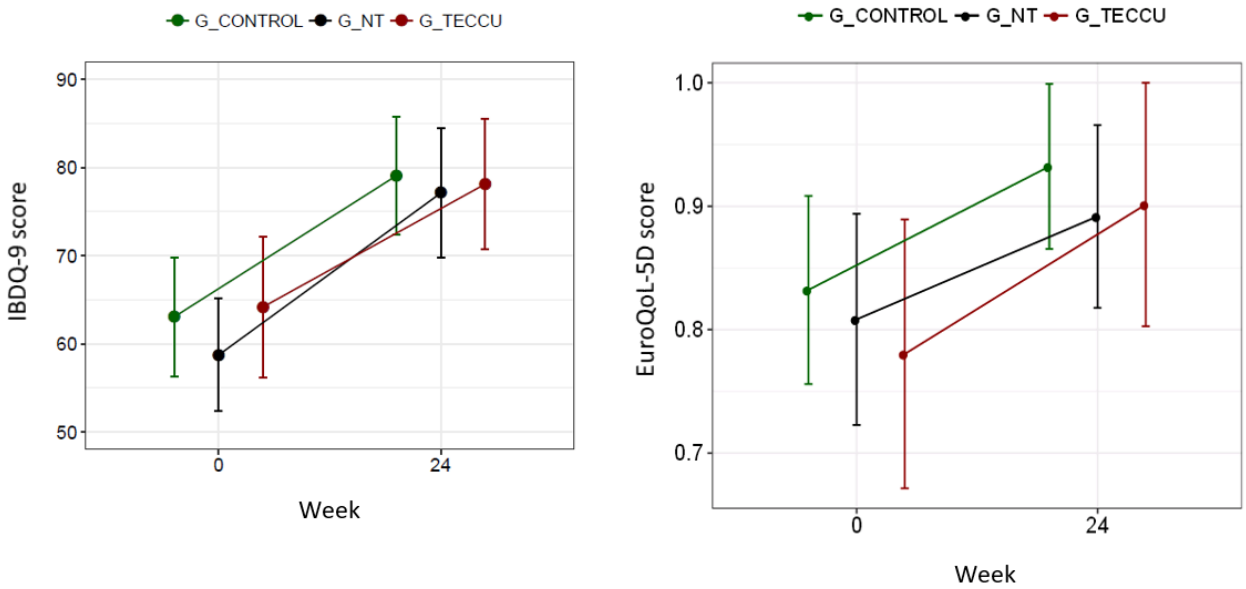
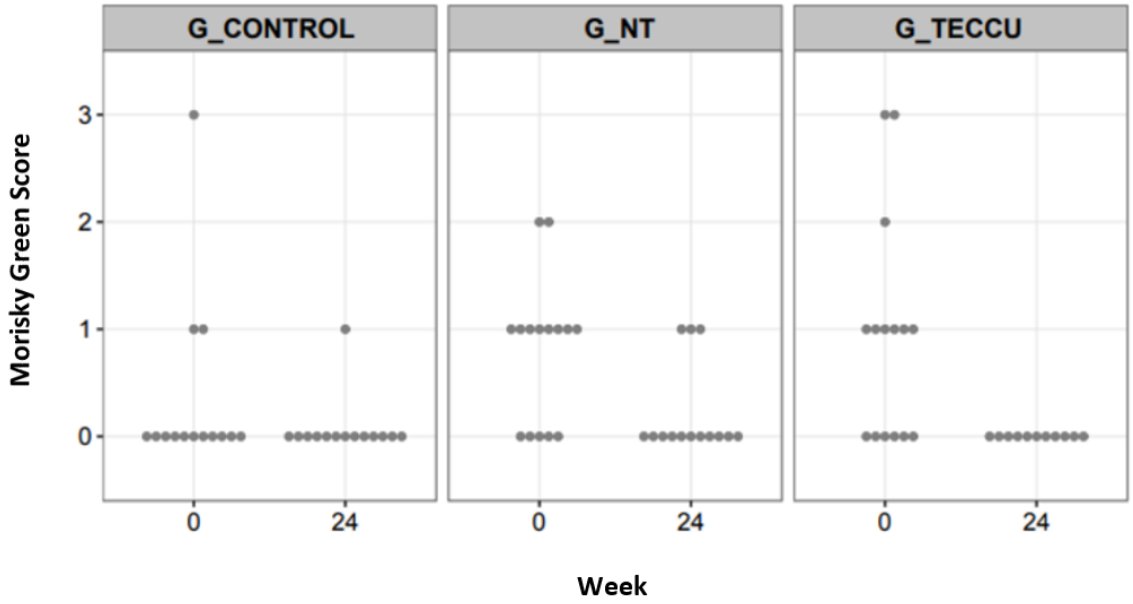


Figure 7. Evolution of the Morisky-Green score over the study period in the 3 groups. G_CONTROL: group receiving standard care with in-person visits; G_NT: group receiving nurse-assisted telephone care; G_TECCU: group receiving remote monitoring.



Patient Satisfaction

Patient satisfaction improved from a median score of 52 to 57 in G_TECCU and from 49.5 to 55 in G_control (overall intervention effect: OR=8.93, 95% CI=2.97-26.84, *P*<.001) at 24 weeks; however, the satisfaction score remained unchanged at 53 points in G_NT. Satisfaction with previous care was high in the mobile phone subgroup of G_TECCU, with a median score of 54.25, although this score improved to 56.25 after 24 weeks of follow-up via mobile phone.

In G_TECCU, neither the patients nor the researchers perceived privacy breaches while using the Web platform. Only two patients in G_TECCU reported minor technical problems with temporary unavailability of the webpage over the follow-up period, but technical assistants resolved the issue remotely. The 3 patients who were noncompliant to the follow-up schedule were contacted by telephone, and 2 of them who were insufficiently compliant specified that they forgot to follow some remote controls because they were in acceptable health conditions. We subsequently scheduled in-person visits, but these patients only attended them once. The only patient who was lost to follow-up mentioned that neither the previous standard care nor the Web platform met her expectations; therefore, she decided to move to another center.

All staff, including 5 gastroenterologists and 2 nurses specializing in IBD, showed good acceptance of the platform and considered it easy to use. However, 2 physicians thought that the calendar included in the platform for health care providers to survey the next controls for each patient was not completely intuitive and would benefit from some change in its appearance. Furthermore, 1 physician reported minor technical problems with temporary unavailability of the webpage over the follow-up period, but the technical assistants resolved the issue remotely after 1 day.

Discussion

Principal Findings

We performed a randomized controlled clinical trial to compare the impact of a Web-based telemanagement system (TECCU), nurse-assisted telephone care, and standard face-to-face visits on health outcomes and outpatient visits in patients with complex IBD. Patients with IBD who start treatment with systemic corticosteroids, immunosuppressants, and biological agents for control of inflammatory activity are considered to have moderate-to-high complexity of disease. Our results showed that TECCU was safe and effective for improving health outcomes and the use of health care resources in this setting.

Comparison With Prior Work

Although telemonitoring apps are well accepted and seem to be a safe approach for follow-up of patients in remission or with mild activity [6,18,41], the reported efficacy of telemonitoring on disease activity and QoL is inconsistent between studies [14,42], and no specific trials have evaluated the impact of telemedicine in patients with complex IBD. Therefore, we designed a controlled 3-arm clinical trial to assess the impact of the TECCU Web program on disease outcomes and health care use in comparison with the main strategies applied to date for the follow-up of patients with complex IBD (standard or telephone support provided by a nurse).

This program was designed in collaboration with patients and researchers to address the particular needs of patients with complex IBD, according to national and European clinical guidelines. As detailed elsewhere [27], this information was structured and filtered using an intelligent prioritization system, with generation of alerts and push notifications according to an integrated intervention protocol, which facilitated a rapid response from nurses and physicians for different events occurring during follow-up in a time period according to the severity of each alert. Furthermore, the Web platform allowed continuous communication between patients and health

providers via electronic messaging, and we incorporated educational elements in the platform to improve disease knowledge among patients and empowerment through interactive materials. With the use of the TECCU Web platform, we found a significant improvement in HRQoL, medication adherence, social activities, and satisfaction; in addition, disease activity improved at the end of follow-up in each group. Moreover, a greater improvement in disease activity and social activities was reported with TECCU than with standard care, but the differences were not significant, probably due to the small sample size and the relatively short follow-up time.

In agreement with previous data [6,8], medication adherence was better in G_TECCU compared to the other groups and reached 100% among completers in G_TECCU, probably because care was continuously adapted at each stage of the disease and communication with physicians was better in this group. Therefore, we were able to easily identify eventual problems related to poor adherence [2]. In addition, constant tailored monitoring with Web TECCU could reduce interference with daily activities, as social impairment due to disease tended to improve in G_TECCU compared with the other two arms. Not surprisingly, the improvement in remission rates and social functioning was associated with high patient satisfaction and correlated directly with a significant improvement in HRQoL, measured by the specific IDBQ-9 and the EQ-5D.

The Web TECCU and nurse-assisted telephone care were associated with fewer outpatient visits than standard care, but the Web-monitoring system reduced the number of telephone calls compared with the other two groups. Our study was performed in a referral hospital with an accredited, well-structured IBD Unit and accessible outpatient clinics with specialized nurses and e-mail or telephone consultations. Therefore, the reduction of face-to-face visits suggests that the effect of the TECCU remote-monitoring system on the frequency of outpatient visits could be favorable in hospitals with referral IBD Units. Although this reduction in office visits could be influenced by the follow-up schedule used in our study, the frequency of visits for all the 3 arms was designed according to the standard clinical practice in our center and based on national and European guidelines. Furthermore, our Web-based telemanagement system was safe and well accepted by patients, with no adverse effects related to the study intervention and no differences among groups in terms of emergency visits, surgeries, hospitalizations, corticosteroid courses, or medication adverse effects, which is consistent with the findings of previous studies [6,8].

Despite different attrition rates in the use of telemedicine apps, our results indicate that it is important to emphasize the feasibility of telemedicine for providing solutions for the medical and social impact of IBD. In our study, only 3 patients did not complete the follow-up schedule in G_TECCU (1 patient was lost to follow-up and 2 patients were insufficiently compliant); this low proportion is likely a result of the reminder system integrated in the platform and the short follow-up period. However, previous clinical trials reported high withdrawal rates in telemedicine groups [6,8,18,22], despite adaptations in the platform design over recent years. Consequently, telemedicine is not suitable for all patients, and improvements in telemonitoring apps and the patient-selection criteria are necessary. Nevertheless, this is not necessarily a barrier to the application of telemedicine, because the majority of patients are adherent to remote monitoring, and the efficacy of these follow-up methods has been reported in different populations including patients with complex IBD, as reported in our trial. It is important to evaluate the disease phenotype, discuss preferences with the patient, and assess the ability to use ICTs. Some patients such as those in the immediate postoperative period and those with perianal disease, who require physical examination, may not benefit from remote monitoring with the technological tools available thus far.

Strengths

The main strength of our study was its randomized controlled design, which allowed us to evaluate the impact of a Web telemanagement system on disease outcomes and compare the approach with the main strategies applied in the follow-up of IBD patients in daily practice. In addition, allocation concealment ensured that researchers were blinded to group assignment during the randomization process. Furthermore, the selection of patients with complex IBD from a referral center provides important data on the effect of telemedicine in patients with moderate-to-severe disease activity or initiating administration of immunosuppressants and biological agents, which is a specific population that has not been sufficiently represented in previous studies. Another advantage was the selection of the activity index and biological tools for measuring disease activity and other outcomes remotely. The Walmsley index was recently validated for self-administration via an online tool [43] and correlates well with other, more complex-activity indexes based on endoscopy [44]. Similarly, the reduced 6- and 9-point versions of the Mayo index, which are not based on endoscopy, show clinical response to treatment, similar to the full Mayo score [30]. Finally, as HBI may not completely reflect the inflammatory activity of the disease [45], the use of biological markers (C-reactive protein and FC) provided added value to remote monitoring, because they are sufficiently sensitive for detecting mucosal inflammation [46], and FC levels can be easily recorded at home.

Limitations

Our study was subject to a series of limitations. First, considering the specific study population, the sample was not large enough to detect statistically significant differences in clinical activity and the types of interventions assessed; as such, neither the patients nor the researchers were masked to the intervention. It is possible that the differences estimated among groups in the scarce literature on the impact on health outcomes could interfere with the sample size calculation in this pilot trial, and the lack of blinding could improve the usual attention provided to patients in G_control, which could also attenuate differences in these outcomes among groups. Nonetheless, our results were analyzed by an independent statistician who was blinded to group identification. Second, the selection of patients with complex IBD in a tertiary referral hospital prevented extrapolation of our results to the entire IBD population. Third, we did not perform colonoscopy in all cases, because it increases costs and risks to patients and is not performed routinely in all

patients during daily practice. Instead, we used biological markers that were highly sensitive for endoscopic activity. Finally, the 24-week follow-up period could be considered a short duration, because both patients and physicians need time to learn how to interact with the platform, thus leading to interference with the efficacy of the Web system in terms of disease outcome in the experimental group. Therefore, trials with longer follow-up periods are required to confirm the efficacy of telemedicine in improving long-term disease outcomes.

Conclusions

TECCU is a safe tool for the control of disease activity and improves HRQoL, medication adherence, and impairment in social activities in patients with complex IBD. In addition, it reduces the use of health care resources and the number of outpatient visits and telephone calls. It is important to select patients who are willing and able to use Web programs, since the results of this study suggest greater efficacy in health outcomes with a consequent improvement in the HRQoL and satisfaction of patients who adhere to telemanagement. Web-based systems are a safe and feasible option for IBD monitoring and could play a key role in reorganizing the structure of health systems if they prove to be cost-effective in the long-term. On the other hand, although telemonitoring platforms are well accepted, thus far, they have not clearly demonstrated their efficacy on health outcomes in patients with complex IBD and those in remission or with mild-to-moderate disease. Future studies with longer follow-up periods are necessary to confirm our findings in order to determine whether Web-based programs can improve the long-term course of IBD in a more-complex setting.

Acknowledgments

This study was supported by grants from the Instituto de Salud Carlos III-Fondo de Investigaciones Sanitarias (FIS PI12/00277) and cofunded by FEDER (Fondo Europeo de Desarrollo Regional).

Conflicts of Interest

DD is the general manager of Connected Health Services.

Multimedia Appendix 1

CONSORT-EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 511KB - jmir_v20i11e11602_app1.pdf]

References

1. Kappelman MD, Porter CQ, Galanko JA, Rifas-Shiman SL, Ollendorf DA, Sandler RS, et al. Utilization of healthcare resources by U.S. children and adults with inflammatory bowel disease. Inflamm Bowel Dis 2011 Jan;17(1):62-68 [FREE Full text] [doi: 10.1002/ibd.21371] [Medline: 20564532]
2. Jackson CA, Clatworthy J, Robinson A, Horne R. Factors associated with non-adherence to oral medication for inflammatory bowel disease: a systematic review. Am J Gastroenterol 2010 Mar;105(3):525-539. [doi: 10.1038/ajg.2009.685] [Medline: 19997092]
3. van Deen WK, van Oijen MGH, Myers KD, Centeno A, Howard W, Choi JM, et al. A nationwide 2010-2012 analysis of U.S. health care utilization in inflammatory bowel diseases. Inflamm Bowel Dis 2014 Oct;20(10):1747-1753. [doi: 10.1097/MIB.0000000000000139] [Medline: 25137415]
4. Høivik ML, Moum B, Solberg IC, Henriksen M, Cvancarova M, Bernklev T, IBSEN Group. Work disability in inflammatory bowel disease patients 10 years after disease onset: results from the IBSEN Study. Gut 2013 Mar;62(3):368-375. [doi: 10.1136/gutjnl-2012-302311] [Medline: 22717453]
5. Hoivik ML, Moum B, Solberg IC, Cvancarova M, Hoie O, Vatn MH, IBSEN Study Group. Health-related quality of life in patients with ulcerative colitis after a 10-year disease course: results from the IBSEN study. Inflamm Bowel Dis 2012 Aug;18(8):1540-1549. [doi: 10.1002/ibd.21863] [Medline: 21936030]
6. Elkjaer M, Shuhaibar M, Burisch J, Bailey Y, Scherfig H, Laugesen B, et al. E-health empowers patients with ulcerative colitis: a randomised controlled trial of the web-guided 'Constant-care' approach. Gut 2010 Dec;59(12):1652-1661. [doi: 10.1136/gut.2010.220160] [Medline: 21071584]
7. Krier M, Kaltenbach T, McQuaid K, Soetikno R. Potential use of telemedicine to provide outpatient care for inflammatory bowel disease. Am J Gastroenterol 2011 Dec;106(12):2063-2067. [doi: 10.1038/ajg.2011.329] [Medline: 22138934]
8. de Jong MJ, van der Meulen-de Jong AE, Romberg-Camps MJ, Becx MC, Maljaars JP, Cilissen M, et al. Telemedicine for management of inflammatory bowel disease (myIBDcoach): a pragmatic, multicentre, randomised controlled trial. Lancet 2017 Sep 02;390(10098):959-968. [doi: 10.1016/S0140-6736(17)31327-2] [Medline: 28716313]
9. Pedersen N, Elkjaer M, Duricova D, Burisch J, Dobrzanski C, Andersen NN, et al. eHealth: individualisation of infliximab treatment and disease course via a self-managed web-based solution in Crohn's disease. Aliment Pharmacol Ther 2012 Nov;36(9):840-849 [FREE Full text] [doi: 10.1111/apt.12043] [Medline: 22971016]

10. Pedersen N, Thielsen P, Martinsen L, Bennedsen M, Haaber A, Langholz E, et al. eHealth: individualization of mesalazine treatment through a self-managed web-based solution in mild-to-moderate ulcerative colitis. Inflamm Bowel Dis 2014 Dec;20(12):2276-2285. [doi: 10.1097/MIB.0000000000000199] [Medline: 25248002]
11. Clarke M, Shah A, Sharma U. Systematic review of studies on telemonitoring of patients with congestive heart failure: a meta-analysis. J Telemed Telecare 2011;17(1):7-14. [doi: 10.1258/jtt.2010.100113] [Medline: 21097564]
12. 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]
13. Bartoli L, Zanaboni P, Masella C, Ursini N. Systematic review of telemedicine services for patients affected by chronic obstructive pulmonary disease (COPD). Telemed J E Health 2009 Nov;15(9):877-883. [doi: 10.1089/tmj.2009.0044] [Medline: 19919194]
14. Aguas PM, Del HJ, Bebia P, Faubel R, Barrios A, Bastida G, et al. Telemedicine in inflammatory bowel disease: opportunities and approaches. Inflamm Bowel Dis 2015 Feb;21(2):392-399. [doi: 10.1097/MIB.0000000000000241] [Medline: 25437818]
15. Helsel BC, Williams JE, Lawson K, Liang J, Markowitz J. Telemedicine and Mobile Health Technology Are Effective in the Management of Digestive Diseases: A Systematic Review. Dig Dis Sci 2018 Jun;63(6):1392-1408. [doi: 10.1007/s10620-018-5054-z] [Medline: 29663265]
16. Cross RK, Cheevers N, Finkelstein J. Home telemanagement for patients with ulcerative colitis (UC HAT). Dig Dis Sci 2009 Nov;54(11):2463-2472. [doi: 10.1007/s10620-008-0640-0] [Medline: 19104937]
17. Cross RK, Arora M, Finkelstein J. Acceptance of telemanagement is high in patients with inflammatory bowel disease. J Clin Gastroenterol 2006 Mar;40(3):200-208. [Medline: 16633120]
18. Cross RK, Cheevers N, Rustgi A, Langenberg P, Finkelstein J. Randomized, controlled trial of home telemanagement in patients with ulcerative colitis (UC HAT). Inflamm Bowel Dis 2012 Jun;18(6):1018-1025 [FREE Full text] [doi: 10.1002/ibd.21795] [Medline: 21688350]
19. Cross RK, Jambaulikar G, Langenberg P, Tracy JK, Collins JF, Katz J, et al. TELEmedicine for Patients with Inflammatory Bowel Disease (TELE-IBD): Design and implementation of randomized clinical trial. Contemp Clin Trials 2015 May;42:132-144. [doi: 10.1016/j.cct.2015.03.006] [Medline: 25812483]
20. Carlsen K, Jakobsen C, Houen G, Kallemose T, Paerregaard A, Riis LB, et al. Self-managed eHealth Disease Monitoring in Children and Adolescents with Inflammatory Bowel Disease: A Randomized Controlled Trial. Inflamm Bowel Dis 2017 Dec;23(3):357-365. [doi: 10.1097/MIB.0000000000001026] [Medline: 28221247]
21. Carlsen K, Houen G, Jakobsen C, Kallemose T, Paerregaard A, Riis LB, et al. Individualized Infliximab Treatment Guided by Patient-managed eHealth in Children and Adolescents with Inflammatory Bowel Disease. Inflamm Bowel Dis 2017 Dec;23(9):1473-1482. [doi: 10.1097/MIB.0000000000001170] [Medline: 28617758]
22. Heida A, Dijkstra A, Muller KA, Rossen JW, Kindermann A, Kokke F, et al. Efficacy of Home Telemonitoring versus Conventional Follow-up: A Randomized Controlled Trial among Teenagers with Inflammatory Bowel Disease. J Crohns Colitis 2018 Mar 28;12(4):432-441. [doi: 10.1093/ecco-jcc/jjx169] [Medline: 29228230]
23. Casellas-Jordá F, Borruel-Sainz N, Torrejón-Herrera A, Castells I. Effect upon hospital activity of the application of a continued care model centered on patients with inflammatory bowel disease. Rev Esp Enferm Dig 2012 Feb;104(1):16-20. [Medline: 22300112]
24. Casey M, Hayes PS, Heaney D, Dowie L, Ólaighin G, Matero M, et al. Implementing transnational telemedicine solutions: a connected health project in rural and remote areas of six Northern Periphery countries Series on European collaborative projects. Eur J Gen Pract 2013 Mar;19(1):52-58. [doi: 10.3109/13814788.2012.761440] [Medline: 23432039]
25. Van Assche G, Dignass A, Panes J, Beaugerie L, Karagiannis J, Allez M, European Crohn's Colitis Organisation (ECCO). The second European evidence-based Consensus on the diagnosis and management of Crohn's disease: Definitions and diagnosis. J Crohns Colitis 2010 Feb;4(1):7-27 [FREE Full text] [doi: 10.1016/j.crohns.2009.12.003] [Medline: 21122488]
26. Dignass A, Eliakim R, Magro F, Maaser C, Chowers Y, Geboes K, et al. Second European evidence-based consensus on the diagnosis and management of ulcerative colitis part 1: definitions and diagnosis. J Crohns Colitis 2012 Dec;6(10):965-990 [FREE Full text] [doi: 10.1016/j.crohns.2012.09.003] [Medline: 23040452]
27. Aguas M, Del Hoyo J, Faubel R, Muñoz D, Domínguez D, Bastida G. A web-based telemanagement system for patients with complex inflammatory bowel disease: protocol for a randomized controlled clinical trial. JMIR Res Protoc 2018 Dec:1-12. [doi: 10.2196/resprot.9639]
28. Harvey RF, Bradshaw JM. A simple index of Crohn's-disease activity. Lancet 1980 Mar 8;1(8167):514. [Medline: 6102236]
29. Walmsley RS, Ayres RC, Pounder RE, Allan RN. A simple clinical colitis activity index. Gut 1998 Jul;43(1):29-32 [FREE Full text] [Medline: 9771402]
30. Lewis JD, Chuai S, Nessel L, Lichtenstein GR, Aberra FN, Ellenberg JH. Use of the noninvasive components of the Mayo score to assess clinical response in ulcerative colitis. Inflamm Bowel Dis 2008 Dec;14(12):1660-1666 [FREE Full text] [doi: 10.1002/ibd.20520] [Medline: 18623174]

31. D'Haens G, Sandborn WJ, Feagan BG, Geboes K, Hanauer SB, Irvine EJ, et al. A review of activity indices and efficacy end points for clinical trials of medical therapy in adults with ulcerative colitis. Gastroenterology 2007 Feb;132(2):763-786. [doi: 10.1053/j.gastro.2006.12.038] [Medline: 17258735]
32. Alcalá MJ, Casellas F, Fontanet G, Prieto L, Malagelada J. Shortened questionnaire on quality of life for inflammatory bowel disease. Inflamm Bowel Dis 2004 Jul;10(4):383-391. [Medline: 15475746]
33. Badia X, Roset M, Montserrat S, Herdman M, Segura A. [The Spanish version of EuroQol: a description and its applications. European Quality of Life scale]. Med Clin (Barc) 1999;112 Suppl 1:79-85. [Medline: 10618804]
34. Badia X, Roset M, Herdman M, Kind P. A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. Med Decis Making 2001;21(1):7-16. [doi: 10.1177/0272989X0102100102] [Medline: 11206949]
35. Vergara M, Montserrat A, Casellas F, Villoria A, Suarez D, Maudsley M, et al. A new validation of the Spanish Work Productivity and Activity Impairment Questionnaire-Crohn's disease version. Value Health 2011;14(6):859-861 [FREE Full text] [doi: 10.1016/j.jval.2011.02.1179] [Medline: 21914506]
36. Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. Med Care 1986 Jan;24(1):67-74. [Medline: 3945130]
37. Larsen DL, Attkisson CC, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfaction: development of a general scale. Eval Program Plann 1979;2(3):197-207. [Medline: 10245370]
38. TECCU. URL: http://82.223.37.103/teccu [accessed 2018-11-03] [WebCite Cache ID 73eLybxRj]
39. Gomollón F, García-López S, Sicilia B, Gisbert JP, Hinojosa J, en representación del Grupo Español de Trabajo de Enfermedad de Crohn y Colitis Ulcerosa or Spanish Group for Working on Crohn'sDiseaseUlcerative Colitis (GETECCU). [Therapeutic guidelines on ulcerative colitis: a GRADE methodology based effort of GETECCU]. Gastroenterol Hepatol 2013 Oct;36(8):e1-47. [Medline: 24215088]
40. R Core Team (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing URL: http://www.r-project.org/ [accessed 2018-11-07] [WebCite Cache ID 73knzJR3d]
41. Robinson A, Thompson DG, Wilkin D, Roberts C, Northwest GRG. Guided self-management and patient-directed follow-up of ulcerative colitis: a randomised trial. Lancet 2001 Sep 22;358(9286):976-981. [doi: 10.1016/S0140-6736(01)06105-0] [Medline: 11583752]
42. Aguas M, Del HJ, Faubel R, Valdivieso B, Nos P. Telemedicine in the treatment of patients with inflammatory bowel disease. Gastroenterol Hepatol 2017 Nov;40(9):641-647. [doi: 10.1016/j.gastrohep.2017.07.001] [Medline: 28797518]
43. Marín-Jiménez I, Nos P, Domènech E, Riestra S, Gisbert JP, Calvet X, et al. Diagnostic Performance of the Simple Clinical Colitis Activity Index Self-Administered Online at Home by Patients With Ulcerative Colitis: CRONICA-UC Study. Am J Gastroenterol 2016 Feb;111(2):261-268. [doi: 10.1038/ajg.2015.403] [Medline: 26753886]
44. Higgins PDR, Schwartz M, Mapili J, Zimmermann EM. Is endoscopy necessary for the measurement of disease activity in ulcerative colitis? Am J Gastroenterol 2005 Feb;100(2):355-361. [doi: 10.1111/j.1572-0241.2005.40641.x] [Medline: 15667493]
45. Zittan E, Kabakchiev B, Kelly OB, Milgrom R, Nguyen GC, Croitoru K, et al. Development of the Harvey-Bradshaw Index-pro (HBI-PRO) Score to Assess Endoscopic Disease Activity in Crohn's Disease. J Crohns Colitis 2017 May 01;11(5):543-548. [doi: 10.1093/ecco-jcc/jjw200] [Medline: 28453763]
46. Lin J, Chen J, Zuo J, Yu A, Xiao Z, Deng F, et al. Meta-analysis: fecal calprotectin for assessment of inflammatory bowel disease activity. Inflamm Bowel Dis 2014 Aug;20(8):1407-1415. [doi: 10.1097/MIB.0000000000000057] [Medline: 24983982]

Abbreviations

CD: Crohn disease
EQ-5D: EuroQol-5D
FC: fecal calprotectin
G_control: group receiving standard care with in-person visits
G_NT: group receiving nurse-assisted telephone care
G_TECCU: group receiving remote monitoring
HBI: Harvey-Bradshaw index
HRQoL: health-related quality of life
IBD: inflammatory bowel disease
IBDQ-9: Inflammatory Bowel Disease Questionnaire 9
ICT: information and communication technologies
SCCAI: Simple Clinical Colitis Activity Index
TECCU: Telemonitoring of Crohn's Disease and Ulcerative Colitis
UC: ulcerative colitis
VAS: visual analog scale

WPAI: Work Productivity and Activity Impairment

Edited by G Eysenbach; submitted 23.07.18; peer-reviewed by V Traver Salcedo, R Cross, P Haubruck; comments to author 16.08.18; revised version received 09.10.18; accepted 22.10.18; published 27.11.18.

Please cite as:
Del Hoyo J, Nos P, Faubel R, Muñoz D, Domínguez D, Bastida G, Valdivieso B, Correcher M, Aguas M
A Web-Based Telemanagement System for Improving Disease Activity and Quality of Life in Patients With Complex Inflammatory Bowel Disease: Pilot Randomized Controlled Trial
J Med Internet Res 2018;20(11):e11602
URL: http://www.jmir.org/2018/11/e11602/
doi:10.2196/11602
PMID:30482739

Original Paper

Guided Internet-Based Cognitive Behavioral Therapy for Adult Depression and Anxiety in Routine Secondary Care: Observational Study

Kim Mathiasen[1,2], MSc; Heleen Riper[3,4], PhD; Tonny E Andersen[1], PhD; Kirsten K Roessler[1], PhD

[1]Department of Psychology, Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark

[2]Centre for Telepsychiatry, Mental Health Services of Southern Denmark, Odense, Denmark

[3]Department of Mental Health, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[4]Department of Research and Innovation, Specialized Mental Health Care, GGZ InGeest, Amsterdam, Netherlands

Corresponding Author:
Kim Mathiasen, MSc
Department of Psychology
Faculty of Health Sciences
University of Southern Denmark
Campusvej 55
Odense, 5230
Denmark
Phone: 45 61677747
Email: kim.mathiasen@rsyd.dk

Abstract

Background: Internet-based cognitive behavioral therapy (iCBT) is a promising new treatment method for depression and anxiety. However, it is important to determine whether its results can be replicated in routine care before its implementation on a large scale. Although many studies have demonstrated the efficacy of iCBT under controlled conditions, only a few studies have investigated its effectiveness in routine care. Furthermore, several effects of iCBT such as treatment effects in routine care are unclear.

Objective: This study aimed to evaluate the clinical effectiveness of iCBT for depression and anxiety in routine secondary care.

Methods: n a retrospective cohort study, we analysed patients treated for depression or anxiety in a dedicated iCBT clinic in secondary care in Denmark. Patients were examined before treatment and weekly thereafter by using the Patient Health Questionnaire-9 and the Generalized Anxiety Disorder-7 scales for the diagnoses of depression and anxiety, respectively. Primary analyses were conducted using a linear mixed-effects model with random slope and intercept. Secondary analyses were conducted using baseline characteristics as predictors (gender, age, highest level of education, occupational status, marital status, psychotropic medication use, consumption of alcohol, and leisure drugs). Additionally, logistic regression analyses were used to predict noncompletion of treatment.

Results: A total of 203 (depression, N=60; anxiety, N=143) patients were included. Participants were mainly female (78.3% with depression and 65.7% with anxiety), with a mean age of 36.03 (SD 10.97) years (range, 19-67 years) for patients with depression and 36.80 (SD 13.55) years (range, 19-69 years) for patients with anxiety. The completion rates were 62% (37) and 40% (57) for depression and anxiety treatments, respectively. The primary analyses revealed large and significant reductions in the symptom levels of depression (beta=-6.27, SE 0.83, *P*<.001, d=1.0) and anxiety (beta=-3.78, SE 0.43, *P*<.001, d=1.1). High baseline severity of the primary disorder was associated with high treatment gains (r=-0.31 for depression; r=-0.41 for anxiety). In patients with anxiety, high baseline severity also predicted a high risk of noncompletion (odds ratio=1.08, CI=1.01-1.16, *P*=.03). An increase in the baseline severity of the comorbid disorder slightly increased the risk of noncompletion for both disorders (depression: odds ratio=1.03, CI=1.01-1.06, *P*=.02; anxiety: odds ratio=1.08, CI=1.01-1.16, *P*=.03).

Conclusions: iCBT can be clinically effective in routine care. Since depression and anxiety are costly and debilitating disorders that are vastly undertreated, this finding is important. Additionally, iCBT may help bridge the gap between the need for treatment and its provision. Our results are comparable to the within-group results of efficacy and effectiveness studies. Our noncompletion

rates are similar to those observed in psychotherapy but are higher than those reported in similar clinics. Multiple factors predicted outcome and noncompletion. However, all predictor effects were statistically weak.

(J Med Internet Res 2018;20(11):e10927) doi:10.2196/10927

KEYWORDS

anxiety; cognitive therapy; cohort studies; depression; Internet; secondary care

Introduction

Background

Anxiety and depression are highly prevalent and costly disorders [1-4]. Although effective treatments for these disorders exist [5], only about half of those in need receive professional help, and only approximately 16% receive minimally adequate evidence-based treatments [6,7]. Many studies have demonstrated the efficacy of guided internet-based cognitive behavioral therapy (iCBT) [8-13], which may help bridge the gap between the need for treatment and its provision.

Although effectiveness studies such as randomized controlled trials (RCTs) conducted in routine care settings [14] are limited in number, they show effect sizes comparable to those obtained under almost-ideal conditions of efficacy studies [15]. Owing to the increasing number of persons seeking help for mental disorders in Denmark, the Danish Agency for Digitisation included iCBT in a national action plan for telemedicine in 2012, which led to the establishment of the first iCBT clinic in Denmark. The clinic, named Internetpsykiatrien, has been in operation since 2014 and a part of routine care since 2015. Funded by the Danish healthcare system, it is free at the point of use.

Internetpsykiatrien is a part of secondary care in the Mental Health Services of Southern Denmark. It delivers guided iCBT [15] for adult patients with anxiety and depression and offers selfreferral; as such, patients can apply for treatment directly at the clinic from a secure and dedicated website, and no general practitioner is required [16]. Herein, we present the results of an observational study of the first routine outcome measurements of a cohort of adult patients who were treated at the clinic with iCBT for depression or anxiety.

This observational cohort study with a pre-post design based on routine outcome measurements is important, as treatments and patients in routine practice may differ from those in RCTs. Patients may differ with regard to their sociodemographic and baseline characteristics as well as their motivation or need for treatment. Additionally, adherence to strict research protocols and treatment manuals and continuous monitoring of clinician adherence are not part of the standard clinical practice in most settings. Thus, although effectiveness trials have higher external validity than efficacy trials, their clinical effect may differ from that in actual routine practice, for example, in the case of RCTs, even in a routine care setting [14,17]. Therefore, studies of the clinical effect of routine care are needed to increase knowledge on the issue in real-world settings [18]. Although efficacy studies are important, the external validity of an intervention should be the prime focus in order to convince decision makers and clinicians of its effectiveness, thereby ensuring its implementation and uptake [19]. In addition to efficacy and effectiveness studies, observational cohort studies can provide important insights into the impact of iCBT in real-life routine care of populations.

The literature on cohort studies in routine practice of comparable iCBT clinics is limited. However, some published studies from the Internetpsykiatrien clinic in Sweden, the eMeistring clinic in Norway, the MindSpot clinic in Australia, and a centralized clinic in Canada [16,20-25] describe dedicated iCBT clinics that deliver guided iCBT and are operated as routine mental health care services. In the Swedish clinic, large, significant within-group effect sizes were found in cohort studies of major depression (n=1203, *d*=1.27) [20], panic disorder (n=570, *d*=0.91) [21], and social anxiety (n=547, *d*=0.86) [22]. In Norway, results for panic disorder showed a large pre-post effect size (n=124, *d*=1.24) [23]. In the Sydney clinic, a cohort study from the first 30 months of operation revealed large, significant effect sizes (n=6149, *d*=1.3-1.4) [24] for transdiagnostic iCBT for anxiety and depression. The centralized unit in Canada offers guided iCBT with selfreferral and provides assistance to community mental health services; the clinic showed large, significant reductions in symptom levels of both anxiety and depression (n=260, *d*=1.2-1.4) [25]. The literature, although scarce, seems to suggest that guided iCBT may be a promising treatment format in routine practice.

Primary Aim

This study aims to improve knowledge of the clinical effect of guided iCBT in routine practice by presenting and discussing analyses of routine outcome measurements in a cohort of adults treated for anxiety or depression at the iCBT clinic Internetpsykiatrien in Odense, Denmark.

Secondary Aim

Our secondary aim was to determine predictors of treatment outcome and noncompletion for iCBT.

Methods

Design

This uncontrolled retrospective cohort study included all patients who had initiated treatment for depression (N=60) or anxiety (N=143) at the iCBT clinic Internetpsykiatrien and provided consent for analysis of their data. Data for all patients were collected from the time reliable data were available until the time of data extraction (June 1, 2017). For patients with anxiety, reliable data was available from March 1, 2015, and for depressed patients, from August 1, 2016. The study was observational in nature and did not interfere with the normal operation of the clinic. All data were collected as routine outcome measurements and part of continuous quality assurance

by the clinic by using online questionnaires. This article adheres to the Strengthening the Reporting of Observational Studies in Epidemiology guidelines for reporting observational studies [26,27] and the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) [28].

Interventions

Internetpsykiatrien operates as part of routine care in Southern Denmark and is funded by the Region of Southern Denmark, which is a tax-funded public authority. The clinic uses self-referral via a website [29], where patients fill out a secure application form. They do not need referral from other sources such as a general practitioner. Treatment including the use of the programs is free of charge for the patients. After an initial screening of their application, patients are contacted for clarification by telephone and, if appropriate (eg, in cases of increased suicidal risk), provided access to more relevant sources of assistance or invited to a video-based assessment interview with a licensed psychologist or a psychologist under the supervision of a licensed psychologist. Eligibility criteria for patients at the clinic are age ≥ 18 years; meeting the diagnostic criteria of the International Classification of Diseases and Related Health Problems, 10th edition, for major depressive disorder, panic disorder, agoraphobia, social phobia, specific phobia, or generalized anxiety disorder; not currently at high risk of suicide; no comorbid substance dependence, bipolar affective disorder, psychotic illness, or obsessive compulsive disorder; not currently under other psychological treatment for depression or anxiety; access to a personal computer and fast internet connection; and adequate understanding of spoken and written Danish.

After assessment, the patients were provided anxiety or depression treatment. Both treatments included access to an iCBT treatment program and weekly or biweekly clinical support from a licensed clinical psychologist or a psychologist under the supervision of a licensed psychologist. Support was provided via a secure text module for depression treatment and via telephone for anxiety treatment. In addition, support comprised technical assistance, help with interventions included in the programs, and encouragement to continue the treatment. The Internetpsykiatrien has been described in detail elsewhere, and its implementation has been compared to similar clinics internationally [30,31].

For depression treatment, the program NoDep by Context Consulting was used. This program was designed and developed as part of a Public Private Innovation project between the Region of Southern Denmark and Context Consulting when the Internetpsykiatrien was under development. The manuscript was written by Specialist Psychologist Krista Nielsen Straarup, who is a well-known expert in CBT for depression in Denmark. The program includes 6 mandatory modules and 2 optional modules and is based on CBT; its core components include psychoeducation, behavioral activation, cognitive restructuring, and relapse prevention. The optional modules include interventions for reducing rumination and restructuring dysfunctional beliefs [32].

The anxiety treatment uses the Danish version of FearFighter, which is distributed in Denmark by Context Consulting [33] and includes 9 mandatory modules based on CBT. FearFighter was originally developed by Professor Dr. Isaac Marks [34] in England. Its core components comprise psychoeducation, in vivo and interoceptive exposure exercises, cognitive restructuring, and relapse prevention. In both programs, patients are reminded to access the programs by SMS and email on a weekly basis, if they fail to do so. This feature can be switched off by the patient. The programs are accessed via the internet on patients' own devices. Access link, username, and access codes are provided to the patient by the clinic.

In both programs, clinicians are provided with administrative access, which includes a status of the progress of patients, their exercises, their scores on the weekly measures, and a secure text module (for NoDep only). Further, the clinicians are notified of high scores on suicidal ideation and lack of progress. In both cases, patients are contacted personally by phone. In cases of suicidal risk, a standard procedure including risk assessment and possible referral to an acute ward is instigated.

Ethics

All the patients analyzed in the study provided informed consent for use of their data at the beginning of treatment. No participants were approached by the research team. The Regional Committees on Health Research Ethics for Southern Denmark waived the need for full approval owing to the retrospective nature of the study, no new intervention was performed, and no human biological material was used. The routine outcome measurement database was approved by the Danish Data Protection Agency.

Measures

The Patient Health Questionnaire-9 (PHQ-9) [35] was used to evaluate changes in the severity of depressive symptoms. This 9-item questionnaire has good psychometric properties [36,37] and can be used to monitor patients with depressive disorders. All items are scored on a 0- to 3-point scale with a total score of 0-27, and higher scores indicated more severe depression. One study of the Danish version confirmed unidimensionality and reliability but suggests collapsing the two middle response categories [38]. However, the scale used in the Internetpsykiatrien clinic followed the original structure of the questionnaire.

The Generalized Anxiety Disorder-7 (GAD-7) scale [39] was used to measure general symptoms of anxiety. It is a 7-item questionnaire originally developed to measure GAD in primary care; each item is scored 0-3, with a total score of 0-21. This scale has shown good psychometric properties, has been validated in primary care [39] and the general population [40], and has performed well as a measure of anxiety-symptom severity in specialized mental health care settings [41]. Thus far, there are no validation studies of the Danish version of the GAD-7 scale.

The Fear Questionnaire [42] was used to describe baseline levels of comorbid anxiety symptoms for patients with depression. Data were extracted from the questionnaire included in the application form. The Fear Questionnaire measures phobias across 6 subscales and provides a total phobia score of 0-120.

Similar to the GAD-7 scale, this scale has demonstrated good psychometric properties [43,44].

All questionnaires were administered in an electronic format in the treatment program or at the initial application form. This was not an open survey, as data were obtained from patients included in treatment at the Internetpsykiatrien clinic. It was mandatory to fill out the questionnaires with no extra incentives, apart from the added value of the supporting psychologist who monitored patients' progress. Completeness of the questionnaires was checked by the program, such that progress was hindered until all items were completed. Once the questionnaire was submitted, it could not be altered. All responses were unique to the individuals, as its administration was a part of the treatment program or linked to the application form. All participants viewed the questionnaires and participated by completing at least some of the items. All available data were analyzed, even if partly filled questionnaires were submitted.

Description of Participants

Baseline characteristics of patients were extracted from a questionnaire delivered as part of the online application form, which was used when patients applied for treatment, and from a pretreatment questionnaire package. The collected data on patient characteristics included gender, age, highest level of education, occupational status, marital status, baseline symptomatic levels of depression and anxiety (PHQ-9, GAD-7, Fear Questionnaire), units of alcohol consumed per week, and use of recreational drugs. The baseline characteristics were analyzed using descriptive statistics.

Primary Analyses

Outcome measures were extracted from the pretreatment questionnaire package and the last of a series of weekly measures administered online as an integrated part of the iCBT programs. The intermittent weekly measures were unfortunately unavailable to the research team.

Treatment effects were estimated using mixed-effects linear regression. The pre-post measurements were considered as repeated observations nested within patients. The slope estimates for the binary treatment variable (0, pre; 1, post) indicated change in the outcome measurements from baseline to follow-up. The primary outcome variables, that is, symptomatic measures of the primary disorder, were included as a fixed effect, whereas the binary treatment variable and subject were included as random effects (random slope and intercept). The slope-intercept correlations were estimated to control for baseline severity. The primary model estimated the overall treatment effect with one slope parameter, one intercept, and one correlation between slope and intercept. No control variables were included. The standardized effect sizes were calculated by taking into account the correlation between the pre-post measurement and expressed as *d* [45]. Multivariate imputation by chained equations was applied to manage missing data. All inferences assumed normally distributed error terms and heteroscedasticity, which were substantiated by visual inspection of the Q-Q normality plot and the plot of fitted values versus standardized residuals.

Predictor Analyses

Predictors of change in symptom severity were determined in subsequent univariate models; specifically, we employed one model per predictor by using the same model type as that in the primary analyses. Regression estimates for the interaction between the predictor variables and the binary treatment variable indicated the extent to which the overall treatment slope was affected by a particular predictor variable. Additionally, two multivariate models were performed. One model included all variables, and the other model included only significant variables from the univariate analyses.

Completion

Analyses of treatment effect among completers and non-completers were conducted using the same model type as that of the primary analyses. Logistic regression was used to predict the risk of noncompletion on the basis of all available baseline variables for each of the two patient groups, which was expressed as odds ratios (ORs). Treatment completion was defined as completion of at least 5 of 6 mandatory modules in the depression program and 8 of 9 modules in the anxiety program. These thresholds were applied because patients at these stages of the programs had completed all core active treatment interventions aimed at symptom reduction.

Statistical Analyses

All data were analyzed using R ver. 3.4.2 [46]. Mixed-effects linear regression was performed using the nlme package [47]. The glm function [46] was used for logistic regression, whereas the mice package was used for imputation [48]. $P<.05$ was set as the threshold value for significance in inferential statistics.

Results

Description of Participants

Among patients with depression, 60 patients started treatment and consented to the analysis, of which 59 (98.3%) patients provided followup data. A mean of 5.4 (SD 2.2) of the 6 mandatory and 2 optional modules were completed. In addition, 143 patients with anxiety commenced treatment, of which 133 (93.0%) patients provided followup data. A mean of 5.9 (SD 2.6) modules were (Figure 1). Table 1 presents the characteristics of all the patients who commenced treatment in the inclusion periods and had quit the program or ended treatment by the point of data extraction.

Figure 1. Patient flow chart.

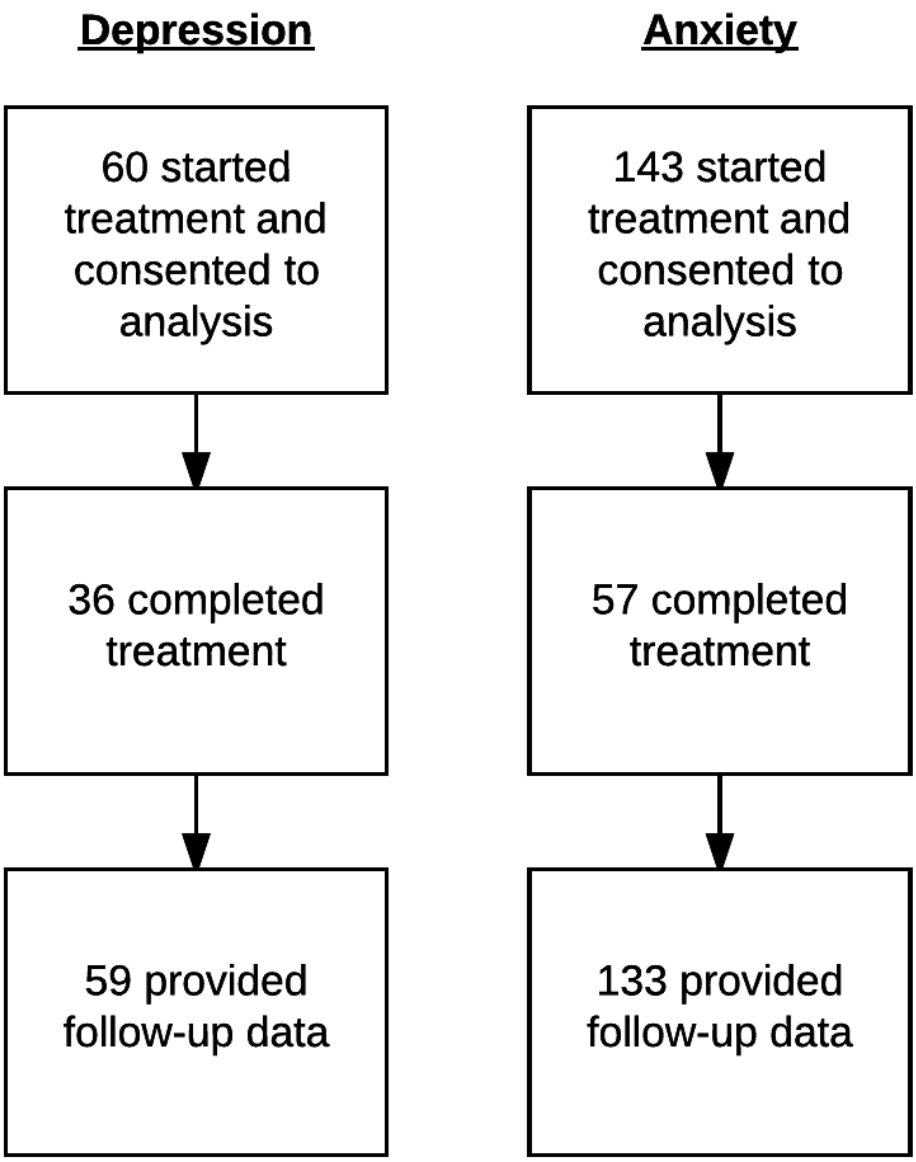


Table 1. Description of patients.

Variable	Depression (N=60)	Anxiety (N=143)
Gender, n (%)		
Female	47 (78.3)	94 (65.7)
Male	13 (21.7)	52 (34.3)
Age (years)		
Mean (SD)	36.03 (10.97)	36.80 (13.55)
Range	19-67	19-69
Education, n (%)	**N=41**	**N=48**
Primary: 9 years in school	6 (14.6)	6 (12.5)
Gymnasium: 12 years in school	5 (12.2)	11 (22.9)
College or vocational school	26 (63.4)	31 (64.6)
University: 17 years in school	0 (0.0)	0 (0.0)
Other	4 (9.8)	0 (0.0)
Occupational status, n (%)	**N=41**	**N=48**
Employed/student	23 (56.1)	39 (81.3)
Unemployed	5 (12.2)	7 (14.6)
Sick leave	5 (12.2)	2 (4.2)
Retired	0 (0.0)	0 (0.0)
Other	8 (19.5)	0 (0.0)
Marital status, n (%)	**N=41**	**N=48**
Single	11 (26.8)[a]	18 (37.5)[b]
Married or de facto not living together	5 (12.2)[a]	2 (6.25)[b]
Married or de facto living together	25 (61.0)[a]	27 (56.3)[b]
Depression severity		
PHQ-9[c] mean (SD)	15.5 (5.7)	9.94 (6.06)
PHQ-9[a], range	5-26	0-26
Anxiety severity		
FQ[b] score, total mean (SD)	22 (22.8)[c]	—[d]
FQ total score, range	0-81[c]	—
GAD-7[e] score, mean (SD)	—	10.42 (5.11)
GAD-7 score, range	—	0-21
Patients taking psychotropic medication, n (%)		**N=136**
No medication	45 (75.0)	95 (69.9)
<1 month	1 (1.7)	5 (3.7)
<2 months	3 (5.0)	2 (1.5)
>2 months	11 (18.3)	34 (25.0)
Units of alcohol comsumed weekly, n (%)		**N=136**
0	32 (53.3)	60 (44.1)
0-5	25 (41.7)	51 (37.5)
5-10	1 (2.7)	17 (12.5)
10-20	2 (4.1)	7 (5.1)

Variable	Depression (N=60)	Anxiety (N=143)
>20	0 (0.0)	1 (0.7)
Recreational drugs, n (%)	3 (5.0)	0 (0.0)[f]

[a]PHQ-9: Patient Health Questionnaire-9.

[b]FQ: Fear Questionnaire.

[c]Taken from the application form.

[d]Not applicable.

[e]GAD-7: Generalized Health Anxiety questionnaire.

[f]n=136 because some data were not collected for all patients at the clinic.

Primary Analyses

Table 2 shows the results of the linear mixed-effects model for depression and anxiety treatments including beta-coefficients, standard errors, and *P*-values for intention-to-treat analyses, completer analysis, and non-completer analysis.

Primary analysis of patients with depression revealed a significant reduction in depressive symptoms on the PHQ-9 (beta=–6.27, SE 0.83, *P*<.001), with a large effect size (*d*=1.0). For patients with anxiety, the linear mixed-effects model demonstrated a significant reduction in the symptom level of anxiety on the GAD-7 (beta=–3.78, SE 0.43, *P*<.001) with a large effect size (*d*=1.1).

Predictor Analyses

Exploratory predictor analyses for interaction effects revealed that some baseline characteristics predicted a change in symptom severity of the primary disorder from pretreatment to posttreatment (Table 2). For patients with depression, comorbid anxiety predicted slightly lesser symptom reduction with a small, but significant, positive addition to the slope (beta=.08, SE 0.04, *P*=.03). Additionally, a very small, but significant, interaction between treatment and time spent in the program was observed (beta=–0.03, SE 0.01, *P*=.02), indicating that for each additional day spent in the program, the PHQ-9 score reduced by 0.03. Furthermore, a college education as the highest education level predicted a steeper slope from pretreatment to posttreatment on the PHQ-9 (beta=–5.06, SE 2.02, *P*=.02) as compared to primary school education. The baseline symptom level of depression (intercept) was negatively correlated with the slope of change (*r*=–0.31), indicating a larger symptomatic reduction if the baseline level of depressive symptoms was higher. However, when all variables were included in one multivariate model, none of the interactions were significant. Similar results were observed when only the significant interactions from the univariate analyses were included in one multivariate model.

For patients with anxiety, two baseline characteristics predicted outcome. Inclusion in the “Other” category significantly increased symptom reduction (beta=–3.49, SE 1.10, *P*=.002) as compared to inclusion in the “Employed/student” category. Additionally, a higher level of comorbid depressive symptoms at baseline predicted a slightly larger reduction in the symptom level of anxiety (beta=–0.18, SE 0.07, *P*=.01). Time spent in the program was not associated with the treatment outcome in patients with anxiety. Finally, similar to patients with depression, the baseline level of anxiety was negatively correlated with slope (*r*=–0.41), indicating a larger improvement with higher baseline levels. Multivariate analysis of all variables indicated that marriage or de facto not living together was significant (beta=–2.64, SE 1.32, *P*=.047), whereas marriage and de facto living together was not significant (beta=1.63, SE 1.12, *P*=.15). Occupational status included in the “Other” category (beta=–2.96, SE 0.083, *P*=.02) and baseline comorbidity (beta=–0.259, SE 0.083, *P*=.002) remained significant in the multivariate analysis. Similar results were obtained when only the significant interactions from the univariate analyses were included; however, in this case, no category under the variable “marital status” was significant.

Completion

Of all patients with depression, 37 (61.7%) completed the treatment in the depression program. Analysis of treatment effect showed that for patients with depression, the level of depressive symptoms on the PHQ-9 reduced significantly (beta=–7.41, SE 1.00, *P*<.001), with a large effect size (*d*=1.1). Similarly, noncompleters showed a significant decrease in symptom severity (beta=–4.43, SE 1.39, *P*=.004), albeit with a lower yet moderate-to-large effect size (*d*=0.7). A nonsignificant trend was noted in the interaction between treatment effect and noncompletion (beta=2.97, SE 1.68, *P*=.08), indicating a nonsignificant decrease in symptom change in cases of noncompletion.

Among patients with anxiety, 57 (39.9%) completed treatment. A significant reduction in the severity of anxiety was observed among patients who completed the program (beta=–4.72, SE 0.65, *P*<.001), with a large effect size (*d*=1.1). Similar results were observed for 86 (60.1%) patients who did not complete the program, albeit with a moderate effect size (beta=–3.22, SE 0.56, *P*<.001, *d*=0.6). Furthermore, no significant difference was observed in the interaction between severity of anxiety and noncompletion (beta=1.49, SE 0.88, *P*=.09), but a trend toward better outcome with completion was noted.

Table 2. Univariate linear mixed-effects model outcomes for severity of depression and anxiety symptoms. Baseline severity correlated with the main effect for depression (r=–0.31) and anxiety (r=–0.41).

Variables	Depression (PHQ-9)[a], N=60			Anxiety (GAD-7)[b], N=143		
	Mean beta (SE)	*P* value[c]	*d*	Mean beta (SE)	*P* value[c]	*d*
Treatment effect[d]						
Main effect	–6.27 (0.83)	<.001	1.0	–3.78 (0.43)	<.001	1.1
Completers	–7.41 (1.00)	<.001	1.1	–4.72 (0.65)	<.001	1.1
Noncompleters	–4.43 (1.39)	.004	0.7	–3.22 (0.56)	<.001	0.6
Predicting outcome[e]						
Gender						
Time point	–4.77 (1.79)	.01		–3.90 (0.73)	<.001	
Gender x time point	–1.91 (2.02)	.35		0.19 (0.91)	.84	
Age						
Time point	–3.89 (2.88)	.18		–3.77 (1.25)	.003	
Age x time point	–0.07 (0.08)	.39		0.00 (0.03)	.99	
Highest education level						
Time point	–3.75 (1.73)	.03		–3.00 (1.54)	.05	
Education level x time point						
Gymnasium vs primary	1.08 (2.99)	.72		0.28 (1.76)	.88	
College vs primary	–5.06 (2.02)	.02		–0.50 (1.69)	.77	
University vs primary				–2.07 (2.03)	.31	
Other vs primary	0.45 (2.56)	.86		–2.21 (1.81)	.22	
Marital status						
Time point	–4.86 (1.36)	<.001		–4.33 (0.76)	<.001	
Marital status x time point						
Married or de facto not living together vs single	1.46 (2.39)	.55		–0.95 (1.09)	.39	
Married or de facto living together vs single	–3.42 (1.78)	.06		1.94 (1.00)	.05	
Occupational status						
Time point	–6.46 (1.24)	<.001		–2.54 (0.71)	<.001	
Income x time point						
Unemployed vs employed or student	1.46 (2.62)	.58		–1.20 (1.07)	.27	
Sick leave vs employed or student	1.21 (2.26)	.59		–0.46 (1.32)	.73	
Other vs employed or student	–1.20 (2.26)	.60		–3.49 (1.10)	.002	
Baseline comorbidity						
Anxiety (FQ[e] total)						
Time point	–8.16 (1.18)	<.001				
Time point x FQ total	0.08 (0.04)	.03				
Depression (PHQ-9)						
Time point				–2.02 (0.81)	.01	
Time point x PHQ-9				–0.18 (0.07)	.01	
Time spent in the program						
Time point	–3.95 (1.26)	.003		–3.34 (0.63)	<.001	
Time spend x time point	–0.03 (0.01)	.02		–0.00 (0.00)	.35	

[a]PHQ-9: Patient Health Questionnaire.

[b]GAD-7: Generalized Anxiety questionnaire.

[c]Two-tailed *P* value.

[d]Pretreatment to last observation.

[e]*d* values cannot meaningfully be calculated for the interaction effects.

[f]FQ: Fear Questionnaire.

For patients with anxiety, a higher level of baseline anxiety predicted a slightly higher risk of noncompletion (OR=1.08, CI=1.01-1.16, *P*=.03). Similar results were noted for higher levels of comorbid depression (OR=1.12, CI=1.05-1.20, *P*<.001). A college education as the highest education level decreased the risk of noncompletion compared to primary school education (OR=0.14, CI=0.02-0.61, *P*=.02). A complete overview of all analyses of predictors for treatment completion is presented in Multimedia Appendix 1.

Discussion

Aim

The aim of the present study was to evaluate the clinical effect of guided iCBT for depression and anxiety in routine care by analyzing routine outcome measurements from the dedicated iCBT clinic Internetpsykiatrien in Odense, Denmark. From March 1, 2015, to June 1, 2017, 203 patients commenced treatment, including 60 patients who underwent treatment for depression and 143 patients who underwent treatment for anxiety, which demonstrates the need for and motivation to seek guided iCBT.

Participants

Of the patients with depression, 78.3% were female, which corresponds to the findings of previous epidemiological research [49-52]. Despite the new format of treatment delivery, the gender difference may be biased by treatment-seeking behavior, as previously shown by a male to female ratio of 2:1. The prevalence of anxiety disorders is not extensively researched in Denmark, but in general practice, more women experience anxiety disorders than men [52]. The mean age, marital status, and employment status of our sample resembles those in primary care [52], which is reasonable, as Internetpsykiatrien offers selfreferral and is thus likely to receive a patient population of patients with depression and anxiety in primary care. The distribution of the highest education level resembles that of the general Danish population: primary, 25.9%; gymnasium, 10.2%; college or vocational school, 51.7%; university, and 9.5; other, 2.65% [53]. However, there are some noticeable differences. In the present study, fewer patients had primary school and university education as the highest level of education and more patients had college or vocational school as the highest level of education. Nonetheless, overall, our sample seemed representative of the Danish population.

Primary Analyses

The main analyses of change in the symptom levels of primary disorder from pretreatment to posttreatment showed large and significant reductions in both depression (beta=–6.27, SE 0.83, *P*<.001, *d*=1.0), and anxiety (beta=–3.78, SE 0.43, *P*<.001, *d*=1.1). These results resemble within-group effect sizes of previous studies on guided iCBT for adult patients with anxiety and depression, including both efficacy studies conducted under controlled research conditions [54-56] and effectiveness studies conducted under routine care conditions [15]. Notably, the results of studies in clinics akin to the one under investigation, which employed similar study designs examining cohorts in routine care (*d*=0.9-1.4) [20-25], are in line with our results. Therefore, this study concludes that the clinic under investigation yields results similar to those seen in other comparable clinics, and the effects of iCBT are satisfactory as compared to within-group effect sizes of previous efficacy and effectiveness studies.

Predictor Analyses

Baseline severity of primary disorders influenced the effect of treatment. For anxiety and depression, a higher baseline level of primary disorder correlated with larger treatment gains, which is a correlation commonly observed in studies on guided iCBT [57-59]. This effect is interesting, because it is often assumed that guided selfhelp should be aimed primarily at milder symptom levels of the disorders, although factors other than severity should be examined as indicators for stratification of patients for different treatments. Some studies found substantial evidence that treatment effect was unrelated to the treatment outcome [60].

There was a difference in how the symptomatic level of a comorbid disorder affected treatment gain. For patients with depression, comorbid anxiety slightly hampered the effect of treatment; in contrast, for patients with anxiety, high levels of comorbid depression increased the effect of treatment to a small extent. Thus, comorbidity might influence the effect of treatment differently for the two disorders, which may add nuance to the findings of a previous study, which concluded that the severity of comorbid disorders predicts poor treatment outcome [59]. Our finding may also support the use of transdiagnostic treatment programs or personalized treatment approaches.

For patients with anxiety, occupational status in the “Other” category rather than the “Employed/student” category positively affected the outcome (beta=–3.49, SE 1.10, *P*=.002). Analysis of the reasons for selection of the “Other” category showed that this category was extensively heterogeneous and included several entries such as apprenticeships, other forms of benefit, or no income, which made it difficult to interpret the result. Neither this category nor any other category of occupational status had an effect on the treatment in patients with depression. However, compared to the employed or student status, the unemployed status was associated with a higher risk of noncompletion for patients with depression. A previous study showed that a poorer employment status was associated with worse treatment outcome in unguided iCBT [61], which was not observed in this study. This difference in results may have

resulted from the different formats of iCBT delivery. The present study investigated guided iCBT; therefore, the guidance may serve as a protective factor for vulnerable patients. This assumption could be substantiated by emerging findings that indicate the importance of the therapeutic alliance in iCBT [62,63]. Further study of the influences of and processes involved in the guidance is needed.

We found that more time spent in the program predicted a greater reduction in symptom severity for patients with depression but not for patients with anxiety, which seems to indicate a dose-response effect of depression treatment, but not anxiety treatment. Multivariate analyses showed different results for both anxiety and depression. For depression, no significant interactions were observed, which might be a result of the small sample size. These predictions should therefore be interpreted carefully.

Completion

Of the patients who commenced treatment, 37 of 60 patients with depression and 57 of 143 patients with anxiety completed treatment, which yields noncompletion rates of 38.3% and 60.1%, respectively. A review of 152 studies on traditional face-to-face psychotherapy reported an average dropout rate of 46.9% [64], which is comparable to our result. However, the dropout rates of iCBT treatment usually range from 0% to 78%, with a weighted mean of 26.5% [65]. This finding is in line with the results from routine-care iCBT clinics in Sweden, Norway, and Canada, where rates of noncompletion ranged from 15.9% to 29.5% [66-68]. Therefore, in the Internetpsykiatrien clinic, the rate of noncompletion seems relatively high, particularly for patients with anxiety. However, it is possible that the cutoff chosen to signify completion of treatment in our study was overly conservative, requiring completion of 83% and 88% of the modules of depression and anxiety, respectively, which are considerably higher than the thresholds used in other studies [23], where merely 56% of the modules needed to be completed. This hypothesis may be substantiated by the fact that significant and moderate-to-large effect sizes were seen among patients who did not complete the programs in the present study, indicating that some patients may terminate their treatment because they are satisfied with their progress. In addition, the difference in effect size between intention-to-treat analyses and completer analyses was not significant. For both disorders, a higher baseline symptom level of comorbid disorders slightly increased the risk of noncompletion. Interestingly, the dropout rates differed between the two treatment groups, with higher dropout rates observed for patients with anxiety than for patients with depression (60.1% vs 38.3%). However, this does not appear to be the case in the literature [66-68], and the underlying reasons are unclear. This difference may be caused by technical variations such as those between programs or the format of delivery of support, for example, written support for depression treatment and telephone support for anxiety treatment. However, it could also be a result of the different symptomatology of the disorders; for example, anxiety is characterized by avoidant behavior. Nonetheless, it is important to determine how to improve adherence, particularly for patients with anxiety, and whether this goal can be attained by different methods for the two different disorders. Further studies should focus on examination of different user behavior and experiences from the programs during and after provision of support. Furthermore, examination of motivations for ending treatment is needed. Finally, research investigating process factors involved in completion or noncompletion is needed to further clarify individual differences in response patterns among patients receiving iCBT.

Limitations

Most of the results from the exploratory predictor analyses were weak and subject to type 1 error due to their exploratory nature; therefore, they should be interpreted with caution. Prediction of treatment outcomes from baseline characteristics is generally difficult, as reported in previous iCBT studies showing varying results [58-61]. Our results therefore need repetition, particularly in routine care, with a larger study sample.

Conclusion

The Internetpsykiatrien clinic has demonstrated its ability to implement iCBT in routine care as well as the existence of, need for, and motivation among patients to seek iCBT treatment for anxiety and depression in the Region of Southern Denmark. The clinic yields results comparable to those of similar international clinics and efficacy and effectiveness studies. As such, the Danish clinic shows good clinical effectiveness. These findings are important, because depression and anxiety are highly prevalent and costly disorders that are substantially undertreated. Our results support the hypothesis that iCBT can help bridge the gap between the need for treatment and its provision even in routine care. Further research is needed to investigate the processes of change and individual response patterns to iCBT in order to improve the ability to personalize treatment for the individual and understand the reasons for premature cessation of treatment.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Univariate logistic mixed-effects model interactions for prediction of noncompletion.

[PDF File (Adobe PDF File), 25KB - jmir_v20i11e10927_app1.pdf]

References

1. Alonso J, Angermeyer MC, Bernert S, Bruffaerts R, Brugha TS, Bryson H, et al. Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. Acta Psychiatr Scand Suppl 2004(420):21-27. [doi: 10.1111/j.1600-0047.2004.00327.x] [Medline: 15128384]
2. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
3. Olesen J, Gustavsson A, Svensson M, Wittchen H, Jönsson B, CDBE2010 SG, et al. The economic cost of brain disorders in Europe. Eur J Neurol 2012 Jan;19(1):155-162. [doi: 10.1111/j.1468-1331.2011.03590.x] [Medline: 22175760]
4. Ustün TB. The global burden of mental disorders. Am J Public Health 1999 Sep;89(9):1315-1318. [Medline: 10474545]
5. Cuijpers P. Four decades of outcome research on psychotherapies for adult depression: An overview of a series of meta-analyses. Canadian Psychology/Psychologie canadienne 2017;58(1):7-19. [doi: 10.1037/cap0000096]
6. Alonso J, Codony M, Kovess V, Angermeyer MC, Katz SJ, Haro JM, et al. Population level of unmet need for mental healthcare in Europe. Br J Psychiatry 2007 Apr;190:299-306 [FREE Full text] [doi: 10.1192/bjp.bp.106.022004] [Medline: 17401035]
7. Thornicroft G, Chatterji S, Evans-Lacko S, Gruber M, Sampson N, Aguilar-Gaxiola S, et al. Undertreatment of people with major depressive disorder in 21 countries. Br J Psychiatry 2017 Dec;210(2):119-124 [FREE Full text] [doi: 10.1192/bjp.bp.116.188078] [Medline: 27908899]
8. Olthuis JV, Watt MC, Bailey K, Hayden JA, Stewart SH. Therapist-supported Internet cognitive behavioural therapy for anxiety disorders in adults. Cochrane Database Syst Rev 2015;3:CD011565. [doi: 10.1002/14651858.CD011565] [Medline: 25742186]
9. Arnberg FK, Linton SJ, Hultcrantz M, Heintz E, Jonsson U. Internet-delivered psychological treatments for mood and anxiety disorders: a systematic review of their efficacy, safety, and cost-effectiveness. PLoS One 2014;9(5):e98118 [FREE Full text] [doi: 10.1371/journal.pone.0098118] [Medline: 24844847]
10. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
11. 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]
12. 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]
13. 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]
14. Streiner DL. The 2 "Es" of research: efficacy and effectiveness trials. Can J Psychiatry 2002 Aug;47(6):552-556. [doi: 10.1177/070674370204700607] [Medline: 12211883]
15. Andersson G, Hedman E. Effectiveness of Guided Internet-Based Cognitive Behavior Therapy in Regular Clinical Settings. Verhaltenstherapie 2013;23(3):140-148. [doi: 10.1159/000354779]
16. Folker A, Mathiasen K, Lauridsen SM, Stenderup E, Dozeman E, Folker MP. Implementing internet-delivered cognitive behavior therapy for common mental health disorders: A comparative case study of implementation challenges perceived by therapists and managers in five European internet services. Internet Interv 2018 Mar;11:60-70 [FREE Full text] [doi: 10.1016/j.invent.2018.02.001] [Medline: 30135761]
17. Shadish WR, Matt GE, Navarro AM, Siegle G, Crits-Christoph P, Hazelrigg MD, et al. Evidence that therapy works in clinically representative conditions. J Consult Clin Psychol 1997 Jun;65(3):355-365. [Medline: 9170759]
18. Andersson G, Carlbring P, Cuijpers P. Internet interventions: Moving from efficacy to effectiveness. E-Journal of Applied Psychology 2009;5(2):18-24 [FREE Full text]
19. Glasgow RE, Green LW, Klesges LM, Abrams DB, Fisher EB, Goldstein MG, et al. External validity: we need to do more. Ann Behav Med 2006 Apr;31(2):105-108. [doi: 10.1207/s15324796abm3102_1] [Medline: 16542124]
20. Hedman E, Ljótsson B, Kaldo V, Hesser H, El AS, Kraepelien M, et al. Effectiveness of Internet-based cognitive behaviour therapy for depression in routine psychiatric care. J Affect Disord 2014 Feb;155:49-58. [doi: 10.1016/j.jad.2013.10.023] [Medline: 24238951]
21. Hedman E, Ljótsson B, Rück C, Bergström J, Andersson G, Kaldo V, et al. Effectiveness of internet-based cognitive behaviour therapy for panic disorder in routine psychiatric care. Acta Psychiatr Scand 2013 Dec;128(6):457-467. [doi: 10.1111/acps.12079] [Medline: 23406572]
22. El AS, Hedman E, Kaldo V, Hesser H, Kraepelien M, Andersson E, et al. Effectiveness of Internet-based cognitive-behavior therapy for social anxiety disorder in clinical psychiatry. J Consult Clin Psychol 2015 Oct;83(5):902-914. [doi: 10.1037/a0039198] [Medline: 26009780]
23. Nordgreen T, Gjestad R, Andersson G, Carlbring P, Havik OE. The implementation of guided Internet-based cognitive behaviour therapy for panic disorder in a routine-care setting: effectiveness and implementation efforts. Cogn Behav Ther 2018 Jan;47(1):62-75. [doi: 10.1080/16506073.2017.1348389] [Medline: 28714775]

24. Titov N, Dear BF, Staples LG, Bennett-Levy J, Klein B, Rapee RM, et al. The first 30 months of the MindSpot Clinic: Evaluation of a national e-mental health service against project objectives. Aust N Z J Psychiatry 2017 Dec;51(12):1227-1239. [doi: 10.1177/0004867416671598] [Medline: 27733709]
25. Hadjistavropoulos HD, Nugent MM, Alberts NM, Staples L, Dear BF, Titov N. Transdiagnostic Internet-delivered cognitive behaviour therapy in Canada: An open trial comparing results of a specialized online clinic and nonspecialized community clinics. J Anxiety Disord 2016 Dec;42:19-29 [FREE Full text] [doi: 10.1016/j.janxdis.2016.05.006] [Medline: 27244278]
26. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med 2007 Oct 16;4(10):e297 [FREE Full text] [doi: 10.1371/journal.pmed.0040297] [Medline: 17941715]
27. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med 2007 Oct 16;4(10):e296 [FREE Full text] [doi: 10.1371/journal.pmed.0040296] [Medline: 17941714]
28. 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]
29. Mental HSOSD. Internetpsykiatrien. 2018. URL: http://www.internetpsykiatrien.dk/wm507341 [accessed 2018-08-21] [WebCite Cache ID 71pvmdSc8]
30. Folker AP, Mathiasen K, Lauridsen SM, Stenderup E, Dozeman E, Folker MP. Implementing internet-delivered cognitive behavior therapy for common mental health disorders: A comparative case study of implementation challenges perceived by therapists and managers in five European internet services. Internet Interv 2018 Mar;11:60-70 [FREE Full text] [doi: 10.1016/j.invent.2018.02.001] [Medline: 30135761]
31. Titov N, Dear B, Nielssen O, Staples L, Hadjistavropoulos H, Nugent M, et al. ICBT in routine care: A descriptive analysis of successful clinics in five countries. Internet Interv 2018 Sep;13:108-115 [FREE Full text] [doi: 10.1016/j.invent.2018.07.006] [Medline: 30206525]
32. Mathiasen K, Andersen TE, Riper H, Kleiboer AAM, Roessler KK. Blended CBT versus face-to-face CBT: a randomised non-inferiority trial. BMC Psychiatry 2016 Dec 05;16(1):432 [FREE Full text] [doi: 10.1186/s12888-016-1140-y] [Medline: 27919234]
33. Mathiasen K, Riper H, Ehlers LH, Valentin JB, Rosenberg NK. Internet-based CBT for social phobia and panic disorder in a specialised anxiety clinic in routine care: Results of a pilot randomised controlled trial. Internet Interv 2016 May;4:92-98 [FREE Full text] [doi: 10.1016/j.invent.2016.03.001] [Medline: 30135794]
34. Marks I, Kenwright M, McDonough M, Whittaker M, Mataix-Cols D. Saving clinicians' time by delegating routine aspects of therapy to a computer: a randomized controlled trial in phobia/panic disorder. Psychol Med 2004 Jan;34(1):9-17. [Medline: 14971623]
35. 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]
36. Wittkampf KA, Naeije L, Schene AH, Huyser J, van Weert HC. Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. Gen Hosp Psychiatry 2007;29(5):388-395. [doi: 10.1016/j.genhosppsych.2007.06.004] [Medline: 17888804]
37. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. Med Care 2004 Dec;42(12):1194-1201. [Medline: 15550799]
38. Pedersen SS, Mathiasen K, Christensen KB, Makransky G. Psychometric analysis of the Patient Health Questionnaire in Danish patients with an implantable cardioverter defibrillator (The DEFIB-WOMEN study). J Psychosom Res 2016 Dec;90:105-112. [doi: 10.1016/j.jpsychores.2016.09.010] [Medline: 27772556]
39. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
40. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
41. Beard C, Björgvinsson T. Beyond generalized anxiety disorder: psychometric properties of the GAD-7 in a heterogeneous psychiatric sample. J Anxiety Disord 2014 Aug;28(6):547-552. [doi: 10.1016/j.janxdis.2014.06.002] [Medline: 24983795]
42. Marks IM, Mathews AM. Brief standard self-rating for phobic patients. Behav Res Ther 1979;17(3):263-267. [Medline: 526242]
43. Gillis MM, Haaga DAF, Ford GT. Normative values for the Beck Anxiety Inventory, Fear Questionnaire, Penn State Worry Questionnaire, and Social Phobia and Anxiety Inventory. Psychological Assessment 1995;7(4):450-455. [doi: 10.1037/1040-3590.7.4.450]
44. Oei T, Moylan A, Evans L. Validity and clinical utility of the Fear Questionnaire for anxiety-disorder patients. Psychological Assessment 1991;3(3):391-397. [doi: 10.1037/1040-3590.3.3.391]
45. Dunlap WP, Cortina JM, Vaslow JB, Burke MJ. Meta-analysis of experiments with matched groups or repeated measures designs. Psychological Methods 1996;1(2):170-177. [doi: 10.1037/1082-989X.1.2.170]

46. R Foundation for Statistical Computing. R: A language and environment for statistical computing URL: https://www.r-project.org/ [accessed 2018-10-30] [WebCite Cache ID 73Yc6FD5O]
47. Pinheiro J, Bates D, DebRoy S, Sarkar D. R Core Team. 2017. URL: https://cran.r-project.org/web/packages/nlme/index.html [accessed 2018-10-30] [WebCite Cache ID 73YcEfuYB]
48. Yucel RM. State of the Multiple Imputation Software. J Stat Softw 2011 Dec;45(1) [FREE Full text] [Medline: 22289957]
49. Weissman MM, Bland RC, Canino GJ, Faravelli C, Greenwald S, Hwu HG, et al. Cross-national epidemiology of major depression and bipolar disorder. JAMA 1996;276(4):293-299. [Medline: 8656541]
50. Olsen LR, Mortensen EL, Bech P. Prevalence of major depression and stress indicators in the Danish general population. Acta Psychiatr Scand 2004 Feb;109(2):96-103. [Medline: 14725589]
51. Christensen Sven Bundgaard Per Bech O. Prevalence of clinical (major) depression in general practice using the DSM-IV version of PRIME-MD. Int J Psychiatry Clin Pract 2001;5(1):49-54. [doi: 10.1080/136515001300225196] [Medline: 24936998]
52. Toft T, Fink P, Oernboel E, Christensen K, Frostholm L, Olesen F. Mental disorders in primary care: prevalence and co-morbidity among disorders. results from the functional illness in primary care (FIP) study. Psychol Med 2005 Aug;35(8):1175-1184. [Medline: 16116943]
53. Befolkningens højst fuldførte uddannelse.: Danmarks Statistik URL: https://dst.dk/da/Statistik/emner/uddannelse-og-viden/befolkningens-uddannelsesstatus/befolkningens-hoejst-fuldfoerte-uddannelse [accessed 2018-10-30] [WebCite Cache ID 73YTdvQd7]
54. Paxling B, Almlöv J, Dahlin M, Carlbring P, Breitholtz E, Eriksson T, et al. Guided internet-delivered cognitive behavior therapy for generalized anxiety disorder: a randomized controlled trial. Cogn Behav Ther 2011;40(3):159-173. [doi: 10.1080/16506073.2011.576699] [Medline: 21770848]
55. Andersson G, Hesser H, Veilord A, Svedling L, Andersson F, Sleman O, et al. Randomised controlled non-inferiority trial with 3-year follow-up of internet-delivered versus face-to-face group cognitive behavioural therapy for depression. J Affect Disord 2013 Dec;151(3):986-994. [doi: 10.1016/j.jad.2013.08.022] [Medline: 24035673]
56. Berger T, Caspar F, Richardson R, Kneubühler B, Sutter D, Andersson G. Internet-based treatment of social phobia: a randomized controlled trial comparing unguided with two types of guided self-help. Behav Res Ther 2011 Mar;49(3):158-169. [doi: 10.1016/j.brat.2010.12.007] [Medline: 21255767]
57. Bower P, Kontopantelis E, Sutton A, Kendrick T, Richards DA, Gilbody S, et al. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. BMJ 2013;346:f540 [FREE Full text] [Medline: 23444423]
58. El Alaoui S, Ljótsson B, Hedman E, Kaldo V, Andersson E, Rück C, et al. Predictors of Symptomatic Change and Adherence in Internet-Based Cognitive Behaviour Therapy for Social Anxiety Disorder in Routine Psychiatric Care. PLoS One 2015;10(4):e0124258 [FREE Full text] [doi: 10.1371/journal.pone.0124258] [Medline: 25893687]
59. Warmerdam L, Van SA, Twisk J, Cuijpers P. Predicting outcome of Internet-based treatment for depressive symptoms. Psychother Res 2013;23(5):559-567. [doi: 10.1080/10503307.2013.807377] [Medline: 23848944]
60. Høifødt RS, Mittner M, Lillevoll K, Katla S, Kolstrup N, Eisemann M, et al. Predictors of Response to Web-Based Cognitive Behavioral Therapy With High-Intensity Face-to-Face Therapist Guidance for Depression: A Bayesian Analysis. J Med Internet Res 2015 Sep 02;17(9):e197 [FREE Full text] [doi: 10.2196/jmir.4351] [Medline: 26333818]
61. Karyotaki E, Kleiboer A, Smit F, Turner DT, Pastor AM, Andersson G, et al. Predictors of treatment dropout in self-guided web-based interventions for depression: an 'individual patient data' meta-analysis. Psychol Med 2015 Oct;45(13):2717-2726. [doi: 10.1017/S0033291715000665] [Medline: 25881626]
62. Bergman Nordgren L, Carlbring P, Linna E, Andersson G. Role of the working alliance on treatment outcome in tailored internet-based cognitive behavioural therapy for anxiety disorders: randomized controlled pilot trial. JMIR Res Protoc 2013 Jan 18;2(1):e4 [FREE Full text] [doi: 10.2196/resprot.2292] [Medline: 23612437]
63. Mol M, Dozeman E, Provoost S, van Schaik A, Riper H, Smit JH. Behind the Scenes of Online Therapeutic Feedback in Blended Therapy for Depression: Mixed-Methods Observational Study. J Med Internet Res 2018 May 03;20(5):e174 [FREE Full text] [doi: 10.2196/jmir.9890] [Medline: 29724708]
64. Wierzbicki M, Pekarik G. A meta-analysis of psychotherapy dropout. Professional Psychology: Research and Practice 1993;24(2):190-195. [doi: 10.1037/0735-7028.24.2.190]
65. Melville KM, Casey LM, Kavanagh DJ. Dropout from Internet-based treatment for psychological disorders. Br J Clin Psychol 2010 Nov;49(Pt 4):455-471. [doi: 10.1348/014466509X472138] [Medline: 19799804]
66. El Alaoui S, Ljótsson B, Hedman E, Svanborg C, Kaldo V, Lindefors N. Predicting Outcome in Internet-Based Cognitive Behaviour Therapy for Major Depression: A Large Cohort Study of Adult Patients in Routine Psychiatric Care. PLoS One 2016;11(9):e0161191 [FREE Full text] [doi: 10.1371/journal.pone.0161191] [Medline: 27618548]
67. Hadjistavropoulos HD, Pugh NE, Hesser H, Andersson G. Predicting Response to Therapist-Assisted Internet-Delivered Cognitive Behavior Therapy for Depression or Anxiety Within an Open Dissemination Trial. Behav Ther 2016 Mar;47(2):155-165. [doi: 10.1016/j.beth.2015.10.006] [Medline: 26956649]

68. Nordgreen T, Havik OE, Ost LG, Furmark T, Carlbring P, Andersson G. Outcome predictors in guided and unguided self-help for social anxiety disorder. Behav Res Ther 2012 Jan;50(1):13-21. [doi: 10.1016/j.brat.2011.10.009] [Medline: 22134140]

Abbreviations

FQ: Fear Questionnaire
GAD-7: Generalized Anxiety Disorder-7
iCBT: Internet-based cognitive behavioral therapy
OR: odds ratio
PHQ-9: Patient Health Questionnaire-9
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 01.05.18; peer-reviewed by J Apolinário-Hagen, L Li; comments to author 17.08.18; revised version received 11.09.18; accepted 12.09.18; published 28.11.18.

Please cite as:
Mathiasen K, Riper H, Andersen TE, Roessler KK
Guided Internet-Based Cognitive Behavioral Therapy for Adult Depression and Anxiety in Routine Secondary Care: Observational Study
J Med Internet Res 2018;20(11):e10927
URL: http://www.jmir.org/2018/11/e10927/
doi:10.2196/10927
PMID:30487118

Original Paper

Defining and Predicting Pain Volatility in Users of the Manage My Pain App: Analysis Using Data Mining and Machine Learning Methods

Quazi Abidur Rahman[1], PhD; Tahir Janmohamed[2], PEng, MBA; Meysam Pirbaglou[3], MSc; Hance Clarke[4], MD, PhD, FRCPC; Paul Ritvo[3,5], PhD; Jane M Heffernan[1], PhD; Joel Katz[3,4,5], PhD

[1]Centre for Disease Modelling, Department of Mathematics and Statistics, York University, Toronto, ON, Canada

[2]ManagingLife, Inc, Toronto, ON, Canada

[3]School of Kinesiology & Health Science, York University, Toronto, ON, Canada

[4]Department of Anesthesia and Pain Management, Toronto General Hospital, Toronto, ON, Canada

[5]Department of Psychology, York University, Toronto, ON, Canada

Corresponding Author:
Tahir Janmohamed, PEng, MBA
ManagingLife, Inc
Unit 4, 850 Richmond Street W
Toronto, ON, M6J1C9
Canada
Phone: 1 4169103760
Email: tahir@managinglife.com

Abstract

Background: Measuring and predicting pain volatility (fluctuation or variability in pain scores over time) can help improve pain management. Perceptions of pain and its consequent disabling effects are often heightened under the conditions of greater uncertainty and unpredictability associated with pain volatility.

Objective: This study aimed to use data mining and machine learning methods to (1) define a new measure of pain volatility and (2) predict future pain volatility levels from users of the pain management app, Manage My Pain, based on demographic, clinical, and app use features.

Methods: Pain volatility was defined as the mean of absolute changes between 2 consecutive self-reported pain severity scores within the observation periods. The *k*-means clustering algorithm was applied to users' pain volatility scores at the first and sixth month of app use to establish a threshold discriminating low from high volatility classes. Subsequently, we extracted 130 demographic, clinical, and app usage features from the first month of app use to predict these 2 volatility classes at the sixth month of app use. Prediction models were developed using 4 methods: (1) logistic regression with ridge estimators; (2) logistic regression with Least Absolute Shrinkage and Selection Operator; (3) Random Forests; and (4) Support Vector Machines. Overall prediction accuracy and accuracy for both classes were calculated to compare the performance of the prediction models. Training and testing were conducted using 5-fold cross validation. A class imbalance issue was addressed using a random subsampling of the training dataset. Users with at least five pain records in both the predictor and outcome periods (N=782 users) are included in the analysis.

Results: *k*-means clustering algorithm was applied to pain volatility scores to establish a threshold of 1.6 to differentiate between low and high volatility classes. After validating the threshold using random subsamples, 2 classes were created: low volatility (n=611) and high volatility (n=171). In this class-imbalanced dataset, all 4 prediction models achieved 78.1% (611/782) to 79.0% (618/782) in overall accuracy. However, all models have a prediction accuracy of less than 18.7% (32/171) for the high volatility class. After addressing the class imbalance issue using random subsampling, results improved across all models for the high volatility class to greater than 59.6% (102/171). The prediction model based on Random Forests performs the best as it consistently achieves approximately 70% accuracy for both classes across 3 random subsamples.

Conclusions: We propose a novel method for measuring pain volatility. Cluster analysis was applied to divide users into subsets of low and high volatility classes. These classes were then predicted at the sixth month of app use with an acceptable degree of

accuracy using machine learning methods based on the features extracted from demographic, clinical, and app use information from the first month.

(J Med Internet Res 2018;20(11):e12001) doi:10.2196/12001

KEYWORDS

chronic pain; pain volatility; data mining; cluster analysis; machine learning; prediction model; Manage My Pain; pain app

Introduction

Background

Digital health apps, both natively developed or Web based, are transforming how people monitor, manage, and communicate health-related information [1]. This trend has been documented in medicine [2], nursing [3], psychology [4], kinesiology [5], and nutrition [6], and multiple health concerns and diseases are being addressed [1].

Pain is one of the most prevalent health-related concerns and is among the top 3 most common reasons for seeking medical help [7]. Scientific publications of data collected from pain management apps add academic credibility to the value of digital health tools and can help both consumers and health care professionals select the right app to support their treatment plans. In a previous study [8], we applied data mining (clustering) methods to understand the engagement patterns of users from a pain management app called Manage My Pain (MMP). In that study, we divided users into 5 clusters based on their level of engagement with the app and then applied statistical methods to characterize each cluster using 6 different user attributes (eg, gender, age, number of pain conditions, number of medications, pain severity, and opioid use).

In an extension of previous work, our aim is to develop prediction models that can be used to identify and predict groups of users who report improvements or decrements in their pain experience. An important question in this effort pertains to the most appropriate statistics to use when measuring change in pain severity over time. The use of average or mean pain intensity or severity scores over time as an index of change in chronic pain has been criticized on empirical and theoretical grounds. Empirically, pain intensity among people with chronic pain tends not to change appreciably over time, given that the pain is, by definition, chronic. This is evident in treatment trials where one would expect the largest magnitude of change. For example, in a study of 1894 chronic pain patients enrolled in the Quebec Pain Registry who received state-of-the-art multidisciplinary pain treatment, a trajectory analysis showed that three-quarters of patients with moderate to severe pain intensity and pain interference scores at the start of treatment showed little to no change over a 2-year period [9]. Their pain remained relatively constant and severe (between 6/10 and 7/10) across the 24-month period. Use of average pain scores has also been criticized from a theoretical perspective in that such an approach does not account for intra- and interindividual differences over time [9,10]. To overcome these limitations, one proposed solution is to adopt different data analytical approaches such as growth mixture modeling for multivariate latent classes [9]. However, as noted above in a study that used such an approach under the ideal conditions for detecting change (ie, multidisciplinary pain treatment), the vast majority of patients did not show a change in mean pain intensity over time [9]. Similarly, within our own evaluations of the MMP database, the mean pain severity levels of most of the users did not change significantly over a 6-month period in the dataset used in this study [11]. It is important to note that the stability of mean pain scores [9,11] does not preclude the possibility that there is substantial daily intraindividual variability.

Another solution is to use a measure of change that captures fluctuation or variability in pain scores over time rather than the typical measures of central tendency (ie, mean and median) that currently dominate the pain literature. Pain volatility is an important contributor to pain experience for people with chronic pain, particularly because of its linkage with the initiation of opioid addiction [12,13]. Moreover, pain perception and consequent disability are heightened under conditions of greater uncertainty and unpredictability [14], and greater pain volatility is one of the contributors to uncertainty and unpredictability. However, as no standard definition for pain volatility exists, studies are required to evaluate the best measure of volatility and to determine the extent to which pain volatility can predict chronic pain outcomes.

Objectives

Accordingly, this study has 2 main objectives. The first is to define a new measure of pain volatility. We apply data mining (clustering) methods on this newly defined measure to differentiate between 2 levels of volatility: high and low. The second objective is to predict users' pain volatility level in the future based on the information extracted from their profile and the pain records created early in the *app* usage history. Logistic regression with ridge estimations, logistic regression with Least Absolute Shrinkage and Selection Operator (LASSO), Random Forests and Support Vector Machines (SVM) are employed to develop prediction models. The issue of class imbalance in the dataset is addressed through subsampling. Training and testing are conducted using standard 5-fold cross validation. Accuracy for the low and high volatility classes and overall accuracy are calculated to measure and compare the performance of prediction models developed in our experiments.

Methods

Manage My Pain

MMP [15], developed by ManagingLife, helps people living with pain to track their pain and functioning on a daily basis using an Android mobile phone app. As MMP was launched in 2011, >28,900 people have created an account and recorded their pain. In total, >810,000 pain episodes have been documented by users.

The central feature of MMP is the *pain record* that enables users to enter details about their pain experience. Each record contains only 1 mandatory item, a rating of pain severity using a slider on a visual analogue scale. Users have the option of completing 7 more items to more comprehensively describe their pain experience. The app issues daily reminders and prompts users to reflect on their daily accomplishments through a *daily reflection*. The completion of the daily reflection and a pain record typically takes less than 1 min to complete. With regular use, users are empowered and gain self-awareness through charts and graphs that provide insight into their pain and functioning and how it changes over time.

The information collected by the app can be summarized into a report intended for clinical use. These reports present information collected by the app in a concise fashion, primarily focusing on changes in the self-reported outcome data between clinical visits. Output is structured on a single page and tends to be more accurate than a patient's recollection of pain since the last clinical visit, as it captures pain closer to the time of experience and is less influenced by recency and recall biases that plague existing methods for capturing pain information [9]. To supplement the information presented in the reports, users can add pain conditions, gender, age, and medications to their profile in the app. Users have the ability to use MMP without creating an account in which case data do not leave the device and are, therefore, not accessible for research such as this study.

Procedure

This study was reviewed and approved by the research ethics board at York University (Human Participants Review Committee, Certificate #e2015-160). The users' database was accessed and downloaded in 2 separate files (using plain text format): (1) user information and (2) pain records. The user information file contains the following fields: user ID, age at date of registration, gender, self-reported pain conditions, and self-reported medications. The pain record file contains the following fields: user ID of creator, date, severity, locations, other associated symptoms, characteristics, effective factors, ineffective factors, aggravating factors, environments, pain type, and pain duration. All fields in the text files are delimited using special characters. The files used in this study were downloaded on July 19, 2018. This study covers pain records entered by users between January 01, 2013 and July 19, 2018.

Data

The primary dataset includes 812,548 pain records from 28,952 users. The outcome period for predicting pain volatility is the sixth month of app usage. The sixth month was chosen as the outcome period as pain lasting at least 6 months meets most generally accepted definitions of chronic pain [16]. In this study, we used the first month as the predictor period, and thus, we collected features from the first month of engagement with MMP to predict pain volatility during the sixth month of engagement with MMP. The mathematical minimum for calculating pain volatility is 2 pain records with severity ratings. However, to increase the reliability of prediction results, users with at least five pain records in both the predictor and outcome periods were considered for prediction experiments in this study. The number of users in the primary dataset that meet this criterion is 795. However, 13 users who reported *other* as their gender were excluded because of the small sample size. Thus, 782 users are selected for this study, and they have 329,070 pain records in the dataset.

Pain Volatility

The most intuitive definition of pain volatility is the SD of pain severity ratings over time. We propose a new definition of pain volatility in this study. We define pain volatility as the mean of absolute changes between 2 consecutive pain severity ratings within each of the 2 observation periods regardless of elapsed time between pain ratings. Therefore, for a series of pain severity ratings $R=<R_1, R_2,\ldots, R_n>$, volatility, $V(R)$ is defined as:

$$V(R) = (|R_2-R_1|+|R_3-R_2|+\ldots+|R_n-R_n-1|) / n \quad [1]$$

The differences between the mean of absolute changes as a measure of volatility and the SD measure of volatility are demonstrated in Figure 1 and Table 1 using 4 different pain scenarios. We expect the measure of volatility to demonstrate reductions (of volatility) in the following order: sample 1 (big changes), sample 2 (small changes), sample 3 (steady upward), and sample 4 (consistent and unchanging). Pain volatility defined as the mean of absolute changes conforms to this order. However, when pain volatility is defined using the SD, sample 3 (steady upward) has a higher value than that of sample 2 (small changes). From a conceptual perspective, a steady upward pattern, although conceivably distressing for a person in pain, does not conform to what we mean by pain volatility, which, by definition, involves fluctuations in pain whether from a consistent baseline (sample 1 and sample 2) or superimposed on an upward or downward trend.

Volatility can be experienced as particularly troublesome when pain severity fluctuates over time and the mean of absolute changes approximates a *saw-tooth* pattern of volatility. This pattern has previously been identified as significant in illness conditions such as atrial flutter, where hemodynamic instability can progress to ventricular fibrillation (when the heart quivers irregularly leading to an elevated risk of potentially life-threatening cardiac events) [17]. Although it remains unclear whether a saw-tooth pattern of pain volatility is more debilitating than a steady upward pattern, researchers require measurement methods elucidating both patterns to better explore associated effects on functionality and quality of life.

The next step in testing this volatility measure was to divide users into 2 distinct classes: high volatility and low volatility using a threshold on the pain volatility measure. We applied a clustering method to identify this threshold. Clustering involves partitioning a set of objects or members of a defined population into 2 or more subgroups such that the members of 1 subgroup are similar to each other but dissimilar to members of the other subgroup(s). Each object or subgroup member is represented using 1 or more variables for the purpose of clustering, which are typically referred to as features or attributes. The similarity or dissimilarity between pairs of objects (or subgroup members) is measured as the distance between the feature vectors representing them.

Figure 1. Demonstration of 4 different patterns of pain severity over time.

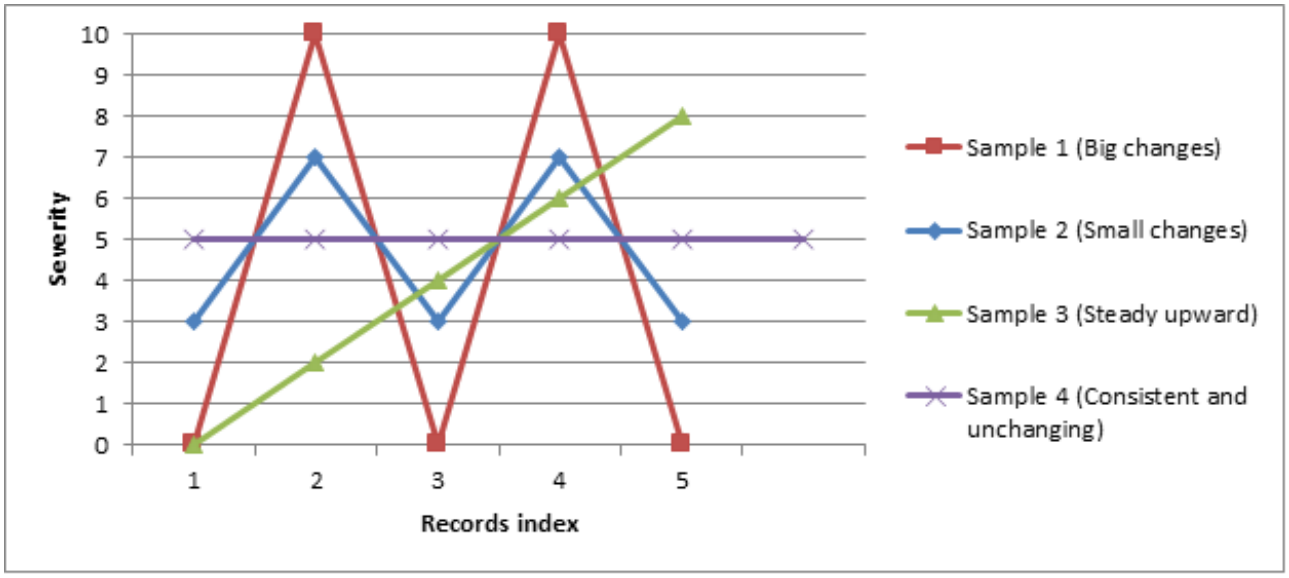


Table 1. Comparing SD of severity ratings and mean of absolute changes as pain volatility measures.

Volatility trajectory	SD of severity ratings	Mean of absolute changes
Sample 1 (Big changes)	5.48	10
Sample 2 (Small changes)	2.19	4
Sample 3 (Steady upward)	3.16	2
Sample 4 (Consistent and unchanging)	0.00	0

The output of a successful clustering process is a set of clusters where each object is assigned membership in one of the candidate clusters. We used the method known as *k*-means [18] as our primary data analytic approach to clustering users. Under the *k*-means clustering method, the number of clusters is set a priori to some constant *k*, and the dataset is partitioned into *k* clusters. In the initialization stage, the *k*-means are selected at random. Each item in the dataset is assigned to the mean closest to it. In each subsequent iteration, for each cluster, the mean is calculated based on the current members of that cluster. Each data point is then reassigned to the cluster whose mean is the closest. The iterative process stops when the cluster membership does not change between iterations.

In our experiments, the feature for clustering users is the pain volatility measure (ie, the mean of absolute changes in pain severity). We clustered the users into 2 clusters, and the volatility measure that divides the 2 groups of users was used as the threshold for defining 2 distinct classes of users: high volatility and low volatility.

Features for Prediction Model

To develop the prediction model, we extracted the following 130 features from each of 782 users:

1. *Gender (1 feature)*: The options for entering gender in the app are male, female, or other. Users who did not include their gender information were coded as unknown. In all, 25% of users belong to this category. There were only 13 users who reported *other* as the gender. They were excluded from further analysis because of small sample size, as mentioned before in the Data subsection.
2. *Age (1 feature)*: The age (in years) recorded is the age of the user on the date of the first record and not as of the date of the analysis. We categorized the age values to facilitate the analysis and added a category to account for users with missing information. Moreover, 31% of users did not provide their date of birth. The age values are divided into 8 categories: (1) unknown, (2) >0 and ≤20, (3) >20 and ≤30, (4) >30 and ≤40, (5) >40 and ≤50, (5) >50 and ≤60, (6) >60 and ≤70, and (7) >70.
3. *Number of self-reported pain conditions (1 feature)*: Users can add 1 or more pain conditions to their profile from a centralized list of over 2500 pain conditions. They can also choose to define their own pain condition if they are unable to find one from the centralized list. Some users did not choose to add a pain condition to their profile. The number of self-reported pain conditions was divided into 5 categories: (1) unknown, (2) 1 condition, (3) 2 conditions, (4) 3 conditions, and (5) more than 3 conditions.
4. *Categories of self-reported pain conditions (5 features)*: Many of the self-reported pain conditions fit into 1 of the following 5 categories: fibromyalgia, headaches, back pain, arthritis, and depression-anxiety. Each self-reported pain condition was mapped to the appropriate category as applicable, and the mapping was reviewed for clinical correctness. For each of these 5 categories, a flag feature was created to indicate if the user has self-reported a pain condition in their profile that corresponds to the category.

5. *Pain record entries (2 features)*: A total of 2 features were used to record number of pain records in the predictor period and the number of days in the predictor period when a user has recorded at least one pain record.
6. *Pain severity rating (3 features)*: The app user must choose a pain severity rating (0-10) for each pain record created. For each user, we calculated the mean and SD of pain severity ratings from the user's records in their predictor period. All users were also assigned to 1 of the following 3 groups based on their mean pain ratings: mild (average pain rating <4), moderate (average pain rating ≥4 to ≤7), or severe (average pain rating >7) [11]. The mean and SD of severity ratings and the severity level grouping (mild or moderate or Severe) were used as features.
7. *Change in pain trend (1 feature)*: A trend line was fitted through the pain severity ratings using linear regression. The difference in pain severity ratings between the end point and the starting point of this trend line was used as a feature.
8. *Pain volatility (2 features)*: Pain volatility in the predictor period, that is, the mean of the absolute changes between each 2 consecutive pain ratings, was used as a feature. Each user was also assigned a level of pain volatility (low or high) based on the threshold established using the clustering approach described in the previous section. This volatility level in the predictor period was used as a feature.
9. *Pain descriptors (64 features)*: For each pain record created in the app, users can report pain locations (eg, the head, abdomen, and back), associated symptoms (eg, dizziness and fever), pain characteristics (eg, burning and cramping), and environment (eg, home and school). Users can choose from a list of default values in each section: 24 pain locations, 20 associated symptoms, 13 characteristics, and 7 environments. For each of these default values, we created a flag feature indicating its presence in any of the pain records in the predictor period. Thus, there are total 64 features in this category. Only 2% of users did not report any of these pain descriptors.
10. *Factors impacting pain (43 features)*: Users in the app can report factors that may have an impact on their pain experience. A total of 3 types of factors are listed in the app: aggravating (eg, sitting and exercise), alleviating (eg, rest and sleep), and ineffective (eg, rest and sleep). Users can choose from a list of default factors in each section: 15 aggravating, 14 alleviating, and 14 ineffective. For each of these default factors, we created a flag feature indicating its presence in any of the pain records in the predictor period, resulting in 43 features in this category. In our dataset, 8% of users did not include any factor impacting their pain.
11. *Medication (5 features)*: Users can add medications to their profile from a standardized list of over 1130. Any medication in a user's profile can be added to a pain record as an aggravating, effective, or ineffective factor. A total of 5 common categories of pain medication are identified: opioids, tricyclic antidepressants, anticonvulsants, cannabinoids, and serotonin-norepinephrine reuptake inhibitors. Medications from the standardized list are mapped to the appropriate categories. For each of these 5 categories, we created a flag feature indicating the presence of any medication that belongs to the category in any of the pain records in the predictor period. Thus, 5 features are added from the medication category.
12. *Neuropathic pain (1 feature)*: We added a flag feature as the indicator of neuropathic pain. Neuropathic pain is indicated if a user has at least two of the following in a pain record's characteristics: pins and needles or tingling, burning, numbness, electric shocks, and light touch or clothing (aggravating factor).
13. *Mental health issues (1 feature)*: Mental health issues are indicated if a user has reported at least one of the following symptoms in a pain record: anxiety or depression (associated symptom) or negative mood or stress (aggravating factors) [19]. A flag feature was created to indicate if at least one pain record in the predictor period meets this criterion.

Prediction Models

We first developed a logistic regression model with ridge estimators for prediction [20]. We then modified the model using LASSO [21]. These 2 logistic regression methods aim to shrink large regression coefficients to avoid overfitting. By constraining the sum of the absolute values of the coefficients, LASSO forces some coefficients to be 0 and, as such, the number of features used in the model reduces. The R package glmnet was used for training and testing logistic regression models [22,23].

We then employed 2 machine learning classifiers to build prediction models for pain volatility: Random Forests [24] and SVM [25]. Random Forests and SVM have been widely used in biomedicine for classification and prediction [26-29]. Random Forests forms an ensemble classifier based on a collection of decision trees learned from multiple random samples taken from the training set. Decision tree classifiers are constructed using the information content of each attribute; thus, the decision tree learning algorithms first select the most informative attributes for classification. Random samples from the training dataset are selected uniformly, with replacement, such that the total size of each random sample is the same as the size of the whole training set. To predict the class of a new instance, each decision tree is applied to the instance, and the final classification decision is made by taking a majority vote over all the decision trees. We applied the standard Random Forests classification package in Weka [30] using 100 trees in the Random Forests implementation. The number of features selected at random at each tree node was set to $2\sqrt{n}$, where n is the total number of features.

The other method, SVM, is primarily a binary linear classifier. A hyperplane is learned from the training dataset in the feature space to separate the training instances for classification. The hyperplane is constructed such that the margin, that is, the distance between the hyperplane and the data points nearest to it, is maximized. If the training instances are not linearly separable, these can be mapped into a high-dimensional space to find a suitable separating hyperplane. In our experiments, we used the Weka libsvm, employing the Gaussian radial basis function kernel.

Measuring Prediction Performance

We used the stratified 5-fold cross-validation procedure for training the models and then testing the prediction performance. In this procedure, both low and high pain volatility users are partitioned into 5 equal-sized groups. One of these groups is used as a test set, whereas the other 4 were used to train the models and classifiers. This is repeated for each of these 5 groups. Thus, we conducted the prediction experiments 5 times, and each time, the training and test sets were completely separate. Through this cross-validation procedure, each user's pain volatility class is tested exactly once. We measure the prediction performance of the methods used in this study by the following 3 measures:

Accuracy of the low volatility class = (Number of correctly predicted low volatility users/Total number of low volatility users) × 100% [2]

Accuracy of the high volatility class = (Number of correctly predicted high volatility users/Total number of high volatility users) × 100% [3]

Overall accuracy = (Number of correctly predicted low and high volatility users/Total number of users) × 100% [4]

Class Imbalance

After defining the low and high volatility classes using the clustering approach, the number of low volatility users is much higher (almost 3 times) than that of high volatility users, as discussed in the Results section. This class imbalance in the dataset produces high accuracy for the majority class (low volatility) in the prediction experiments, whereas the accuracy of the minority class (high volatility) remains very low. We used the procedure of subsampling from the majority class to create a balanced dataset for training prediction models. Under the subsampling method, instances are chosen at random from the majority class to make the size of the 2 classes equal. We repeated the subsampling procedure 3 times to ensure stability of the results. We conducted prediction experiments both on the original and the balanced dataset.

Results

Pain Volatility Classes

We combined the pain volatility measures of all users from the predictor period (ie, the first month of app usage) and the outcome period (ie, the sixth month of app usage) and then divided these data into 2 clusters using the *k*-means algorithm. Figure 2 shows the clustering output. There are a total of 1564 data points as each user has 2 values: 1 from the predictor and 1 from the outcome period. The first 782 data points (indices 1-782) are volatility values from the predictor period and the next 782 are from the outcome period. The black and red colors indicate 2 distinct classes (low and high volatility, respectively), and the numerical threshold dividing these 2 volatility classes is approximately 1.6.

To further validate this threshold, we randomly chose subsamples of 782 values from the total of 1564 and reapplied the clustering algorithm. We repeated this procedure 4 times. Figure 3 shows these 4 clustering results. The threshold of 1.6 is consistent across all these 4 random subsamples. Hence, users having a volatility measure greater than 1.6 are assigned the class of *high* in our prediction experiments. All other users belong to the volatility class of *low*.

Figure 2. Clustering pain volatility measures. Total number of data points is 1564. Each user has 2 data points, 1 each from the predictor and outcome periods. Data points with index (x-axis) 1 to 782 are volatility values from the predictor period and 783 to 1564 are from the outcome period. Black and red colors indicate low and high volatility levels, respectively.

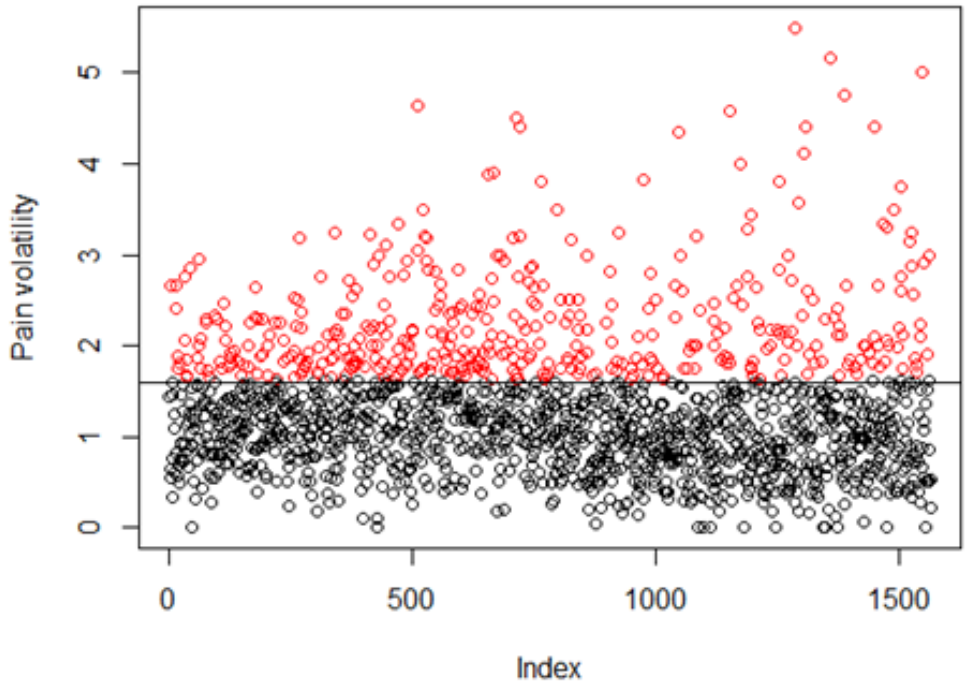
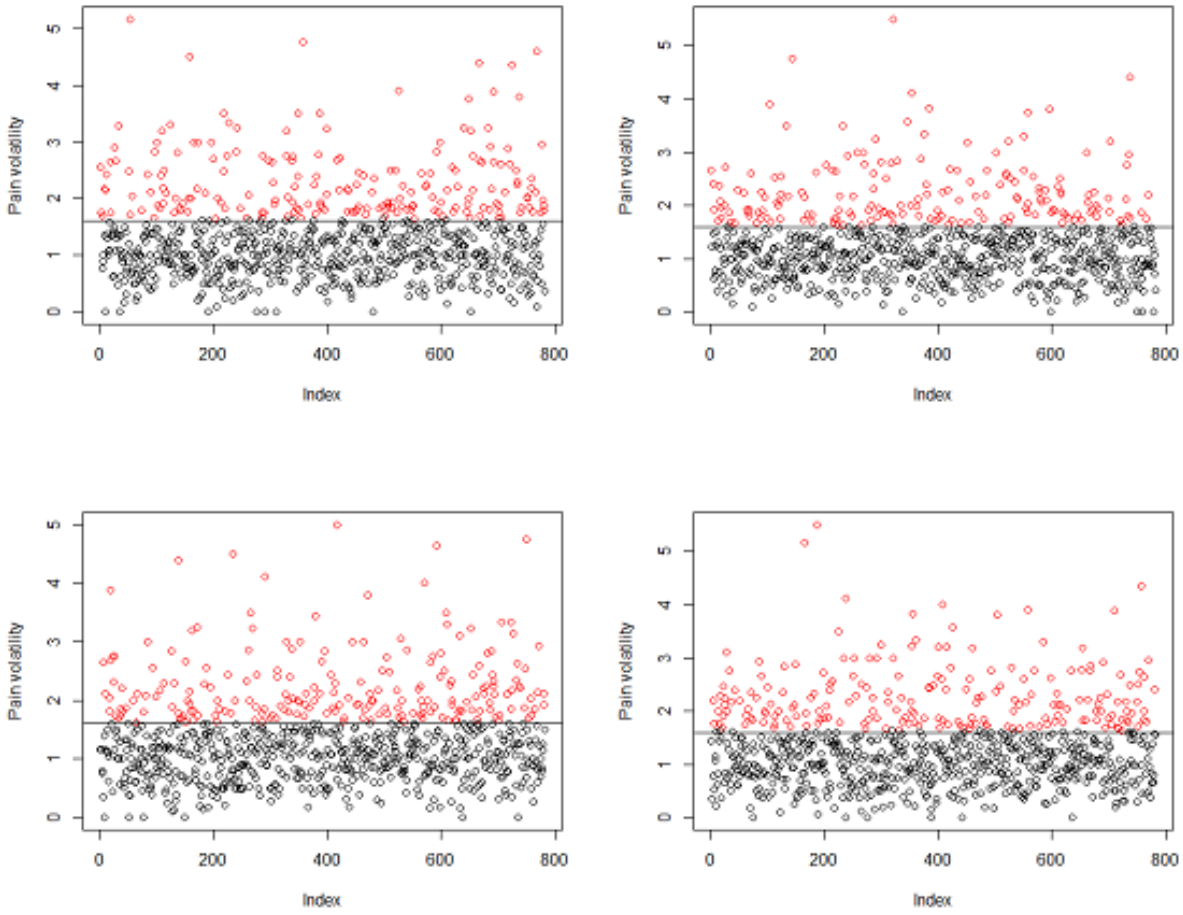


Figure 3. Clustering randomly selected subsets of pain volatility measures.



Prediction Results

Using the pain volatility threshold of 1.6 resulted in the following division of users in the outcome period: 611 had low volatility and 171 had high volatility. There is an obvious class imbalance in the dataset as the number of low volatility users is more than 3 times the number of high volatility users. We first applied logistic regression with ridge estimators and LASSO, Random Forests, and SVM on the original dataset of 782 users. The prediction performance of 4 methods using 5-fold cross validation is presented in Table 2.

All 4 methods achieved 78.1% (611/782) to 79.0% (618/782) overall accuracy. However, in all methods, the accuracy of the high volatility class is significantly low. Although the accuracy for the majority class (low volatility) is more than 95.9% (586/611) across the methods, the accuracy for the minority class (high volatility) is less than 18.7(32/171). We hypothesize that the lower accuracy in the high volatility class is a result of the class imbalance. To address this, as discussed in the Methods section, we randomly subsampled the low volatility class to create training sets such that the number of instances from both classes is the same. We conducted the random subsampling 3 times and reapplied all 4 methods for prediction. The results are shown in Table 3 and Figure 4.

Table 2. Prediction performance using the original dataset of 782 users.

Performance measure	Logistic regression (ridge), n (%)	Logistic regression (LASSO[a]), n (%)	Random Forests, n (%)	SVM[b], n (%)
Accuracy (low volatility class; N=611)	610 (99.8)	607 (99.3)	587 (96.1)	605 (99.0)
Accuracy (high volatility class; N=171)	1(0.6)	10 (5.3)	31 (18.1)	2 (1.2)
Overall accuracy (N=782)	611 (78.1)	617 (79.0)	618 (79.0)	607 (77.6)

[a]LASSO: Least Absolute Shrinkage and Selection Operator.

[b]SVM: Support Vector Machines.

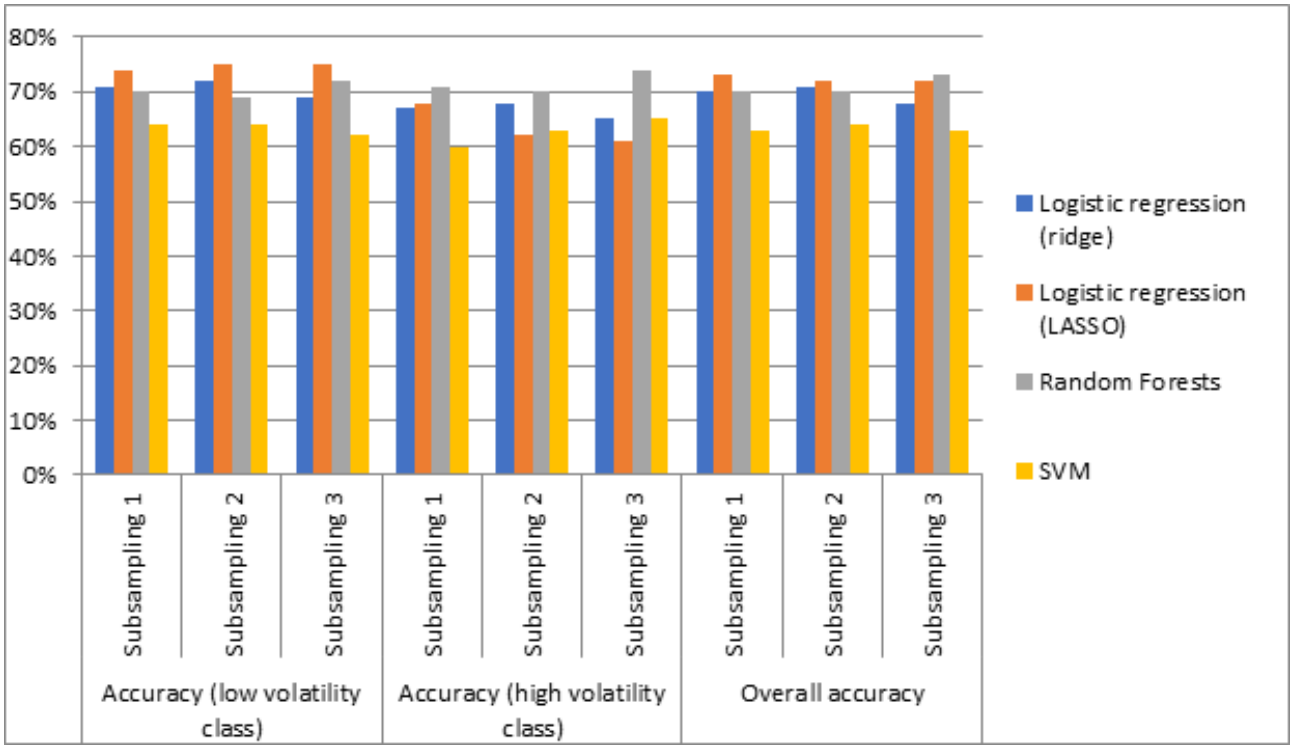
Table 3. Prediction performance using the balanced dataset where random subsampling of the majority class (low volatility) was applied to make class sizes equal in the training dataset.

Performance Measure	Logistic regression (ridge), n (%)	Logistic regression (LASSO[a]), n (%)	Random Forests, n (%)	SVM[b], n (%)
Accuracy (low volatility class; N=611)				
Subsampling 1	433 (70.9)	455 (74.5)	428 (70.0)	391 (64.0)
Subsampling 2	442 (72.3)	460 (75.3)	424 (69.4)	391 (64.0)
Subsampling 3	424 (69.4)	456 (74.6)	440 (72.0)	379 (62.0)
Accuracy (high volatility class; N=171)				
Subsampling 1	115 (67.3)	116 (67.8)	121 (70.8)	103 (60.2)
Subsampling 2	116 (67.8)	106 (62.0)	120 (70.2)	103 (63.2)
Subsampling 3	111 (64.9)	105 (61.4)	127 (74.3)	111 (64.9)
Overall accuracy (N=782)				
Subsampling 1	548 (70.1)	571 (73.0)	549 (70.2)	494 (63.2)
Subsampling 2	558 (71.4)	566 (72.4)	544 (69.6)	499 (63.8)
Subsampling 3	535 (68.4)	561 (71.7)	567 (72.5)	490 (62.7)

[a]LASSO: Least Absolute Shrinkage and Selection Operator.

[b]SVM: Support Vector Machines.

Figure 4. Prediction performance using the balanced dataset. LASSO: Least Absolute Shrinkage and Selection Operator; SVM: Support Vector Machines.



The overall accuracy is between 68.4% (535/782) and 73% (571/782) for Random Forests and logistic regression models. These 3 methods perform much better than SVM. Although the overall accuracy of the prediction model is reduced to some extent after balancing the dataset (Table 2 vs Table 3), the accuracy for the high volatility class is significantly improved. All methods have less than 18.7% (32/171) accuracy for the high volatility class using the original class imbalanced dataset. However, after random subsampling of the majority class, this improves to at least 60.2% (103/171). All 3 accuracy measures are approximately 70% across 3 subsamples using Random Forests. Although logistic regression models perform slightly better than Random Forests for the low volatility class, Random Forests performs better for the high volatility class. This is the only method that achieves approximately 70% accuracy consistently for both volatility classes across different subsamples. Thus, Random Forests performs the best in predicting the level of pain volatility of MMP users at the sixth

month of usage based on the features collected from profile information and pain records in the first month of usage.

Discussion

Principal Findings

In this study, we defined a new pain volatility measure. We employed clustering methods on this measure to distinguish between low and high levels of pain volatility. Subsequently, we predicted pain volatility levels in users of MMP, a digital health app for recording pain experiences. We extracted 130 features from the users' profile information and pain history in the first month of their app usage. These features were used to build prediction models where the outcome was the level of pain volatility in the sixth month. A total of 4 methods were used to develop prediction models: logistic regression with ridge estimators, logistic regression with LASSO, Random Forests, and SVM. We addressed the issue of class imbalance by random subsampling of the training dataset and repeated this procedure 3 times. The prediction model developed using Random Forests performs the best, and the accuracy level achieved for both low and high volatility classes is approximately 70%.

Major Contributions

Although recent years have seen increased interest in applying machine learning methodologies in the study of chronic pain [31,32], this is the first study of its kind that aims to define and predict chronic pain volatility using data mining and machine learning methods. The results of our study are important for several reasons. First, the study involved the use of a large dataset based on real-world data from people with pain who autonomously use the app. This contrasts with the typical ways in which data are gathered by pain researchers, namely, through randomized clinical trials, surveys, and prospective trials, during which the researchers actively seek out participants. Data gathered from real-world sources have an important and complementary role to play in outcomes research and health care delivery [33]. Second, use of data from MMP, a digital health app for monitoring and tracking pain, is consistent with a recent trend in mobile health showing that similar apps are transforming how people monitor, manage, and communicate health-related information [1,2]. Third, and perhaps the most important, results from this study show that by using features of the dataset extracted over a 1-month period, we could predict pain volatility 6 months later, with a reasonably high degree of accuracy. Although the typical approach of using average pain scores may seem adequate for evaluating patterns of incrementally increasing or decreasing pain, these methods are not useful for evaluating a saw-tooth-like volatility pattern. In this study, we have explored a method that appears to reflect the quantitative levels of an important volatility pattern.

There are clinical implications to this study. Should this study on pain volatility be corroborated and shown to be a valid and reliable concept, we will be in a position to begin to identify risk factors for heightened volatility and, therefore, to potentially prevent the development of high pain volatility through effective interventions. That is, we will be able to predict patients at high risk of developing high pain volatility and the downstream negative consequences of such volatility (eg, poorer quality of life, psychosocial distress, and increased pain disability). At present, the MMP app is used for tracking and monitoring pain, and users are able to plot their pain scores as a function of time. Should pain volatility be shown to be an important, valid, and reliable construct, the app might be modified to allow users to track and plot pain volatility.

Future Work

In future, we shall focus on selecting a subset of features that are significant predictors of pain volatility. Reducing the size of the feature set will make the prediction model easier to interpret. Furthermore, in consultation with pain experts on our team and in the broader pain community, we will validate this reduced feature set. This validated subset of features may lead to an improvement in the accuracy of prediction models as redundant features are removed. The remaining set of predictors of heightened pain volatility will be evaluated for modifiability and causality and targeted through clinical trials aimed at reducing pain volatility.

Acknowledgments

QAR is supported by Mitacs. JK is supported by a Tier 1 Canadian Institutes of Health Research Canada Research Chair in Health Psychology at York University. HC is supported by a Merit award from the University of Toronto, Department of Anesthesia. JMH is a York University Research Chair.

Conflicts of Interest

TJ is the founder and CEO of ManagingLife, Inc. JK and HC are unpaid members of the ManagingLife Advisory Board, providing guidance on the product and the company's research initiatives.

References

1. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: a systematic review. J Telemed Telecare 2016 Oct 18:22-30. [doi: 10.1177/1357633X16673538] [Medline: 27760883]
2. Ramirez V, Johnson E, Gonzalez C, Ramirez V, Rubino B, Rossetti G. Assessing the use of mobile health technology by patients: an observational study in primary care clinics. JMIR Mhealth Uhealth 2016;4(2):e41 [FREE Full text] [doi: 10.2196/mhealth.4928] [Medline: 27095507]

3. Stinson JN, Jibb LA, Nguyen C, Nathan PC, Maloney AM, Dupuis LL, et al. Construct validity and reliability of a real-time multidimensional smartphone app to assess pain in children and adolescents with cancer. Pain 2015 Dec;156(12):2607-2615. [doi: 10.1097/j.pain.0000000000000385] [Medline: 26580680]
4. 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]
5. 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 [FREE Full text] [doi: 10.2196/jmir.4871] [Medline: 26441467]
6. Rollo ME, Aguiar EJ, Williams RL, Wynne K, Kriss M, Callister R, et al. eHealth technologies to support nutrition and physical activity behaviors in diabetes self-management. Diabetes Metab Syndr Obes 2016;9:381-390 [FREE Full text] [doi: 10.2147/DMSO.S95247] [Medline: 27853384]
7. St Sauver JL, Warner DO, Yawn BP, Jacobson DJ, McGree ME, Pankratz JJ, et al. Why patients visit their doctors: assessing the most prevalent conditions in a defined American population. Mayo Clin Proc 2013 Jan;88(1):56-67 [FREE Full text] [doi: 10.1016/j.mayocp.2012.08.020] [Medline: 23274019]
8. Rahman QA, Janmohamed T, Pirbaglou M, Ritvo P, Heffernan JM, Clarke H, et al. Patterns of user engagement with the mobile app, Manage My Pain: results of a data mining investigation. JMIR Mhealth Uhealth 2017 Jul 12;5(7):e96 [FREE Full text] [doi: 10.2196/mhealth.7871] [Medline: 28701291]
9. Pagé MG, Romero Escobar EM, Ware MA, Choinière M. Predicting treatment outcomes of pain patients attending tertiary multidisciplinary pain treatment centers: a pain trajectory approach. Can J Pain 2017 Aug 4;1(1):61-74. [doi: 10.1080/24740527.2017.1325715]
10. Pillai Riddell R, Flora DB, Stevens SA, Stevens B, Cohen LL, Greenberg S, et al. Variability in infant acute pain responding meaningfully obscured by averaging pain responses. Pain 2013 May;154(5):714-721. [doi: 10.1016/j.pain.2013.01.015] [Medline: 23531475]
11. Rahman QA, Janmohamed T, Azam MA, Clarke H, Heffernan JM, Ritvo P, et al. F1000research. Predicting users' future pain experience based on the engagement with Manage My Pain, an mHealth app URL: https://f1000research.com/posters/7-967 [accessed 2018-11-01] [WebCite Cache ID 73bm4bFmp]
12. Worley MJ, Heinzerling KG, Shoptaw S, Ling W. Pain volatility and prescription opioid addiction treatment outcomes in patients with chronic pain. Exp Clin Psychopharmacol 2015 Dec;23(6):428-435 [FREE Full text] [doi: 10.1037/pha0000039] [Medline: 26302337]
13. Worley MJ, Heinzerling KG, Shoptaw S, Ling W. Volatility and change in chronic pain severity predict outcomes of treatment for prescription opioid addiction. Addiction 2017 Jul;112(7):1202-1209 [FREE Full text] [doi: 10.1111/add.13782] [Medline: 28164407]
14. Bélanger C, Blais Morin B, Brousseau A, Gagné N, Tremblay A, Daigle K, et al. Unpredictable pain timings lead to greater pain when people are highly intolerant of uncertainty. Scand J Pain 2017 Oct;17:367-372. [doi: 10.1016/j.sjpain.2017.09.013] [Medline: 29033299]
15. ManagingLife. Manage My Pain URL: https://managinglife.com/ [accessed 2018-07-29] [WebCite Cache ID 71HKRTjFq]
16. Merskey H, Bogduk N, editors. Classification of chronic pain. Descriptions of chronic pain syndromes and definitions of pain terms. Seattle, WA: IASP Press; 1994.
17. Cosío FG. Atrial flutter, typical and atypical: a review. Arrhythm Electrophysiol Rev 2017 Jun;6(2):55-62 [FREE Full text] [doi: 10.15420/aer.2017.5.2] [Medline: 28835836]
18. Kaufman L, Rousseeuw P. Finding Groups in Data: An Introduction to Cluster Analysis. Hoboken, NJ: Wiley-Interscience; 2005.
19. Portenoy R. Development and testing of a neuropathic pain screening questionnaire: ID Pain. Curr Med Res Opin 2006 Aug;22(8):1555-1565. [doi: 10.1185/030079906X115702] [Medline: 16870080]
20. Le Cessie S, Van Houwelingen CV. Ridge estimators in logistic regression. Applied Statistics 1992;41(1):191. [doi: 10.2307/2347628]
21. Tibshirani R. Regression shrinkage and selection via the Lasso. J R Statist Soc B 1996;58(1):267-288. [doi: 10.2307/2346178]
22. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33(1):1-22 [FREE Full text] [Medline: 20808728]
23. Friedman J, Hastie T, Tibshirani R, Simon N, Narasimhan B, Qian J. CRAN.r-project. 2018. Package 'glmnet': Lasso and Elastic-Net Regularized Generalized Linear Models URL: https://CRAN.R-project.org/package=glmnet [accessed 2018-07-26] [WebCite Cache ID 71CTJZrEC]
24. Breiman L. Random Forests. Mach learn 2001 Oct;45(1):5-32 [FREE Full text] [doi: 10.1023/A:1010933404324]
25. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995 Sep;20(3):273-297. [doi: 10.1007/BF00994018]
26. Rahman QA, Tereshchenko LG, Kongkatong M, Abraham T, Abraham MR, Shatkay H. Utilizing ECG-based heartbeat classification for hypertrophic cardiomyopathy identification. IEEE Trans Nanobioscience 2015 Jul;14(5):505-512 [FREE Full text] [doi: 10.1109/TNB.2015.2426213] [Medline: 25915962]

27. Ellis K, Kerr J, Godbole S, Lanckriet G, Wing D, Marshall S. A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. Physiol Meas 2014 Nov;35(11):2191-2203 [FREE Full text] [doi: 10.1088/0967-3334/35/11/2191] [Medline: 25340969]
28. Park E, Chang HJ, Nam HS. Use of machine leaning classifiers and sensor data to detect neurological deficit in stroke patients. J Med Internet Res 2017 Apr 18;19(4):e120 [FREE Full text] [doi: 10.2196/jmir.7092] [Medline: 28420599]
29. Lee J. Patient-specific predictive modeling using random forests: an observational study for the critically ill. JMIR Med Inform 2017 Jan 17;5(1):e3 [FREE Full text] [doi: 10.2196/medinform.6690] [Medline: 28096065]
30. Frank E, Hall MA, Witten IH. Weka Workbench. 2016. The WEKA workbench, online appendix for Data Mining: Practical Machine Learning Tools and Techniques URL: https://www.cs.waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf [accessed 2018-07-29] [WebCite Cache ID 71HQt2a5T]
31. Cheng JC, Rogachov A, Hemington KS, Kucyi A, Bosma RL, Lindquist MA, et al. Multivariate machine learning distinguishes cross-network dynamic functional connectivity patterns in state and trait neuropathic pain. Pain 2018 Sep;159(9):1764-1776. [doi: 10.1097/j.pain.0000000000001264] [Medline: 29708944]
32. Rogachov A, Cheng JC, Hemington KS, Bosma RL, Kim JA, Osborne NR, et al. Abnormal low-frequency oscillations reflect trait-like pain ratings in chronic pain patients revealed through a machine learning approach. J Neurosci 2018 Aug 15;38(33):7293-7302. [doi: 10.1523/JNEUROSCI.0578-18.2018] [Medline: 30012686]
33. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-world evidence -What is it and what can it tell us? N Engl J Med 2016 Dec 8;375(23):2293-2297. [doi: 10.1056/NEJMsb1609216] [Medline: 27959688]

Abbreviations

LASSO: Least Absolute Shrinkage and Selection Operator
MMP: Manage My Pain
SVM: Support Vector Machines

Edited by G Eysenbach; submitted 23.08.18; peer-reviewed by G Page, E Park, F Li; comments to author 21.09.18; revised version received 04.10.18; accepted 22.10.18; published 15.11.18.

Please cite as:
Rahman QA, Janmohamed T, Pirbaglou M, Clarke H, Ritvo P, Heffernan JM, Katz J
Defining and Predicting Pain Volatility in Users of the Manage My Pain App: Analysis Using Data Mining and Machine Learning Methods
J Med Internet Res 2018;20(11):e12001
URL: http://www.jmir.org/2018/11/e12001/
doi:10.2196/12001
PMID:30442636

Original Paper

An Unsupervised Smart App–Optimized HIV Self-Testing Program in Montreal, Canada: Cross-Sectional Study

Nitika Pant Pai[1,2*], MD, MPH, PhD; Megan Smallwood[2*], MSc; Laurence Desjardins[3], BA, OPSQ; Alexandre Goyette[3], BSc, RN; Krisztian G Birkas[2], BCompSci, MSc CompSci; Anne-Fanny Vassal[3], MA; Lawrence Joseph[2], PhD; Réjean Thomas[3], MD, CM, OQ, DHC

[1]Department of Medicine, McGill University, Montreal, QC, Canada

[2]Division of Clinical Epidemiology, Research Institute of McGill University Health Centre, Montreal, QC, Canada

[3]Clinique Médicale L'Actuel, Montreal, QC, Canada

[*]these authors contributed equally

Corresponding Author:
Nitika Pant Pai, MD, MPH, PhD
Division of Clinical Epidemiology
Research Institute of McGill University Health Centre
5252 Boul de Maisonneuve O
Montreal, QC, H4A 3S5
Canada
Phone: 1 5149341934 ext 44729
Email: nitika.pai@mcgill.ca

Abstract

Background: Although HIV self-testing strategies have been recommended by the World Health Organization, HIV self-tests are not yet approved in Canada. Currently approved HIV self-tests offer toll-free lines that are insufficient for initiating expedited linkages to counseling and care, accurate interpretation, and support during HIV self-testing. We developed an innovative, multilingual software app called HIVSmart! to plug these gaps.

Objective: This study aimed to test our app-optimized oral HIV self-testing strategy for feasibility in men who have sex with men (MSM) who presented to test at a large sexual health clinic (Clinique Médicale L'Actuel) in Montreal.

Methods: Between July 2016 and February 2017, we offered a strategy consisting of the OraQuick In-Home HIV Test (an investigational device) and a tablet installed with the HIVSmart! app to study participants, who presented at a private office in the clinic, mimicking an unsupervised home environment. We evaluated the strategy for its feasibility, acceptability, and preference. Using the HIVSmart! app, participants were guided through the self-testing process. We determined feasibility with a metric defined as the completion rate, which consisted of the following 3 steps: (1) self-test conduct; (2) self-test interpretation; and (3) linkages to care. Participants independently performed, interpreted, recorded their self-test and result, engaged in pre- and posttest counseling, and sought linkages to care. Laboratory tests (p24, Western Blot, and RNA), as per country algorithms, were expedited, and linkages based on the rapid test status were arranged.

Results: Mean age of the 451 participants enrolled was 34 (range, 18-73) years. Of all participants, 97.1% (438/451) completed and submitted the survey through the HIVSmart! app. In total, 84.7% (371/438) of the participants were well educated (beyond high school) and 52.5% (230/438) had been tested within the past 6 months. Of the 451, 11.5% (52/451) were on pre-exposure prophylaxis. Feasibility (completion rate), an average proportion of the 3 steps, was computed to be 96.6% (419/451). The acceptability of the strategy was high at 98.5% (451/458). A majority of the participants (448/451, 99.3%) were found to be self-tested and lab-confirmed negative and were counseled after self- and rapid tests. In total, 0.7% (3/451) of the participants who self-tested positive and were lab-confirmed positive were linked to a physician within the same day. Furthermore, 98.8% (417/422) of the participants found the app to be useful and 94.0% (424/451) were willing to recommend it to a friend or partner.

Conclusions: The HIVSmart! app-optimized strategy was feasible, accepted, and preferred by an educated, urban MSM population of Montreal. With the app, participants were able to perform, interpret, store results, and get rapidly linked to care. The HIVSmart!-optimized, self-testing strategy could be adapted and contextualized to many at-risk populations within Canada and worldwide, thereby maximizing its public health impact.

(J Med Internet Res 2018;20(11):e10258) doi:10.2196/10258

KEYWORDS

feasibility; HIV; impact; mobile phone; MSM; self-testing

Introduction

In 2014, the Joint United Nations Programme on HIV/AIDS released its 90-90-90 targets, calling for 90% of those living with HIV to be tested, 90% of those tested positive to be on treatment, and 90% on those on treatment to remain virologically suppressed [1]. To reach the first 90 by 2020, it is imperative that we identify those who are living unaware of their HIV status. One such strategy that has the potential to reach the undiagnosed is HIV self-testing. In 2016, the World Health Organization recommended the scale-up of self-testing as an additional alternative to the conventional HIV testing services [2,3]. HIV self-tests are now available for use, sale, and distribution in many countries. More than 59 countries have HIV self-testing policies in place [4-7]. The benefits of self-testing include privacy, confidentiality, convenience, engagement in self-care, empowerment, and proactivity in seeking health [8]. The potential concerns with HIV self-testing are the rapid establishment of linkages for self-testers to counseling and care, rapid reporting of false-negative results in the acute window period, and affordability of self-tests. Nonetheless, benefits of HIV self-testing are generally understood to outweigh the associated risks [8,9].

Evidence and data on self-testing in Canada are sparse [10-12]. Neither oral nor blood-based tests are approved yet by Health Canada [4]. Canadian policies on HIV self-testing are in development. Recently, we surveyed Canadian stakeholders involved in HIV testing initiatives across the country and found that many were in favor of self-testing but were concerned about linkages to care and accuracy of self-tests. Furthermore, HIV self-testing strategies were perceived to pose an economic threat to the prevailing HIV testing models in Canada [13]. While global research on the implementation of HIV self-testing has increased exponentially, few studies have explored self-testing in the Canadian context. To date, a hypothetical self-testing study used focus groups or surveys of attitudes and acceptability, while another evaluated a self-testing strategy in a low-risk student population [11,12]. Thus, implementation research evidence on HIV self-testing is needed for Canada.

In any setting, understanding the context in which self-testing should be implemented is critical to its success. In Canada, the HIV epidemic is disproportionally represented in key populations, such as men who have sex with men (MSM), injection drug users (IDUs), Aboriginal populations, and immigrants from HIV endemic countries. Approximately 18%-25% of Canadian MSM populations are unaware of their HIV-positive status [14], and the number may be proportionally higher for IDUs, Aboriginals, and immigrants, which underscores the need for accessible HIV self-testing services.

Methods

HIVSmart!

HIVSmart! (Figure 1), a Canadian innovation funded by Grand Challenges Canada [15], (which is funded by the Government of Canada), is a multiplatform smartphone-, tablet-, or Web-based (Android, iPhone, and iPad) confidential software app. This study was funded by the Canadian Institutes of Health Research [16]. HIVSmart! plugs gaps in the self-testing process, works with any approved HIV self-test, engages, and proactively informs any intended user of an HIV self-test. It interprets and stores data confidentially, links users to counseling or care within a rapid turnaround time, and encourages them to stay in care. For this study, we adapted the HIVSmart! app-based program for Canadians. Reverse innovation entailed language adaptations (French-Canadian) and customizations to the US Food and Drug Administration-approved oral self-test products and obtaining Health Canada's Investigational Testing Authorization for research.

HIVSmart! is unique in that it is a complete app-based solution. It is an improvement from the Web prototype of HIVSmart!, which was initially evaluated in health care professionals of South Africa and students of McGill University [12,17]. This novel app-based solution is currently being tested at scale in South Africa [18].

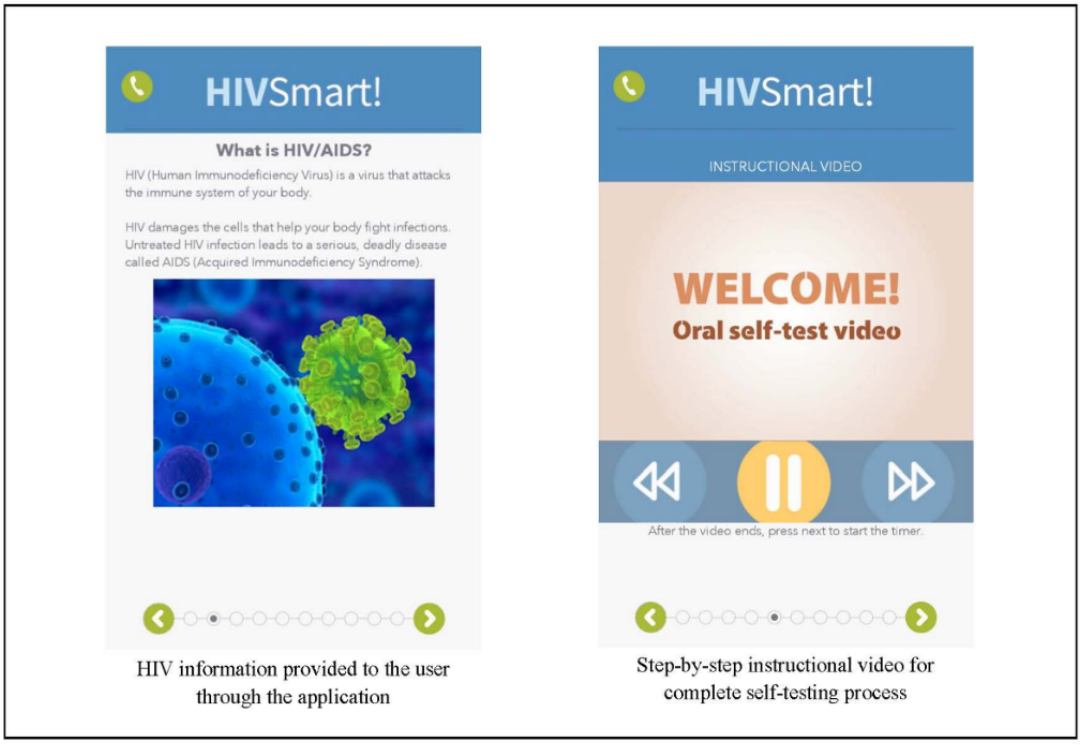
Study Participants and Design

Between July 2016 and February 2017, we conducted a cross-sectional feasibility study at Clinique Médicale L'Actuel, a private, sexual health clinic specializing in testing and treatment of HIV and sexually transmitted infections, located in urban Montreal. Ethical approvals were obtained from Clinique L'Actuel's independent review board (Veritas Institutional Review Board), and the McGill University Health Centre Research Ethics Board. Study procedures were duly followed and complied with the regulations stipulated by the Institutional Review Boards of McGill University Health Centre and Veritas.

In high risk, MSM populations presenting to the clinic, we set out to achieve the following objectives. We aimed to evaluate the feasibility, acceptability, and preference for an unsupervised HIV self-testing strategy that involved a self-test (OraQuick In-Home HIV Test) and an accompanying optimized app (HIVSmart!) in an unsupervised setting that mimicked a home environment.

Participants were recruited by convenience sampling; clinic staff recruited participants during routine and drop-in clinic visits. The study was advertised via flyers, social media, and through e-newsletters.

Figure 1. Screenshots of the HIVSmart! app.



Participants were deemed eligible to join the study if they were 18 years or older, self-identified as an MSM, of unknown HIV status, and comfortable using smartphones or tablets. Participants receiving pre- or postexposure prophylaxis (PrEP or PEP) antiretrovirals for prevention (ARVs) were also invited to participate so as to not exclude these high-risk groups. Participants were excluded if they self-reported a previously confirmed positive HIV diagnosis.

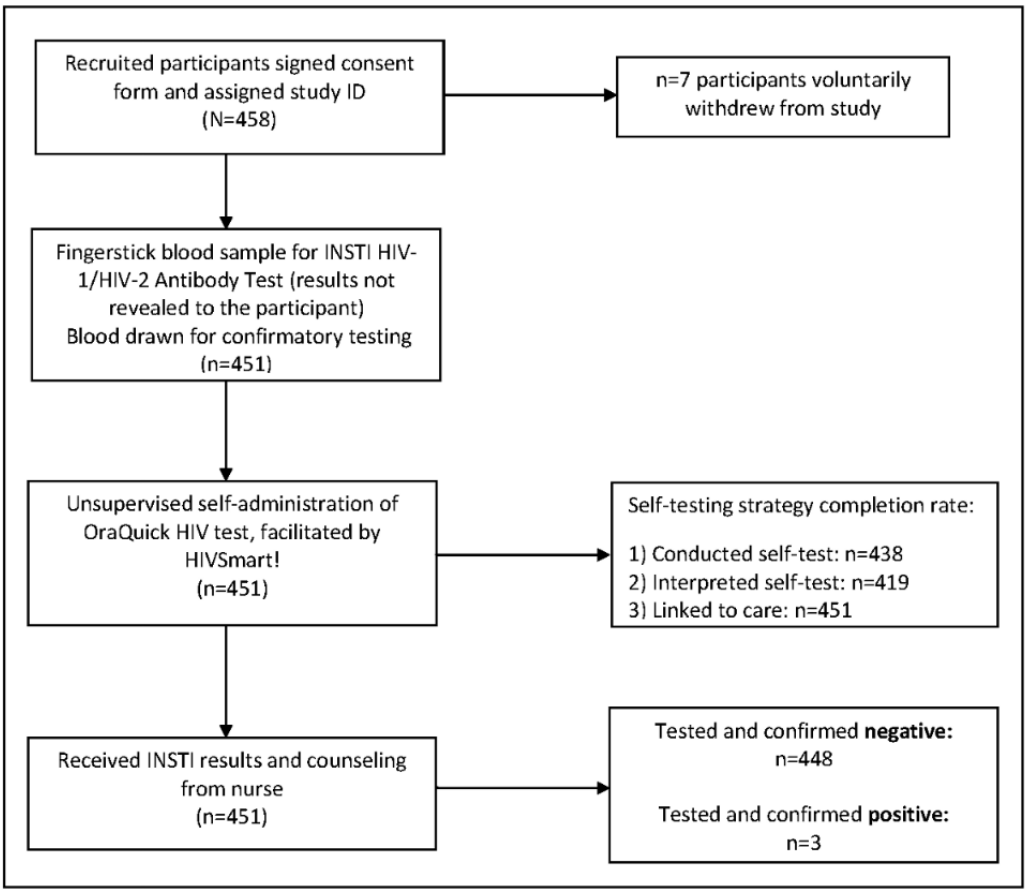
A flow chart outlining the study flow is represented in Figure 2. Once deemed eligible to participate, the study procedure was explained to participants. Informed consent (written) was obtained from all participants before beginning the study. Participants could withdraw consent at any point throughout the study. At this time, participants were informed about the strategy that involved deidentified data collection on a confidential, compliant secure server.

Study Procedure

As per Health Canada's product monographs and Investigational Testing Authorization recommendations, the usual clinic procedures were followed in the study. Following participants' informed consent and agreement to participate, participants were explained about data storage and confidentiality. Following this, a rapid test (INSTI HIV-1/HIV-2 Antibody Test; bioLytical Laboratories) was performed by the research nurse. However, the rapid test result was not revealed to participants at this time. Blood was drawn for laboratory-based confirmatory testing, following the clinic's protocol (Centre hospitalier de l'Université de Montréal): p24 antigen testing and anti-HIV-1/2 testing for all samples, confirmed by Western Blot if abnormal. Samples underwent RNA testing in case of suspected acute HIV infection. Lab testing conformed to the national testing algorithms.

Participants were then brought to a private office in the clinic and provided the OraQuick Test kit along with a tablet installed with HIVSmart!. The process of self-testing went thus: participants were left alone (unsupervised) to navigate the app on the tablet, stage their personal risk for HIV, perform the self-test as per instructions on HIVSmart!, and conduct the test unsupervised, mimicking a home environment. All participants followed instructions on pretest counseling, staging, conducting results, and storing their results on screen. Participants were asked to wait for their test result using a timer built into the app. They were encouraged to use the phone number provided in the app to call for counseling or assistance at any time and respond to a research questionnaire within the app. Upon completion of the self-testing process, participants were offered the choice of calling the provided phone number to receive posttest counseling or receiving face-to-face posttest counseling with a research nurse. Phone-based counseling was rapidly followed by face-to-face counseling with the research nurse. Upon meeting with the research nurse, who interpreted and recorded the self-test results in real-time separately, the INSTI rapid tests results were provided to participants and they were offered posttest counseling based on their INSTI results. Participants with positive HIV results were seen by a physician immediately for follow-up care. Participants with negative self-tests were encouraged to return for retesting in 3 months. As per the Institutional Review Board recommendations, all participants were offered a modest compensation for their time (Can $20).

Figure 2. Flow of study participants. ID: identifier.



Outcomes and Metrics

The primary outcome of this study was the feasibility of the HIVSmart! self-testing strategy. Feasibility was documented by the metric completion rate, defined as the number of participants who successfully completed all steps of the self-testing procedure, over the total number of consenting participants. Steps were defined as follows: (1) self-test conduct; (2) self-test interpretation; and (3) seeking linkage to care. It was computed as an average proportion of the completion for each step in the procedure.

Secondary outcomes were patient-centered outcomes (acceptability of the strategy, preference for counseling, and costs) that were collected using the metrics listed below:

1. *Acceptability*: the number of participants who consented to evaluate the strategy, over the total number of people approached for recruitment. Refusals were documented by staff at the clinic.
2. *Preference for (counseling)*: participants were asked to respond to a questionnaire, regarding preferences for suitable posttest counseling options, in a hypothetical scenario where self-tests were available for purchase in pharmacies.
3. *Preference for (cost)*: participants were asked in a questionnaire to select a price point at which self-tests should be made available in Canada (<Can $10, Can $10-20, Can $20-40, and >Can $40).
4. *Usefulness of the app*: participants were asked if the app was found to be useful in assisting them in the self-testing process.
5. *Recommendations to friends or partners*: participants were asked if the app would be recommended by them to their friends or partners.

Data Analyses and Sample Size Estimations

Our sample size was estimated using the completion rate as the main outcome. Estimating a conservative completion rate of 70% with a CI width of 10% (SD 0.0565%), we calculated a target sample size of 400.

Basic proportions for demographics and outcomes obtained through the HIVSmart! app were computed with their 95% CIs. A stratified analysis comparing participants receiving ARVs (PrEP or PEP) versus no ARVs was conducted for the main feasibility outcome variables with Fisher's exact test, which determines significant differences between groups (StataCorp, 2013, *Stata Statistical Software: Release 13*, StataCorp LP).

Results

Participant Characteristics

A total of 451 participants met the eligibility criteria and accepted to evaluate the HIVSmart! self-testing strategy (Table 1); 97.1% (438/451) of the participants submitted the HIVSmart! survey (Table 2).

Table 1. Descriptive characteristics of study participants.

Characteristics	Value (N=451)
Age (years), mean (range)	33.6 (32.6-34.7)
Antiretrovirals for prevention status, n (%)	
None	394 (87.4)
Postexposure prophylaxis	5 (1.1)
Pre-exposure prophylaxis	52 (11.5)

Participants were well educated, with 84.7% (371/438) educated beyond high school; 79.5% (348/438) were employed, and 52.5% (230/438) had been tested in the past 6 months. Participants self-identified themselves as male (98.6%, 432/438) and homosexual (91.3%, 400/438); 96.1% (421/438) of participants were sexually active, with 25.1% (110/438) stating 6-10 partners in the past 6 months; 21.9% (96/438) stating ≥11; and 70.1% (307/438) engaged in condom-less sex. In total, 11.5% (52/451) participants were currently taking PrEP and 1.1% (5/451) were taking PEP.

Acceptability was high at 98.5% (451/458); 7 participants refused to participate or withdrew themselves from the study.

Feasibility, as documented by the completion rate of the HIVSmart! self-testing strategy, was computed by taking an average of the 3 steps of the self-testing strategy—(1) self-test conduct; (2) self-test interpretation; and (3) linkages to care.

1. *Self-test conduct:* 97.1% (438/451) of the participants conducted the self-test successfully.
2. *Self-test interpretation*: 92.9% (419/451) of participants interpreted their self-test successfully.
3. *Linkages to care*: 100% (451/451) of the participants sought linkages to care successfully.

To compute feasibility, we took an average proportion of the 3 steps highlighted above, resulting in the overall feasibility of 96.7%.

Regarding Test Interpretations

Incomplete test conduct occurred in the initial set-up stages of the study when participants were unable to submit their results through the app because of Wi-Fi connectivity issues. These were later resolved by resolving incompatibility issues of the app and the clinic's Wi-Fi server.

Regarding test interpretations, a few participants mistakenly interpreted their negative result (the control line) as their positive result, despite instructions. An invalid result was also recorded that was truly negative.

Regarding test results, 3 participants tested positive for HIV (0.7% seropositivity) with both the self-test and rapid test, which were rapidly confirmed by laboratory results, and the participants were linked to a physician within the same day (linkage: 3/3, 100%) and returned for a follow-up appointment. All negative rapid and self-test results (448/451, 99.3%) were confirmed negative through laboratory testing, and all were linked to counseling (448/448, 100%). Lab testing conformed to the national testing algorithms.

Regarding linkages, all participants (451/451, 100%) were linked to in-person counseling following the self-testing procedure. All participants used the phone line and later met the research nurse. The average turnaround time to linkage to counseling ranged from 2 to 6 hours.

Regarding the preference for counseling, participants were in favor of counseling over the phone, followed by face-to-face counseling in a clinic. Some favored counseling over the internet (chat or website; 132/421, 31.4%) or in a pharmacy (121/421, 28.7%), both followed by face-to-face counseling in clinic and counseling in clinics only (132/421, 31.4%). Participants were generally not in favor of no face-to-face counseling (28/421, 6.7%) or no counseling at all (3/421, 0.7%).

Regarding cost preferences, half of the participants (206/421, 48.8%) selected Can $10-20 and 27.4% (115/421) selected <Can $10. In terms of the usefulness of the app, 98.8% (417/422) of the participants found the app helpful in guiding them through the self-testing process. Finally, 94.3% (395/419) of the participants said that they would recommend this self-testing strategy to their partner.

Table 2. Participant information collected through the HIVSmart! survey (n=438).

Information	Value, n (%)
Gender (self-identified)	
Male	432 (98.6)
Transgender	2 (0.5)
I do not wish to answer	3 (0.7)
Other	1 (0.2)
Sexual orientation	
Homosexual	400 (91.3)
Bisexual	32 (7.3)
Heterosexual	4 (0.9)
I do not wish to answer	1 (0.2)
Other	1 (0.2)
Highest level of education	
High school degree not completed	13 (3.0)
High school	51 (11.6)
College or technical school	125 (28.5)
Undergraduate degree	169 (38.6)
Graduate degree (master's or PhD)	77 (17.6)
Other	1 (0.2)
I do not wish to answer	2 (0.5)
Employment status	
Unemployed	46 (10.5)
Employed	348 (79.5)
Other	36 (8.2)
I do not wish to answer	8 (1.8)
Tested for HIV in the past 6 months	
No, I've never been tested	25 (5.7)
Yes, in the last 6 months	230 (52.5)
Yes, more than 6 months ago	178 (40.6)
I have been tested, but did not want to receive my result	1 (0.2)
I do not know if I have ever been tested	3 (0.7)
I do not wish to answer	1 (0.2)
Sexually active	
No	16 (3.7)
Yes	421 (96.1)
I do not wish to answer	1 (0.2)
Number of different sexual partners in the past 6 months	
0	20 (4.6)
1	38 (8.7)
2-5	174 (39.7)
6-10	110 (25.1)
≥11	96 (21.9)
In the past 6 months:	

Information	Value, n (%)
Had sex without a condom	307 (70.1)
Had sex with an HIV-infected partner	74 (16.9)
Had sex with a sex worker	18 (4.1)
Had sex under the influence of alcohol	195 (44.5)
Had sex under the influence of drugs	83 (18.9)
Had sex with multiple partners	219 (50.5)
Injected drugs (excluding medicine)	5 (1.1)
Exposure to HIV (eg, needles) in the workplace in the past 6 months	
No	426 (97.3)
Yes	10 (2.3)
I do not wish to answer	2 (0.5)

Discussion

Principal Findings

In this Canadian Institutes of Health Research funded innovative Canadian study, we investigated the feasibility of implementing an app-optimized unsupervised HIV self-testing strategy in a clinical setting.

Our unsupervised self-testing strategy was found to be feasible to operationalize (419/451, 96.6%), was well accepted (451/458, 98.5%), and preferred by participants. HIV was detected, and all participants were linked to care within a working day. All of our self-testers were linked to counseling or directed to a physician within hours, an essential service to offer with self-testing. Participants found the app-based approach to be a useful (417/422, 98.8%) in completing the HIV self-testing procedure, and a majority (395/419, 94.3%) wanted to recommend it to their friend.

This app-optimized, self-testing strategy was aimed to plug gaps in the self-testing process that are associated with an accurate detection, self-test interpretation, and rapid initiation of linkages to counseling and care. We were limited by the lack of approved HIV self-tests in Canada, which restricted our evaluation to a clinic environment [19-21] instead of homes. To stimulate a home environment, we set up kiosks in clinics where participants could test unsupervised, yet a nurse was always available to offer counseling and support [22]. Regarding costs, we found that a majority of participants believed that an acceptable price to pay for an HIV self-test in Canada was between Can $10 and $20. This finding is in line with 3 studies from New York City, where participants felt that an affordable and accessible price point was US $15-$25 (compared with the prevailing US $40) [23-25].

This is the first Canadian study to report data on the use of an app-based strategy. Many digital innovations (ie, Web-based programs, kiosk-based tablets, and short message service [SMS] text messaging services) are available, yet a complete, portable, and patient-friendly app-based solution for self-testing from engagement to linkage to counseling and care is novel [26].

In 5 studies that have evaluated some digital innovations, we observed a few limitations in their offer of services that impacted the process. A Dutch study evaluated a Web-based strategy, where participants could purchase self-tests and access self-test instructions and counseling, but they did not provide data or information on linkages [27]. Of 4 US studies, 2 reported positive findings on the feasibility of use of a kiosk tablet for HIV self-testing in emergency rooms but reported poor engagement (50%) of participants and limited data on detection and linkages [28,29]. Another US study evaluating the use of SMS text messaging self-test results only by participants found that it was preferred by participants [30]. The fourth study provided written and Web-based video instructions to participants to choose a picture that resembled their self-test result and reported comparable results to ours, with 100% of positive OraQuick results and 98% of negative results being interpreted correctly [31]. None of these studies evaluated linkages to care.

The HIVSmart! app is an integrated innovation. It offers a personalized experience of self-testing from access to linkage, which is a step up from a website-, tablet-, or SMS text messaging-only service and is housed in a secure Health Insurance Portability and Accountability Act-compliant cloud-based platform. Integrated innovations [32], like HIVSmart!, are a new trend in the digital innovations space, as reported in a recent systematic review [33]. In 2011-2013, we evaluated a Web-based HIVSmart! strategy successfully in South African health care professionals [17]. We are currently testing the strategy at scale in South African townships in a project funded by both the Governments of South Africa and Canada [18]. A prototype of this strategy was evaluated in students way back in 2009 [12].

With the increasing availability of PrEP in Canada, and the desire to self-test frequently expressed by those on PrEP, we included a subsample of participants on PrEP in this study. However, neither our main outcomes nor HIV status differed by the PrEP status. Our study was underpowered to detect subgroup differences (small number of PrEP participants). Studies suggest that PrEP increases the window period for seroconversion, taking longer to get a positive test result [34,35]. Yet, many of our self-test results were consistent with the rapid

test results. Concordant with a study in PrEP users from Kenya, participants were in favor of self-testing [36].

Limitation

A limitation of this study included convenience sampling that raises a concern of selection bias.

Conclusions

In Canada, future research with HIV self-tests from provinces with high rates of undocumented HIV infection (Saskatchewan) and marginalized populations with undiagnosed HIV infection is warranted. Future research that incorporates digital strategies to plug service delivery gaps important for HIV self-testing will make it easier to offer and document self-testing.

Rapid approvals by Food and Drug Administration, Conformité Européenne, and World Health Organization Prequalification of self-tests, both oral and blood-based HIV self-tests, will help expand options to self-test in Canada. It will also increase the visibility of HIV self-tests in pharmacies, clinics, and outreach settings and democratize the process of HIV self-testing. Finally, the adoption of proven digital solutions will help improve engagement and expedite rapid linkages to care. Doing so will help address the last mile problem of detecting undiagnosed HIV infection in marginalized Canadians, thereby accelerating progress toward Joint United Nations Programme on HIV/AIDS 90-90-90 targets in Canada.

We conclude that the HIVSmart! app-optimized strategy is feasible, accepted, and preferred by an educated, urban MSM population of Montreal. With the app, participants were able to perform, interpret, store results, and rapidly link to care. The HIVSmart!-optimized, self-testing strategy could be adapted and contextualized to many at-risk populations within Canada and worldwide, thereby maximizing its public health impact.

Acknowledgments

The authors are grateful to all the staff at Clinique Médicale L'Actuel Montreal. We would also like to acknowledge Dr Marc Steben, Dr Bertrand Lebouché, and Dr Jean Pierre Routy for supporting the study. This work was funded by an operating grant from the Canadian Institutes of Health Research (grant #HHP-137872) and the Fonds de recherche du Québec - Santé Research-Scholar Junior 2 and Senior, awarded to NPP.

Authors' Contributions

NPP was involved in the concept, execution, write-up, critique, and overall responsibility of data for the project; MS was involved in execution, data analyses, write-up, and critique; LD was involved in execution, data collection, write-up, and critique; AG was involved in execution, data collection, write-up, and critique; KB was involved in execution, data collection, and write-up; AFV was involved in execution, data collection, write-up, and critique; LJ was involved in data analyses, write-up, and critique; and RT was involved in execution, data collection, write-up, critique, and responsibility for the project implementation.

Conflicts of Interest

None declared.

References

1. Joint United Nations Programme on HIV/AIDS. UNAIDS.org. Geneva: UNAIDS; 2014. 90-90-90: An ambitious treatment target to help end the AIDS epidemic URL: http://www.unaids.org/sites/default/files/media_asset/90-90-90_en.pdf [accessed 2017-08-16]
2. Consolidated Guidelines on HIV Testing Services: 5Cs: Consent, Confidentiality, Counselling, Correct Results and Connection. Geneva: World Health Organization; 2016 Nov 29. URL: https://www.who.int/hiv/pub/self-testing/hiv-self-testing-guidelines/en/ [accessed 2018-11-07] [WebCite Cache ID 73lDUcrSc]
3. Consolidated Guidelines on HIV Testing Services: 5Cs: Consent, Confidentiality, Counselling, Correct Results and Connection. Geneva: World Health Organization; 2015 Jul 15. URL: https://www.who.int/hiv/pub/guidelines/hiv-testing-services/en/ [accessed 2018-11-07] [WebCite Cache ID 73lDRKQQm]
4. Epstein JS. July 3 Approval Letter, OraQuick In-Home HIV Test. Silver Spring: U.S. Food and Drug Administration,; 2012 Jul 03. URL: https://www.fda.gov/BiologicsBloodVaccines/BloodBloodProducts/ApprovedProducts/PremarketApprovalsPMAs/ucm310592.htm [accessed 2018-10-29] [WebCite Cache ID 73X4YoqQN]
5. Pebody R. First HIV home test approved for sale in UK. London: NAM Publications; 2015 Apr 24. URL: http://www.aidsmap.com/First-HIV-home-test-approved-for-sale-in-UK/page/2964537/ [accessed 2018-10-25] [WebCite Cache ID 73RJjAcug]
6. HIVST. HIVST.org: HIV self-testing research and policy hub. 2017. URL: http://www.hivst.org/ [accessed 2017-08-15]
7. Unitaid, World Health Organization. Market and technology landscape: HIV rapid diagnostic tests for self-testing, 4th edition. Geneva: Unitaid; 2018 Jul. URL: https://unitaid.org/assets/HIVST-landscape-report.pdf [accessed 2018-09-20]
8. Wood BR, Ballenger C, Stekler JD. Arguments for and against HIV self-testing. HIV AIDS (Auckl) 2014;6:117-126 [FREE Full text] [doi: 10.2147/HIV.S49083] [Medline: 25114592]

9. Brown AN, Djimeu EW, Cameron DB. A review of the evidence of harm from self-tests. AIDS Behav 2014 Jul;18 Suppl 4:S445-S449 [FREE Full text] [doi: 10.1007/s10461-014-0831-y] [Medline: 24989129]
10. Health Canada. Draft Guidance Document on HIV Simple/Rapid Test Kits. 2017 Sep 14. Consultation on HIV Simple/Rapid Test Kits URL: https://tinyurl.com/yc487qw2
11. Haig TA, Otis J, Veillette-Bourbeau L, Caruso J, Jolimore J, Ferlatte O, et al. HIV self-testing for MSM: acceptability among community members and service providers in Vancouver, Toronto, and Montreal. Canadian Journal of Infectious Diseases & Medical Microbiology (Supplement B) 2015;26:13B.
12. Pant Pai N, Bhargava M, Joseph L, Sharma J, Pillay S, Balram B, et al. Will an unsupervised self-testing strategy be feasible to operationalize in Canada? Results from a pilot study in students of a large canadian university. AIDS Res Treat 2014;2014:747619 [FREE Full text] [doi: 10.1155/2014/747619] [Medline: 24511392]
13. Pai NP, Smallwood M, Gulati D, Lapczak N, Musten A, Gaydos C, et al. What do Key Stakeholders Think About HIV Self-Testing in Canada? Results from a Cross-Sectional Survey. AIDS Behav 2018 Feb;22(2):606-615 [FREE Full text] [doi: 10.1007/s10461-017-1764-z] [Medline: 28439755]
14. Public Health Agency of Canada. Summary: Etimates of HIV incidence, prevalence and proportion undiagnosed in Canada, 2014. Ottawa; 2015 Nov 27. URL: https://www.canada.ca/content/dam/canada/health-canada/migration/healthy-canadians/publications/diseases-conditions-maladies-affections/hiv-aids-estimates-2014-vih-sida-estimations/alt/hiv-aids-estimates-2014-vih-sida-estimations-eng.pdf
15. Pai N. Grand Challenges Canada. 2011. To develop a synergsitic, innovative, implementation strategy for self-testing for HIV in South Africa URL: http://www.grandchallenges.ca/grantee-stars/0027-01/
16. Pai N. Candian Institutes of Health Research. 2017. Feasibility of an innovative AideSmart! app-based multiplexed, point-of-care screening/counselling strategy for HIV-associated co-infections (HCV, Syphilis, HBV) in key Canadian populations URL: https://tinyurl.com/y9v4ep8x
17. Pant Pai N, Behlim T, Abrahams L, Vadnais C, Shivkumar S, Pillay S, et al. Will an unsupervised self-testing strategy for HIV work in health care workers of South Africa? A cross sectional pilot feasibility study. PLoS One 2013;8(11):e79772 [FREE Full text] [doi: 10.1371/journal.pone.0079772] [Medline: 24312185]
18. Pai N. Grand Challenges Canada. 2015. HIVSmart! A smartphone app-based HIV self-testing strategy to identify undiagnosed cases of HIV URL: http://www.grandchallenges.ca/grantee-stars/0710-05/
19. Choko AT, MacPherson P, Webb EL, Willey BA, Feasy H, Sambakunsi R, et al. Uptake, Accuracy, Safety, and Linkage into Care over Two Years of Promoting Annual Self-Testing for HIV in Blantyre, Malawi: A Community-Based Prospective Study. PLoS Med 2015 Sep;12(9):e1001873 [FREE Full text] [doi: 10.1371/journal.pmed.1001873] [Medline: 26348035]
20. Thirumurthy H, Masters SH, Mavedzenge SN, Maman S, Omanga E, Agot K. Promoting male partner HIV testing and safer sexual decision making through secondary distribution of self-tests by HIV-negative female sex workers and women receiving antenatal and post-partum care in Kenya: a cohort study. Lancet HIV 2016 Dec;3(6):e266-e274 [FREE Full text] [doi: 10.1016/S2352-3018(16)00041-2] [Medline: 27240789]
21. Maheswaran H, Petrou S, MacPherson P, Choko AT, Kumwenda F, Lalloo DG, et al. Cost and quality of life analysis of HIV self-testing and facility-based HIV testing and counselling in Blantyre, Malawi. BMC Med 2016 Feb 19;14:34 [FREE Full text] [doi: 10.1186/s12916-016-0577-7] [Medline: 26891969]
22. Wray T, Chan PA, Simpanen E, Operario D. eTEST: Developing a Smart Home HIV Testing Kit that Enables Active, Real-Time Follow-Up and Referral After Testing. JMIR Mhealth Uhealth 2017 May 08;5(5):e62 [FREE Full text] [doi: 10.2196/mhealth.6491] [Medline: 28483744]
23. Frye V, Wilton L, Hirshfied S, Chiasson MA, Usher D, Lucy D, All About Me Study Team. "Just Because It's Out There, People Aren't Going to Use It." HIV Self-Testing Among Young, Black MSM, and Transgender Women. AIDS Patient Care STDS 2015 Nov;29(11):617-624 [FREE Full text] [doi: 10.1089/apc.2015.0100] [Medline: 26376029]
24. Lippman SA, Moran L, Sevelius J, Castillo LS, Ventura A, Treves-Kagan S, et al. Acceptability and Feasibility of HIV Self-Testing Among Transgender Women in San Francisco: A Mixed Methods Pilot Study. AIDS Behav 2016 Apr;20(4):928-938 [FREE Full text] [doi: 10.1007/s10461-015-1236-2] [Medline: 26511864]
25. Myers JE, El-Sadr Davis OY, Weinstein ER, Remch M, Edelstein A, Khawja A, et al. Availability, Accessibility, and Price of Rapid HIV Self-Tests, New York City Pharmacies, Summer 2013. AIDS Behav 2017 Feb;21(2):515-524. [doi: 10.1007/s10461-016-1594-4] [Medline: 27804092]
26. LeGrand S, Muessig KE, Horvath KJ, Rosengren AL, Hightow-Weidman LB. Using technology to support HIV self-testing among MSM. Curr Opin HIV AIDS 2017 Sep;12(5):425-431 [FREE Full text] [doi: 10.1097/COH.0000000000000400] [Medline: 28617712]
27. Bartelsman M, Joore IK, van Bergen JE, Hogewoning AA, Zuure FR, van Veen MG, HIV Transmission Elimination AMsterdam (H-TEAM) initiative. HIV testing week 2015: lowering barriers for HIV testing among high-risk groups in Amsterdam. BMC Infect Dis 2017 Dec 01;17(1):529 [FREE Full text] [doi: 10.1186/s12879-017-2617-0] [Medline: 28764661]
28. Gaydos CA, Solis M, Hsieh Y, Jett-Goheen M, Nour S, Rothman RE. Use of tablet-based kiosks in the emergency department to guide patient HIV self-testing with a point-of-care oral fluid test. Int J STD AIDS 2013 Sep;24(9):716-721 [FREE Full text] [doi: 10.1177/0956462413487321] [Medline: 23970610]

29. Haukoos JS, Hopkins E, Bender B, Al-Tayyib A, Long J, Harvey J, Denver Emergency Department HIV Testing Research Consortium. Use of kiosks and patient understanding of opt-out and opt-in consent for routine rapid human immunodeficiency virus screening in the emergency department. Acad Emerg Med 2012 Mar;19(3):287-293 [FREE Full text] [doi: 10.1111/j.1553-2712.2012.01290.x] [Medline: 22435861]
30. Daniels J, Rosengren L, Young S, Klausner JD. Will Men Who Have Sex With Men Use Short-Messaging Services to Send Photos of Completed HIV Self-Tests to Researchers? J Assoc Nurses AIDS Care 2016;27(5):722-726. [doi: 10.1016/j.jana.2016.05.002] [Medline: 27293208]
31. Chavez P, Wesolowski L, Owen M, Gravens L, Sullivan P, MacGowan R. Perceptions and performance of self-administered rapid HIV tests conducted by untrained users in real world settings. 2016 Presented at: 2016 HIV Diagnostics Conference; 21-24 March, 2016; Atlanta, Georgia, USA URL: http://hivtestingconference.org/wp-content/uploads/2016/03/2016-HIV-DX-Program-Book_WEB_FINAL.pdf
32. El-Noush H, Silver KL, Pamba AO, Singer PA. Innovating for women's, children's, and adolescents' health. BMJ 2015 Sep 14;351:h4151 [FREE Full text] [Medline: 26371217]
33. Daher J, Vijh R, Linthwaite B, Dave S, Kim J, Dheda K, et al. Do digital innovations for HIV and sexually transmitted infections work? Results from a systematic review (1996-2017). BMJ Open 2017 Nov 03;7(11):e017604 [FREE Full text] [doi: 10.1136/bmjopen-2017-017604] [Medline: 29101138]
34. Donnell D, Ramos E, Celum C, Baeten J, Dragavon J, Tappero J, Partners PrEP Study Team. The effect of oral preexposure prophylaxis on the progression of HIV-1 seroconversion. AIDS 2017 Dec 10;31(14):2007-2016 [FREE Full text] [doi: 10.1097/QAD.0000000000001577] [Medline: 28692542]
35. Suntharasamai P, Martin M, Choopanya K, Vanichseni S, Sangkum U, Tararut P, et al. Assessment of Oral Fluid HIV Test Performance in an HIV Pre-Exposure Prophylaxis Trial in Bangkok, Thailand. PLoS One 2015;10(12):e0145859 [FREE Full text] [doi: 10.1371/journal.pone.0145859] [Medline: 26717405]
36. Ngure K, Heffron R, Mugo N, Thomson KA, Irungu E, Njuguna N, et al. Feasibility and acceptability of HIV self-testing among pre-exposure prophylaxis users in Kenya. J Int AIDS Soc 2017 Dec 10;20(1):21234 [FREE Full text] [doi: 10.7448/IAS.20.1.21234] [Medline: 28362073]

Abbreviations

ITA: Investigational Testing Authorization
MSM: men who have sex with men
PEP: postexposure prophylaxis
PrEP: pre-exposure prophylaxis
SMS: short service message

Edited by G Eysenbach; submitted 20.03.18; peer-reviewed by M Steben, S Badman, C Figueroa; comments to author 05.08.18; revised version received 23.08.18; accepted 14.09.18; published 27.11.18.

Please cite as:
Pant Pai N, Smallwood M, Desjardins L, Goyette A, Birkas KG, Vassal AF, Joseph L, Thomas R
An Unsupervised Smart App–Optimized HIV Self-Testing Program in Montreal, Canada: Cross-Sectional Study
J Med Internet Res 2018;20(11):e10258
URL: http://www.jmir.org/2018/11/e10258/
doi:10.2196/10258
PMID:30465709

Original Paper

Using Facebook for Health Promotion in "Hard-to-Reach" Truck Drivers: Qualitative Analysis

Marguerite C Sendall[1*], MHSc, PhD; Laura K McCosker[1*], BNurs, BSocSci; Phil Crane[1,2*], GradDipHlthEduc, MA, PhD; Bevan Rowland[1,3*], GradCert, BEd, MSc; Marylou Fleming[1*], DipT, BEd, MA, PhD; Herbert C Biggs[1,3*], BA, PhD

[1]School of Public Health and Social Work, Faculty of Health, Queensland University of Technology, Brisbane, Australia
[2]School of Social Sciences, Faculty of Arts, Business and Law, University of the Sunshine Coast, Sippy Downs, Australia
[3]Centre for Accident Research and Road Safety - Queensland, Queensland University of Technology, Brisbane, Australia
[*]all authors contributed equally

Corresponding Author:
Marguerite C Sendall, MHSc, PhD
School of Public Health and Social Work
Faculty of Health
Queensland University of Technology
Victoria Park Road
Kelvin Grove
Brisbane, 4059
Australia
Phone: 61 +617 3138 3526
Email: m.sendall@qut.edu.au

Abstract

Background: Workers in the road transport industry, and particularly truck drivers, are at increased risk of chronic diseases. Innovative health promotion strategies involving technologies such as social media may engage this "hard-to-reach" group. There is a paucity of evidence for the efficacy of social media technologies for health promotion in the Australian transport industry.

Objective: This study analyzed qualitative data from interviews and focus group discussions to evaluate a social media health promotion intervention, the Truckin' Healthy Facebook webpage, in selected Australian transport industry workplaces.

Methods: We engaged 5 workplace managers and 30 truck drivers from 6 transport industry organizations in developing workplace health promotion strategies, including a social media intervention, within a Participatory Action Research approach. Mixed methods, including a pre- and postintervention manager survey, truck driver survey, key informant semistructured interviews, truck driver focus groups, and focused observation, were used to evaluate the social media intervention. We asked questions about workplace managers' and truck drivers' opinions, engagement, and satisfaction with the intervention. This paper focuses on qualitative data.

Results: Of the workplace managers who reported implementing the social media intervention at their workplace, all (3/3, 100%) reported satisfaction with the intervention and expressed a keen interest in learning more about social media and how it may be used for workplace health promotion and other purposes. Truck drivers were poorly engaged with the intervention because (1) many believed they were the "wrong age" and lacked the necessary skills; (2) the cost of smartphone technology was prohibitive; (3) they confined their use of social media to nonwork-related purposes; and (4) many workplaces had "no Facebook" policies.

Conclusions: The use of social media as a health promotion intervention in transport industry workplaces has potential. Workplace interventions using social media can benefit from a Participatory Action Research approach. Involving managers and workers in the design of social media health promotion interventions and developing strategies to support and deliver the interventions helps to facilitate their success. The workers' profile, including their age and familiarity with social media, and work, workplace, and family context is important to consider in this process. Much more research needs to be undertaken to better understand the effective use of social media to engage "hard-to-reach" groups.

(J Med Internet Res 2018;20(11):e286) doi:10.2196/jmir.9689

KEYWORDS

communication; health promotion interventions; mobile phone; social media; transport industry; truck drivers; workplace health promotion; workplace managers

Introduction

The road transport industry accounts for 2% of Australia's workforce, employing 192,600 Australians [1,2]. Road transport industry workers, in general, and truck drivers, in particular, are identified at increased risk of chronic diseases such as cardiovascular disease and type 2 diabetes [3,4]. Workplaces are increasingly recognized as environments conducive to health promotion. For truck drivers, their workplace is their vehicle and, as a highly mobile workforce, traditional workplace health promotion strategies may be limited in their effectiveness [5-11]. Innovative new health promotion strategies, including social media technologies, may engage and improve the health outcomes of this "hard-to-reach" group.

The use of social media and related technologies is increasingly popular for personal enjoyment [12], as a business tool [13], and in health promotion interventions [14]. This is true for transport industry workplaces and truck drivers. A 2011 study conducted by a UK trucking magazine found that 52% of surveyed truck drivers used social media in their spare time [15]. A similar survey undertaken in 2016 by a US trucking company reported that 93% of their truck driver respondents used smartphones, and 73% reported they checked social media outlets, such as Facebook, Twitter, and Instagram, at least once each day [16]. In addition, US fleet managers are rapidly adopting smartphones and work-related apps to manage their fleets, with 59% of those surveyed tracking their fleet or managing fleet activities or services through mobile apps [17]. The use of information technologies, such as social media, has become a part of the daily life of modern truckies. While few studies have been conducted in Australia, the popularity of social media within the transport industry internationally suggests it is feasible to use it as a health promotion strategy in the Australian context.

A literature review suggests that truck drivers, in Australia and internationally, use social media and related technologies for a variety of purposes. Facebook is a popular social media platform and has been used by truck drivers for a variety of purposes, including to advocate for improved safety at work, report road accidents, identify and stop fatigued drivers, search for missing drivers, and socially connect with other drivers [18-27]. Other social media technologies are used in the trucking industry to map traffic congestion, identify parts dealers, communicate fatigue laws, plan routes, find truck stops, assist with loading, provide voice-guided navigation, and communicate pickup requirements [28-30].

In Australia, social media have been used in the transport industry for health promotion purposes in a limited capacity. Herbert [31] reported an Australian intervention allowing truck drivers to attend medical appointments by videoconferencing on a smartphone or tablet. Gilson et al [32,33] described smartphone interventions to monitor nutrition and physical activity in Australian truck drivers. Kenny [34] reported an intervention providing counseling and peer support through Facebook and Skype to truck drivers who have been involved in workplace accidents or work-related losses, while an Australian trucking magazine described the use of smartphone apps by drivers to share health and physical activity ideas [35]. Of these interventions, only 2 [32,33] included an evaluation of effectiveness. These studies found that technological barriers with smartphone technology prevented a number of drivers from participating in the intervention [32]; however, positive health behavior change was measured as improved driver self-regulation of healthy choices and was observed in participants [33]. While this suggests potential, with so few studies conducted, there remains a paucity of evidence for the efficacy of social media for health promotion in the Australian road transport industry.

The *Queensland Transport Industry Workplace Health Intervention* was a settings-based, mixed-methods Participatory Action Research (PAR) project. The project identified health promotion interventions for transport industry workplaces to support truck drivers to improve their health knowledge and behavior [36]. Over a two-year period, between July 2012 and June 2014, the project worked with 6 transport industry workplaces in south-east Queensland to develop, implement, and evaluate 7 health promotion interventions. One of these interventions was a Facebook page called *Truckin' Healthy*, which aimed to communicate physical activity and nutrition health promotion messages to truck drivers. An evaluation of this social media intervention, based on the qualitative data from interviews and focus group discussions, is the focus of this paper.

Methods

Approach: Participatory Action Research

The project used a range of methods located within a PAR approach to engage workplaces and, specifically, workplace managers and truck drivers, in the development, implementation, and evaluation of workplace health promotion interventions. PAR is a recognized public health research methodology, which relies on the stakeholders' involvement in decision making, planning processes, and implementation to produce meaningful change [37,38]. For this project, a mixed-methods framework was utilized with PAR. Truck drivers and workplace managers were engaged in the PAR process through interviews, focus groups, and paper-based evaluation surveys. This paper focuses on the qualitative data from interviews and focus group discussions.

At the commencement of the project in July 2012, workplace managers and truck drivers were invited to participate in interviews and focus groups. These were anonymous, confidential, and audiorecorded with consent. We conducted them at a time and place convenient to the managers or drivers, usually a location at the workplace, such as a manager's office or tea room. They lasted between 20 and 70 minutes. Interview

and focus group questions were open-ended and semistructured, developed by the research team with the aim of encouraging participants to share their experiences, attitudes, and values about workplace health promotion and health promotion interventions.

During the initial PAR process, each of the 5 workplace managers was interviewed, and one focus group was conducted with truck drivers at each workplace (involving approximately 30 truck drivers). We asked questions about the various factors which positively and negatively influence truck drivers' workplace health behavior and their engagement in health promotion activities. This information was organized into an action plan for each workplace. The action plans were used to prompt workplace managers and truck drivers to provide feedback about interventions proposed by the research team and encourage their own ideas to improve their workplace health behavior and engagement in health promotion activities.

Social Media Intervention—The Truckin' Healthy Facebook Page

Through this PAR process, 7 workplace health promotion interventions were developed collaboratively by workplaces and the project team. One of these interventions was the *Truckin' Healthy* Facebook webpage (Figure 1). During the intervention phase, social media was identified by a number of the workplace managers as a potentially effective way of communicating health promotion messages to their "hard-to-reach" mobile workforce. The *Truckin' Healthy* Facebook page was developed throughout the intervention period.

Figure 1. Screenshot of the Truckin' Healthy Facebook page. Source: Created by the project team at the Queensland University of Technology for the Queensland Transport Industry Workplace Health Intervention.

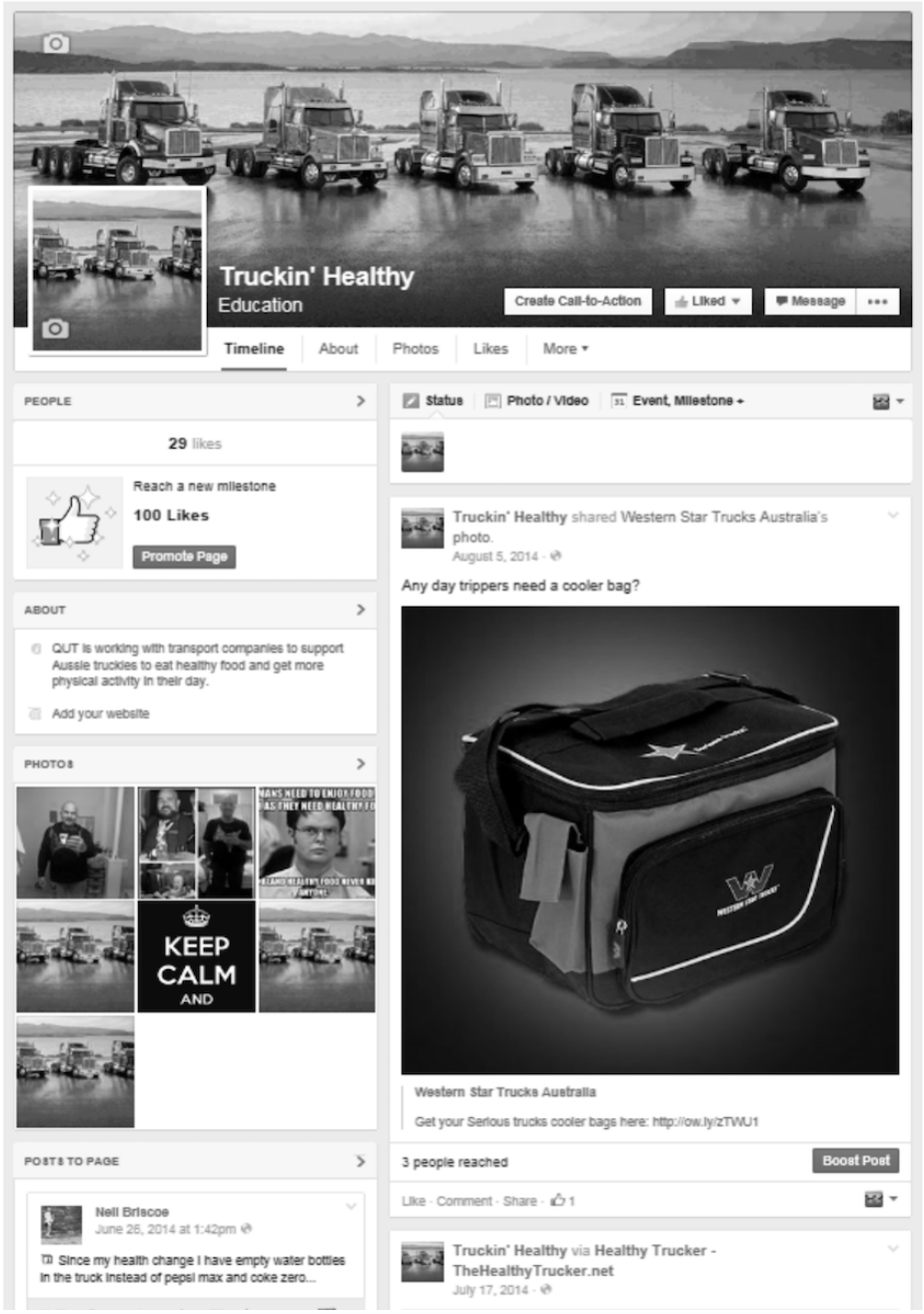


Figure 2. Poster advertising the Truckin' Healthy Facebook page.

The *Truckin' Healthy* Facebook page presented content intended to improve participants' health-related knowledge and behavior. This content included factual health messages, practical exercise tips, simple recipes, health-related videos, and product ideas (eg, lunchbox coolers). Messages were underpinned by the principles of health communication—change-oriented, audience-centered, and situated within a broad ecological perspective. These principle-based messages were integrated with other, multilevel communication strategies and interventions [39]. The information provided was consistent with current national nutrition and physical activity guidelines and delivered in a language targeted to truck drivers. The health-related content was integrated with general interest content related to trucks and the trucking industry. The page was updated at least twice each week throughout the 3-month intervention period. At the end of the intervention period in May 2014, the page was "liked" by 29 people, including truck drivers and 2 transport industry workplace managers from the participating workplaces. Each post had been "shared" or "liked" at least once by page participants.

Workplace managers decided how to promote the *Truckin' Healthy* Facebook webpage. A poster advertising the page (Figure 2) was developed by the project team, and workplace managers decided where to display the poster in the workplace. Workplaces promoted the Facebook page in other ways. For example, one workplace manager explained:

> *We [promoted] that on our drivers' toolbox [talk] as well, for them to go and like it if they wanted to or have a look at it.* [TWP120141217]

Data Collection

Data about all the interventions, including the social media intervention, was collected at two time-points— postintervention (3 months after the implementation of interventions commenced) and follow-up (6 months after the implementation of interventions commenced). We collected quantitative and qualitative data. This paper focuses on reporting qualitative data evaluating the social media intervention, with other findings reported elsewhere [40]. At postintervention and follow-up, PAR methods, including semistructured interviews, focus groups, and paper-based evaluation surveys, were used to collect data. Questions were asked about workplace managers' and truck drivers' opinions, engagement, and satisfaction with social media intervention.

The project team encountered challenges in facilitating drivers' and, to a lesser extent, workplace managers' completion of written surveys. These challenges may have been underpinned by difficulties associated with workplaces prioritizing long-term changes to workplace culture. For this reason, the project team took the approach of collecting the data contained in the surveys using a face-to-face, semistructured interview or focus group format. During postintervention and follow-up, each of the 5 workplace managers participated in at least 1 and up to 2 face-to-face semistructured interviews. A focus group involving approximately 10 truck drivers was conducted at one workplace.

A question in the semistructured interviews, "Did your workplace promote the *Truckin' Healthy* Facebook page?" produced responses which were rich and textured. Additional questions, including "Did you press 'like' to add the *Truckin' Healthy* Facebook page to your feed?" and "Overall, how satisfied were you with the information provided on the *Truckin' Healthy* Facebook page?," produced other detailed responses. Workplace managers commented on the social media intervention when asked questions such as "Which of the initiatives do you consider to be most beneficial for your workplace/employees?" and "What factors made it easy/difficult for you to implement any of the strategies?"

Data Analysis

A rigorous process of coding-and-theming was used to analyze the data obtained in the interviews and focus groups [41]. After the audiorecordings were transcribed, quotes or significant statements made by truck drivers and workplace managers about the social media intervention were separated from information about other interventions. Using an open and axial coding-and-theming approach [41], these quotes were read several times and then organized into groups of ideas by one member of the project team. This process was repeated until each of the quotes settled into a discrete theme. These themes were then discussed critically, and agreed upon, with other

project team members. These themes represent the key aspects of drivers', workplace managers', and workplaces' engagement with social media as a health promotion intervention. In this study, 6 themes were identified.

Ethical Considerations

This research was approved by the Queensland University of Technology Human Research Ethics Committee (approval number 1300000412).

Trustworthiness

We applied 4 elements of trustworthiness to this research to ensure rigor [41]. First, the researchers used purposive sampling, ensured privacy and confidentiality, established rapport with participants to create trust and honesty, and asked open-ended questions to ensure authenticity. Next, the sample included a range of truckies (day, pickup and delivery, two-up down, and line haul drivers) to ensure applicability. The research is dependable because the researchers systematically and thoroughly described and documented all phases of data collection and analysis. Finally, to ensure confirmability, all findings are traceable and grounded in the raw data.

Results

Overview

In the postintervention interviews, most workplace managers (3/4, 75%) reported they promoted the Facebook page at some time during the intervention period. Of managers whose workplaces participated in this activity, all (3/3, 100%) reported they were "satisfied" with the information contained on the Facebook page, and 2 reported they "liked" the page to add it to their personal Facebook feed. Based on the data provided by workplace managers in the postintervention interviews, of all project interventions, the Facebook page had the second highest rate of implementation and one of the highest rates of satisfaction. However, this did not translate into drivers' engagement with the intervention.

All participating truck drivers were males. At postintervention, the mean age of participating truck drivers was 43 (range 22-67) years. Less than half of the drivers (10/22, 46%) reported they were aware their workplace had promoted the Facebook page. Of these 10 drivers, most (6/10, 60%) reported they were "satisfied" or "very satisfied" with the Facebook page, but a significant proportion (4/10, 40%) recorded a "neutral" response. Very few drivers reported they "liked" the page to add it to their Facebook feed. Of drivers who responded to the survey question "Where do you get information about your health?", the proportion of drivers reporting the internet as a source of health information was unchanged between pre- and postintervention (7/44, 15% vs 3/22, 14%, respectively). Furthermore, workplace managers identified a number of reasons for drivers' lack of engagement with the social media intervention.

Six themes emerged, which represent the key aspects of drivers', workplace managers', and workplaces' engagement with social media as a health promotion intervention.

Lack of Engagement With Social Media: "There Wouldn't be One Driver Who Gets on Facebook"

Workplace managers identified drivers' poor engagement with social media as a key barrier to the effectiveness of the intervention. The managers reported only a small percentage of drivers used social media technologies, such as Facebook, in a private capacity. None of the workplaces used social media specifically, or innovative technology generally, for work purposes. One workplace manager said:

> *There wouldn't be one truck driver [in the company] who gets on Facebook. They don't get on Facebook. I mean, who does get on Facebook?* [TWP420]

Another workplace manager commented as follows:

> *I don't think a lot of our drivers are on Facebook. I would say, out of our truck drivers, maybe twenty percent use social media...A lot of them would have mobile phones, for sure. But I'd say the ones that were 'tech savvy' and who would jump on social media, a very small group... A lot of them are over fifty years old.* [TWP620]

Impact of Age on Social Media Use: "My Drivers Are the Wrong Age Group"

Workplace managers perceived their predominately middle-aged workforce was the "wrong age" to engage with social media. In a postintervention focus group, drivers discussed this issue, highlighting they were "too old" to be familiar or comfortable with social media. One workplace manager said:

> *My drivers are the wrong age-group in most cases; they're not young enough for it, they're old 'sticks-in-the-mud'… It's technology they don't use… [Most] of the guys are sixty-plus.* [TWP120]

This was reinforced by a truck driver, who commented:

> *We're all sixty years old; we don't change our thoughts on much.* [BTP]

Other Barriers to Social Media Use: "How Would You Get on Facebook?"

There are a number of other barriers to drivers' use of social media. The theme of "other barriers to social media use" can be organized into 2 subthemes—the prohibitive cost of smartphone technology and drivers' tendency to confine the use of social media to nonwork-related purposes.

The perceived cost of social media was identified by one workplace manager as a key barrier to the intervention. This manager commented:

> *And how would you get on Facebook? You'd have to have a [smart]phone... but a lot of fellas wouldn't have those... They cost $700, a lot of drivers wouldn't have $700, you've got to remember that.* [TWP420]

Another key barrier to the intervention was many drivers confined the use of social media to nonwork-related purposes. In the focus groups, drivers reflected the primary use of social media technologies like Facebook was to connect with family

and friends. For example, one manager explained why he did not "like" the page to add it to his Facebook feed:

> *I keep my Facebook strictly for personal things.* [TWP120]

Effect of Workplace Policy on Social Media Use: "We Had a No-Facebook Policy"

A number of workplace managers explained the implementation of the intervention and, specifically, their promotion of the Facebook page was limited by a "no-Facebook policy" at the workplace. Many workplaces had guidelines and policies about social media technologies, such as Facebook, at work and about posting work-related content on such sites. The managers at these workplaces felt the rules limited their capacity to promote the Facebook page to drivers. One workplace manager commented:

> *No, we didn't [promote it]… That was when we had a no-Facebook policy… Like a lot of places, there was too much rubbish getting on it, ex-employees and things like that... It becomes a bit of an issue, and there are also privacy issues.* [TWP220]

The workplace managers identified those drivers who did engage with social media technologies, such as Facebook, were most likely to do so using smartphones. The use of smartphones on the job is illegal while operating trucks. A number of workplace managers reflected on this problem. Additionally, when implementing the intervention, some workplace managers did report attempting to overcome the limitations associated with rules about the use of Facebook and phones in the workplace. One workplace manager said:

> *I can't stop them using Facebook in their private time. I'd encourage them to go on and say, 'Listen, when you're at home you might like to have a look at it.'* [TWP320]

Influence of Others on Social Media Use: "His Wife Runs His Facebook Page for Him"

Drivers were poorly engaged with social media, but their partners and friends were engaged. This was, particularly, true for drivers' wives, who were identified as being responsible for connecting drivers with social media technologies such as Facebook. For example, one workplace manager said:

> *There's one [driver] - his wife runs his Facebook page for him. She just updates the photos of all the grandkids and that sort of stuff. He doesn't have anything to do with it.* [TWP120]

Workplace managers reflected on the potential reach of a social media health promotion intervention. This included the potential to reach others in a truck driver's social circle. One workplace manager commented:

> *I think [Truckin' Healthy] sort of worked because I said 'like' and then I pressed 'share'… Even people that I'm associated with outside, that aren't drivers, liked that… The message is still getting out there… If you want anyone to know anything, just put it on Facebook* [TWP520]

Another manager commented on the visual immediacy of Facebook for health promotion:

> *You can see if people like Facebook…* [GBFF]

Increased Interest in Social Media Use: "Show Me What's on There"

The potential of social media in workplace health promotion was identified by many of the workplace managers. Some workplace managers demonstrated a keen interest in learning more about how the technology worked. For example, one workplace manager commented:

> *My daughter's got me on Facebook…I knew [the project team] were coming up and I said, "Show me what's on there"…She said all the trucks have been on…all the pictures of the flood on the farm. I said, 'You're joking!'* [TWP420]

Some workplace managers reflected about new ways of using technology in workplaces, which may be harnessed for health promotion purposes. One workplace manager explained:

> *We're rebuilding our Web face and we're going to do a lot more messaging over [that]. And we're going to get some Windows tablets for the drivers to use in their trucks, so they'll have an intranet in their trucks…From there we'll link all the policies, procedures, all the internal information on the intranet page* [TWP320]

Discussion

Principal Findings

The results of this project reveal a range of barriers to the use of social media as a health promotion intervention in transport industry workplaces. These barriers are (1) truck drivers believe they are the "wrong age" and lack the necessary skills; (2) the cost of smartphone technology is prohibitive; (3) truck drivers confine their use of social media to nonwork-related purposes; and (4) many workplaces have "no Facebook" policies. Other people around truck drivers, particularly partners, do use social media. Many workplace managers identified the potential for social media in workplace health promotion. This section will explore these complex, interrelated concepts in detail.

The key barrier to the use of social media as a health promotion intervention in transport industry workplaces was that drivers do not engage with it for a variety of reasons including demographic constraints and the prohibitive cost of smartphone technology. Further complicating this problem is drivers who do use social media do so primarily for nonwork-related purposes. These issues were not identified during the PAR process undertaken at the beginning of the project. Drivers' lack of engagement with social media is reflected in other similar studies of health promotion interventions in transport industry workplaces, including those in Australia. For example, Gilson et al [33] reported a 55% uptake in an intervention using smartphone technology to monitor nutrition and physical activity in Australian truck drivers, but concluded, "smartphone technology prohibited a number of drivers from progressing to intervention." The issue of drivers' lack of engagement with

social media is likely to resolve as smartphone technology becomes more common, in workplaces and wider society, and people of all ages become more accustomed. At present, drivers' lack of engagement with social media significantly obstructs it as a potential health promotion intervention in transport industry workplaces [42,43].

This project found drivers were poorly engaged with social media, but their partners and friends were engaged. This raises the novel possibility of reaching truck drivers through social media interventions targeted at partners and friends. The importance of family and peer support for positive health outcomes has been demonstrated by other studies on health interventions in blue-collar workplaces [44-48]. This idea was not explored in this research, but it identifies an important area for future research. This research suggests if social media interventions in older male populations are to be effective, it is important to consider—and, where possible, utilize—their family context.

Another key barrier to the effectiveness of social media in health promotion interventions in transport industry workplaces is workplace policy. Many workplaces had "no-Facebook" policies, implemented because of drivers' inappropriate use of smartphones at work. The literature and popular media reveal these as common problems. For example, there are numerous reports of drivers' dismissals after breaches of workplace social media policies [7,49,50] and of serious road accidents caused by drivers' irresponsible use of smartphones while operating heavy vehicles [51-53]. The use of any mobile phone while driving is illegal in Australia. A potential solution to this barrier, as suggested by one of the workplace managers during an informal "corridor conversation," is to implement social media health promotion interventions in which truck drivers are encouraged or, perhaps, incentivized in their spare time. This may be problematic considering issues such as long working hours, fatigue, and limited family time, anecdotally reported by many drivers.

A number of other barriers have been discussed in the literature with regard to social media in workplace health promotion interventions. These barriers include a lack of knowledge and understanding of social media among workplace managers and workplace managers' failure to accept new ways of thinking related to the use of social media [54]. These barriers were not apparent in this project. Many of the workplace managers in this research demonstrated a keen interest in social media. Some had ideas about how social media could be used in their workplace in future, for health promotion and other applications. The intervention implemented was suggested by one of the workplace managers, demonstrating workplace managers can play an active role in developing local social media strategies to engage their "hard-to-reach" workforce in health promotion.

Despite considerable barriers, this project and the literature suggest there is great potential for social media in health promotion interventions in transport industry workplaces. Much more needs to be known about how to best use social media to engage with the target audience and, particularly, "hard-to-reach" groups such as truck drivers [8]. Hudson and Hall [55] recommended creating a "trial period" prior to the implementation of a social media intervention. The aim of a trial period is to develop an understanding of the target audience and build a "credible and engaging content strategy" focusing on the audience's unique interests and needs [55].

The contextualization of interventions in this way is a fundamental aspect of the PAR process, which underpinned this project [37]. The findings paper for this project (published elsewhere) demonstrates the contextualization of interventions is pivotal to achieving workplace culture change and improvements in health knowledge and behavior in the participating truck drivers. This is an effect observed in other health promotion interventions in blue-collar workplaces [47,48,55-58]. Future projects should seek to engage participants to a greater extent in the design of a social media health promotion intervention. This would have the effect of fostering a more integrated Web-based network beyond simply "joining a Facebook group," a key aspect of engaging "hard-to-reach" groups in social media health promotion interventions [59,60].

In addition to the contextualization of a social media workplace health promotion intervention, the contextualization of strategies to support and deliver this intervention needs to be considered. In this project, workplace managers used posters displayed in depots to advertise the Facebook page. Some mentioned the page in toolbox talks with drivers. Other studies using social media to engage "hard-to-reach" groups in health promotion interventions have used more rigorous advertising, which is integrated with other interventions, including posters, personal health messages, and public announcements [59]. This delivers a more cohesive intervention strategy and provides opportunities to "piggyback" the advertisement of less-popular interventions on more-popular ones with higher levels of engagement. Other recent studies seeking to engage "hard-to-reach" groups, such as blue-collar workers, using social media health promotion interventions, deliver interventions through a range of different channels, including Facebook, Twitter, Skype, Instagram, and other popular apps [59,61-64]. This multipronged approach significantly increases the reach of the intervention and provides additional opportunities to tailor strategies to each workplace's unique context.

Limitations and Future Research

There are a number of limitations to this project. First, the convenience sample of truck drivers may create self-selection bias. Second, the project relied on the self-reporting of data in interview and focus groups, which may have resulted in socially desirable responses. Lastly, the small sample size and highly contextual nature of the social media intervention limits the generalizability of findings. However, other organizations with a small-to-medium sized, blue-collar, mobile workforce may find them of relevance.

Future projects investigating social media in "hard-to-reach" workplaces should consider a larger sample from a wider demographic, including rural and regional workplaces, to improve the generalizability. Consideration should be given to improving engagement with social media interventions by integrating these with other interventions that do not use social media, by using a variety of social media technologies and engaging workplace managers, workers, and their families in

the intervention design. Furthermore, future workplace health promotion interventions should consider reaching truck drivers through social media interventions targeted at partners and friends.

Conclusions

Transport industry workplaces, in Australia and other countries, are increasingly technology-centric, and truck drivers are engaging with social media in their workplaces on a day-to-day basis. This project found that truck drivers participating in a workplace health promotion intervention involving Facebook were generally poorly engaged because of their older age, lack of skills with social media, the cost of smartphone technology, or their workplace "no Facebook" policy. Truck drivers' family and friends were engaged with this intervention, which suggests if social media interventions in older male populations are to be effective, it is important to consider—and, where possible, utilize—their family context. The findings showed a high level of engagement from workplace managers in developing local social media strategies, which attests to the potentially important role of managers in engaging "hard-to-reach" workforce in health promotion. Therefore, it is important to gain greater insights into social media in workplace health promotion interventions and understand the enablers and inhibitors to potential use in the transport industry.

Despite some barriers evident in truck drivers, it is important to explore the potential of social media in workplace health promotion interventions in the transport industry. Much more needs to be known about how to best use social media to engage with "hard-to-reach" groups such as truck drivers. This can be achieved by methods of inquiry, such as PAR, which allow interventions to fit the sociocultural context of workplaces. Strategies for the contextualization include the involvement of occupational groups in the design of social media workplace health promotion interventions and strategies to support and deliver interventions. The context of workers themselves, including their age and familiarity with social media, and work, workplace, and family context is important to consider in this process. The "context matters" because it enables the development of more nuanced, integrated, and extensive Web-based networks, key aspects of engaging "hard-to-reach" groups in social media interventions for health promotion.

Acknowledgments

The authors would like to thank all participating workplaces. The authors thank Emily Mann for her contribution to the project. This project was supported by the Queensland Government under the *Healthier. Happier. Initiative*.

Conflicts of Interest

None declared.

References

1. Australian Bureau of Statistics. 2018 Aug. 6202.0 - Labour Force, Australia, September 2018 URL: http://www.abs.gov.au/ausstats/abs@.nsf/mf/6202.0 [WebCite Cache ID 6vV4ED27R]
2. National Transport Insurance. 2016. NTI's guide to the trucking industry URL: https://www.nti.com.au/document/nti-guide-to-the-trucking-industry-2016.pdf [accessed 2017-12-10] [WebCite Cache ID 6vbJJIdPm]
3. Sieber WK, Robinson CF, Birdsey J, Chen GX, Hitchcock EM, Lincoln JE, et al. Obesity and other risk factors: the national survey of U.S. long-haul truck driver health and injury. Am J Ind Med 2014 Jun;57(6):615-626 [FREE Full text] [doi: 10.1002/ajim.22293] [Medline: 24390804]
4. Gilson N, Pavey T, Vandelanotte C, Duncan M, Wright O, Gomersall S, et al. Shifting Gears: Physical inactivity, unhealthy diet choices and chronic disease risks in Australian truck drivers. Journal of Science and Medicine in Sport 2014 Dec;18:e98 [FREE Full text] [doi: 10.1016/j.jsams.2014.11.369]
5. Ng M, Yousuf B, Bigelow P, Eerd D. Effectiveness of health promotion programmes for truck drivers: A systematic review. Health Educ J 2015 May;74(3):270-286. [doi: 10.1177/0017896914533953]
6. Boeijinga A, Hoeken H, Sanders J. Health promotion in the trucking setting: Understanding Dutch truck drivers' road to healthy lifestyle changes. Work 2016 Oct 17;55(2):385-397. [doi: 10.3233/WOR-162409] [Medline: 27689592]
7. Boeijinga A, Hoeken H, Sanders J. Health Communication Risk Versus Planning Health Narratives Targeting Dutch Truck Drivers: Obtaining Impact Via Different Routes? Int J Commun-US 2017;11:5007-5026 [FREE Full text]
8. Boeijinga A, Hoeken H, Sanders J. An analysis of health promotion materials for Dutch truck drivers: Off target and too complex? Work 2017 Apr;56(4):539-549. [doi: 10.3233/WOR-172503] [Medline: 28269799]
9. Bennetts A. Workplace Health and Safety Queensland. Queensland Workplaces for Wellness Initiative URL: https://www.worksafe.qld.gov.au/__data/assets/pdf_file/0004/83056/qld-workplaces-for-wellness-initiative-osit-paper.pdf [accessed 2018-10-03] [WebCite Cache ID 72uEugDGd]
10. Linde JA, Nygaard KE, MacLehose RF, Mitchell NR, Harnack LJ, Cousins JM, et al. HealthWorks: results of a multi-component group-randomized worksite environmental intervention trial for weight gain prevention. Int J Behav Nutr Phys Act 2012 Feb 16;9:14 [FREE Full text] [doi: 10.1186/1479-5868-9-14] [Medline: 22340088]
11. Rongen A, Robroek SJW, van Lenthe FJ, Burdorf A. Workplace health promotion: a meta-analysis of effectiveness. Am J Prev Med 2013 Apr;44(4):406-415. [doi: 10.1016/j.amepre.2012.12.007] [Medline: 23498108]

12. Wang Z, Tchernev JM, Solloway T. A dynamic longitudinal examination of social media use, needs, and gratifications among college students. Computers in Human Behavior 2012 Sep;28(5):1829-1839. [doi: 10.1016/j.chb.2012.05.001]
13. Kaplan AM. If you love something, let it go mobile: Mobile marketing and mobile social media 4x4. Business Horizons 2012 Mar;55(2):129-139. [doi: 10.1016/j.bushor.2011.10.009]
14. Shepherd A, Sanders C, Doyle M, Shaw J. Using social media for support and feedback by mental health service users: thematic analysis of a twitter conversation. BMC Psychiatry 2015;15:29 [FREE Full text] [doi: 10.1186/s12888-015-0408-y] [Medline: 25881089]
15. Druce C. Commercial Motor. 2011 Aug 25. Meet today's trucker URL: http://archive.commercialmotor.com/article/25th-august-2011/14/meet-todays-trucker [accessed 2017-11-30] [WebCite Cache ID 6vMM4TU32]
16. Atlas Van Lines. King of the Road survey results 2016. URL: https://www.atlasvanlines.com/AtlasVanLines/media/General/2016King-of-the-Road-Survey.pdf [accessed 2017-12-11] [WebCite Cache ID 6vbUTjVSy]
17. Capital Lease Group. Survey finds fleet managers rapidly adopting smartphones and apps URL: http://www.capitalleasegroup.com/telematics/survey-finds-fleet-managers-rapidly-adopting-smartphones-and-apps [accessed 2017-12-10] [WebCite Cache ID 6vbWPLH52]
18. Big Rigs. 2017 May 19. Big brother is watching the good and the bad URL: https://www.bigrigs.com.au/news/big-brother-is-watching-the-good-and-bad/3179829/ [accessed 2017-12-06] [WebCite Cache ID 6vUuBhSFz]
19. Moody S. Big Rigs. 2017 Aug 14. OUTRAGEOUS! How angry drivers rule our roads URL: https://www.bigrigs.com.au/news/outrageous-how-angry-drivers-rule-our-roads/3212036/ [accessed 2017-12-07] [WebCite Cache ID 6vWuPRmGb]
20. Big Rigs. 2016 Jun 28. Isolation, long hours risky business URL: https://www.bigrigs.com.au/news/isolation-long-hours-risky-business/3053136/ [accessed 2017-12-07] [WebCite Cache ID 6vWuCaOnp]
21. Big Rigs. 2017 Jun 17. Truckie stood down after dash cam captures horror drive URL: https://www.bigrigs.com.au/news/driver-forced-road-truck-attempts-overtake/3186871/ [accessed 2017-12-07] [WebCite Cache ID 6vWtwwyWN]
22. Bartels C. Big Rigs. 2016 Aug 16. Were authorities too harsh with lost truckie? URL: https://www.bigrigs.com.au/news/were-authorities-too-harsh-with-lost-truckie/3077832/ [accessed 2017-12-07] [WebCite Cache ID 6vWtFaASo]
23. Morcombe J. Big Rigs. 2017 May 4. Truck clocked doing 70km/h in crash spot URL: https://www.bigrigs.com.au/news/truck-clocked-doing-70kmh-crash-spot/3173844/ [accessed 2018-10-04] [WebCite Cache ID 72uHEuLx1]
24. Mortimer L. Big Rigs. 2017 Apr 5. Confusion reigns as oversized trucks banned from Rocky URL: https://www.bigrigs.com.au/news/chaos-truck-stuck-near-power-lines-bruce-hwy-detou/3163241/ [accessed 2017-12-07] [WebCite Cache ID 6vWtS8aLM]
25. Dills T. Overdrive. 2016 Nov 15. Grow a mustache for men's health URL: https://www.overdriveonline.com/grow-a-mustache-for-mens-health/ [accessed 2017-12-07] [WebCite Cache ID 6vWscUrPB]
26. Dills T. Overdrive. 2016 Oct 28. Of personal bests, the St. Christopher fund and truckin', writin' and runnin' URL: https://www.overdriveonline.com/of-personal-bests-the-st-christopher-fund-and-truckin-writin-and-runnin/ [accessed 2017-12-07] [WebCite Cache ID 6vWsFTZU3]
27. Parker W. Overdrive. 2016 Oct 14. Truckers for Truckers and TruckersFinalMile.org unite URL: https://www.overdriveonline.com/truckers-for-truckers-and-truckersfinalmile-org-unite/ [accessed 2017-12-07] [WebCite Cache ID 6vWroOO4a]
28. Jones K. Fleet Owner. 2016 Oct 04. New app locates truck parking spaces URL: http://www.fleetowner.com/fleet-management/new-app-locates-truck-parking-spaces [accessed 2017-11-30] [WebCite Cache ID 6vLg8OysJ]
29. Dills T. Overdrive. 2017 Nov 13. New app puts tracking delivery confirmation in the hands of independents URL: http://www.overdriveonline.com/new-app-puts-trackingdelivery-confirmation-in-the-hands-of-independents/ [accessed 2017-11-30] [WebCite Cache ID 6vLhnoxeo]
30. Williams A. CMO. 2017 Jun 23. Uber for the truckies: How one Aussie startup is disrupting the freight industry URL: https://www.cmo.com.au/article/620949/uber-truckies-how-one-aussie-startup-disrupting-freight-industry/ [accessed 2017-12-07] [WebCite Cache ID 6vX1Hb9cA]
31. Herbert L. Australian Broadcasting Commission Rural News. 2014 Aug 8. Truck drivers get help in virtual waiting room URL: http://www.abc.net.au/news/2014-04-08/truck-health-app/5374828 [accessed 2017-11-27] [WebCite Cache ID 6vHnvba9W]
32. Gilson N, Pavey T, Gomersall S, Vandelanotte C, Duncan M, Wright O, et al. Shifting gears: Process evaluation of an activity tracker and smart phone application to promote healthy lifestyle choices in Australian truck drivers. Journal of Science and Medicine in Sport 2014 Dec 1;18:e124. [doi: 10.1016/j.jsams.2014.11.097]
33. Gilson ND, Pavey TG, Wright OR, Vandelanotte C, Duncan MJ, Gomersall S, et al. The impact of an m-Health financial incentives program on the physical activity and diet of Australian truck drivers. BMC Public Health 2017 May 18;17(1):467 [FREE Full text] [doi: 10.1186/s12889-017-4380-y] [Medline: 28521767]
34. Kenny M. Stock Journal. 2013 Jun 22. Support for SA trucks URL: https://www.stockjournal.com.au/story/3589137/support-for-sa-truckies/ [accessed 2018-10-03] [WebCite Cache ID 72uL83qBU]
35. Big Rigs. 2013 Mar 31. Truckies to step it up URL: https://www.bigrigs.com.au/news/truckies-to-step-it-up/1805336/ [accessed 2017-12-06] [WebCite Cache ID 6vV17wpua]

36. Big Rigs. 2014 Jun 10. QUT Health Intervention to get drivers to eat better URL: https://www.bigrigs.com.au/news/qut-health-intervention-get-drivers-eat-better/2284951/ [accessed 2018-03-14] [WebCite Cache ID 6xttIybUX]
37. Crane P, O'Regan M. On PAR: Using Participatory Action Research to Improve Early Intervention. Canberra: Department of Families, Housing, Community Services and Indigenous Affairs, Australian Government; Aug 2010.
38. Judd J, Keleher H. Reorienting health services in the Northern Territory of Australia: a conceptual model for building health promotion capacity in the workforce. Glob Health Promot 2013 Jun;20(2):53-63. [doi: 10.1177/1757975913486685] [Medline: 23797940]
39. Bernhardt JM. Communication at the core of effective public health. Am J Public Health 2004 Dec;94(12):2051-2053. [Medline: 15569948]
40. Sendall M, Crane P, McCosker L, Biggs H, Fleming M, Rowland B. Workplace interventions to improve truck drivers' health knowledge, behaviours and self-reported outcomes. Road Transp Res 2016 Mar;25(1):31-43.
41. Liamputtong P. Qualitative research methods. Fourth edition. South Melbourne, Victoria: Oxford University Press Australia; Nov 20, 2012.
42. Heaton K, Combs B, Griffin R. Truck Drivers' Use of the Internet: A Mobile Health Lifeline. Workplace Health Saf 2017 Jun;65(6):240-247. [doi: 10.1177/2165079916665401] [Medline: 27807256]
43. Solenhill M, Grotta A, Pasquali E, Bakkman L, Bellocco R, Trolle LY. The Effect of Tailored Web-Based Feedback and Optional Telephone Coaching on Health Improvements: A Randomized Intervention Among Employees in the Transport Service Industry. J Med Internet Res 2016 Aug 11;18(8):e158 [FREE Full text] [doi: 10.2196/jmir.4005] [Medline: 27514859]
44. Heinen L, Darling H. Addressing obesity in the workplace: the role of employers. Milbank Q 2009 Mar;87(1):101-122 [FREE Full text] [doi: 10.1111/j.1468-0009.2009.00549.x] [Medline: 19298417]
45. Escoffery C, Kegler MC, Alcantara I, Wilson M, Glanz K. A qualitative examination of the role of small, rural worksites in obesity prevention. Prev Chronic Dis 2011 Jul;8(4):A75 [FREE Full text] [Medline: 21672399]
46. Staats U, Lohaus D, Christmann A, Woitschek M. Fighting against a shortage of truck drivers in logistics: Measures that employers can take to promote drivers' work ability and health. Work 2017;58(3):383-397. [doi: 10.3233/WOR-172626] [Medline: 29036873]
47. Oliffe JL, Bottorff JL, Sharp P, Caperchione CM, Johnson ST, Healy T, et al. Healthy Eating and Active Living: Rural-Based Working Men's Perspectives. Am J Mens Health 2017 Nov;11(6):1664-1672 [FREE Full text] [doi: 10.1177/1557988315619372] [Medline: 26669775]
48. Bottorff J, Seaton C. The Conversation, Australia. 2017 Oct 4. Friendly competition motivates men to get healthy on the job URL: https://theconversation.com/friendly-competition-motivates-men-to-get-healthy-on-the-job-83867 [accessed 2017-12-13] [WebCite Cache ID 6vg1Aq64M]
49. Carson V. The Courier Mail. 2014 Dec 29. Sacked BHP Billiton truck drivers get jobs back after Fair Work Commission hearing URL: http://www.couriermail.com.au/news/queensland/sacked-bhp-billiton-truck-drivers-get-jobs-back-after-fair-work-commission-hearing/news-story/81d9d5e5bcbab058e756a24eff759970?nk=aa03614e30e3663ac7493d15f11eec36-1513150172 [accessed 2017-12-13] [WebCite Cache ID 6vfr5HC3s]
50. Cullen S. Australian Broadcasting Commission News. 2012 Oct 3. Linfox loses appeal over truckie's Facebook comments URL: http://www.abc.net.au/news/2012-10-03/linfox-loses-appeal-over-truckies-facebook-comments/4293644 [accessed 2017-11-27] [WebCite Cache ID 6vHombaku]
51. Cook C. Australian Broadcasting Commission News. 2013 Aug 28. Truck driver pleads guilty over fatal Singleton crash URL: http://www.abc.net.au/news/2013-08-28/truck-driver-pleads-guilty-over-fatal-singleton-crash/4917864 [accessed 2017-11-27] [WebCite Cache ID 6vHoroIRD]
52. Stampler L. Time. 2013 Nov 4. Trucker in deadly crash was ogling women's pics URL: http://nation.time.com/2013/11/04/trucker-in-deadly-crash-was-ogling-womens-pics/ [accessed 2017-11-27] [WebCite Cache ID 6vHjn1Did]
53. BBC News. 2016 Oct 31. Lorry driver Tomasz Kroker jailed over four-death crash URL: http://www.bbc.com/news/uk-england-37823457 [accessed 2017-12-13] [WebCite Cache ID 6vfs6qca9]
54. Muuraiskangas S, Harjumaa M, Kaipainen K, Ermes M. Process and Effects Evaluation of a Digital Mental Health Intervention Targeted at Improving Occupational Well-Being: Lessons From an Intervention Study With Failed Adoption. JMIR Ment Health 2016 May 11;3(2):e13 [FREE Full text] [doi: 10.2196/mental.4465] [Medline: 27170553]
55. Hudson H, Hall J. Value of social media in reaching and engaging employers in Total Worker Health. J Occup Environ Med 2013 Dec;55(12 Suppl):S78-S81. [doi: 10.1097/JOM.0000000000000035] [Medline: 24284750]
56. Kolmet M, Mariño R, Plummer D. Anglo-Australian male blue-collar workers discuss gender and health issues. International Journal of Men's Health 2006 Jan 1;5(1):81-91. [doi: 10.3149/jmh.0501.81]
57. Rasmussen K, Glasscock DJ, Hansen ON, Carstensen O, Jepsen JF, Nielsen KJ. Worker participation in change processes in a Danish industrial setting. Am J Ind Med 2006 Sep;49(9):767-779. [doi: 10.1002/ajim.20350] [Medline: 16804911]
58. Silverman M. Silverman Research. 2012. The social media garden: A digital era research study into social media at work URL: http://www.silvermanresearch.com/wp-content/uploads/2012/06/The-Social-Media-Garden-Report1.pdf [accessed 2017-11-27] [WebCite Cache ID 6vHawprWY]

59. Du Plessis K, Cronin D, Corney T, Green E. Australian blue-collar men's health and well-being: contextual issues for workplace health promotion interventions. Health Promot Pract 2013 Sep;14(5):715-720. [doi: 10.1177/1524839912464046] [Medline: 23159997]
60. Lingard H, Turner M. Improving the health of male, blue collar construction workers: a social ecological perspective. Construction Management and Economics 2015 Mar 06;33(1):18-34. [doi: 10.1080/01446193.2014.997259]
61. Martinez O, Wu E, Shultz AZ, Capote J, López RJ, Sandfort T, et al. Still a hard-to-reach population? Using social media to recruit Latino gay couples for an HIV intervention adaptation study. J Med Internet Res 2014;16(4):e113 [FREE Full text] [doi: 10.2196/jmir.3311] [Medline: 24763130]
62. Gold J, Pedrana AE, Stoove MA, Chang S, Howard S, Asselin J, et al. Developing health promotion interventions on social networking sites: recommendations from The FaceSpace Project. J Med Internet Res 2012;14(1):e30 [FREE Full text] [doi: 10.2196/jmir.1875] [Medline: 22374589]
63. Korda H, Itani Z. Harnessing social media for health promotion and behavior change. Health Promot Pract 2013 Jan;14(1):15-23. [doi: 10.1177/1524839911405850] [Medline: 21558472]
64. Haines-Saah RJ, Kelly MT, Oliffe JL, Bottorff JL. Picture Me Smokefree: a qualitative study using social media and digital photography to engage young adults in tobacco reduction and cessation. J Med Internet Res 2015;17(1):e27 [FREE Full text] [doi: 10.2196/jmir.4061] [Medline: 25624064]

Abbreviations

PAR: Participatory Action Research

Edited by G Eysenbach; submitted 19.12.17; peer-reviewed by R Pankomera, JR Bautista; comments to author 22.02.18; revised version received 12.07.18; accepted 22.07.18; published 01.11.18.

Please cite as:
Sendall MC, McCosker LK, Crane P, Rowland B, Fleming M, Biggs HC
Using Facebook for Health Promotion in "Hard-to-Reach" Truck Drivers: Qualitative Analysis
J Med Internet Res 2018;20(11):e286
URL: https://www.jmir.org/2018/11/e286/
doi:10.2196/jmir.9689
PMID:30389653

Short Paper

Correlations Between Hospitals' Social Media Presence and Reputation Score and Ranking: Cross-Sectional Analysis

Justin D Triemstra[1], MD; Rachel Stork Poeppelman[2], MD; Vineet M Arora[3], MD, MAPP

[1]Department of Pediatrics, Helen DeVos Children's Hospital, Michigan State University College of Human Medicine, Grand Rapids, MI, United States
[2]Department of Pediatrics, Nationwide Children's Hospital, The Ohio State University College of Medicine, Columbus, OH, United States
[3]Department of Medicine, Section of General Internal Medicine, University of Chicago, Chicago, IL, United States

Corresponding Author:
Justin D Triemstra, MD
Department of Pediatrics
Helen DeVos Children's Hospital
Michigan State University College of Human Medicine
330 Barclay Avenue
Suite 203
Grand Rapids, MI,
United States
Phone: 1 616 391 2123
Fax: 1 616 391 250
Email: justin.triemstra@helendevoschildrens.org

Abstract

Background: The US News and World Report reputation score correlates strongly with overall rank in adult and pediatric hospital rankings. Social media affects how information is disseminated to physicians and is used by hospitals as a marketing tool to recruit patients. It is unclear whether the reputation score for adult and children's hospitals relates to social media presence.

Objective: The objective of our study was to analyze the association between a hospital's social media metrics and the US News 2017-2018 Best Hospital Rankings for adult and children's hospitals.

Methods: We conducted a cross-sectional analysis of the reputation score, total score, and social media metrics (Twitter, Facebook, and Instagram) of hospitals who received at least one subspecialty ranking in the 2017-2018 US News publicly available annual rankings. Regression analysis was employed to analyze the partial correlation coefficients between social media metrics and a hospital's total points (ie, rank) and reputation score for both adult and children's hospitals while controlling for the bed size and time on Twitter.

Results: We observed significant correlations for children's hospitals' reputation score and total points with the number of Twitter followers (total points: r=.465, P<.001; reputation: r=.524, P<.001) and Facebook followers (total points: r=.392, P=.002; reputation: r=.518, P<.001). Significant correlations for the adult hospitals reputation score were found with the number of Twitter followers (r=.848, P<.001), number of tweets (r=.535, P<.001), Klout score (r=.242, P=.02), and Facebook followers (r=.743, P<.001). In addition, significant correlations for adult hospitals total points were found with Twitter followers (r=.548, P<.001), number of tweets (r=.358, P<.001), Klout score (r=.203, P=.05), Facebook followers (r=.500, P<.001), and Instagram followers (r=.692, P<.001).

Conclusions: A statistically significant correlation exists between multiple social media metrics and both a hospital's reputation score and total points (ie, overall rank). This association may indicate that a hospital's reputation may be influenced by its social media presence or that the reputation or rank of a hospital drives social media followers.

(J Med Internet Res 2018;20(11):e289) doi:10.2196/jmir.9713

KEYWORDS

social media; hospitals; benchmarking; hospital ranking

Introduction

The *US News and World Report* Best Hospital Rankings for adult and children's hospitals have been released annually since 1990 [1,2]. These rankings affect the public reputation of a hospital and have been predicted to affect nonemergent volumes and total revenue by 5% in year-to-year changes in rankings [3]. Therefore, these rankings are highly anticipated by physicians and hospital administrators and are commonly used as marketing tools by hospitals to attract patients [4].

The final ranking of a hospital or subspecialty is determined by objective quality data in addition to subjective reputation data. The reputation score for each subspecialty is derived from annual surveys asking physicians where they would send their sickest patients, ignoring cost or location [1,2]. Owing to this subjective component of the total score, the *US News* hospital rankings have been criticized in recent years [5]. Therefore, *US News* has adjusted the composition of the final score, relying more on objective data in recent years [1,2]. However, Bush et al and Cua et al recently reported that the reputation of a hospital still disproportionately affects the final ranking in children's [6] and adult hospitals by showing that the reputation component has a more significant influence on the total *US News* score than the objective components [7].

In this digital age, it is apparent that social media can influence how information is disseminated and the reputation of organizations; for example, *US News* uses a social media platform, Doximity, to distribute their annual survey to obtain the peer score for their annual rankings [2]. In addition, physicians in all fields have become increasingly engaged with social media [8]. Most hospitals have a Web-based social media presence [9] and use it for hospital promotion, education, community partnership, and fundraising purposes [10]. Therefore, because it has been shown that the reputation of a hospital contributes to *US News* rankings [6,7] and that social media presence can affect an organization's reputation in today's society, it would be important to understand whether social media presence correlates with *US News* hospital rankings. Interestingly, Ciprut et al found that the Twitter activity of urology departments was associated with *US News* rankings in Urology [11], although this analysis was not expanded to include other social media platforms or specialties. To address these gaps, this study aims to conduct a cross-sectional analysis of the adult and children's hospitals' social media presence over multiple platforms and a hospital's *US News* rank to determine whether these correlations exist.

Methods

Hospital Ranking Data

We conducted a cross-sectional analysis of data from the 2017-2018 *US News and World Report* publicly available annual rankings for adult and children's hospitals. *US News* hospital rankings are composed of both objective and subjective data that determine the total *US News* score; this score then determines the overall rank of the hospital within each subspecialty. The objective data are primarily based on survival outcomes, patient safety indices, and other care-related indicators, including nursing ratios, patient volume, and other subspecialty-specific quality of care data [1,2]. The subjective component consists of an annual survey of board-certified physicians who list up to 5 adult hospitals (up to 10 for children's hospitals) that have the highest expertise in that subspecialty or where they would send their sickest patients, ignoring cost or location. This survey determines a hospital's reputation score for that subspecialty [1,2]. Although the significance of the reputation component has been decreasing over time, it still composes 27.5% of the score for adult hospitals and 15% (8.5% in cardiology and heart surgery) for children's hospitals.

Using these data, we determined a hospital's total reputation score, as previously described in the literature, as the sum of each subspecialty reputation score for an individual hospital [6]. Although the total raw *US News* score is not reported for every hospital, the formula is reported to the public and was used to calculate each hospital's total *US News* score [1,2]. These publicly available data were manually extracted from the 2017-2018 *US News* Adult and Children's hospital reports. The descriptive statistics for these data can be viewed in those publicly available reports [1,2] with a summary of the statistics available in Multimedia Appendix 1. We used Microsoft Excel software (Microsoft Corporation) to store the data for this study.

Statistical Analysis

The primary objective of this study was to analyze the association between social media metrics and *US News* rankings. Twitter, Facebook, and Instagram metrics for all hospitals that received a ranking in at least one subspecialty in *US News* were obtained on August 30, 2017, for children's hospitals and October 12, 2017, for adult hospitals (Multimedia Appendix 1). For Twitter metrics, we obtained the number of followers, number of tweets, and Klout score (max=100), a metric that quantifies the Web-based social influence over the past 90 days, for each adult and children's hospital with active accounts. All hospitals had active social media accounts for >90 days, ensuring that the Klout score represents an accurate estimation of the influence for all hospitals. Furthermore, the number of Facebook and Instagram followers were obtained for each adult and children's hospital with an active account.

In this study, multiple regression analyses were used with social media metrics as the dependent variable. In one set of analyses, the independent variables were the total reputation score and the number of hospital beds at each institution. For the other set of analyses, the independent variables were the total points and the number of hospital beds at each institution. For analyses of tweets and Twitter followers, the number of months on Twitter was also included as one of the independent variables. Of note, we did not include a control for the active months of activity on Facebook and Instagram because they are not easily identifiable or reported. From these analyses, partial correlation coefficients were calculated between each social media metric and both the reputation score and total points. We defined statistical significance as $P<.05$. Analyses were performed using Stata v.15.0 (StataCorp).

Results

Children's Hospitals

Table 1 summarizes the social media and correlation statistics for children's hospitals when controlling for time on Twitter for Twitter metrics and bed size. Of the 79 children's hospitals with at least one subspecialty ranked, 69, 68, and 46 hospitals had active Twitter, Facebook, or Instagram accounts, respectively. Children's hospitals without an independent account were removed from the analysis. Twitter and Facebook followers correlated significantly with the total points (ie, higher rank) as well as with the reputation for children's hospitals (P<.001).

Table 1. Social media metrics for US Children's Hospitals and correlation with *US News* ranking and reputation score.

Social media metrics	Number of hospitals	Median (range)	Partial correlation coefficient for total points (*P* value)	Partial correlation coefficient for total reputation (*P* value)
Twitter followers	69	8306 (163-63,400)	.465 (<.001)	.524 (<.001)
Number of tweets	69	4714 (116-23,400)	.0181 (.93)	−.036 (.80)
Klout score	62	58 (23-81)	.117 (.38)	.067 (.61)
Facebook followers	68	59,141 (772-649,407)	.392 (.002)	.518 (<.001)
Instagram followers	46	4879 (260-5600)	.197 (.21)	.177 (.26)

Adult Hospitals

Table 2 summarizes the social media and correlation statistics for adult hospitals when controlling for the bed size and time on Twitter for Twitter metrics. Of the 128 adult hospitals with, at least, one subspecialty ranked, 106, 105, and 62 had active Twitter, Facebook, or Instagram accounts, respectively. For adult hospitals, Twitter followers, the number of tweets, Klout score, Facebook followers, and Instagram followers correlated significantly with the total points (Table 2). Except for Instagram followers, these metrics also correlated significantly with the total reputation (Table 2). Hospitals with scores of 0 in the number of tweets and Klout score had active accounts with followers but no posts.

Table 2. Social media metrics for US Adult Hospitals and correlation with *US News* ranking and reputation score.

Social media metrics	Number of hospitals	Median (range)	Partial correlation coefficient for total points (*P* value)	Partial correlation coefficient for total reputation (*P* value)
Twitter followers	106	8074 (125-1,710,000)	.548 (<.001)	.848 (<.001)
Number of tweets	106	6087 (0-39,400)	.358 (<.001)	.535 (<.001)
Klout score	98	60 (0-92)	.203 (.05)	.242 (.02)
Facebook followers	105	18,582 (205-1,995,180)	.500 (<.001)	.743 (<.001)
Instagram followers	62	1615 (153-49,700)	.692 (<.001)	.831 (.26)

Discussion

Principal Findings

In this first study comparing social media metrics for adult and pediatric hospitals with *US News* rankings and reputation score, we found that both adult and children's hospitals with more Twitter and Facebook followers had a higher 2017-2018 *US News* reputation score and total points (ie, overall rank). Although this correlation suggests that social media presence may influence the reputation of a hospital, it is also possible that the reputation and rank of a hospital drive social media followers. We also found that adult hospitals with more tweets and higher Klout scores, a measure of active social media participation, were associated with a higher reputation score and overall rank. Although it is possible that more prestigious hospitals have more followers because of prestige, it is also possible that this correlation may reflect that active engagement with social media may affect a hospital's *US News* reputation score. Therefore, we have found marked correlations between a hospital's social media metrics and its hospital rankings; however, further longitudinal studies will need to be conducted to determine causality.

Limitations

There are a few limitations to our initial study. First, because social media metrics are cumulative, it is possible that hospitals which adopted social media earlier have larger followings and therefore, more posts. Furthermore, larger hospitals may have superior social media metrics because of increased resources for marketing and social media engagement. When controlling for time on Twitter and bed size of hospitals, these correlations remained significant, although other external factors, such as the size of the marketing department and total revenue of a hospital, may contribute as well. Second, owing to the cross-sectional analysis used in our study, we were unable to assess whether the increased social media presence over time correlates directly with an increase in reputation; this type of analysis could be conducted as a future study to confirm this finding. Third, because the subjective component of the rankings is based on physician voting, it would be helpful to know how many followers of a hospital are physicians; however, these data are not readily available, and usernames do not always

identify the profession of their owner. Finally, Instagram correlated with both scores for adult and children's hospitals, but it was only significant for the adult total score; this is likely because of Instagram being a newer platform with fewer hospitals active on Instagram.

Conclusions

We believe that this study is the first to report the association between social media metrics and *US News and World Report* annual hospital rankings for both adult and children's hospitals. This study establishes a correlation between increased social media engagement and *US News* rankings and reputation score for hospitals. Therefore, because previous studies have shown that physicians are increasingly engaged in large networks on social media [8] and *US News* rankings are disproportionately influenced by reputation scores [7], increasing a hospital's social media presence could be a potential method of improving the reputation of hospitals and their rank in *US News* annual Best Hospital Rankings. Future studies should include a content analysis of Twitter, Facebook, and Instagram posts to determine the type of content being posted by hospitals and an assessment of the relationship between the use of social media and changes in hospital rankings over time to confirm whether increasing social media presence correlates with an increase in the hospital rank.

Acknowledgments

We thank Alan T Davis, PhD, Associate Professor, Department of Surgery, Michigan State University, for his assistance in statistical analysis and review of the paper. He did not receive compensation.

Conflicts of Interest

VMA served on the US News & World Report Patient Safety Panel in 2015-2016.

Multimedia Appendix 1

Summary of the total points, reputation score, bed size for adult and children's hospitals, and descriptive statistics for each children's hospital's and adult hospital's social media data.

[XLSX File (Microsoft Excel File), 25KB - jmir_v20i11e289_app1.xlsx]

References

1. Murrey OE, Powell R, Murphy J, Bell D, Stanley M. U.S. News & World Report. 2017. U.S. News & World Report 2017-18 Best Children's Hospitals 2017-2018 URL: http://static.usnews.com/documents/health/best-hospitals/BCH_Methodology_2017-18.pdf [accessed 2017-12-19] [WebCite Cache ID 6vofAVHb9]
2. Murrey OE, Murphy J, Bell D, Morley M, Stanley M, Powell R. US News & World Report. 2017. Methodology U.S. News & World Report 2017-18 Best Hospitals: Specialty Rankings URL: http://static.usnews.com/documents/health/best-hospitals/BH_Methodology_2017-18.pdf [accessed 2017-12-19] [WebCite Cache ID 6vofULoSd]
3. Pope DG. Reacting to rankings: evidence from "America's Best Hospitals". J Health Econ 2009 Dec;28(6):1154-1165. [doi: 10.1016/j.jhealeco.2009.08.006] [Medline: 19818518]
4. Larson RJ, Schwartz LM, Woloshin S, Welch HG. Advertising by academic medical centers. Arch Intern Med 2005 Mar 28;165(6):645-651. [doi: 10.1001/archinte.165.6.645] [Medline: 15795340]
5. Sehgal AR. The role of reputation in U.S. News & World Report's rankings of the top 50 American hospitals. Ann Intern Med 2010 Apr 20;152(8):521-525. [doi: 10.7326/0003-4819-152-8-201004200-00009] [Medline: 20404383]
6. Bush RA, Quigley EJ, Fox L, Garcia-Bassets I. Role of reputation in top pediatric specialties rankings. Pediatrics 2011 Dec;128(6):1168-1172. [doi: 10.1542/peds.2011-1027] [Medline: 22123878]
7. Cua S, Moffatt-Bruce S, White S. Reputation and the Best Hospital Rankings: What Does It Really Mean? Am J Med Qual 2017;32(6):632-637. [doi: 10.1177/1062860617691843] [Medline: 28693335]
8. Mishori R, Singh LO, Levy B, Newport C. Mapping physician Twitter networks: describing how they work as a first step in understanding connectivity, information flow, and message diffusion. J Med Internet Res 2014;16(4):e107 [FREE Full text] [doi: 10.2196/jmir.3006] [Medline: 24733146]
9. Griffis HM, Kilaru AS, Werner RM, Asch DA, Hershey JC, Hill S, et al. Use of social media across US hospitals: descriptive analysis of adoption and utilization. J Med Internet Res 2014;16(11):e264 [FREE Full text] [doi: 10.2196/jmir.3758] [Medline: 25431831]
10. Wong CA, Ostapovich G, Kramer-Golinkoff E, Griffis H, Asch DA, Merchant RM. How U.S. children's hospitals use social media: A mixed methods study. Healthc (Amst) 2016 Mar;4(1):15-21. [doi: 10.1016/j.hjdsi.2015.12.004] [Medline: 27001094]
11. Ciprut S, Curnyn C, Davuluri M, Sternberg K, Loeb S. Twitter Activity Associated With U.S. News and World Report Reputation Scores for Urology Departments. Urology 2017 Oct;108:11-16. [doi: 10.1016/j.urology.2017.05.051] [Medline: 28669746]

Edited by G Eysenbach; submitted 21.12.17; peer-reviewed by S Loeb, A Olszewski, J Calvert, J Percival; comments to author 18.03.18; revised version received 01.06.18; accepted 11.07.18; published 08.11.18.

Please cite as:
Triemstra JD, Poeppelman RS, Arora VM
Correlations Between Hospitals' Social Media Presence and Reputation Score and Ranking: Cross-Sectional Analysis
J Med Internet Res 2018;20(11):e289
URL: http://www.jmir.org/2018/11/e289/
doi:10.2196/jmir.9713
PMID:30409768

Original Paper

Health Care Professionals' Social Media Behavior and the Underlying Factors of Social Media Adoption and Use: Quantitative Study

Joe Hazzam[1], BSc, MBA; Abdelmounaim Lahrech[1], PhD

Faculty of Business and Law, The British University in Dubai, Dubai, United Arab Emirates

Corresponding Author:
Joe Hazzam, BSc, MBA
Faculty of Business and Law
The British University in Dubai
Dubai Academic City
Dubai,
United Arab Emirates
Phone: 971 504278354
Email: joe_hazzam@hotmail.com

Abstract

Background: In the last decade, social media has emerged as a newer platform for knowledge dissemination, information exchange, and interpersonal communication for health care professionals (HCPs). However, the underlying behaviors of HCPs and the ethical use of social media for productivity enhancement and a sustainable health care system remain ambiguous.

Objective: This study seeks to understand the factors that relate to the frequency use of social media in the health care discipline. It also aims to explore the underlying online behaviors of HCPs, which include the exchange of medical information with peers, interpersonal communication, and productivity enhancement in their daily practice.

Methods: This study adopted the quantitative method in collecting and analyzing data. A survey instrument based on the behavioral and technology acceptance theories was developed for this purpose. The survey was distributed via social media platforms to 973 participants that included physicians, pharmacists, and allied HCPs working in the United Arab Emirates. The responses from 203 completed questionnaires (response rate 20.3%) were analyzed.

Results: Of 203 respondents, 133 HCPs used WhatsApp (65.5%); therefore, WhatsApp had the highest number of users compared to Facebook and YouTube, with 101 users out of 203 (49.7%). Overall, 109 of 203 (53.6%) HCPs used social media platforms for the exchange of peer medical information and 108 of 203 (53.2%) used social media several times during the day to improve their interpersonal communication with colleagues. However, only 71 of 203 (34.9%) utilized social media to enhance their productivity in general. The structural model equation showed that behavioral intention (beta=.47; $P<.001$), habit (beta=.26; $P=.001$), attitude (beta=.20; $P=.002$), and perceived usefulness (beta=.12; $P=.09$) were positively and significantly related to frequency of use. The model explained a rate of 45% variance in the frequency of use and a rate of 17% variance in the social media intention of use.

Conclusions: The research highlights the significant factors that relate to the adoption of social media platforms in health care practice. Based on the findings of this study, the use of online platforms facilitates the exchange of medical information among peers and enhances the share of experiences that support HCP's learning and development. Moreover, social media platforms foster a higher level of communication among practitioners and might improve daily productivity. Future researchers might explore other variables such as training and external factors. For instance, they may draw on areas related to guidelines and policies. From this standpoint, the health care discipline can benefit from highly interactive platforms and adopt them for development, collaboration, and better health outcomes.

(J Med Internet Res 2018;20(11):e12035) doi:10.2196/12035

KEYWORDS

social media; health care professionals and social media; integrated behavioral model; technology acceptance theories

Introduction

Overview

With the advent of the internet, the health care sector—like any other discipline—has dramatically improved in terms of the speed of information exchange and the introduction of newer communication media. In fact, this advance in technology supports practitioners in developing better skills to yield better health outcomes and inform behavioral changes. The introduction of eHealth heralds a new era in developing advanced and innovative strategies to support the decision-making process in the health care sector. The advance of online technologies will also support practitioners in collecting, managing, and interacting for higher quality service. The growing use of social media platforms in the health care disciplines might also inform health behaviors in real time and support the exchange of information for better clinical outcomes [1]. They also support health campaign programs and patient education and provide an opportunity for information exchange and expansion of physician networks. However, poor quality of information, patient confidentiality, and legal issues are some risks and challenges that could impact the effective and beneficial integration of online platforms [2].

Social media usage differs in various health care settings because of the disparity between different medical tasks and responsibilities. Researchers have started to explore the rate of social media adoption and usability from the point of view of different stakeholders. In medical education, a study found that the degree of popularity and awareness is higher in undergraduate than in postgraduate circles. However, the study revealed that both groups had an interest in the use of new technologies [3]. Other researchers have studied the rate and usage of social media in different health care populations. For instance, pharmacists' excessive use of social media is for the purpose of expanding their network and keeping in touch with old friends rather than for the purpose of education or professional development [4]. These findings are not in line with physician's behaviors using social media. Adopting online platforms by doctors is increasing; however, guidelines and policies may act as barriers and reduce the quantity of information posted or shared in an effort to protect patient confidentiality [5]. The new online media are changing the communication behaviors of various stakeholders and increasing the exchange of information among health care professionals (HCPs) for an optimal decision process. However, the value of patient privacy should be respected on highly interactive and open platforms [6].

The main factor that affects the intention of professionals to participate in social media is the creation of virtual communities of practice. In fact, there is improved knowledge sharing among colleagues, but because these online groups are private, this confines knowledge to specific users and prevents the dissemination of information within a multidisciplinary environment to help improve performance and outcomes [7]. The use of social media within a health care context is affected by ethical dilemmas and privacy concerns that could prevent users from benefiting from this highly interactive means of communication. Research and trials are trying to explain the mechanism of information transfer and interaction by different users [8]. Practitioners in different health care disciplines can benefit from the significant data generated from the patients' interaction within the social media groups. Moreover, interpersonal communication can be improved and evidence-based knowledge can be diffused faster than traditional channels. Research might explore the safe and the cost-effective use of social media to complement the evidence-based practice; moreover, the policies and guidelines can support the efficient adoption for personal development and knowledge updates [9].

The identification of the factors that promote practical use along with the effective type of platforms represents a gap in knowledge and requires further research [10]. The preliminary evidence of adopting social media in different health care practices reveals divergent opinions on the benefits, challenges, level of information exchange, communication, and productivity achieved. The study of users' characteristics, behaviors, and external factors that relate to usage intention and frequency of use represent the potential to advance knowledge on the optimal integration of social media in health care practice. Additionally, it supports the extension of technology and information acceptance theories in the highly sophisticated settings of health care [11].

The primary research question is to understand the relationships between the influential factors and the adoption and usage of social media by HCPs. The research also aims to understand the strength of relationships between the underlying factors and the adoption of social media by HCPs. It also seeks to investigate the type and the frequency use of platforms by HCPs.

This research first attempts to define the theoretical research background. Next, it discusses, through a literature review, the relationships between the essential factors, the intention of social media adoption, and the actual use. Then, it presents the proposed conceptual model and the pilot study to test this conceptual model.

Theoretical Background

The primary goal of the research is to understand the underlying behaviors, norms, and control factors that influence the intention and usage of social media by HCPs. The research proposes the presence of considerable complexities within a regulated and highly ethical discipline. Therefore, the study of the factors that inhibit or enhance the adoption of online platforms could inform future research to explore the intervention strategies that support active usage designed to improve health outcomes.

The theories of sociology, behavior change, and psychology are utilized to understand the acceptance of technology and usage in different organizational and consumer contexts. The Technology Acceptance Model (TAM) is adopted as a theoretical background due to the high level of prediction in adopting new technology [12]. The health care discipline is a highly regulated behavioral discipline that is governed by policies and guidelines. Therefore, the constructs of the Integrated Behavioral Model [13] will support the model and could inform future research to explore the intervention strategies that support active use for improving health outcomes.

The constructs of the TAM emphasize the role of perceived usefulness and ease of use as main predictors of user behavior [12]. The TAM is identified as a theoretical background for the study of the perceived usefulness and ease of use as significant predictors of using social media technologies by two specialties of physicians in sharing and receiving information [14]. The TAM and the behavioral theories of reasoned actions [15] and planned behaviors [16] are employed as the theoretical background to study the acceptance of the personal digital assistant by HCPs for practical tasks achievements and easy access to medical records [17]. The research investigates the crucial factors that relate to the user's behavioral intention. The professional image of practitioners within their social network influences their intention to use new technologies in the belief that a new system might enhance their image and social status. Therefore, the image of professionals is integrated into the research model as an important factor in the health care discipline. The acceptance of Web 2.0 tools for information sharing and interaction between nurses underlies the correlation between the independent variables of the behavior theories and the adoption of new technologies [18].

Research Model and Hypothesis Development

Health care professionals face daily challenges and issues that entail a high level of communication and information exchange among colleagues. With the addition of social media, the decision-making process will be enhanced and will yield useful health outcomes and contribute to delivering outstanding performance. Public health professionals share a positive attitude toward the usage of online media for informing health change behaviors [19]. The HCPs communicate continuously to improve the performance of their daily tasks. Their use of Twitter has positively impacted their productivity and professional development. In fact, social media platforms are positively perceived for being an effective communication tool involving different stakeholders and for promoting further education and professional gain [20]. The dissemination of information and the update of knowledge are essential predictors of practitioners' daily performance; therefore, social media enhances the processes involved in decision making. For example, dentists exhibit a positive attitude toward participation in social media platforms such as blogs and podcasts. They stress the practical role of online channels in sharing clinical outcomes and other relevant information, such as treatment options, that enhance the participants' practical knowledge [21]. Health care professionals can analyze the content of the discussions and identify the best solution. This understanding will support the decision process and enhance patient awareness about the changes required [22]. The social media strategy of using different online tools and platforms such as blogs and tweets supports the evidence-based dissemination of child and youth diagnosis and treatment options. Therefore, information exchange and an organized communication strategy support pediatrics HCP's personal development and productivity [23].

The attitude toward social media benefits and perceived outcomes are important predictors of HCP's use of social media platforms [14,24]. Hypothesis 1 states that attitude will positively relate to the frequency of social media use. Hypothesis 2 states that perceived usefulness will positively relate to the frequency of social media use.

Social media opens up an opportunity for practitioners to interact, share, and develop knowledge through communication. The accessibility of such platforms reinforces the spread and transfer of knowledge [25]. User-friendly platforms that support timely responses, easy access, and easy navigation enhance the knowledge exchange of HCPs. In addition, communication supports a faster achievement of tasks [26]. Easy access to social media is a facilitator for HCP's intention of use [8,14,24]. Hypothesis 3 states that ease of use will positively relate to the behavioral intention of social media use. Hypothesis 4 states that ease of use will positively relate to perceived usefulness.

The use of social media in health care practice might be useful for medical information exchange and interpersonal communication with peers. However, confidentiality within a highly controlled discipline raises many questions about the efficient use of social media without undermining the HCPs professional status. Without privacy control to protect the personal information of HCPs, the professional relationships with the patients could be altered as well as the effective communication of their health behavior messages [26]. Health care professionals and students are open to integrating social media in teaching content; however, the legal concerns and the technical issues remain major challenges for users and nonusers of online platforms [27]. The eHealth platforms are effective in addressing the patient's problems; however, the privacy of intervention, reliability, and security are the main inhibitors of daily usage [28]. Social networking sites are useful for public health campaigns, but the management of privacy and personal data remains a significant concern for practitioners [29]. Privacy challenges, ethical concerns, and access to platforms are the essential concerns of the HCP's intention of using social media [8,18]. Hypothesis 5 states that perceived control will positively relate to the behavioral intention of social media use. Hypothesis 6 states that perceived control will positively relate to ease of use.

Collaboration between colleagues, departments, and other stakeholders is fundamental to the achievement of health outcomes. Social status and relational identity with peers affect the HCP's reinforced or varied use of social media [30]. The majority of pharmacy teachers joined Facebook's online platform at the request of their friends or family members. They expressed their need to be socially connected with others [31]. Doctor-patient interaction on social media platforms strongly affects knowledge, efficacy, and perceived outcomes [32]. Experts and key opinion leaders are well respected in health care communities; moreover, they are the source of knowledge and act as advisers for juniors, residents, and newly graduated HCPs. The participation of qualified HCPs in online forums evokes a higher engagement rate by the practitioners [33]. The interactions between HCPs increases factual knowledge for higher productivity. Also, the presence of experts and key opinion leaders facilitates higher participation rates [34]. Hypothesis 7 states that perceived norms will positively relate to the behavioral intention of social media use. Hypothesis 8 states that image will positively relate to the behavioral intention of social media use. Hypothesis 9 states that image will

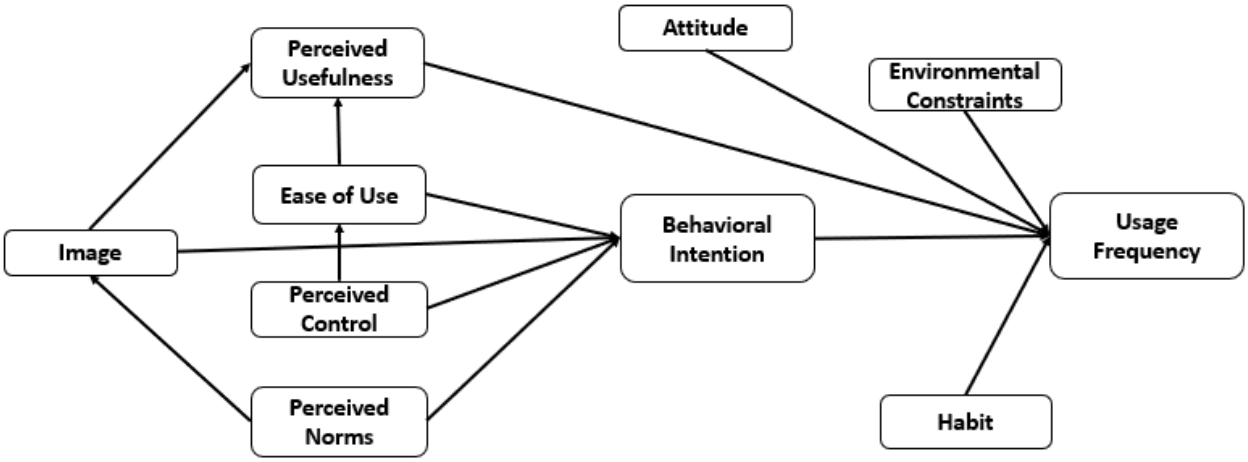
positively relate to the perceived usefulness. Hypothesis 10 states that perceived norms will positively relate to the image.

Environmental constraints can inhibit actual behavior despite the presence of the intention [13]. Legal policies are emphasized as a critical factor in not sharing clinical images and content on social media channels by health care students [27,35]. The perceived usefulness and social interaction are the primary motivators for HCPs to join a virtual community of practice; however, time limitations are regarded as a barrier against use [36]. Hypothesis 11 states that environmental constraints will negatively relate to the frequency of social media use.

Health care practitioners use social media in their private life but using it in the workplace is affected by many factors that could facilitate or inhibit such behavior. The activities of HCP communities on the social media platform Twitter reveal the differences between the group's followers and their habits of engagement through the retweet metric [37]. Previous positive experiences support a repetitive behavior that may become a habit [13,14,38]. Hypothesis 12 states that habit will positively relate to the frequency of social media use.

The theory of reasoned actions, the theory of planned behaviors, and the technology acceptance theory explain that the best predictor of behavior change or the adoption of innovation is the intention toward the action [12,15,16,38]. The impact of behavioral intention of use has been tested empirically in different organizations and individual settings. Hypothesis 13 states that behavioral intention will positively relate to the frequency of social media use. Figure 1 demonstrates the proposed research model.

Figure 1. The research model.



Methods

Measurement

Before designing the questionnaire, the researcher discussed the constructs of the proposed conceptual model with 10 experts in the field. The respondents in this elicitation phase [13] consisted of one dermatologist, one plastic surgeon, one dentist, one nurse, four pharmacists, one health insurance specialist, and one member from a hospital administration. Choosing different specialties during the design of the questionnaire was based on the researcher's knowledge of the varied tasks and responsibilities within the health care context. Therefore, the opinions of the different stakeholders supported a holistic understanding of social media usage in health care practices.

The research model and expert validation informed the development of the study's survey questions. The items and scales were adapted from previous researchers (see Multimedia Appendix 1). The questions about attitudes toward usage of social media were measured using a 10-point semantic scale adapted from McGowan et al [14]. The perceived usefulness, ease of use, and environmental constraint questions were adapted from Venkatesh et al [38], using a seven-point scale ranging from "strongly agree" to "strongly disagree." The perceived norms, image, and perceived control questions were adapted from Yi et al [17] and measured as per the previous constructs scale. The questions about intention of use and the frequency of use were adapted from McGowan et al [14] and measured using a scale from "not aware" to "current user" and "never" to "many times daily," respectively.

The questionnaire included 14 sections. The first section introduced the survey and consisted of the consent form. The second section concentrated on the demographics of the participants, including age, gender, specialty, years of experience, and organization type. The other parts concentrated on the 10 variables of the research model. Experts were consulted about the clarity of the questions and the appropriate time length needed for completing the survey.

Data Collection

The three stratified target populations in this study were physicians, pharmacists, and allied employees working in any health care discipline located in the United Arab Emirates. The survey link was first sent via a LinkedIn account to the pharmacist population, which consisted of 518 participants. Next, the survey was administered to a WhatsApp group consisting of 235 medical doctors and 220 endodontists and other allied employees working in the United Arab Emirates. The survey generated 211 responses, with eight incomplete surveys or missing data. The number of completed

questionnaires was 203, representing a response rate of 20.3% of the total survey views on the social media platforms of WhatsApp and LinkedIn.

Data Analysis

Stata version 13.0 software was used in the data analysis of the research questions. The construct, convergent, and discriminant validity, as well as the reliability of the questionnaire items, were measured and explained. Confirmatory factor analysis emphasized the model fit. A structural equation model explained the relationships between the research model constructs.

Results

Participant Characteristics and Descriptive Statistics

Characteristics of participants are shown in Table 1.

Of the 203 respondents, 145 (71.4%) were aged between 25 and 45 years and 130 (64.0%) were female HCPs. Most worked in the private sector (147/203, 72.4%). The sample included 101 (49.8%) physicians, 35 (17.2%) pharmacists, and 67 (33.0%) allied HCPs. The social media platform that had the highest current users was WhatsApp with 133 of 203 (65.5%), followed by Facebook and YouTube with 101 of 203 (49.7%). Of the 203 respondents, 109 (53.6%) HCPs used social media platforms for the exchange of medical information with their peers (Figure 2) and 108 of 203 (53.2%) used them several times a day to improve their interpersonal communication with colleagues (Figure 3). In addition, 71 of 203 respondents (34.9%) used these platforms to improve their overall productivity (Figure 4). The respondents' areas of expertise correlated significantly with frequency of use (beta=.91, *P*=.02). Overall, 50 of 101 (49.5%) physicians, 15 of 35 (43%) pharmacists, and 43 of 67 (64%) allied employees used social media platforms for the exchange of medical knowledge with their peers. In addition, 47 of 101 (46.5%) physicians, 15 of 35 (43%) pharmacists, and 47 of 67 (70%) allied employees used social media platforms to improve their interpersonal communication with their peers. Finally, 31 of 101 (30.6%) physicians, 13 of 35 (37%) pharmacists, and 27 of 67 (40%) allied employees used social media to enhance their productivity in general.

Table 1. Participant characteristics (N=203).

Variable and category	Frequency, n (%)
Age in years	
<25	35 (17.2)
25-35	90 (44.3)
36-45	55 (27.1)
46-55	17 (8.4)
56-65	5 (2.5)
66-75	1 (0.5)
Gender	
Male	73 (36)
Female	130 (64)
Type of organization	
Private	147 (72.4)
Public	56 (27.6)
Specialties	
Physician	101 (49.8)
Pharmacist	35 (17.2)
Allied	67 (33)

Figure 2. Frequency that the health care professionals used social media for the exchange of medical knowledge with peers.

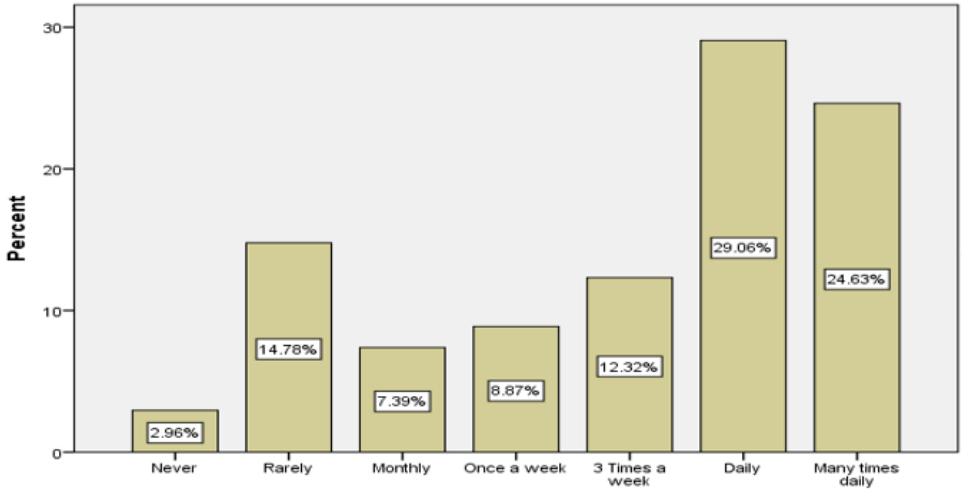


Figure 3. Frequency that the health care professionals used social media for improving their interpersonal communication with peers.

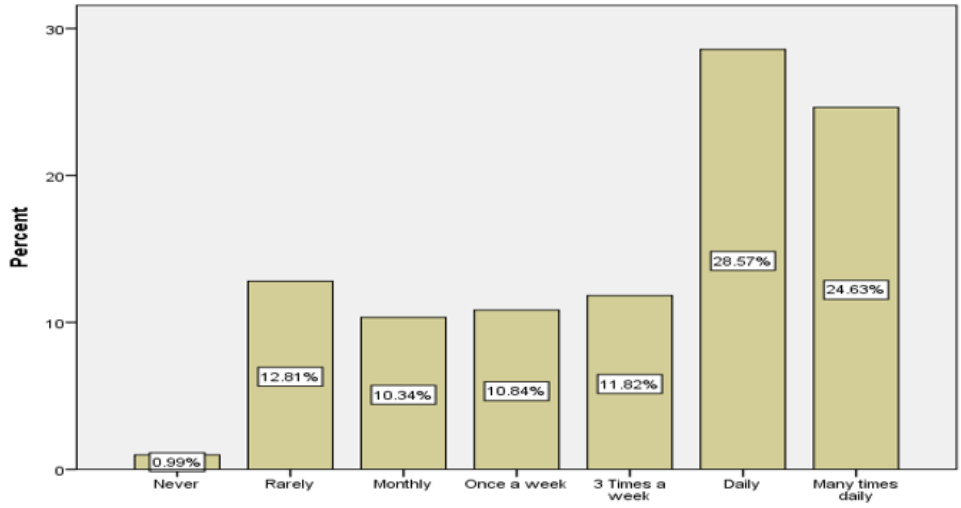
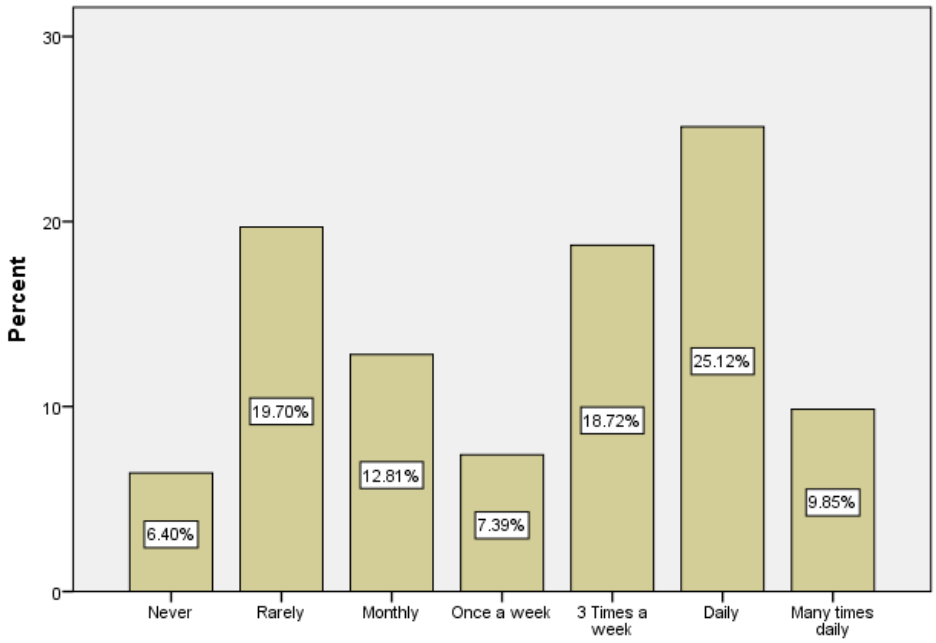


Figure 4. Frequency that the health care professionals used social media for increasing their overall productivity.



Common Method Bias

The data were collected using the same questionnaire during the same period; therefore, the existence of common method bias was tested using the Harman single factor test which could inflate or deflate the estimates between the latent constructs [39]. Common method bias exists if only one factor explains 50% of the variance among the measures during the exploratory factor analysis with unrotated factor solutions [40]. In this study, the variance explained by the single factor was 24.88%, which is less than the 50% cutoff criteria, thus emphasizing the absence of common method bias.

Measurement Model

Table 2 shows the results of the measurement model. To evaluate the construct validity and to confirm that the variable measurements represented what they were supposed to represent, an exploratory factor analysis was carried out on the 41 items [41]. The main purpose was to confirm that the questions of the survey reflected the constructs of the research model. Ten factors were extracted along with the items loaded on their intended constructs. Items BI1, BI3, and BI8 (see survey questions in Multimedia Appendix 1) were removed because their outer standardized loading was lower than .40 [42]. The study proceeded with reliability tests for the 10 factors included. Cronbach alpha for the total items was .891; hence, the score was higher than the recommended .7 [42]. Cronbach alpha for each factor extracted showed values higher than .7 (see Table 2). To assess the convergent validity, the composite reliability was calculated and the average variance extracted (AVE). The composite reliability should be greater than 0.70, and the AVE should be greater than 0.5 [43,44]. As shown in Table 2, the composite reliability and AVE values complied with these criteria.

The discriminant validity was evaluated using the Fornell-Larcker criterion [44]. The variance between the latent variables and their indicators should be higher than the variance explained with the other latent variables; therefore, the square root of the AVEs should be greater than the correlation between the constructs [44]. All diagonal square roots of the AVEs were greater than the correlation between the constructs represented in the off-diagonal elements (see Table 3).

Table 2. Cronbach alpha, composite reliability, and average variance extracted for the constructs.

Construct	Cronbach alpha	Composite reliability coefficient	Average variance extracted
Attitude	.87	0.90	0.60
Perceived usefulness	.85	0.84	0.57
Ease of use	.90	0.89	0.68
Image	.90	0.85	0.66
Perceived norms	.80	0.74	0.50
Perceived control	.89	0.87	0.65
Habit	.86	0.79	0.56
Environmental constraints	.72	0.83	0.63
Behavioral intention	.84	0.87	0.58
Use frequency	.86	0.82	0.61

Table 3. Correlations and the square roots of the average variance extracted. Off-diagonal elements are correlations. Diagonal elements are square roots of average variance extracted.

	AT[a]	PU[b]	EU[c]	IM[d]	PN[e]	PC[f]	HA[g]	EC[h]	BI[i]	UF[j]	Age	G[k]	P[l]	OT[m]
AT	.77													
PU	.12	.75												
EU	.11	.40	.82											
IM	.06	.48	.18	.81										
PN	.10	.55	.21	.57	.70									
PC	.10	.40	.29	.40	.35	.80								
HA	.10	.44	.48	.43	.48	.49	.74							
EC	.01	–.10	–.11	.05	–.07	–.16	.12	.79						
BI	.09	.20	.35	.01	.17	.13	.25	–.02	.76					
UF	.22	.33	.38	.19	.29	.35	.39	.09	.48	N/A[n]				
Age	–.04	–.03	–.28	.03	.04	–.05	–.18	.00	–.22	–.18	N/A			
G	.12	.02	–.08	.08	.12	.12	–.03	–.05	–.06	–.01	.23	N/A		
P	.08	.08	.13	.00	–.04	–.07	.06	.16	–.09	.17	–.04	–.15	N/A	
OT	.00	–.09	.00	–.30	–.20	–.12	–.04	–.04	–.05	.13	–.26	–.12	.03	N/A

[a]AT: attitude.

[b]PU: perceived usefulness.

[c]EU: ease of use.

[d]IM: image.

[e]PN: perceived norms.

[f]PC: perceived control.

[g]HA: habit.

[h]EC: environmental constraints.

[i]BI: behavioral intention.

[j]UF: use frequency.

[k]G: gender.

[l]P: profession.

[m]OT: organization type.

[n]N/A: not applicable because they were not reflective constructs.

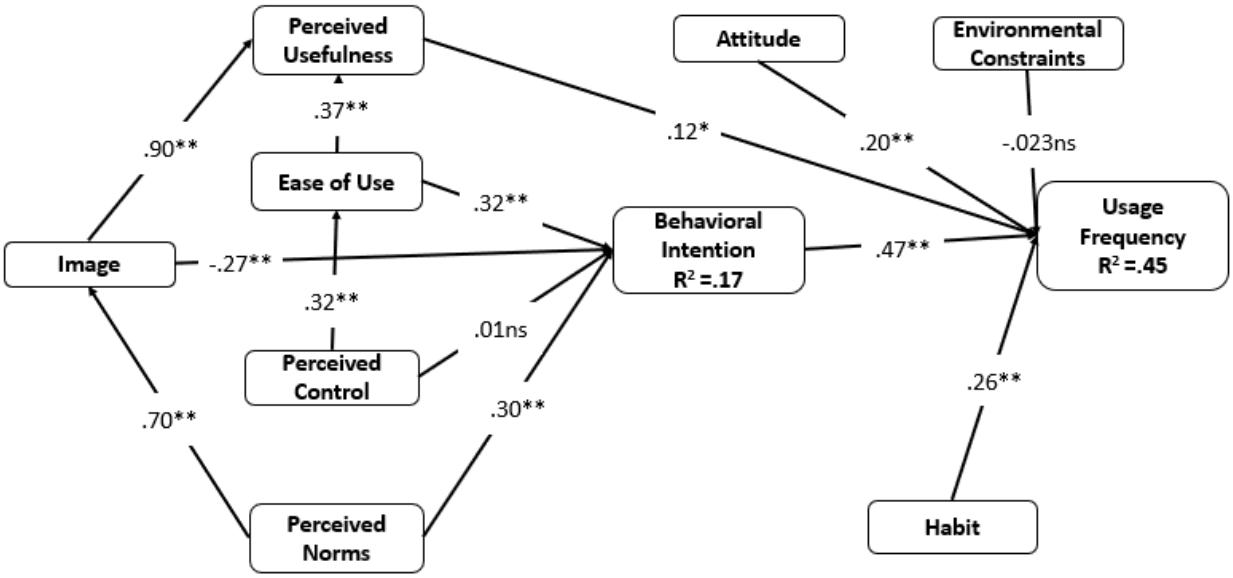
Structural Equation Modeling

In the first stage of structural equation modeling, the measurement model was specified and the goodness of fit was assessed. A covariance-based structural equation model was used to test the fit of the confirmatory factor analysis model to the research data [45]. The result of the confirmatory factor analysis for goodness of fit generated acceptable fit indexes (χ^2_{639}=1083.6, *P*<.001; confirmatory fit index [CFI]=0.907; root mean square error of approximation [RMSEA]=0.059). The second stage was the structural equation analysis. The indexes of the goodness of fit of the linear model were acceptable (χ^2_{614}=1094.7, *P*<.001; CFI=0.900; RMSEA=0.062]. The study also used path analysis and hypothesis testing. Table 4 presents a summary of the hypothesis statistical tests. Overall, the model explained 45% of the variance in the use frequency and 17% of the variance in the behavioral intention (see Figure 5). The predictors of the use frequency were the perceived usefulness (beta=.12, *P*=.09), the attitude (beta=.20, *P*=.002), the habit (beta=.26, *P*=.001), and the behavioral intention (beta=.47, *P*<.001). The predictors of the behavioral intention were the ease of use (beta=.32, *P*<.001), the image (beta=–.27, *P*=.02), and the norms (beta=.30, *P*=.02). To analyze if the ease of use mediated perceived control on behavioral intention, the indirect effect of perceived control on the behavioral intention was tested and was significant (beta=.09, *P*=.002), unlike the direct effect (beta=.01, *P*=.90). Therefore, it can be concluded that the ease of use mediated the relationship between perceived control and behavioral intention.

Table 4. Summary of findings regarding hypotheses.

Hypothesis	Beta	*z*	*P* value	Results
H1: The attitude will positively relate to the frequency of social media use	.20	3.15	.002	Accepted
H2: The perceived usefulness will positively relate to the frequency of social media use	.12	7.67	.09	Accepted
H3: The ease of use will positively relate to the behavioral intention of social media use	.32	4.42	<.001	Accepted
H4: The ease of use will positively relate to the perceived usefulness	.37	6.34	<.001	Accepted
H5: The perceived control will positively relate to the behavioral intention of social media use	.01	0.13	.99	Not accepted
H6: The perceived control will positively relate to ease of use	.32	5.26	<.001	Accepted
H7: The perceived norms will positively relate to the behavioral intention of social media use	.30	2.44	.02	Accepted
H8: The image will positively relate to behavioral intention of social media use	–.27	–2.35	.02	Not accepted
H9: The image will positively relate to the perceived usefulness	.90	9.66	<.001	Accepted
H10: The perceived norms will positively relate to the image	.70	13.64	<.001	Accepted
H11: The environmental constraints will negatively relate to the frequency of social media use	–.03	–0.51	.61	Not accepted
H12: The habit will positively relate to the frequency of social media use	.26	3.36	.001	Accepted
H13: The behavioral intention will positively relate to the frequency of social media use	.47	7.67	<.001	Accepted

Figure 5. Structural model results. *P<.10; **P<.05; ns: nonsignificant.



Discussion

Principal Results

Health care professionals are exposed to an overwhelming amount of information that needs to be managed and applied effectively for task achievement. Social media platforms can foster a higher level of communication for knowledge and experience sharing. This study has explored the impact of different factors on social media use by HCPs. The frequency of use was measured through three main dimensions related to medical information exchange with peers, interpersonal communication, and productivity enhancement. The results reveal that the most used platforms are WhatsApp, Facebook, and YouTube. Previous research in a similar area concluded that HCPs use Twitter mostly for personal development and exchange of knowledge [46].

The results show that the frequency of use differs from one specialty to another. Such variation is explained by the users' characteristics, the various tasks, and job responsibilities. The study confirms the role of attitude, perceived usefulness, habit, and behavioral intention as the major predictors of HCP's frequency of social media use. The attitude variable was studied in previous research as the main factor that relates to the physicians' frequency of social media use [14]. Moreover, behavioral intention is highly related to the frequency of use; hence, the theoretical construct of intention [12,13,15,16] was empirically tested in online settings within the HCP population. The results of this study support the role of habit in predicting the frequency of use, but the role of environmental constraints as an inhibitor of social media frequency of use cannot be confirmed [13]. The results may be explained by HCP's

characteristics as online users or the absence of policies regulating social media usage.

Ease of use positively relates to behavioral intention, which highlights easy access and navigation through social media platforms [23,24]. The norms predict the HCP's intention of using the social media platforms and positively impact the image of the practitioners. The results reveal the positive role of the social network in the user level of engagement [47]. Moreover, the social relationships influence the participation of the HCPs in online Web-based platforms for better daily outcomes [33]. It was hypothesized that the image of HCPs would impact the behavioral intention of using social media platforms; however, this hypothesis remains unsupported.

The findings suggest that the image of HCPs relates negatively to the intention of use. Confidentiality, personal information, and the openness of social media platforms might explain such a negative relationship. Image highly predicts perceived usefulness. The findings also confirm that HCPs refer to experts or key opinion leaders for better decision processes and outcomes [34,48,49]. The perceived control does not relate significantly to behavioral intention. Rather, it positively and significantly relates to ease of use. The analysis of the indirect relationship between perceived control and behavioral intention is significant; therefore, the study confirms the mediation effect of ease of use.

Limitations and Future Research

This research may be the first to explain variation in the frequency of use by different health care specialties. It also confirms the role of behavior and technology acceptance theories in predicting social media use by HCPs. However, there are limitations in the research. One is the convenience sample, which includes HCPs who are current users of social media platforms. The offline practitioners might provide a better understanding of the other important variables that influence frequency of use. The other is the absence of public health practitioners and researchers. Their presence would influence the perceived usefulness of the platforms since previous studies have dealt with cost reductions for recruiting patients through social media apps [50,51]. Future researchers might explore other variables such as training, for example. Hence, the health care discipline can benefit from the highly interactive platforms and employ it for the purpose of development, collaboration, and better health outcomes.

Conclusion

New knowledge and experiences for HCPs are critical for career development. They need medical knowledge in real time to take decisions, as well effective communication with peers to resolve problems that arise in practice. Knowledge created and shared with colleagues and medical communities enhances performance and supports effective interventions for behavioral health changes. Social media platforms provide an opportunity for HCPs to interact, share information, and disseminate knowledge for better decision making. The importance of this study lies in the medical knowledge and practices shared by HCPs on social networks with their peers. This research suggests that the efficient use of social media platforms might significantly reduce the challenges raised within traditional health care settings. Furthermore, it can support practitioners' career advancement by keeping them current and well-informed. This research adopts the theories of behavior changes and technology acceptance to build a model that can explain underlying user perceptions and other variables that contribute to the successful use of social media platforms by HCPs. The study model was tested, and the results confirm a rate of 45% variance in the frequency of use and 17% variance in the behavioral intention of using social media platforms. The attitude, perceived usefulness, behavioral intention, and habit significantly impact use frequency. The perceived norms, ease of use, and image have a considerable effect on behavioral intention. On this basis, this research has practical implications and efforts should be expended to create social media platforms that take into account the HCP's characteristics, concerns, and objectives of use.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey instrument.
[PDF File (Adobe PDF File), 28 KB - jmir_v20i11e12035_app1.pdf]

References

1. 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]
2. Ventola CL. Social media and health care professionals: benefits, risks, and best practices. P T 2014 Jul;39(7):491-520 [FREE Full text] [Medline: 25083128]
3. Sandars J, Schroter S. Web 2.0 technologies for undergraduate and postgraduate medical education: an online survey. Postgrad Med J 2007 Dec;83(986):759-762 [FREE Full text] [doi: 10.1136/pgmj.2007.063123] [Medline: 18057175]
4. Alkhateeb FM, Clauson KA, Latif DA. Pharmacist use of social media. Int J Pharm Pract 2011 Apr;19(2):140-142. [doi: 10.1111/j.2042-7174.2010.00087.x] [Medline: 21385246]

5. von Muhlen M, Ohno-Machado L. Reviewing social media use by clinicians. J Am Med Inform Assoc 2012;19(5):777-781 [FREE Full text] [doi: 10.1136/amiajnl-2012-000990] [Medline: 22759618]
6. Gholami-Kordkheili F, Wild V, Strech D. The impact of social media on medical professionalism: a systematic qualitative review of challenges and opportunities. J Med Internet Res 2013 Aug 28;15(8):e184 [FREE Full text] [doi: 10.2196/jmir.2708] [Medline: 23985172]
7. Rolls K, Hansen M, Jackson D, Elliott D. How health care professionals use social media to create virtual communities: an integrative review. J Med Internet Res 2016 Dec 16;18(6):e166 [FREE Full text] [doi: 10.2196/jmir.5312] [Medline: 27328967]
8. Dumas A, Lapointe A, Desroches S. Users, uses, and effects of social media in dietetic practice: scoping review of the quantitative and qualitative evidence. J Med Internet Res 2018 Feb 20;20(2):e55 [FREE Full text] [doi: 10.2196/jmir.9230] [Medline: 29463487]
9. Grajales III IF, Sheps S, Ho KN, Novak-Lauscher H, Eysenbach G. Social media: a review and tutorial of applications in medicine and health care. J Med Internet Res 2014 Feb 11;16(2):e13 [FREE Full text] [doi: 10.2196/jmir.2912] [Medline: 24518354]
10. Moorhead S, Hazlett D, Harrison L, Carroll J, 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 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
11. Li J, Talaei-Khoei A, Seale H, Ray P, Macintyre CR. Health care provider adoption of eHealth: systematic literature review. Interact J Med Res 2013 Apr 16;2(1):e7 [FREE Full text] [doi: 10.2196/ijmr.2468] [Medline: 23608679]
12. Davis F. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quart 1989 Sep;13(3):319. [doi: 10.2307/249008]
13. Fishbein M. The role of theory in HIV prevention. AIDS Care 2000 Jun;12(3):273-278. [doi: 10.1080/09540120050042918] [Medline: 10928203]
14. McGowan B, Wasko M, Vartabedian B, Miller R, Freiherr D, 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 Sep 24;14(5):e117 [FREE Full text] [doi: 10.2196/jmir.2138] [Medline: 23006336]
15. Fishbein M. Readings in Attitude Theory and Measurement. New York: Wiley; 1967.
16. Ajzen I. The theory of planned behavior. Organ Behav Hum Dec 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
17. Yi M, Jackson J, Park J, Probst J. Understanding information technology acceptance by individual professionals: Toward an integrative view. Inform Manage 2006 Apr;43(3):350-363. [doi: 10.1016/j.im.2005.08.006]
18. Lau A. Hospital-based nurses' perceptions of the adoption of Web 2.0 tools for knowledge sharing, learning, social interaction and the production of collective intelligence. J Med Internet Res 2011 Nov 11;13(4):e92 [FREE Full text] [doi: 10.2196/jmir.1398] [Medline: 22079851]
19. Franco M, Tursunbayeva A, Pagliari C. Social media for e-Government in the public health sector: protocol for a systematic review. JMIR Res Protoc 2016 Mar 11;5(1):e42 [FREE Full text] [doi: 10.2196/resprot.5421] [Medline: 26969199]
20. Hart M, Stetten N, Islam S, Pizarro K. Twitter and public health (Part 1): how individual public health professionals use Twitter for professional development. JMIR Public Health Surveill 2017 Sep 20;3(3):e60 [FREE Full text] [doi: 10.2196/publichealth.6795] [Medline: 28931499]
21. Melkers J, Hicks D, Rosenblum S, Isett K, Elliott J. Dental blogs, podcasts, and associated social media: descriptive mapping and analysis. J Med Internet Res 2017 Jul 26;19(7):e269 [FREE Full text] [doi: 10.2196/jmir.7868] [Medline: 28747291]
22. Merolli M, Gray K, Martin-Sanchez F, Lopez-Campos G. Patient-reported outcomes and therapeutic affordances of social media: findings from a global online survey of people with chronic pain. J Med Internet Res 2015 Jan 22;17(1):e20 [FREE Full text] [doi: 10.2196/jmir.3915] [Medline: 25616273]
23. Dyson M, Newton A, Shave K, Featherstone R, Thomson D, Wingert A, et al. Social media for the dissemination of Cochrane Child Health Evidence: evaluation study. J Med Internet Res 2017 Sep 01;19(9):e308 [FREE Full text] [doi: 10.2196/jmir.7819] [Medline: 28864427]
24. Roland D, Spurr J, Cabrera D. Preliminary evidence for the emergence of a health care online community of practice: using a netnographic framework for Twitter hashtag analytics. J Med Internet Res 2017 Jul 14;19(7):e252 [FREE Full text] [doi: 10.2196/jmir.7072] [Medline: 28710054]
25. Demiris G. The diffusion of virtual communities in health care: concepts and challenges. Patient Educ Couns 2006 Aug;62(2):178-188. [doi: 10.1016/j.pec.2005.10.003] [Medline: 16406472]
26. MacDonald J, Sohn S, Ellis P. Privacy, professionalism and Facebook: a dilemma for young doctors. Med Educ 2010 Aug;44(8):805-813. [doi: 10.1111/j.1365-2923.2010.03720.x] [Medline: 20633220]
27. D'Souza K, Henningham L, Zou R, Huang J, O'Sullivan E, Last J, et al. Attitudes of health professional educators toward the use of social media as a teaching tool: global cross-sectional study. JMIR Med Educ 2017 Aug 04;3(2):e13 [FREE Full text] [doi: 10.2196/mededu.6429] [Medline: 28778841]

28. Ariens L, Schussler-Raymakers F, Frima C, Flinterman A, Hamminga E, Arents B, et al. Barriers and facilitators to eHealth use in daily practice: perspectives of patients and professionals in dermatology. J Med Internet Res 2017 Sep 05;19(9):e300 [FREE Full text] [doi: 10.2196/jmir.7512] [Medline: 28874336]
29. Balatsoukas P, Kennedy C, 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 Jun 11;17(6):e141 [FREE Full text] [doi: 10.2196/jmir.3662] [Medline: 26068087]
30. Pan Z, Lu Y, Wang B, Chau P. Who do you think you are? Common and differential effects of social self-identity on social media usage. J Manage Inform Syst 2017;34(1):71-101.
31. Metzger AH, Finley KN, Ulbrich TR, McAuley JW. Pharmacy faculty members' perspectives on the student/faculty relationship in online social networks. Am J Pharm Educ 2010 Dec 15;74(10):188 [FREE Full text] [Medline: 21436929]
32. Wu T, Deng Z, Feng Z, Gaskin D, Zhang D, Wang R. The effect of doctor-consumer interaction on social media on consumers' health behaviors: cross-sectional study. J Med Internet Res 2018 Feb 28;20(2):e73 [FREE Full text] [doi: 10.2196/jmir.9003] [Medline: 29490892]
33. Poirier J, Cobb N. Social influence as a driver of engagement in a web-based health intervention. J Med Internet Res 2012 Feb 22;14(1):e36 [FREE Full text] [doi: 10.2196/jmir.1957] [Medline: 22356829]
34. Stewart S, Abidi S. Applying social network analysis to understand the knowledge sharing behaviour of practitioners in a clinical online discussion forum. J Med Internet Res 2012 Dec 04;14(6):e170 [FREE Full text] [doi: 10.2196/jmir.1982] [Medline: 23211783]
35. O'Sullivan E, Cutts E, Kavikondala S, Salcedo A, D'Souza K, Hernandez-Torre M, et al. Social media in health science education: an international survey. JMIR Med Educ 2017 Jan 04;3(1):e1 [FREE Full text] [doi: 10.2196/mededu.6304] [Medline: 28052842]
36. Barnett S, Jones S, Bennett S, Iverson D, Bonney A. Perceptions of family physician trainees and trainers regarding the usefulness of a virtual community of practice. J Med Internet Res 2013 May 10;15(5):e92 [FREE Full text] [doi: 10.2196/jmir.2555] [Medline: 23666237]
37. Mishori R, Singh L, Levy B, Newport C. Mapping physician Twitter networks: describing how they work as a first step in understanding connectivity, information flow, and message diffusion. J Med Internet Res 2014 Apr 14;16(4):e107 [FREE Full text] [doi: 10.2196/jmir.3006] [Medline: 24733146]
38. Venkatesh, Thong, Xu. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Quart 2012;36(1):157-178. [doi: 10.2307/41410412]
39. Bagozzi R, Yi Y. On the evaluation of structural equation models. JAMS 1988 Mar;16(1):74-94. [doi: 10.1007/BF02723327]
40. Podsakoff PM, MacKenzie SB, Podsakoff NP. Sources of method bias in social science research and recommendations on how to control it. Annu Rev Psychol 2012;63:539-569. [doi: 10.1146/annurev-psych-120710-100452] [Medline: 21838546]
41. Field A. Discovering Statistics Using IBM SPSS Statistics. 4th edition. London: SAGE Publications; 2013.
42. Churchill G. A paradigm for developing better measures of marketing constructs. J Mark Res 1979 Feb;16(1):64-73. [doi: 10.2307/3150876]
43. MacKenzie S, Podsakoff P, Podsakoff N. Construct measurement and validation procedures in MIS and behavioral research: integrating new and existing techniques. MIS Quart 2011;35(2):293-334. [doi: 10.2307/23044045]
44. Fornell C, Larcker D. Structural equation models with unobservable variables and measurement error: algebra and statistics. J Mark Res 1981 Aug;18(3):382. [doi: 10.2307/3150980]
45. Anderson JC, Gerbing DW. Structural equation modeling in practice: a review and recommended two-step approach. Psychol Bull 1988;103(3):423.
46. Alsobayel H. Use of social media for professional development by health care professionals: a cross-sectional web-based survey. JMIR Med Educ 2016 Sep 12;2(2):e15 [FREE Full text] [doi: 10.2196/mededu.6232] [Medline: 27731855]
47. Centola D. The spread of behavior in an online social network experiment. Science 2010 Sep 03;329(5996):1194-1197 [FREE Full text] [doi: 10.1126/science.1185231] [Medline: 20813952]
48. Francke A, Smit M, de Veer AJ, Mistiaen P. Factors influencing the implementation of clinical guidelines for health care professionals: a systematic meta-review. BMC Med Inform Decis Mak 2008 Sep 12;8(1):38 [FREE Full text] [doi: 10.1186/1472-6947-8-38] [Medline: 18789150]
49. Bennett NL, Casebeer LL, Zheng S, Kristofco R. Information-seeking behaviors and reflective practice. J Contin Educ Health Prof 2006;26(2):120-127. [doi: 10.1002/chp.60] [Medline: 16802313]
50. Davies B, Kotter M. Lessons from recruitment to an internet-based survey for degenerative cervical myelopathy: comparison of free and fee-based methods. JMIR Research Protocols 2018;7(2):18. [doi: 10.2196/resprot.6567]
51. Christensen T, Riis A, Hatch E, Wise L, Nielsen M, Rothman K, et al. Costs and efficiency of online and offline recruitment methods: a web-based cohort study. J Med Internet Res 2017 Dec 01;19(3):e58 [FREE Full text] [doi: 10.2196/jmir.6716] [Medline: 28249833]

Abbreviations

AVE: average variance extracted

CFI: confirmatory fit index
CR: composite reliability coefficient
HCP: health care professional
RMSEA: root mean square error of approximation
TAM: Technology Acceptance Model

Edited by G Eysenbach; submitted 25.08.18; peer-reviewed by B Boateng, H Miller; comments to author 13.09.18; revised version received 25.09.18; accepted 12.10.18; published 07.11.18.

Please cite as:
Hazzam J, Lahrech A
Health Care Professionals' Social Media Behavior and the Underlying Factors of Social Media Adoption and Use: Quantitative Study
J Med Internet Res 2018;20(11):e12035
URL: http://www.jmir.org/2018/11/e12035/
doi:10.2196/12035
PMID:30404773

Review

Assessing the Methods, Tools, and Statistical Approaches in Google Trends Research: Systematic Review

Amaryllis Mavragani[1], BSc, MSc; Gabriela Ochoa[1], BSc, MSc, PhD; Konstantinos P Tsagarakis[2], DipEng, PhD

[1]Department of Computing Science and Mathematics, University of Stirling, Stirling, Scotland, United Kingdom

[2]Department of Environmental Engineering, Democritus University of Thrace, Xanthi, Greece

Corresponding Author:
Amaryllis Mavragani, BSc, MSc
Department of Computing Science and Mathematics
University of Stirling
Stirling, Scotland, FK94LA,
United Kingdom
Phone: 44 7523782711
Email: amaryllis.mavragani1@stir.ac.uk

Abstract

Background: In the era of information overload, are big data analytics the answer to access and better manage available knowledge? Over the last decade, the use of Web-based data in public health issues, that is, infodemiology, has been proven useful in assessing various aspects of human behavior. Google Trends is the most popular tool to gather such information, and it has been used in several topics up to this point, with health and medicine being the most focused subject. Web-based behavior is monitored and analyzed in order to examine actual human behavior so as to predict, better assess, and even prevent health-related issues that constantly arise in everyday life.

Objective: This systematic review aimed at reporting and further presenting and analyzing the methods, tools, and statistical approaches for Google Trends (infodemiology) studies in health-related topics from 2006 to 2016 to provide an overview of the usefulness of said tool and be a point of reference for future research on the subject.

Methods: Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines for selecting studies, we searched for the term "Google Trends" in the Scopus and PubMed databases from 2006 to 2016, applying specific criteria for types of publications and topics. A total of 109 published papers were extracted, excluding duplicates and those that did not fall inside the topics of health and medicine or the selected article types. We then further categorized the published papers according to their methodological approach, namely, visualization, seasonality, correlations, forecasting, and modeling.

Results: All the examined papers comprised, by definition, time series analysis, and all but two included data visualization. A total of 23.1% (24/104) studies used Google Trends data for examining seasonality, while 39.4% (41/104) and 32.7% (34/104) of the studies used correlations and modeling, respectively. Only 8.7% (9/104) of the studies used Google Trends data for predictions and forecasting in health-related topics; therefore, it is evident that a gap exists in forecasting using Google Trends data.

Conclusions: The monitoring of online queries can provide insight into human behavior, as this field is significantly and continuously growing and will be proven more than valuable in the future for assessing behavioral changes and providing ground for research using data that could not have been accessed otherwise.

(J Med Internet Res 2018;20(11):e270) doi:10.2196/jmir.9366

KEYWORDS

big data; health assessment; infodemiology; Google Trends; medicine; review; statistical analysis

Introduction

Big data are characterized by the 8 Vs [1]: volume (exponentially increasing volumes) [2], variety (wide range of datasets), velocity (high processing speed) [3], veracity, value [4,5], variability, volatility, and validity [1]. Big data have shown great potential in forecasting and better decision making [1]; though handling these data with conventional ways is inadequate [6], they are being continuously integrated in research [7] with novel approaches and methods.

The analysis of online search queries has been of notable popularity in the field of big data analytics in academic research [8,9]. As internet penetration is continuously increasing, the use of search traffic data, social media data, and data from other Web-based sources and tools can assist in facilitating a better understanding and analysis of Web-based behavior and behavioral changes [10].

The most popular tool for analyzing behavior using Web-based data is *Google Trends* [11]. Online search traffic data have been suggested to be a good analyzer of internet behavior, while Google Trends acts as a reliable tool in predicting changes in human behavior; subject to careful selection of the searched-for terms, Google data can accurately measure the public's interest [12]. Google Trends provides the field of big data with new opportunities, as it has been shown to be valid [13] and has been proven valuable [14,15], accurate [16], and beneficial [17] for forecasting. Therefore, great potential arises from using Web-based queries to examine topics and issues that would have been difficult or even impossible to explore without the use of big data. The monitoring of Web-based activity is a valid indicator of public behavior, and it has been effectively used in predictions [18,19], nowcastings [20], and forecasting [17,21,22].

Google Trends shows the changes in online interest for time series in any selected term in any country or region over a selected time period, for example, a specific year, several years, 3 weeks, 4 months, 30 days, 7 days, 4 hours, 1 hour, or a specified time-frame. In addition, different terms in different regions can be compared simultaneously. Data are downloaded from the Web in ".csv" format and are adjusted as follows: "*Search results are proportionate to the time and location of a query: Each data point is divided by the total searches of the geography and time range it represents, to compare relative popularity. Otherwise places with the most search volume would always be ranked highest. The resulting numbers are then scaled on a range of 0 to 100 based on a topic's proportion to all searches on all topics. Different regions that show the same number of searches for a term will not always have the same total search volumes* " [23].

Healthcare is one of the fields in which big data are widely applied [24,25], with the number of publications in this field showing a high increase [26]. Researchers have placed a significant focus on examining Web-based search queries for health and medicine related topics [27]. Data from Google Trends have been shown to be valuable in predictions, detection of outbreaks, and monitoring interest, as detailed below, while such applications could be analyzed and evaluated by government officials and policy makers to deal with various health issues and disease occurrence.

The monitoring and analysis of internet data fall under the research field of infodemiology, that is, employing data collected from Web-based sources aiming at informing public health and policy [28]. These data have the advantage of being real time, thus tackling the issue of long periods of delay from gathering data to analysis and forecasting. Over the past decade, the field of infodemiology has been shown to be highly valuable in assessing health topics, retrieving web-based data from, for example, Google [29,30], Twitter [31-34], social media [35,36], or combinations of ≥2 Web-based data sources [37,38].

As the use of Google Trends in examining human behavior is relatively novel, new methods of assessing Google health data are constantly arising. Up to this point, several topics have been examined, such as epilepsy [39,40], cancer [41], thrombosis [42], silicosis [43], and various medical procedures including cancer screening examinations [44,45], bariatric surgery [46], and laser eye surgery [47].

Another trend rising is the measurement of the change in interest in controversial issues [48,49] and in drug-related subjects, such as searches in prescription [50] or illicit drugs [51,52]. In addition, Google Trends data have been used in examining interest in various aspects of the health care system [53-55].

Apart from the above, Google Trends data have also been useful in measuring the public's reaction to various outbreaks or incidents, such as attention to the epidemic of Middle East Respiratory Syndrome [56], the Ebola outbreak [57], measles [58], and Swine flu [59], as well as the influence of media coverage on online interest [60]. Google queries for the respective terms have been reported to increase or peak when a public figure or celebrity is related [61-65].

Google Trends has also been valuable in examining seasonal trends in various diseases and health issues, such as Lyme disease [66], urinary tract infection [67], asthma [30], varicose vein treatment [68], and snoring and sleep apnea [69]. Furthermore, Deiner et al [70] showed that indeed there exists the same seasonality in Google Trends and clinical diagnoses. What has also been reported is that seasonality in Google searches on tobacco is correlated with seasonality in Google searches on lung cancer [71], while online queries for allergic rhinitis have the same seasonality as in real life cases [72]. Thus, we observe that, apart from measuring public interest, Google Trends studies show that the seasonality of online search traffic data can be related to the seasonality of actual cases of the respective diseases searched for.

As mentioned above, Google queries have been used so far to examine general interest in drugs. Taking a step further, Schuster et al [73] found a correlation between the percentage change in the global revenues in Lipitor statin for dyslipidemia treatment and Google searches, while several other studies have reported findings toward this direction, that is, correlations of Web-based searches with prescription issuing [74-76]. The detection and monitoring of flu has also been of notable popularity in health assessment. Data from Google Flu Trends have been shown to correlate with official flu data [77,78], and Google data on the relevant terms correlate with cases of influenza-like illness [79].

In addition, online search queries for suicide have been shown to be associated with actual suicide rates [80,81], while other examples indicative of the relationship between Web-based data and human behavior include the correlations between official data and internet searches in veterinary issues [82], sleep deprivation [83], sexually transmitted infections [84], Ebola-related searches [85], and allergies [86,87].

Furthermore, Zhou et al [88] showed how the early detection of tuberculosis outbreaks can be improved using Google Trends

data; while suicide rates and Google data seem to be related, the former are suggested to be a good indicator for developing suicide prevention policies [89]. In addition, methamphetamine criminal behavior has been shown to be related to meth searches [90]. Finally, recent research on using Google Trends in predictions and forecasting include the development of predictive models of pertussis occurrence [91], while online search queries have been employed to forecast dementia incidence [92] and prescription volumes in ototopical antibiotics [93].

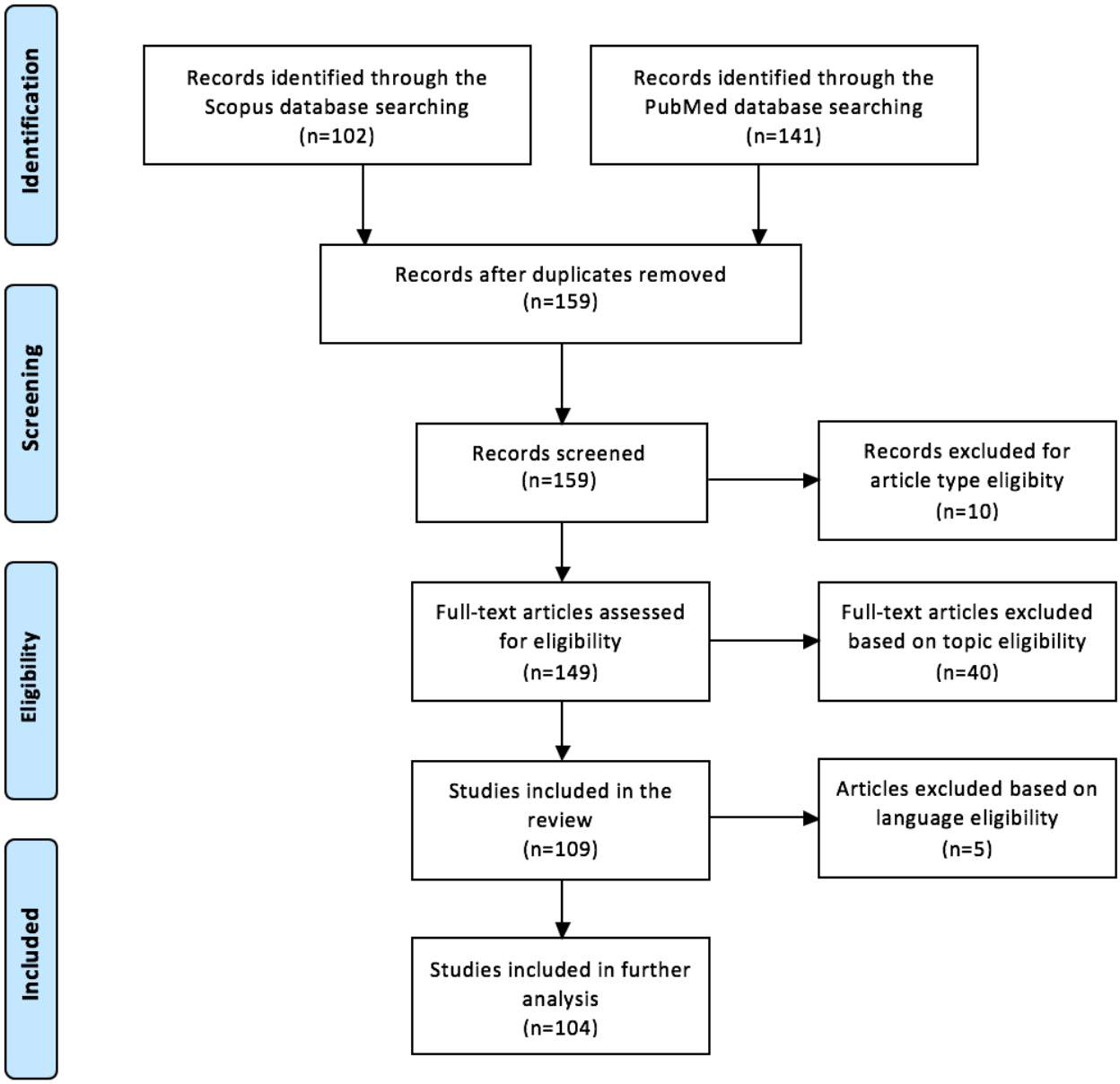
Given the diversity of subjects that Google Trends data have been used up for until this point to examine changes in interest and the usefulness of this tool in assessing human behavior, it is evident that the analysis of online search traffic data is indeed valuable in exploring and predicting behavioral changes.

In 2014, Nuti et al [27] published a systematic review of Google Trends research including the years up to 2013. This review was of importance as the first one in the field, and it reported Google Trends research up to that point. The current review differs from Nuti et al's in two ways. First, it includes 3 more full years of Google Trends research, that is, 2014, 2015, and 2016, which account for the vast majority of the research conducted in this field for the examined period based on our selection criteria. Second, while the first part of our paper is a systematic review reporting standard information, that is, authors, country, region, keywords, and language, the second part offers a detailed analysis and categorization of the methods, approaches, and statistical tools used in each of this paper. Thus, it serves as a point of reference in Google Trends research not only by subject or topic but by analysis or method as well.

Methods

The aim of this review was to include all articles on the topics of health and medicine that have used Google Trends data since its establishment in 2006 through 2016. We searched for the term "Google Trends" in the Scopus [94] and PubMed [95] databases from 2006 to 2016, and following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (Figure 1), the total number of publications included in this review was 109.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram of the selection procedure for including studies.



First, we conducted a search in Scopus for the keyword "Google Trends" in the "Abstract-Title-Keywords" field for "Articles," "Articles in press," "Reviews," and "Conference papers" from 2006 to 2016. Out of the available categories, we selected "Medicine," "Biochemistry Genetics and Molecular Biology," "Neuroscience," "Immunology and Microbiology," "Pharmacology, Toxicology, and Pharmaceuticals," "Health Profession," "Nursing," and "Veterinary." The search returned 102 publications. Second, we searched for the keyword "Google Trends" in PubMed from 2006 to 2016, which provided a total of 141 publications. Excluding the duplicates, which numbered 84 in total, 159 publications met our criteria. Excluding the ones that did not match the criteria for article type (10 publications) and the ones that did not fall inside the scope of health and medicine (40 publications), a total of 109 studies were included in this review. Note that 5 studies were written in a language other than English and were therefore not included in the quantitative part or in the detailed analysis of the methods of each study. Figure 2 depicts the number of publications by year from 2009 to 2016: 2 in 2009, 3 in 2010, 2 in 2011, 1 in 2012, 12 in 2013, 21 in 2014, 28 in 2015, and 40 in 2016.

The selected studies are further analyzed according to their methodologies, and the gaps, advantages, and limitations of the tool have been discussed so as to assist in future research. Thus, we provide a more detailed categorization of the examined papers according to the main category that they belong to, that is, visualization and general time series analysis, seasonality, correlations, predictions or forecasting, modeling, and statistical method or tool employed. Note that a study can fall into >1 category. The categorization by individual medical field is not applicable due to the high number of individual topics. Table 1 consists of the description of each parameter used to classify each study.

Figure 2. Google Trends' publications per year in health-related fields from 2009 to 2016.

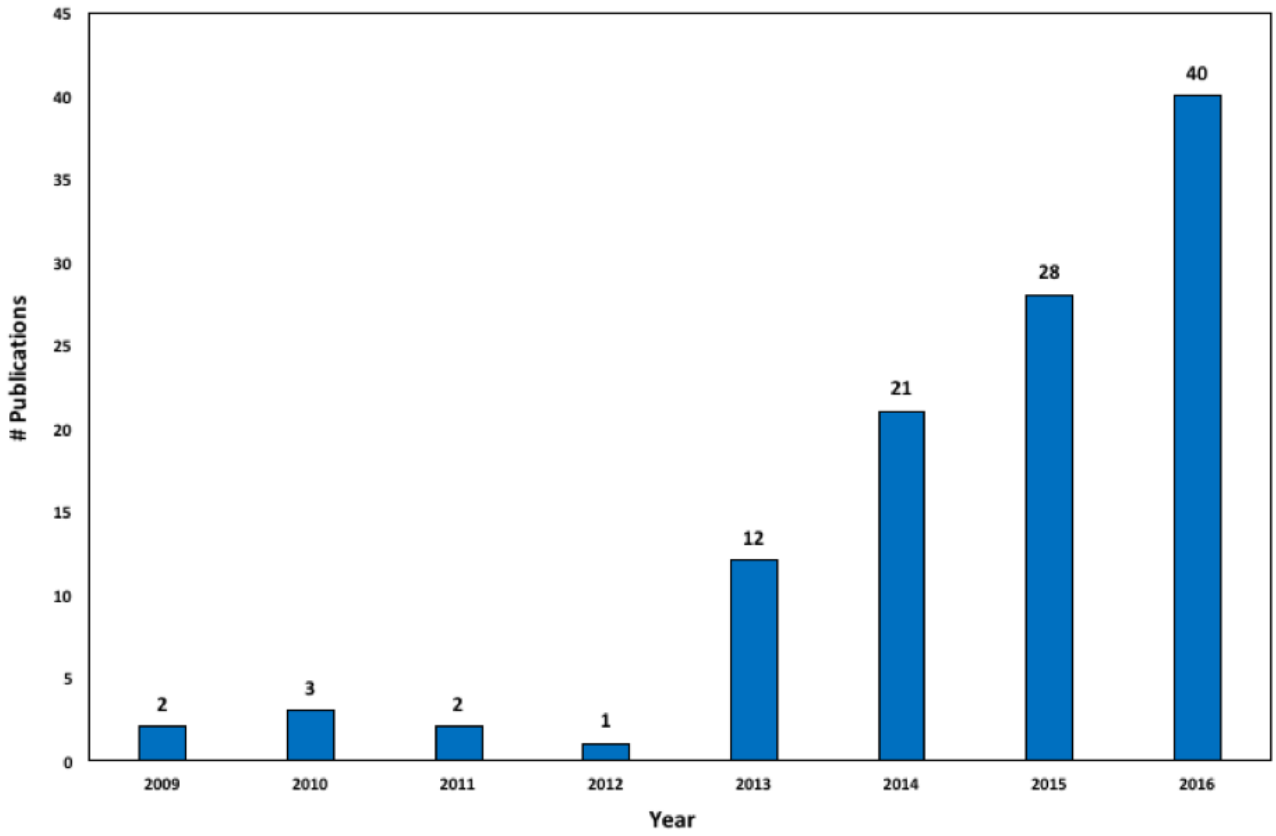


Table 1. Description of the parameters used for classification.

Parameter	Description
Authors	Includes the surname of the authors, date of publication, and link to the reference list (eg, *Smith et al, 2016 [57]*).
Period	Refers to the time-frame for which Google Trends data were retrieved and used in the study (eg, *2004-2015*).
Region	Refers to the country or countries or region (eg, *USA; Worldwide; Oceania*) that Google Trends data were extracted for.
Language	Refers to the language in which the Google Trends search was conducted (eg, search for the Italian word *Si*).
Keywords	Basic keywords are included in this category, mostly referring to the health topic examined and important keywords used to describe it.
Visualization (V)	Includes any form of visualization, that is, figures, maps, and screenshots (eg, screenshots of the Google Trends website).
Seasonality (S)	Studies that have explored the seasonality of the respective topic are included.
Correlations (C)	Studies that have examined correlations are included in this category. Correlations may be between Google Trends data and official data, among Google Trends time series, or between Google Trends and other Web-based sources' time series.
Forecasting (F)	This category includes studies that conducted forecasting of either Google Trends time series or diseases, outbreaks, etc, using Google Trends data, independent of the method used.
Modeling (M)	Studies in this category conducted some form of modeling using Google Trends data.
Statistical Tools (St)	This category includes the studies that used statistical tools or tests, eg, *t* test. Tools and methods for statistical modeling, (eg, *regression*), are not included in this category but only in the category of Modeling.

Results

Multimedia Appendix 1 consists of the first classification of the selected studies [27,39-57,59-93,96-144]; there are 104 in total, as the studies of Kohler et al [145], Orellano et al [146], Cjuno et al [147], Tejada-Llacsa [148], and Yang et al [149] are written in German, Spanish, or Chinese, and thus are not included in the more detailed categorization and analysis.

All the examined papers involve, by definition, time series analysis, and almost all include some form of visualization. Only 8.7% (9/104) studies used Google Trends data for predictions and forecasting, and 23.1% (24/104) used them for examining seasonality, while correlations and modeling were performed in 39.4% (41/104) and 32.7% (34/104) studies, respectively. As the category of forecasting and predictions exhibits the least number of studies, it is evident that a gap exists in the literature for forecasting using Google Trends in health assessment.

As is evident in Multimedia Appendix 1, Google queries have been employed up to this point in many countries and several languages. Figure 3 shows a worldwide map by examined country for assessing health and medicine related issues using Google Trends data up to 2016. Worldwide, the studies that explore topics related to the respective terms number 23 in total. As far as individual countries are concerned, US data have been employed in the most (60) studies, while other countries that have been significantly examined include the United Kingdom (15), Australia (13), Canada (9), Germany (8), and Italy (7).

The four most examined countries are English-speaking ones. The reasons for this could include that Google Trends, though not case-sensitive, does take into account accents and spelling mistakes; therefore, for countries with more complicated alphabets, the analysis of Web-based data should be more careful. In addition, other factors that could play a significant role and are taken into account when choosing the countries to be examined using online search traffic data are the availability of official data, the openness of said data, any internet restrictions or monitoring in countries with lower scores in freedom of press or freedom of speech, and internet penetration.

The rest of the analysis consists of the further breaking down of the initial categorization to include the respective methods that were used for examining seasonality, correlations, forecasting, and performing statistical tests and estimating models, along with a concise introduction to each of these methods and how they were used to assess health issues.

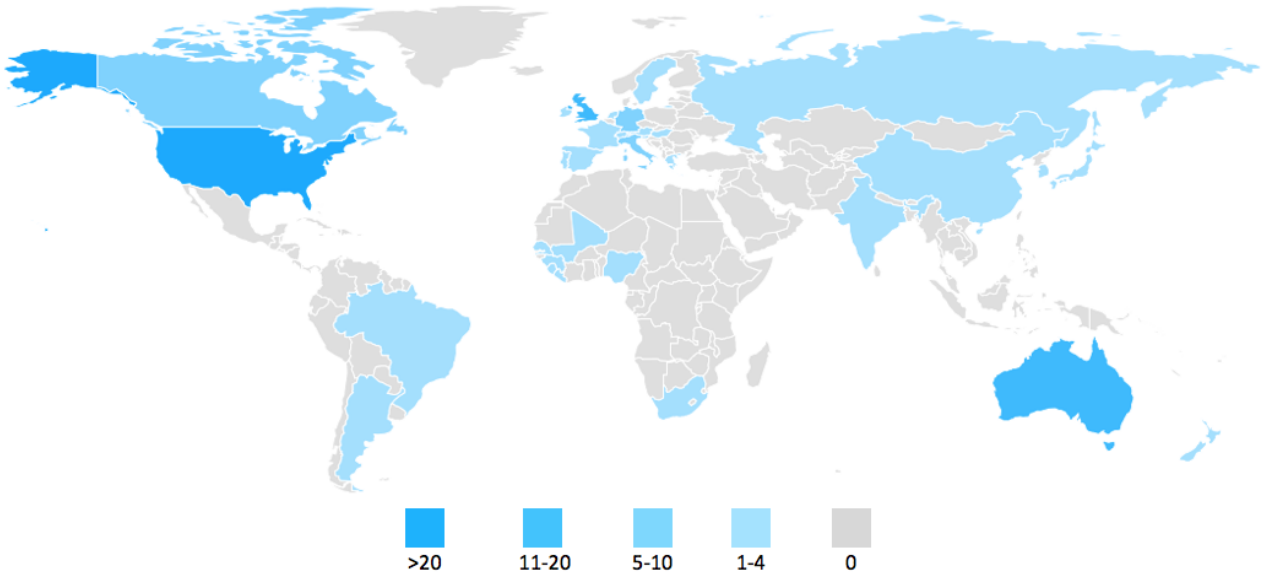
Table 2 shows the methods used to explore seasonality; Tables 3 and 4 present the methods used to examine correlations and perform predictions and forecasting, respectively. Finally, Tables 5 and 6 list the modeling methods and other statistical tools employed in health assessment using Google Trends.

The most popular way to explore seasonality is to use visual evidence and examine and discuss peaks, as shown in Table 2. Furthermore, several studies have used cosinor analysis [8,69,134,138,142], which is a time series analysis method for seasonal data using least squares.

Apart from seasonality [122], analysis of variance (ANOVA) has been also used for geographical comparisons between regions or countries [49,51,68,93] and between differences in monthly data [41]. It is a test used for examining if significant differences between means exist. In the case of 2 means, *t* test is the equivalent to ANOVA.

The Kruskal-Wallis test is also a popular method for examining seasonality using Google Trends [57,68,113]. It is a nonparametric, independent of distribution test, for continuous as well as ordinal-level dependent variables, employed when the one-way ANOVA assumptions do not hold, that is, for examining statistically significant differences between ≥3 groups. It uses random sample with independent observations, with the dependent variable being at least ordinal.

Figure 3. Countries by number of Scopus and PubMed publications using Google Trends.



Other methods of exploring seasonality include the nonparametric tests (independent of distribution) Wilcoxon signed rank [18,113] and Mann-Whitney U test [67], which are used for comparing data in different seasons or time periods when the equivalent parametric *t* tests cannot be used. The latter has been also used by some studies to compare weekly data [105] and differences among regions [113].

For examining correlations (Table 3), the vast majority of the studies used the Pearson correlation coefficient, which examines the strength of association between 2 quantitative, continuous variables, employed when the relationship is linear. The Spearman rho (rank-order) correlation, the second most used method, is the nonparametric version of the Pearson correlation, has also been used to explore seasonality between time series [70]. Spearman correlation coefficient (denoted by ρ or r_s) measures the levels to which 2 ranked variables (ordinal, interval, or ratio) are related to each other.

Cross-correlations are used for examining the relationship of 2 time series, while simultaneously exploring if the data are periodic. It is often employed in correlating Google Trends data with observed data [50,82,90,135] and between different Google search terms [80], while it can be also used for examining linear and temporal associations of seasonal data [71]. Cross-correlations have been also used in forecasting, where Wang et al [92] showed that cross-correlations of new dementia cases with Google Trends data can assist with the forecasting of dementia cases, and Solano et al [80] forecasted the suicide rates 2 years ahead using Google queries. The autocorrelations are basically cross-correlations for one time series, that is, a time series cross-correlated with itself.

The Kendall's tau-b test correlation coefficient is a nonparametric alternative to Pearson and Spearman correlations and is used to measure the strength and direction of the relationship between 2 (at least ordinal) variables. It has been employed by 1 study [138] to examine the correlations between Google Trends data and the results of a paper interview survey.

The Spearman-Brown prediction (or prophecy) formula is used to predict how reliable the test is after changing its length. It has also been employed by only 1 study [65] to explore the relationship between railway suicide and Google hits.

The generalized linear model estimates the linear relationship between a dependent and ≥1 independent variables. It was used by Domnich et al [79] to predict influenza-like illness morbidity, with the exploratory variables being "Influenza," "Fever," and "Tachipirin search volumes," along with the Holt-Winters method and the autoregressive moving average process for the residuals. Holt-Winters is a method employed in exploring the seasonality in time series, and for predictions, the autoregressive moving average (also called the Box-Jenkins model) is a special case of the autoregressive integrated moving average, used for the analysis of time series and predictions.

Autoregressive integrated moving average is a commonly used method for time series analysis and predictions [55,63,86,92,141], the latter having also been assessed by linear regressions and modeling [88,91]. Multivariable regressions are used to estimate the relationship of ≥2 independent variables with a dependent one. In Google Trends, they have been used to relate Ebola searches, reported cases, and the Human Development Index [85] and to study the relationship between climate and environmental variables and Google hits [125].

Hierarchical linear modeling is a regression of ordinary least squares that is employed to analyze hierarchically structured data, that is, units that are grouped together, and it has been employed by 1 study so far [83].

The Mann-Kendall test, which is the nonparametric alternative test to the independent sample, has been used to show the statistical differences of peaks [43] and to detect trends [140]. Finally, the *t* test is used to compare 2 sample means of the same population, and it has been employed for comparing Google searches with the baseline period [105] and to examine the statistical differences of peaks [41].

Table 2. Methods for exploring seasonality with Google Trends in health assessment.

Number	Authors	Method	Description
1	Bakker et al, 2016 [96]	Morlet Wavelet Analysis	To test the seasonality of Google Trends data in the examined countries
2	Braun and Harreus, 2013 [104]	Visual evidence	N/A[a]
3	Crowson et al, 2016 [93]	Seasonal peaks	N/A
4	Deiner et al, 2016 [70]	Spearman correlation	Correlating the seasonality of clinical diagnoses with Google Trends data
5	El-Sheikha, 2015 [113]	Kruskal-Wallis test	To show seasonality for different months
6	Garrison et al, 2015 [116]	Least-squares sinusoidal model	Variability in outcomes (supported also from a comparison with searches in Australia)
7	Harsha et al, 2014 [68]	Kruskal-Wallis test	Seasonal (monthly) comparisons
8	Harsha et al, 2015 [119]	Kruskal-Wallis test	Seasonal (monthly) comparisons
9	Hassid et al, 2016 [120]	Pearson correlation	To examine seasonal variations across symptoms
10	Ingram and Plante, 2013 [122]	Cosinor analysis; analysis of variance	To test the seasonal variation of the normalized Google Trends data; to compare the seasonal increase among the examined countries
11	Ingram et al, 2015 [69]	Cosinor analysis	To test the seasonal variation of the normalized Google Trends data
12	Kang et al, 2015 [72]	Visual observation	N/A
13	Leffler et al, 2010 [125]	Correlations	Showing correlations among the 4 seasons for the 39 examined terms
14	Liu et al, 2016 [127]	Seasonal model and a null model	Seasonality explained the searches significantly better with an F-test
15	Phelan et al, 2016 [133]	Correlograms (autocorrelations plots)	Visual interpretation for exploring seasonal peaks
16	Plante and Ingram, 2014 [134]	Cosinor analysis	To test the seasonal variation of the normalized Google Trends data
17	Rossignol et al, 2013 [67]	Mann-Whitney U test; Harmonic Product Spectrum	Comparison of summer vs winter hits; evaluation of seasonality
18	Seifter et al, 2010 [66]	Visual evidence	N/A
19	Sentana-Lledo et al, 2016 [138]	Cosinor analysis	To test the seasonal variations of the Google Trends data
20	Takada, 2012 [139]	Visual evidence	N/A
21	Telfer and Woodburn, 2015 [140]	Two-way Wilcoxon signed rank test	To explore differences between winter and summer
22	Toosi and Kalia, 2015 [142]	Visual evidence; cosinor analysis	To identify differences in seasonality between countries
23	Willson et al, 2015 [86]	Visual evidence	N/A
24	Zhang et al, 2015 [71]	Periodograms; ideal pass filter	To study the periodograms; to extract seasonal components

[a]N/A: not applicable.

Many studies have employed Google Trends for visualizing the changes in online interest or discussing peaks and spikes [60,62,123,124]. Brigo and Trinka [40] and Brigo et al [39] have studied the search volumes for related terms, Chaves et al [109] and Luckett et al [128] have explored terms related to the studied topic, and Davis et al [110] have examined related internet searches. Other approaches include the reporting of the polynomial trend lines [46] and investigation of statistically significant differences in yearly increases [119]. In addition, "Google Correlate" has been used to explore related terms [91,138].

Finally, several studies have used other sources of big data, namely, Google News [43,63,80], Twitter [43,54,61,63,108], Yandex [52], Baidu [121], Wikipedia [43,63], Facebook and Google+ [54], and YouTube [43,54,63]. Google is the most popular search engine. However, other Web-based sources are used or even preferred to Google in some regions; therefore, many studies use data from these sources to examine general interest in the respective subjects, compare them to Google Trends data, or use them together as variables.

Table 3. Methods of exploring correlations using Google Trends in health assessment.

Number	Authors	Method	Description
1	Alicino et al, 2015 [85]	Pearson correlation	Ebola-related Google Trends data with Ebola cases
2	Arora et al, 2016 [81]	Spearman correlation	Suicide search activity vs official suicide rates (and per age)
3	Bakker et al, 2016 [96]	Correlations	Between Google Trends data and reported cases
4	Bragazzi et al, 2016 [99]	Pearson correlation	Between Google Trends data and epidemiological data
5	Bragazzi, 2013 [98]	Autocorrelation; Pearson correlation	For the time series for multiple sclerosis (MS); between MS terms
6	Bragazzi et al, 2016 [101]	Autocorrelation; Partial Autocorrelation	To compute correlation of the time series with its own values
7	Bragazzi et al, 2016 [102]	Pearson correlation	Status epilepticus terms with etiology and management related terms
8	Bragazzi et al, 2016 [43]	Pearson correlation	Google searches for Silicosis with Normalized Google News, Google Scholar, PubMed Publications, Twitter traffic, Wikipedia
9	Bragazzi et al, 2016 [63]	Pearson correlation	Among Google Trends data and other data generating sources
10	Bragazzi, 2014 [103]	Pearson correlation; autocorrelation and partial autocorrelation	Nonsuicidal self-injury and related terms; nonsuicidal self-injury plots showed regular cyclical pattern
11	Cavazos-Regh et al, 2015 [107]	Pearson correlation	Among Google Trends data for noncigarette tobacco and prevalence
12	Cho et al, 2013 [78]	Pearson correlation	Google flu-related queries with surveillance data for different influenza seasons
13	Crowson et al, 2016 [93]	Pearson correlation	Between the selected keywords. Between medical prescriptions data and Google Trends data
14	Deiner et al, 2016 [70]	Spearman correlation	For correlating seasonality of clinical diagnoses with Google Trends data
15	Domnich et al, 2015 [79]	Pearson correlation	Among the examined search terms and influenza-like illness
16	Foroughi et al, 2016 [115]	Rank correlations; cross-country correlations; Pearson correlations	For search volumes; for the search volumes for cancer; for the weekly search volumes between countries
17	Gahr et al, 2015 [75]	Pearson correlation	Among annual prescription volumes and Google Trends data
18	Gamma et al, 2016 [90]	Cross-correlations	Cross-correlations between search volumes and crime statistics
19	Gollust et al, 2016 [117]	Multinomial Logit Models	To relate health insurance rates
20	Guernier et al, 2016 [82]	Spearman correlation; cross-correlation	Correlating the examined search terms with notifications of tick paralysis cases record; with lag values from −7 to +7 months
21	Hassid et al, 2016 [120]	Pearson correlation	Between Google Trends data and National Inpatient Sample data
22	Johnson et al, 2014 [84]	Pearson correlation	Pearson correlations to explore the relation of Google Trends data and sexually transmitted infection reported rates
23	Kang et al, 2013 [77]	Pearson correlation	To explore the association of (and among) search terms with surveillance data
24	Kang et al, 2015 [72]	Spearman correlation	Google Trends data for allergic rhinitis and related Google Trends terms and real world epidemiologic data for the United States
25	Koburger et al, 2015 [65]	Spearman-Brown correlation	To explore relations among Google Trends data and railway suicides
26	Ling and Lee, 2016 [126]	Pearson correlation	Between disease prevalence and Google Trends data
27	Mavragani et al, 2016 [76]	Pearson correlation	Between Google Trends data and published papers and Google Trends data with prescriptions
28	Phelan et al, 2016 [133]	Linear Regression	To examine if there is significant correlation between searches and time

Number	Authors	Method	Description
29	Poletto et al, 2016 [56]	Pearson correlation	Between Google Trends data and number of alerts published by ProMED mail and the number of Disease Outbreak News published by the World Health Organization
30	Pollett et al, 2015 [91]	Pearson correlation	To shortlist related search terms to pertussis
31	Rohart et al, 2016 [135]	Spearman rank correlations; Spearman correlation; cross-correlations	For the diseases examined; correlations between diseases and the investigated search metrics; to identify best lags
32	Shin et al, 2016 [137]	Spearman correlation	Between Google Trends data and the number of confirmed cases of Middle East Respiratory Syndrome and for quarantined cases of Middle East Respiratory Syndrome
33	Schootman et al, 2015 [45]	Pearson correlation	Between Respiratory Syncytial Virus and Behavioral Risk Factor Surveillance System prevalence data for 5 cancer screening tests
34	Schuster et al, 2010 [73]	Correlations	Lipitor Google Trends data and Lipitor revenues
35	Sentana-Lledo et al, 2016 [138]	Kendall's Tau-b test	To explore the correlation of Google Trends data with paper interview survey results
36	Simmering et al, 2014 [50]	Cross-correlations	Between Google Trends data for drugs and drug utilization, to see changes in search volumes following knowledge events
37	Solano et al, 2016 [80]	Correlations; cross-correlations	Between Google Trends data for suicide and national suicide rates; between different search terms
38	Wang et al, 2015 [92]	Pearson correlation	Between Google Trends data and new dementia cases
39	Willson et al, 2015 [86]	Spearman correlation	Between Google Trends data and observed data for aeroallergens
40	Zhang et al, 2015 [71]	Cross-correlations	To examine linear and temporal associations of the seasonal data
41	Zhang et al, 2016 [51]	Pearson correlation	To study pairwise comparisons among searches for different terms in Google Trends

Table 4. Forecasting and predictions using Google Trends in health assessment.

Number	Authors	Method	Description
1	Bakker et al, 2016 [96]	Statistical model	For forecasting chicken poxforce of infection, that is, monthly per capita rate of infection of children 0-14
2	Domnich et al, 2015 [79]	Generalized least squares (maximum likelihood estimates); Holt-Winters	Query-based models to predict influenza-like illness morbidity, with the exploratory variables: Influenza, Fever, Tachipirin; compared for forecasting power with Holt-Winters based on the real data (hold out set)
3	Parker et al, 2016 [132]	Statistical model	For forecasting deaths for 1 year in advance (2015)
4	Pollett et al, 2015 [91]	Prediction model	Tested the predicted model with a left-out dataset for prediction accuracy
5	Rohart et al, 2016 [135]	Linear models	To forecast with 1 or 2 weeks step
6	Solano et al, 2016 [80]	Cross-Correlations	Forecasting for suicides for 2 years without data (2013-14) based on Google Trends data of those years
7	Wang et al, 2015 [92]	Cross-Correlations	To investigate forecasting with lags of 0-12 months
8	Zhang et al, 2016 [51]	Autoregressive Moving Average	To predict Respiratory Syncytial Virus for "dabbing"
9	Zhou et al, 2011 [88]	Dynamic model	To provide real time estimations by correcting the forecasting with the new morbidity data when published

Table 5. Statistical modeling using Google Trends in health assessment.

Number	Authors	Method	Description
1	Alicino et al, 2015 [85]	Multivariate regression	For relating Ebola Google Trends data, number of Ebola Cases, and the Human Development Index
2	Bakker et al, 2016 [96]	Statistical model	For forecasting chicken poxforce of infection, that is, monthly per capita rate of infection
3	Bentley and Ormerod, 2009 [59]	Maximum likelihood estimation	Established social model for engaging a new behavior for Web-based searching for flu terms
	Barnes et al, 2015 [83]	Hierarchical linear modeling	Three levels: 3 Mondays, 6 years, 47 search terms
4	Bragazzi, 2013 [98]	Multiple linear regression	To confirm multiannual long-term trends
5	Domnich et al, 2015 [79]	Generalized linear model, autoregressive moving average process	Query volume-based models to predict influenza-like illness morbidity
6	El-Sheikha, 2015 [113]	Linear regression	To show the global, regional, and country level interest for the search term
7	Fenichel et al, 2013 [114]	Moving average, generalized linear model	Google Trends data as a variable in predicting loses in flights
8	Garrison et al, 2015 [116]	Seasonal model	Best fit combination of a straight line and a sinusoid
9	Gollust et al, 2016 [117]	Multinomial logit models	To relate health insurance rates
10	Haney et al, 2014 [55]	ARIMA[a]	Radiology residency interest
11	Harsha et al, 2014 [68]	Linear model	Statistical justification of annual increase in search volumes
12	Harsha et al, 2015 [119]	Linear model	Statistical justification of annual increase in search volumes and of the Web-based interest related to applications for interventional radiology
13	Leffler et al, 2010 [125]	Multivariable Linear Regressions	For studying the effect of climatic and environmental variables to internet searches
17	Linkov et al, 2014 [46]	Polynomial trend lines	Fitted spline polynomial trend lines per time without statistical reporting
18	Liu et al, 2016 [127]	Seasonal model	Best fit combination of a straight line and a sinusoid
19	Majumder et al, 2016 [129]	Linear Smoothing	To adjust HealthMap to using Google Trends, model fits
20	Noar et al, 2013 [64]	Linear Regression	To estimate the slope coefficient for changes in the magnitude of the effect size of Google Trends data and media search increases
21	Parker et al, 2016[132]	L1-regularization on Google Trends	To build a model for forecasting deaths in each state
22	Phelan et al, 2014 [49]	Linear Regression	To estimate the relation between news reports and search activity
23	Phelan et al, 2016 [133]	Linear Regression	To examine if there is a significant correlation between searches and time
24	Pollett et al, 2015 [91]	Linear Regression	Prediction model for pertussis cases based on Google Trends data of the most related terms
25	Rohart et al, 2016 [135]	Linear models	To forecast with 1 or 2 weeks step
26	Scatà et al, 2016 [136]	Epidemic model	Google Trends data is a measure of awareness, along with other sources
27	Schuster et al, 2010 [73]	Generalized Linear models	Google Trends data for the examined drugs, Google Trends data and changes in annual revenues, and Google Trends data vs resource utilization
28	Stein et al, 2013 [47]	Regression Fit Lines	To examine differences in queries
29	Telfer and Woodburn, 2015 [140]	Visual decomposition; local regression	Figures 4, 6 and 8; regression-based decomposition of the time series for the search terms
30	Troelstra et al, 2016 [141]	ARIMA	To account for dependency between data points in time series for “quit smoking” searches
31	Willson et al, 2015 [86]	ARIMA	To quantify the effect of the observed (pollen) counts with the levels of search activity

Number	Authors	Method	Description
32	Willson et al, 2015 [87]	ARIMA	To quantify the effect of the observed (pollen) counts with the levels of search activity
33	Yang et al, 2015 [144]	Prediction model (ARGO[b])	To predict influenza-like illness
34	Zhou et al, 2011 [88]	Dynamic Modeling	For forecasting tuberculosis incidents using Google Trends data

[a]ARIMA: autoregressive integrated moving average.

[b]ARGO: autoregression with Google search data.

Table 6. Statistical tests and tools using Google Trends in health assessment.

Number	Authors	Method	Description
1	Bragazzi et al, 2016 [43]	Mann-Kendall test	To show the statistical difference of peaks from the remaining period
2	Bragazzi et al, 2016 [63]	ARIMA[a]	To show increased web searches due to an event, and correct seasonality
3	Campen et al, 2014 [105]	Independent samples *t* test; Mann-Whitney U test with Bonferroni correction	For comparing searches with baseline period; for multiple weekly data comparisons
4	Crowson et al, 2016 [93]	ANOVA[b] (Post-hoc Tukey test)	To compare grouped geographical federal regions of the United States (Northeast, Midwest, South, West)
5	El-Sheikha, 2015 [113]	Wilcoxon rank test; Mann-Whitney	To study the change of interest at different time periods; to compare Web-based interest between the Northern and Southern hemispheres
6	Gahr et al, 2015 [75]	Coefficients of determination	To determine the amount of variability between annual prescription volumes and Google search terms
7	Harsha et al, 2014 [68]	ANOVA (Tukey-Kramer post hot test)	For the comparisons of US regions
8	Murray et al, 2016 [41]	ANOVA; *t* test	To explore differences in months' means per year; for the statistical differences of peaks compared with the remaining hits
9	Noar et al, 2013 [64]	Augmented Dickey-Fuller tests	To test for nonstationarity of the time series
10	Phelan et al, 2014 [49]	ANOVA	To explore differences among countries
11	Rohart et al, 2016 [135]	Mean Square Error for Prediction	To assess prediction accuracy
12	Telfer and Woodburn, 2015 [140]	Mann-Kendall trend tests	To detect trends significantly larger than the variance in the data for search terms
13	Troelstra et al, 2016 [141]	ARIMA	Studied the effect of smoking cessation policies with ARIMA interrupted time series modeling (Multimedia Appendix 1)
14	Zhang et al, 2015 [71]	Augmented Dickey-Fuller test	To detect whether or not the extracted seasonal components of the studied trends were stationary
15	Zhang et al, 2016 [51]	ANOVA	To examine the search interest for dabbing between groups of legal status states in the United States

[a]ARIMA: autoregressive integrated moving average.

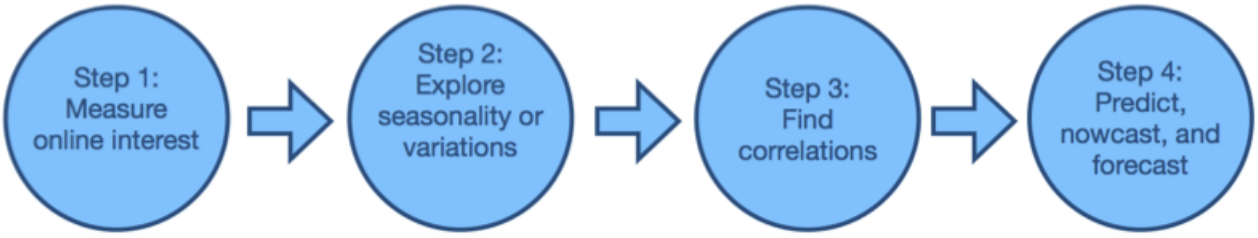
[b]ANOVA: analysis of variance.

Discussion

Principal Findings

With internet penetration constantly growing, users' Web-based search patterns can provide a great opportunity to examine and further predict human behavior. In addressing the challenge of big data analytics, Google Trends has been a popular tool in research over the past decade, with its main advantage being that it uses the revealed and not the stated data. Health and medicine are the most popular fields where Google Trends data have been employed so far to examine and predict human behavior. This review provides a detailed overview and classification of the examined studies (109 in total from 2006 through 2016), which are then further categorized and analyzed by approach, method, and statistical tools employed for data analysis.

Figure 4. The four steps toward employing Google Trends for health assessment.



The vast majority of studies using Google Trends in health assessment so far have included data visualization, that is, figures, maps, or screenshots. As discussed in the analysis, the most popular way of using Google Trends data in this field is correlating them with official data on disease occurrence, spreading, and outbreaks. The assessment of suicide tendencies and (prescription or illegal) drug-related queries has been of notably growing popularity over the course of the last years. As is evident, the gap in the existing literature is the use of Google Trends for predictions and forecasting in health-related topics and issues. Though data on reported cases of various health issues and the respective Google Trends data have been correlated in a large number of studies, only a few have proceeded with forecasting incidents and occurrences using online search traffic data.

In research using Google Trends in health and medicine from 2006 to 2016, the ultimate goal is to be able to use and analyze Web-based data to predict and provide insight to better assess health issues and topics. The four main steps, based on the presentation of the papers published up to this point in assessing health using Google Trends, are as follows (Figure 4):

1. Measure the general Web-based interest.
2. Detect any variations or seasonality of Web-based interest, and proceed with examining any relations between actual events or cases.
3. Correlate Web-based search queries among them or with official or actual data and events.
4. Predict, nowcast, and forecast health-related events, outbreaks, etc.

Limitations

This review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines for selecting the examined papers from the Scopus and PubMed databases. Though this includes the majority of papers published on the topic from 2006 to 2016, the studies that are not indexed in these databases or are not indexed based on the selection criteria used in this review were not included in further analysis. In addition, as is evident in Figure 2, research using Google Trends data has shown a significant increase from each year to the next since 2013. This review included studies published in Google Trends research through 2016. However, there are several studies published in 2017 and 2018 that are not included. This review provides, at first, an overall description of each examined study, which is standard review information. The second part is a classification and assessment of the methodology, tools, and results of each study. Though the first part mainly reports what is included in the methodology of each study, the second part could include a bias, as it is the authors' assessment and categorization of the methods employed based on the results obtained after a very careful and thorough examination of each individual study.

Conclusions

This review consists of the studies published from 2006 to 2016 on Google Trends research in the Scopus and PubMed databases based on the selected criteria. The aim of this review was to serve as a point of reference for future research in health assessment using Google Trends, as each study, apart from the basic information, for example, period, region, language, is also categorized by the method, approach, and statistical tools employed for the analysis of the data retrieved from Google Trends. Google Trends data are being all the more integrated in infodemiology research, and Web-based data have been shown to empirically correlate with official health data in many topics. It is thus evident that this field will become increasingly popular in the future in health assessment, as the gathering of real time data is crucial in monitoring and analyzing seasonal diseases as well as epidemics and outbreaks.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Publication details and categorization.

[PDF File (Adobe PDF File), 256KB - jmir_v20i11e270_app1.pdf]

References

1. Al Nuaimi E, Al Neyadi H, Mohamed N, Al-Jaroodi J. Applications of big data to smart cities. J Internet Serv Appl 2015 Dec 1;6(1):1-15. [doi: 10.1186/s13174-015-0041-5]

2. Hilbert M, López P. The world's technological capacity to store, communicate, and compute information. Science 2011 Apr 01;332(6025):60-65 [FREE Full text] [doi: 10.1126/science.1200970] [Medline: 21310967]
3. Philip Chen C, Zhang C. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences 2014 Aug;275:314-347. [doi: 10.1016/j.ins.2014.01.015]
4. Jin X, Wah BW, Cheng X, Wang Y. Significance and Challenges of Big Data Research. Big Data Research 2015 Jun;2(2):59-64. [doi: 10.1016/j.bdr.2015.01.006]
5. Fosso Wamba S, Akter S, Edwards A, Chopin G, Gnanzou D. How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. International Journal of Production Economics 2015 Jul;165:234-246. [doi: 10.1016/j.ijpe.2014.12.031]
6. Chang RM, Kauffman RJ, Kwon Y. Understanding the paradigm shift to computational social science in the presence of big data. Decision Support Systems 2014 Jul;63:67-80. [doi: 10.1016/j.dss.2013.08.008]
7. Gandomi A, Haider M. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 2015 Apr;35(2):137-144. [doi: 10.1016/j.ijinfomgt.2014.10.007]
8. Preis T, Moat HS, Stanley HE, Bishop SR. Quantifying the advantage of looking forward. Sci Rep 2012 Apr;2:350 [FREE Full text] [doi: 10.1038/srep00350] [Medline: 22482034]
9. Preis T, Moat HS, Stanley HE. Quantifying trading behavior in financial markets using Google Trends. Sci Rep 2013 Apr;3:1684 [FREE Full text] [doi: 10.1038/srep01684] [Medline: 23619126]
10. Burnap P, Rana OF, Avis N, Williams M, Housley W, Edwards A, et al. Detecting tension in online communities with computational Twitter analysis. Technological Forecasting and Social Change 2015 Jun;95:96-108. [doi: 10.1016/j.techfore.2013.04.013]
11. Google Trends. URL: https://trends.google.com/trends/explore [accessed 2017-11-08] [WebCite Cache ID 6uot1LkyX]
12. Scharkow M, Vogelgesang J. Measuring the Public Agenda using Search Engine Queries. International Journal of Public Opinion Research 2011 Mar 01;23(1):104-113. [doi: 10.1093/ijpor/edq048]
13. McCallum ML, Bury GW. Public interest in the environment is falling: a response to Ficetola (2013). Biodivers Conserv 2014 Feb 14;23(4):1057-1062. [doi: 10.1007/s10531-014-0640-7]
14. Jun S, Park D. Consumer information search behavior and purchasing decisions: Empirical evidence from Korea. Technological Forecasting and Social Change 2016 Jun;107:97-111. [doi: 10.1016/j.techfore.2016.03.021]
15. Jun S, Park D, Yeom J. The possibility of using search traffic information to explore consumer product attitudes and forecast consumer preference. Technological Forecasting and Social Change 2014 Jul;86:237-253. [doi: 10.1016/j.techfore.2013.10.021]
16. Han S, Chung H, Kang B. It is time to prepare for the future: Forecasting social trends. Communications in Computer and Information Scienc 2012:e2012-e2031.
17. Vosen S, Schmidt T. Forecasting private consumption: survey-based indicators vs. Google trends. J. Forecast 2011 Jan 13;30(6):565-578. [doi: 10.1002/for.1213]
18. Choi H, Varian H. Predicting the Present with Google Trends. Economic Record. (SUPPL.1) 2012;88:2-9. [doi: 10.1111/j.1475-4932.2012.00809.x]
19. Mavragani A, Tsagarakis KP. YES or NO: Predicting the 2015 GReferendum results using Google Trends. Technological Forecasting and Social Change 2016 Aug;109:1-5. [doi: 10.1016/j.techfore.2016.04.028]
20. Carrière-Swallow Y, Labbé F. Nowcasting with Google Trends in an Emerging Market. J. Forecast 2011 Nov 20;32(4):289-298. [doi: 10.1002/for.1252]
21. Vicente MR, López-Menéndez AJ, Pérez R. Forecasting unemployment with internet search data: Does it help to improve predictions when job destruction is skyrocketing? Technological Forecasting and Social Change 2015 Mar;92:132-139. [doi: 10.1016/j.techfore.2014.12.005]
22. Jun S, Yeom J, Son J. A study of the method using search traffic to analyze new technology adoption. Technological Forecasting and Social Change 2014 Jan;81:82-95. [doi: 10.1016/j.techfore.2013.02.007]
23. Google. How Trends data is adjusted URL: https://support.google.com/trends/answer/4365533?hl=en [accessed 2017-11-08] [WebCite Cache ID 6uot8lARg]
24. Fan J, Han F, Liu H. Challenges of Big Data Analysis. Natl Sci Rev 2014 Jun;1(2):293-314 [FREE Full text] [doi: 10.1093/nsr/nwt032] [Medline: 25419469]
25. Yoo C, Ramirez L, Liuzzi J. Big data analysis using modern statistical and machine learning methods in medicine. Int Neurourol J 2014 Jun;18(2):50-57 [FREE Full text] [doi: 10.5213/inj.2014.18.2.50] [Medline: 24987556]
26. Gu D, Li J, Li X, Liang C. Visualizing the knowledge structure and evolution of big data research in healthcare informatics. Int J Med Inform 2017 Dec;98:22-32. [doi: 10.1016/j.ijmedinf.2016.11.006] [Medline: 28034409]
27. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of google trends in health care research: a systematic review. PLoS One 2014 Oct;9(10):e109583 [FREE Full text] [doi: 10.1371/journal.pone.0109583] [Medline: 25337815]
28. 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;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]

29. Phillips CA, Barz LA, Li Y, Schapira MM, Bailey LC, Merchant RM. Relationship Between State-Level Google Online Search Volume and Cancer Incidence in the United States: Retrospective Study. J Med Internet Res 2018 Jan 08;20(1):e6 [FREE Full text] [doi: 10.2196/jmir.8870] [Medline: 29311051]
30. Mavragani A, Sampri A, Sypsa K, Tsagarakis KP. Integrating Smart Health in the US Health Care System: Infodemiology Study of Asthma Monitoring in the Google Era. JMIR Public Health Surveill 2018 Mar 12;4(1):e24 [FREE Full text] [doi: 10.2196/publichealth.8726] [Medline: 29530839]
31. Chen T, Dredze M. Vaccine Images on Twitter: Analysis of What Images are Shared. J Med Internet Res 2018 Apr 03;20(4):e130 [FREE Full text] [doi: 10.2196/jmir.8221] [Medline: 29615386]
32. Farhadloo M, Winneg K, Chan MS, Hall JK, Albarracin D. Associations of Topics of Discussion on Twitter With Survey Measures of Attitudes, Knowledge, and Behaviors Related to Zika: Probabilistic Study in the United States. JMIR Public Health Surveill 2018 Feb 09;4(1):e16 [FREE Full text] [doi: 10.2196/publichealth.8186] [Medline: 29426815]
33. Simpson SS, Adams N, Brugman CM, Conners TJ. Detecting Novel and Emerging Drug Terms Using Natural Language Processing: A Social Media Corpus Study. JMIR Public Health Surveill 2018 Jan 08;4(1):e2 [FREE Full text] [doi: 10.2196/publichealth.7726] [Medline: 29311050]
34. van Lent LG, Sungur H, Kunneman FA, van de Velde B, Das E. Too Far to Care? Measuring Public Attention and Fear for Ebola Using Twitter. J Med Internet Res 2017 Jun 13;19(6):e193 [FREE Full text] [doi: 10.2196/jmir.7219] [Medline: 28611015]
35. Wongkoblap A, Vadillo MA, Curcin V. Researching Mental Health Disorders in the Era of Social Media: Systematic Review. J Med Internet Res 2017 Jun 29;19(6):e228 [FREE Full text] [doi: 10.2196/jmir.7215] [Medline: 28663166]
36. Huesch M, Chetlen A, Segel J, Schetter S. Frequencies of Private Mentions and Sharing of Mammography and Breast Cancer Terms on Facebook: A Pilot Study. J Med Internet Res 2017 Jun 09;19(6):e201 [FREE Full text] [doi: 10.2196/jmir.7508] [Medline: 28600279]
37. Lu FS, Hou S, Baltrusaitis K, Shah M, Leskovec J, Sosic R, et al. Accurate Influenza Monitoring and Forecasting Using Novel Internet Data Streams: A Case Study in the Boston Metropolis. JMIR Public Health Surveill 2018 Jan 09;4(1):e4 [FREE Full text] [doi: 10.2196/publichealth.8950] [Medline: 29317382]
38. Roccetti M, Marfia G, Salomoni P, Prandi C, Zagari RM, Gningaye KFL, et al. Attitudes of Crohn's Disease Patients: Infodemiology Case Study and Sentiment Analysis of Facebook and Twitter Posts. JMIR Public Health Surveill 2017 Aug 09;3(3):e51 [FREE Full text] [doi: 10.2196/publichealth.7004] [Medline: 28793981]
39. Brigo F, Igwe SC, Ausserer H, Nardone R, Tezzon F, Bongiovanni LG, et al. Why do people Google epilepsy? An infodemiological study of online behavior for epilepsy-related search terms. Epilepsy Behav 2014 Feb;31:67-70. [doi: 10.1016/j.yebeh.2013.11.020] [Medline: 24361764]
40. Brigo F, Trinka E. Google search behavior for status epilepticus. Epilepsy Behav 2015 Aug;49:146-149. [doi: 10.1016/j.yebeh.2015.02.029] [Medline: 25873438]
41. Murray G, O'Rourke C, Hogan J, Fenton JE. Detecting internet search activity for mouth cancer in Ireland. Br J Oral Maxillofac Surg 2016 Feb;54(2):163-165. [doi: 10.1016/j.bjoms.2015.12.005] [Medline: 26774361]
42. Scheres LJJ, Lijfering WM, Middeldorp S, Cannegieter SC. Influence of World Thrombosis Day on digital information seeking on venous thrombosis: a Google Trends study. J Thromb Haemost 2016 Dec;14(12):2325-2328. [doi: 10.1111/jth.13529] [Medline: 27735128]
43. Bragazzi NL, Dini G, Toletone A, Brigo F, Durando P. Leveraging Big Data for Exploring Occupational Diseases-Related Interest at the Level of Scientific Community, Media Coverage and Novel Data Streams: The Example of Silicosis as a Pilot Study. PLoS One 2016;11(11):e0166051 [FREE Full text] [doi: 10.1371/journal.pone.0166051] [Medline: 27806115]
44. Rosenkrantz AB, Prabhu V. Public Interest in Imaging-Based Cancer Screening Examinations in the United States: Analysis Using a Web-Based Search Tool. AJR Am J Roentgenol 2016 Jan;206(1):113-118. [doi: 10.2214/AJR.15.14840] [Medline: 26700342]
45. Schootman M, Toor A, Cavazos-Rehg P, Jeffe DB, McQueen A, Eberth J, et al. The utility of Google Trends data to examine interest in cancer screening. BMJ Open 2015 Jun 08;5(6):e006678 [FREE Full text] [doi: 10.1136/bmjopen-2014-006678] [Medline: 26056120]
46. Linkov F, Bovbjerg DH, Freese KE, Ramanathan R, Eid GM, Gourash W. Bariatric surgery interest around the world: what Google Trends can teach us. Surg Obes Relat Dis 2014 May;10(3):533-538. [doi: 10.1016/j.soard.2013.10.007] [Medline: 24794184]
47. Stein JD, Childers DM, Nan B, Mian SI. Gauging interest of the general public in laser-assisted in situ keratomileusis eye surgery. Cornea 2013 Jul;32(7):1015-1018 [FREE Full text] [doi: 10.1097/ICO.0b013e318283c85a] [Medline: 23538615]
48. Gafson AR, Giovannoni G. CCSVI-A. A call to clinicans and scientists to vocalise in an Internet age. Mult Scler Relat Disord 2014 Mar;3(2):143-146. [doi: 10.1016/j.msard.2013.10.005] [Medline: 25878001]
49. Phelan N, Kelly JC, Moore DP, Kenny P. The effect of the metal-on-metal hip controversy on Internet search activity. Eur J Orthop Surg Traumatol 2014 Oct;24(7):1203-1210. [doi: 10.1007/s00590-013-1399-3] [Medline: 24390041]
50. Simmering JE, Polgreen LA, Polgreen PM. Web search query volume as a measure of pharmaceutical utilization and changes in prescribing patterns. Res Social Adm Pharm 2014;10(6):896-903. [doi: 10.1016/j.sapharm.2014.01.003] [Medline: 24603135]

51. Zhang Z, Zheng X, Zeng DD, Leischow SJ. Tracking Dabbing Using Search Query Surveillance: A Case Study in the United States. J Med Internet Res 2016 Sep 16;18(9):e252 [FREE Full text] [doi: 10.2196/jmir.5802] [Medline: 27637361]
52. Zheluk A, Quinn C, Meylakhs P. Internet search and krokodil in the Russian Federation: an infoveillance study. J Med Internet Res 2014 Sep 18;16(9):e212 [FREE Full text] [doi: 10.2196/jmir.3203] [Medline: 25236385]
53. Kadry B, Chu LF, Kadry B, Gammas D, Macario A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. J Med Internet Res 2011 Nov;13(4):e95 [FREE Full text] [doi: 10.2196/jmir.1960] [Medline: 22088924]
54. Huesch MD, Currid-Halkett E, Doctor JN. Public hospital quality report awareness: evidence from National and Californian Internet searches and social media mentions, 2012. BMJ Open 2014 Mar 11;4(3):e004417 [FREE Full text] [doi: 10.1136/bmjopen-2013-004417] [Medline: 24618223]
55. Haney NM, Kinsella SD, Morey JM. United States medical school graduate interest in radiology residency programs as depicted by online search tools. J Am Coll Radiol 2014 Feb;11(2):193-197. [doi: 10.1016/j.jacr.2013.06.023] [Medline: 24120904]
56. Poletto C, Boëlle P, Colizza V. Risk of MERS importation and onward transmission: a systematic review and analysis of cases reported to WHO. BMC Infect Dis 2016 Aug 25;16(1):448 [FREE Full text] [doi: 10.1186/s12879-016-1787-5] [Medline: 27562369]
57. Hossain L, Kam D, Kong F, Wigand RT, Bossomaier T. Social media in Ebola outbreak. Epidemiol Infect 2016 Jul;144(10):2136-2143. [doi: 10.1017/S095026881600039X] [Medline: 26939535]
58. Mavragani A, Ochoa G. The Internet and the Anti-Vaccine Movement: Tracking the 2017 EU Measles Outbreak. BDCC 2018 Jan 16;2(1):2. [doi: 10.3390/bdcc2010002]
59. Bentley RA, Ormerod P. Social versus independent interest in 'bird flu' and 'swine flu'. PLoS Curr 2009 Sep 3;1:RRN1036. [doi: 10.1371/currents.RRN1036]
60. Kostkova P, Fowler D, Wiseman S, Weinberg JR. Major infection events over 5 years: how is media coverage influencing online information needs of health care professionals and the public? J Med Internet Res 2013 Jul 15;15(7):e107 [FREE Full text] [doi: 10.2196/jmir.2146] [Medline: 23856364]
61. Pandey A, Abdullah K, Drazner MH. Impact of Vice President Cheney on public interest in left ventricular assist devices and heart transplantation. Am J Cardiol 2014 May 01;113(9):1529-1531. [doi: 10.1016/j.amjcard.2014.02.007] [Medline: 24630787]
62. Brigo F, Lochner P, Tezzon F, Nardone R. Web search behavior for multiple sclerosis: An infodemiological study. Multiple Sclerosis and Related Disorders 2014 Jul;3(4):440-443. [doi: 10.1016/j.msard.2014.02.005]
63. Bragazzi NL, Watad A, Brigo F, Adawi M, Amital H, Shoenfeld Y. Public health awareness of autoimmune diseases after the death of a celebrity. Clin Rheumatol 2016 Dec 20:1911-1917. [doi: 10.1007/s10067-016-3513-5] [Medline: 28000011]
64. Noar S, Ribisl K, Althouse B, Willoughby J, Ayers J. Using digital surveillance to examine the impact of public figure pancreatic cancer announcements on media and search query outcomes. Journal of the National Cancer Institute - Monographs 2013:188-194.
65. Koburger N, Mergl R, Rummel-Kluge C, Ibelshäuser A, Meise U, Postuvan V, et al. Celebrity suicide on the railway network: Can one case trigger international effects? J Affect Disord 2015 Oct 01;185:38-46. [doi: 10.1016/j.jad.2015.06.037] [Medline: 26143403]
66. Seifter A, Schwarzwalder A, Geis K, Aucott J. The utility of "Google Trends" for epidemiological research: Lyme disease as an example. Geospat Health 2010 May;4(2):135-137. [doi: 10.4081/gh.2010.195] [Medline: 20503183]
67. Rossignol L, Pelat C, Lambert B, Flahault A, Chartier-Kastler E, Hanslik T. A method to assess seasonality of urinary tract infections based on medication sales and google trends. PLoS One 2013;8(10):e76020 [FREE Full text] [doi: 10.1371/journal.pone.0076020] [Medline: 24204587]
68. Harsha AK, Schmitt JE, Stavropoulos SW. Know your market: use of online query tools to quantify trends in patient information-seeking behavior for varicose vein treatment. J Vasc Interv Radiol 2014 Jan;25(1):53-57. [doi: 10.1016/j.jvir.2013.09.015] [Medline: 24286941]
69. Ingram DG, Matthews CK, Plante DT. Seasonal trends in sleep-disordered breathing: evidence from Internet search engine query data. Sleep Breath 2015 Mar;19(1):79-84. [doi: 10.1007/s11325-014-0965-1] [Medline: 24595717]
70. Deiner MS, Lietman TM, McLeod SD, Chodosh J, Porco TC. Surveillance Tools Emerging From Search Engines and Social Media Data for Determining Eye Disease Patterns. JAMA Ophthalmol 2016 Sep 01;134(9):1024-1030 [FREE Full text] [doi: 10.1001/jamaophthalmol.2016.2267] [Medline: 27416554]
71. Zhang Z, Zheng X, Zeng DD, Leischow SJ. Information seeking regarding tobacco and lung cancer: effects of seasonality. PLoS One 2015 Mar;10(3):e0117938 [FREE Full text] [doi: 10.1371/journal.pone.0117938] [Medline: 25781020]
72. Kang M, Song W, Choi S, Kim H, Ha H, Kim S, et al. Google unveils a glimpse of allergic rhinitis in the real world. Allergy 2015 Jan;70(1):124-128. [doi: 10.1111/all.12528] [Medline: 25280183]
73. Schuster N, Rogers M, McMahon JL. Using search engine query data to track pharmaceutical utilization: a study of statins. The American journal of managed care 2010;16(8):215-219.

74. Skeldon SC, Kozhimannil KB, Majumdar SR, Law MR. The effect of competing direct-to-consumer advertising campaigns on the use of drugs for benign prostatic hyperplasia: time series analysis. J Gen Intern Med 2015 Apr;30(4):514-520 [FREE Full text] [doi: 10.1007/s11606-014-3063-y] [Medline: 25338730]
75. Gahr M, Uzelac Z, Zeiss R, Connemann BJ, Lang D, Schönfeldt-Lecuona C. Linking Annual Prescription Volume of Antidepressants to Corresponding Web Search Query Data: A Possible Proxy for Medical Prescription Behavior? J Clin Psychopharmacol 2015 Dec;35(6):681-685. [doi: 10.1097/JCP.0000000000000397] [Medline: 26355849]
76. Mavragani A, Sypsa K, Sampri A, Tsagarakis K. Quantifying the UK Online Interest in Substances of the EU Watchlist for Water Monitoring: Diclofenac, Estradiol, and the Macrolide Antibiotics. Water 2016 Nov 18;8(11):542. [doi: 10.3390/w8110542]
77. Kang M, Zhong H, He J, Rutherford S, Yang F. Using Google Trends for influenza surveillance in South China. PLoS One 2013;8(1):e55205 [FREE Full text] [doi: 10.1371/journal.pone.0055205] [Medline: 23372837]
78. Cho S, Sohn CH, Jo MW, Shin S, Lee JH, Ryoo SM, et al. Correlation between national influenza surveillance data and google trends in South Korea. PLoS One 2013 Dec;8(12):e81422 [FREE Full text] [doi: 10.1371/journal.pone.0081422] [Medline: 24339927]
79. Domnich A, Panatto D, Signori A, Lai PL, Gasparini R, Amicizia D. Age-related differences in the accuracy of web query-based predictions of influenza-like illness. PLoS One 2015;10(5):e0127754 [FREE Full text] [doi: 10.1371/journal.pone.0127754] [Medline: 26011418]
80. Solano P, Ustulin M, Pizzorno E, Vichi M, Pompili M, Serafini G, et al. A Google-based approach for monitoring suicide risk. Psychiatry Res 2016 Dec 30;246:581-586. [doi: 10.1016/j.psychres.2016.10.030] [Medline: 27837725]
81. Arora VS, Stuckler D, McKee M. Tracking search engine queries for suicide in the United Kingdom, 2004-2013. Public Health 2016 Aug;137:147-153. [doi: 10.1016/j.puhe.2015.10.015] [Medline: 26976489]
82. Guernier V, Milinovich GJ, Bezerra SMA, Haworth M, Coleman G, Soares MRJ. Use of big data in the surveillance of veterinary diseases: early detection of tick paralysis in companion animals. Parasit Vectors 2016 Dec 23;9(1):303 [FREE Full text] [doi: 10.1186/s13071-016-1590-6] [Medline: 27215214]
83. Barnes CM, Gunia BC, Wagner DT. Sleep and moral awareness. J Sleep Res 2015 Apr;24(2):181-188 [FREE Full text] [doi: 10.1111/jsr.12231] [Medline: 25159702]
84. Johnson AK, Mehta SD. A comparison of Internet search trends and sexually transmitted infection rates using Google trends. Sex Transm Dis 2014 Jan;41(1):61-63. [doi: 10.1097/OLQ.0000000000000065] [Medline: 24326584]
85. Alicino C, Bragazzi NL, Faccio V, Amicizia D, Panatto D, Gasparini R, et al. Assessing Ebola-related web search behaviour: insights and implications from an analytical study of Google Trends-based query volumes. Infect Dis Poverty 2015 Dec 10;4:54 [FREE Full text] [doi: 10.1186/s40249-015-0090-9] [Medline: 26654247]
86. Willson TJ, Lospinoso J, Weitzel E, McMains K. Correlating Regional Aeroallergen Effects on Internet Search Activity. Otolaryngol Head Neck Surg 2014 Dec 12;152(2):228-232. [doi: 10.1177/0194599814560149] [Medline: 25505261]
87. Willson TJ, Shams A, Lospinoso J, Weitzel E, McMains K. Searching for Cedar: Geographic Variation in Single Aeroallergen Shows Dose Response in Internet Search Activity. Otolaryngol Head Neck Surg 2015 Nov 02;153(5):770-774. [doi: 10.1177/0194599815601650] [Medline: 26340925]
88. Zhou X, Ye J, Feng Y. Tuberculosis surveillance by analyzing Google trends. IEEE Trans Biomed Eng 2011 Aug;58(8):2247-2254. [doi: 10.1109/TBME.2011.2132132] [Medline: 21435969]
89. Fond G, Gaman A, Brunel L, Haffen E, Llorca P. Google Trends ® : Ready for real-time suicide prevention or just a Zeta-Jones effect? An exploratory study. Psychiatry Research 2015 Aug;228(3):913-917. [doi: 10.1016/j.psychres.2015.04.022]
90. Gamma A, Schleifer R, Weinmann W, Buadze A, Liebrenz M. Could Google Trends Be Used to Predict Methamphetamine-Related Crime? An Analysis of Search Volume Data in Switzerland, Germany, and Austria. PLoS ONE 2016 Nov 30;11(11):e0166566. [doi: 10.1371/journal.pone.0166566]
91. Pollett S, Wood N, Boscardin WJ, Bengtsson H, Schwarcz S, Harriman K, et al. Validating the Use of Google Trends to Enhance Pertussis Surveillance in California. PLoS Curr 2015 Oct 19:1-10. [doi: 10.1371/currents.outbreaks.7119696b3e7523faa4543faac87c56c2]
92. Wang H, Chen D, Yu H, Chen Y. Forecasting the Incidence of Dementia and Dementia-Related Outpatient Visits With Google Trends: Evidence From Taiwan. J Med Internet Res 2015 Nov 19;17(11):e264 [FREE Full text] [doi: 10.2196/jmir.4516] [Medline: 26586281]
93. Crowson MG, Schulz K, Tucci DL. National Utilization and Forecasting of Ototopical Antibiotics. Otology & Neurotology 2016;37(8):1049-1054. [doi: 10.1097/MAO.0000000000001115]
94. Scopus. URL: https://www.scopus.com/home.uri [accessed 2017-11-08] [WebCite Cache ID 6uotA2G5x]
95. PubMed. accessed 18/4/2018. URL: https://www.ncbi.nlm.nih.gov/pubmed/ [accessed 2018-09-07] [WebCite Cache ID 72Fb7LRKh]
96. Bakker KM, Martinez-Bakker ME, Helm B, Stevenson TJ. Digital epidemiology reveals global childhood disease seasonality and the effects of immunization. PNAS 2016;113(24):6689 [FREE Full text] [doi: 10.1073/pnas.1523941113] [Medline: 27247405]

97. Borron SW, Watts SH, Tull J, Baeza S, Diebold S, Barrow A. Intentional Misuse and Abuse of Loperamide: A New Look at a Drug with "Low Abuse Potential". J Emerg Med 2017 Jul;53(1):73-84. [doi: 10.1016/j.jemermed.2017.03.018] [Medline: 28501383]
98. Bragazzi NL. Infodemiology and infoveillance of multiple sclerosis in Italy. Mult Scler Int 2013;2013:924029 [FREE Full text] [doi: 10.1155/2013/924029] [Medline: 24027636]
99. Bragazzi NL, Dini G, Toletone A, Brigo F, Durando P. Infodemiological data concerning silicosis in the USA in the period 2004-2010 correlating with real-world statistical data. Data Brief 2017 Feb;10:457-464 [FREE Full text] [doi: 10.1016/j.dib.2016.11.021] [Medline: 28054008]
100. Bragazzi NL, Barberis I, Rosselli R, Gianfredi V, Nucci D, Moretti M, et al. How often people google for vaccination: Qualitative and quantitative insights from a systematic search of the web-based activities using Google Trends. Hum Vaccin Immunother 2017 Feb;13(2):464-469. [doi: 10.1080/21645515.2017.1264742] [Medline: 27983896]
101. Bragazzi N, Bacigaluppi S, Robba C, Siri A, Canepa G, Brigo F. Infodemiological data of West-Nile virus disease in Italy in the study period 2004-2015. Data Brief 2016:839-845 [FREE Full text]
102. Bragazzi NL, Bacigaluppi S, Robba C, Nardone R, Trinka E, Brigo F. Infodemiology of status epilepticus: A systematic validation of the Google Trends-based search queries. Epilepsy Behav 2016 Feb;55:120-123. [doi: 10.1016/j.yebeh.2015.12.017] [Medline: 26773681]
103. Bragazzi NL. A Google Trends-based approach for monitoring NSSI. Psychol Res Behav Manag 2013 Dec;7:1-8 [FREE Full text] [doi: 10.2147/PRBM.S44084] [Medline: 24376364]
104. Braun T, Harréus U. Medical nowcasting using Google Trends: application in otolaryngology. Eur Arch Otorhinolaryngol 2013 Jul;270(7):2157-2160. [doi: 10.1007/s00405-013-2532-y] [Medline: 23632877]
105. van Campen JS, van Diessen E, Otte WM, Joels M, Jansen FE, Braun KPJ. Does Saint Nicholas provoke seizures? Hints from Google Trends. Epilepsy Behav 2014 Mar;32:132-134. [doi: 10.1016/j.yebeh.2014.01.019] [Medline: 24548849]
106. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. Clin Infect Dis 2009 Nov 15;49(10):1557-1564 [FREE Full text] [doi: 10.1086/630200] [Medline: 19845471]
107. Cavazos-Rehg PA, Krauss MJ, Spitznagel EL, Lowery A, Grucza RA, Chaloupka FJ, et al. Monitoring of non-cigarette tobacco use using Google Trends. Tob Control 2015 May;24(3):249-255 [FREE Full text] [doi: 10.1136/tobaccocontrol-2013-051276] [Medline: 24500269]
108. Cha Y, Stow CA. Mining web-based data to assess public response to environmental events. Environ Pollut 2015 Mar;198:97-99. [doi: 10.1016/j.envpol.2014.12.027] [Medline: 25577650]
109. Chaves JN, Libardi AL, Agostinho-Pesse RS, Morettin M, Alvarenga KDF. Tele-health: assessment of websites on newborn hearing screening in Portuguese Language. Codas 2015 Dec;27(6):526-533 [FREE Full text] [doi: 10.1590/2317-1782/20152014169] [Medline: 26691616]
110. Davis NF, Gnanappiragasam S, Thornhill JA. Interstitial cystitis/painful bladder syndrome: the influence of modern diagnostic criteria on epidemiology and on Internet search activity by the public. Transl Androl Urol 2015 Oct;4(5):506-511 [FREE Full text] [doi: 10.3978/j.issn.2223-4683.2015.06.08] [Medline: 26816850]
111. Fazeli DS, Carlos RC, Hall KS, Dalton VK. Novel data sources for women's health research: mapping breast screening online information seeking through Google trends. Acad Radiol 2014 Sep;21(9):1172-1176 [FREE Full text] [doi: 10.1016/j.acra.2014.05.005] [Medline: 24998689]
112. DeVilbiss E, Lee B. Brief Report: Trends in U.S. National Autism Awareness from 2004 to 2014: The Impact of National Autism Awareness Month. Journal of Autism and Developmental Disorders 2014;44(12):3271-3273. [doi: 10.1007/s10803-014-2160-4] [Medline: 24915931]
113. El-Sheikha J. Global search demand for varicose vein information on the internet. Phlebology 2015 Sep;30(8):533-540. [doi: 10.1177/0268355514542681] [Medline: 24993972]
114. Fenichel EP, Kuminoff NV, Chowell G. Skip the trip: air travelers' behavioral responses to pandemic influenza. PLoS One 2013 Mar;8(3):e58249 [FREE Full text] [doi: 10.1371/journal.pone.0058249] [Medline: 23526970]
115. Foroughi F, Lam AK, Lim MS, Saremi N, Ahmadvand A. “Googling” for Cancer: An Infodemiological Assessment of Online Search Interests in Australia, Canada, New Zealand, the United Kingdom, and the United States. JMIR Cancer 2016 May 04;2(1):e5 [FREE Full text] [doi: 10.2196/cancer.5212] [Medline: 28410185]
116. Garrison SR, Dormuth CR, Morrow RL, Carney GA, Khan KM. Seasonal effects on the occurrence of nocturnal leg cramps: a prospective cohort study. CMAJ 2015 Mar 03;187(4):248-253 [FREE Full text] [doi: 10.1503/cmaj.140497] [Medline: 25623650]
117. Gollust SE, Qin X, Wilcock AD, Baum LM, Barry CL, Niederdeppe J, et al. Search and You Shall Find: Geographic Characteristics Associated With Google Searches During the Affordable Care Act's First Enrollment Period. Med Care Res Rev 2017 Dec;74(6):723-735. [doi: 10.1177/1077558716660944] [Medline: 27457426]
118. Harorli OT, Harorli H. Evaluation of internet search trends of some common oral problems, 2004 to 2014. Community Dental Health 2014;31(3):188-192. [doi: 10.1922/CDH_3330Harorl?05]
119. Harsha AK, Schmitt JE, Stavropoulos SW. Match day: online search trends reflect growing interest in IR training. J Vasc Interv Radiol 2015 Jan;26(1):95-100. [doi: 10.1016/j.jvir.2014.09.011] [Medline: 25541447]

120. Hassid BG, Day LW, Awad MA, Sewell JL, Osterberg EC, Breyer BN. Using Search Engine Query Data to Explore the Epidemiology of Common Gastrointestinal Symptoms. Dig Dis Sci 2017 Dec;62(3):588-592. [doi: 10.1007/s10620-016-4384-y] [Medline: 27878646]
121. Huang J, Zheng R, Emery S. Assessing the impact of the national smoking ban in indoor public places in china: evidence from quit smoking related online searches. PLoS One 2013 Jun;8(6):e65577 [FREE Full text] [doi: 10.1371/journal.pone.0065577] [Medline: 23776504]
122. Ingram DG, Plante DT. Seasonal trends in restless legs symptomatology: evidence from Internet search query data. Sleep Med 2013 Dec;14(12):1364-1368. [doi: 10.1016/j.sleep.2013.06.016] [Medline: 24152798]
123. Jha S, Wang Z, Laucis N, Bhattacharyya T. Trends in Media Reports, Oral Bisphosphonate Prescriptions, and Hip Fractures 1996-2012: An Ecological Analysis. J Bone Miner Res 2015 Dec;30(12):2179-2187 [FREE Full text] [doi: 10.1002/jbmr.2565] [Medline: 26018247]
124. Lawson MAC, Lawson MA, Kalff R, Walter J. Google Search Queries About Neurosurgical Topics: Are They a Suitable Guide for Neurosurgeons? World Neurosurg 2016 Jun;90:179-185. [doi: 10.1016/j.wneu.2016.02.045] [Medline: 26898496]
125. Leffler CT, Davenport B, Chan D. Frequency and seasonal variation of ophthalmology-related internet searches. Can J Ophthalmol 2010 Jun;45(3):274-279. [doi: 10.3129/i10-022] [Medline: 20436544]
126. Ling R, Lee J. Disease Monitoring and Health Campaign Evaluation Using Google Search Activities for HIV and AIDS, Stroke, Colorectal Cancer, and Marijuana Use in Canada: A Retrospective Observational Study. JMIR Public Health Surveill 2016 Oct 12;2(2):e156 [FREE Full text] [doi: 10.2196/publichealth.6504] [Medline: 27733330]
127. Liu F, Allan GM, Korownyk C, Kolber M, Flook N, Sternberg H, et al. Seasonality of Ankle Swelling: Population Symptom Reporting Using Google Trends. Ann Fam Med 2016 Dec;14(4):356-358 [FREE Full text] [doi: 10.1370/afm.1953] [Medline: 27401424]
128. Luckett T, Disler R, Hosie A, Johnson M, Davidson P, Currow D, et al. Content and quality of websites supporting self-management of chronic breathlessness in advanced illness: a systematic review. NPJ Prim Care Respir Med 2016 Dec 26;26:16025 [FREE Full text] [doi: 10.1038/npjpcrm.2016.25] [Medline: 27225898]
129. Majumder MS, Santillana M, Mekaru SR, McGinnis DP, Khan K, Brownstein JS. Utilizing Nontraditional Data Sources for Near Real-Time Estimation of Transmission Dynamics During the 2015-2016 Colombian Zika Virus Disease Outbreak. JMIR Public Health Surveill 2016 Jun 01;2(1):e30 [FREE Full text] [doi: 10.2196/publichealth.5814] [Medline: 27251981]
130. Mattin MJ, Solano-Gallego L, Dhollander S, Afonso A, Brodbelt DC. The frequency and distribution of canine leishmaniosis diagnosed by veterinary practitioners in Europe. Vet J 2014 Jun;200(3):410-419. [doi: 10.1016/j.tvjl.2014.03.033] [Medline: 24767097]
131. Myers L, Jones J, Boesten N, Lancman M. Psychogenic non-epileptic seizures (PNES) on the Internet: Online representation of the disorder and frequency of search terms. Seizure 2016 Aug 01;40:114-122 [FREE Full text]
132. Parker J, Cuthbertson C, Loveridge S, Skidmore M, Dyar W. Forecasting state-level premature deaths from alcohol, drugs, and suicides using Google Trends data. J Affect Disord 2017 Dec 15;213:9-15. [doi: 10.1016/j.jad.2016.10.038] [Medline: 28171770]
133. Phelan N, Davy S, O'Keeffe GW, Barry DS. Googling in anatomy education: Can google trends inform educators of national online search patterns of anatomical syllabi? Anat Sci Educ 2017 Mar;10(2):152-159. [doi: 10.1002/ase.1641] [Medline: 27547967]
134. Plante DT, Ingram DG. Seasonal trends in tinnitus symptomatology: evidence from Internet search engine query data. Eur Arch Otorhinolaryngol 2015 Oct;272(10):2807-2813. [doi: 10.1007/s00405-014-3287-9] [Medline: 25234771]
135. Rohart F, Milinovich GJ, Avril SMR, Lê Cao KA, Tong S, Hu W. Disease surveillance based on Internet-based linear models: an Australian case study of previously unmodeled infection diseases. Sci Rep 2016 Dec 20;6:38522 [FREE Full text] [doi: 10.1038/srep38522] [Medline: 27994231]
136. Scatà M, Di Stefano A, Liò P, La Corte A. The Impact of Heterogeneity and Awareness in Modeling Epidemic Spreading on Multiplex Networks. Sci Rep 2016 Dec 16;6:37105 [FREE Full text] [doi: 10.1038/srep37105] [Medline: 27848978]
137. Shin S, Seo D, An J, Kwak H, Kim S, Gwack J, et al. High correlation of Middle East respiratory syndrome spread with Google search and Twitter trends in Korea. Sci Rep 2016 Sep 06;6:32920 [FREE Full text] [doi: 10.1038/srep32920] [Medline: 27595921]
138. Sentana-Lledo D, Barbu CM, Ngo MN, Wu Y, Sethuraman K, Levy MZ. Seasons, Searches, and Intentions: What The Internet Can Tell Us About The Bed Bug (Hemiptera: Cimicidae) Epidemic. J Med Entomol 2016 Jan;53(1):116-121. [doi: 10.1093/jme/tjv158] [Medline: 26474879]
139. Takada K. Japanese Interest in "Hotaru" (Fireflies) and "Kabuto-Mushi" (Japanese Rhinoceros Beetles) Corresponds with Seasonality in Visible Abundance. Insects 2012 Apr 10;3(2):424-431 [FREE Full text] [doi: 10.3390/insects3020424] [Medline: 26466535]
140. Telfer S, Woodburn J. Let me Google that for you: a time series analysis of seasonality in internet search trends for terms related to foot and ankle pain. J Foot Ankle Res 2015 Jul;8:27 [FREE Full text] [doi: 10.1186/s13047-015-0074-9] [Medline: 26146521]

141. Troelstra SA, Bosdriesz JR, de Boer MR, Kunst AE. Effect of Tobacco Control Policies on Information Seeking for Smoking Cessation in the Netherlands: A Google Trends Study. PLoS One 2016 Feb;11(2):e0148489 [FREE Full text] [doi: 10.1371/journal.pone.0148489] [Medline: 26849567]
142. Toosi B, Kalia S. Seasonal and Geographic Patterns in Tanning Using Real-Time Data From Google Trends. JAMA Dermatol 2016 Feb;152(2):215-217. [doi: 10.1001/jamadermatol.2015.3008] [Medline: 26719968]
143. Warren KE, Wen LS. Measles, social media and surveillance in Baltimore City. J Public Health (Oxf) 2017 Sep 01;39(3):e73-e78. [doi: 10.1093/pubmed/fdw076] [Medline: 27521926]
144. Yang S, Santillana M, Kou SC. Accurate estimation of influenza epidemics using Google search data via ARGO. PNAS 2015;112(47):14473 [FREE Full text] [doi: 10.1073/pnas.1515373112] [Medline: 26553980]
145. Köhler MJ, Springer S, Kaatz M. [On the seasonality of dermatoses: a retrospective analysis of search engine query data depending on the season]. Hautarzt 2014 Sep 14;65(9):814-822. [doi: 10.1007/s00105-014-2848-6] [Medline: 25234631]
146. Orellano PW, Reynoso JI, Antman J, Argibay O. Uso de la herramienta Google Trends para estimar la incidencia de enfermedades tipo influenza en Argentina. Cad. Saúde Pública 2015 Apr;31(4):691-700. [doi: 10.1590/0102-311X00072814]
147. Cjuno J, Taype-Rondan A. Estacionalidad de la cefalea en el hemisferio norte y el hemisferio sur: una aproximación utilizando Google Trends. Rev. méd. Chile 2016 Jul;144(7):947-947. [doi: 10.4067/S0034-98872016000700019]
148. Tejada-Llacsa P. Gaceta Sanitaria. 2016. ¿Qué se busca sobre el aborto en Internet? Una evaluación con Google Trends en Perú URL: https://linkinghub.elsevier.com/retrieve/pii/S0213911116300486 [accessed 2018-09-07] [WebCite Cache ID 72FbxshJz]
149. Yang Y, Zeng Q, Zhao H, Yi J, Li Q, Xia Y. Hepatitis B prediction model based on Google trends. Journal of Shanghai Jiaotong University (Medical Science) 2013;33(2):204-208. [doi: 10.3969/j.issn.1674-8115.2013.02.016]

Abbreviations

ANOVA: analysis of variance
ARIMA: autoregressive integrated moving average.
MS: multiple sclerosis

Edited by G Eysenbach; submitted 08.11.17; peer-reviewed by A Benis, J Bian, C Fincham; comments to author 15.03.18; revised version received 07.05.18; accepted 21.06.18; published 06.11.18.

Please cite as:
Mavragani A, Ochoa G, Tsagarakis KP
Assessing the Methods, Tools, and Statistical Approaches in Google Trends Research: Systematic Review
J Med Internet Res 2018;20(11):e270
URL: https://www.jmir.org/2018/11/e270/
doi:10.2196/jmir.9366
PMID:30401664

Original Paper

Hookah-Related Posts to Twitter From 2017 to 2018: Thematic Analysis

Jon-Patrick Allem[1], PhD, MA; Likhit Dharmapuri[2], BTech; Adam M Leventhal[1], PhD; Jennifer B Unger[1], PhD; Tess Boley Cruz[1], PhD

[1]Keck School of Medicine of University of Southern California, Los Angeles, CA, United States

[2]Department of Computer Science, University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Jon-Patrick Allem, PhD, MA
Keck School of Medicine of University of Southern California
2001 North Soto Street
Los Angeles, CA, 90032
United States
Phone: 1 8586030812
Email: allem@usc.edu

Abstract

Background: Hookah (or tobacco waterpipe) use has recently become prevalent in the United States. The contexts and experiences associated with hookah use are unclear, yet such information is abundant via publicly available hookah users' social media postings.

Objective: In this study, we utilized Twitter data to characterize Twitter users' recent experiences with hookah.

Methods: Twitter posts containing the term "hookah" were obtained from April 1, 2017 to 29 March, 2018. Text classifiers were used to identify clusters of topics that tended to co-occur in posts (n=176,706).

Results: The most prevalent topic cluster was Person Tagging (use of @username to tag another Twitter account in a post) at 21.58% (38,137/176,706) followed by Promotional or Social Events (eg, mentions of ladies' nights, parties, etc) at 20.20% (35,701/176,706) and Appeal or Abuse Liability (eg, craving, enjoying hookah) at 18.12% (32,013/176,706). Additional topics included Hookah Use Behavior (eg, mentions of taking a "hit" of hookah) at 11.67% (20,603/176,706), Polysubstance Use (eg, hookah use along with other substances) at 10.95% (19,353/176,706), Buying or Selling (eg, buy, order, purchase, sell) at 9.37% (16,552/176,706), and Flavors (eg, mint, cinnamon, watermelon) at 1.66% (2927/176,706). The topic Dislike of Hookah (eg, hate, quit, dislike) was rare at 0.59% (1043/176,706).

Conclusions: Social events, appeal or abuse liability, flavors, and polysubstance use were the common contexts and experiences associated with Twitter discussions about hookah in 2017-2018. Considered in concert with traditional data sources about hookah, these results suggest that social events, appeal or abuse liability, flavors, and polysubstance use warrant consideration as targets in future surveillance, policy making, and interventions addressing hookah.

(J Med Internet Res 2018;20(11):e11669) doi:10.2196/11669

KEYWORDS

hookah; waterpipe; Twitter; social media; nicotine; flavors; social smoking; infodemiology

Introduction

Hookah (or tobacco waterpipe) use has recently grown in popularity in the United States, especially among youth and young adults [1,2]. While exposure to hookah smoke has similar health risks to that of combustible cigarettes [3,4], it is perceived as safer than cigarettes in certain vulnerable groups [5] and is subject to fewer regulations [6]. For example, hookah is offered in many flavors, whereas flavored cigarettes are banned in the United States.

Publicly accessible data from individuals who post information on social media websites (eg, Twitter, Instagram, YouTube) can be efficiently harnessed to quickly capture and describe the context of tobacco use [7-9]. Previous analyses of hookah-related posts to social media websites through the year 2017 provide some information about hookah-related contexts, including the importance and appreciation of stylized waterpipes

[10,11], use of hookah in social settings [10], copromotion with alcohol [10], and primarily positive user experiences [12-16]. However, cultural trends, the tobacco product consumer marketplace, and tobacco product health policies are evolving constantly and rapidly. The contexts and experiences associated with hookah use rapidly change as well, making it important to provide up-to-date information on such issues to inform targets for surveillance, policy making, and interventions addressing hookah.

In this study, we demonstrate the utility of collecting data from Twitter to document and describe hookah-related conversations from 2017 to 2018. Our goal was to determine the public's recent experiences with hookah including understanding the social and environmental contexts in which hookah use occurs. Twitter is used by 24% of US adults (23% of men, 24% of women, 24% of white individuals, 26% of African American individuals, and 20% of Hispanic individuals), with 46% of users on the platform daily [17]. Findings from this study should inform tobacco control policy and prevention efforts and demonstrate the utility in using Twitter data for rapid surveillance of health behaviors and tobacco-related products like hookah.

Methods

Data Collection

Twitter posts containing the term "hookah" (or "#hookah") were obtained from Twitter's Streaming Application Program Interface (API; the filtered stream using the Twitter4J library for collecting tweets with no gaps in the collection time) from April 1, 2017 to 29 March, 2018. There were a total of 963,954 posts during this time.

Data Processing

We removed retweets and non-English posts, resulting in 348,834 unique posts that were used for analysis. While the word waterpipe is used in academic papers and presentations to refer to hookah, it is uncommon for individuals to use this term on social media, and it was, therefore, not included in this study [18]. To clean the data, we removed tweets from accounts identified as social bots [19,20] using Botometer (also known as Bot or Not) [21], resulting in a final analytical sample of 176,706 tweets from 90,718 unique users.

The final sample was prepared for analysis, which included the process of basic normalization (eg, remove punctuation, lowercase all text), stop word removal (eg, the words "a" and "the"), normalization of Twitter user mentions (eg, "@janedoe" is converted to "@username"), lemmatization (eg, "cat," "cats," "cat's," are all converted to "cat"), and nonprintable character removal (eg, emojis) [13]. All analyses relied on public, anonymized data; adhered to the terms and conditions, terms of use, and privacy policies of Twitter; and were performed under Institutional Review Board approval from the authors' university. To protect privacy, no tweets were reported verbatim in this report.

Topic Identification Methodology

Initially, we analyzed the tweets using word frequencies (of single words and double-word combinations, also known as one grams and bigrams) and visualized the data through word clouds to identify common topics (Multimedia Appendix 1). From this assessment, the authors came to an expert consensus on several topics including *Person Tagging* (eg, the use of @username to tag another Twitter account in a post), *Buying or Selling* (eg, words indicative of buying, selling, or purchasing hookah), *Appeal* or *Abuse Liability* (eg, words indicative of craving, wanting, needing, enjoying, and loving hookah), *Hookah Use Behavior* (eg, mentions of taking a "hit" of hookah or smoking hookah), *Promotional or Social Events* (eg, mentions of ladies' nights, parties, etc), *Polysubstance Use* (eg, words indicative of alcohol, marijuana, or other substance use along with hookah), and *Flavors* (eg, use of the words "cinnamon," "blueberry," and "watermelon"; Textbox 1). In line with prior research [22,23], we looked for words and phrases that suggested *Dislike of Hookah* (eg, "don't hookah" and "quit hookah").

Next, we used Word2Vec, a language modeling technique developed by Google that allows users to learn text representations for creating text classifiers [24]. Word2Vec creates embeddings (eg, numerical representations of words that help capture meaning, semantic relationships, and context) for text by using each word in a corpus to predict the words that usually surround it. In other words, Word2Vec creates word embeddings where semantic relationships between words are preserved. One advantage of this technique is that words that are synonyms will have similar embeddings, whereas words that are antonyms will have dissimilar embeddings. Similarly, in the Word2Vec representations of words, the relationship between "king" and "queen" is equal to the relationship between "man" and "woman."

We used Word2Vec to find similar words for the one grams and bigrams that we identified per topic in the word cloud stage. This process, along with visual inspection and manual edits, allowed us to expand our word list per topic by identifying words that appeared, in posts, in a similar context as our original keywords. For example, through this process we found that the words "crave," "love," "enjoy," and "need" appeared in posts that were similar to posts that contained the words "want" and "hookah".

Classification was done by checking for the presence of any one of the keywords (one grams and bigrams) in a tweet. If a tweet consisted of any of the keywords associated with a topic, the tweet was classified as part of that topic. In other words, we used a rule-based classification script written in Python where each tweet was checked for the presence of a specified set of n-grams representing a theme. For each analysis, we present findings in a confusion matrix where the diagonal line indicates the prevalence of a topic and the off-diagonal lines indicate topic overlap. For example, a hypothetical post such as "I'm craving hookah and a beer right now" could be classified under *Appeal or Abuse Liability* and *Polysubstance Use*. The number of posts containing both contents would be found at the intersection of the matrix for these 2 topics or at 2.14% (3824/176,706).

Textbox 1. Themes and common words found in posts along with the word “hookah”; these words are meant to provide further context for each theme, are not exhaustive, and are listed in alphabetical order.

Person tagging

- @username

Promotional events

- Bar
- Food
- Friday
- Lounge
- Night
- Party
- Saturday

Appeal or abuse liability

- Crave
- Enjoy
- Everyday
- Get
- Like
- Love
- Need
- Want

Hookah use behavior

- Hit
- Pass
- Puff
- Smoke
- Used

Polysubstance use

- Alcohol
- Beer
- Blunt
- Cigs
- Cocktails
- Drinks
- JUUL
- Liquor
- Margaritas
- Vodka
- Weed
- Wine
- Vape

Buying or Selling

- Bought Buy
- Order
- Paying
- Purchase
- Sell

Flavors

- Flavors
- Mint
- Cinnamon
- Watermelon
- Blueberry
- Guava
- Grape
- Apple
- Fruit
- Peach
- Orange
- Mango
- Candy

Results

The total coverage of the 8 topics that we identified constituted 65.45% (115,658/176,706) of all tweets in the corpus of tweets (Figure 1). The remaining 34.59% (61,048/176,706) of tweets were too varied to be classified into a single topic with meaningful coverage (coverage of each subsequent topic was less than 1% of the total tweets). The most prevalent topic was *Person Tagging* at 21.58% (38,137/176,706), followed by *Promotional or Social Events* at 20.20% (35,701/176,706), *Appeal or Abuse Liability* at 18.12% (32,013/176,706), and *Hookah Use Behavior* at 11.67% (20,603/176,706).

Figure 1. Prevalence of topics.

Person Tagging	**38,137 (21.58%)**							
Promotional Events	7666 (4.34%)	**35,701 (20.20%)**						
Appeal or Abuse Liability	5186 (2.93%)	6191 (3.50%)	**31,614 (17.89%)**					
Hookah Use	3621 (2.05%)	944 (0.53%)	5306 (3.00%)	**20,949 (11.86%)**				
Polysubstance Use	3430 (1.94%)	2824 (1.60%)	3778 (2.14%)	3252 (1.84%)	**19,353 (10.95%)**			
Buying or Selling	3250 (1.84%)	2440 (1.38%)	7257 (4.11%)	1421 (0.80%)	1809 (1.02%)	**16,552 (9.37%)**		
Flavor	538 (0.30%)	194 (0.11%)	507 (0.29%)	223 (0.13%)	196 (0.11%)	443 (0.25%)	**2927 (1.66%)**	
Dislike of Hookah	202 (0.11%)	75 (0.04%)	77 (0.04%)	440 (0.25%)	158 (0.09%)	42 (0.02%)	6 (0.00%)	**1043 (0.59%)**
	Person Tagging	**Promotional Events**	**Appeal or Abuse Liability**	**Hookah Use**	**Polysubstance Use**	**Buying or Selling**	**Flavor**	**Dislike of Hookah**

About 10.95% (19,353/176,706) of the corpus was *Polysubstance Use*, while *Buying or Selling* comprised 9.37% (16,552/176,706) and *Flavors* comprised 1.66% (2927/176,706) of the tweets. The least common topic was *Dislike of Hookah* at 0.59% (1043/176,706). The most common topic overlap was between *Person Tagging* and *Promotional or Social Events* at 4.34% (7666/176,706), followed by *Buying or Selling* and *Appeal or Abuse Liability* at 4.12% (7276/176,706) and *Promotional or Social Events* and *Appeal or Abuse Liability* at 3.52% (6225/176,706).

Discussion

Principal Findings

The topics identified in this study of hookah-related posts to Twitter from 2017 to 2018 provide several insights about the public's recent experience with hookah. The most prevalent topic was *Person Tagging* or an individual Twitter user directly communicating to another user (a follower or friend) about hookah, while the most common topic overlap was *Person Tagging* and *Promotional or Social Events*. These findings demonstrate that Twitter users communicate shared values around, and experiences with, hookah. In other words, such posts may notify others about hookah-related events and align people into a community around hookah. Similarly, recent research characterizing JUUL-related posts to Twitter found instances of *Person Tagging* where posts suggested that people were notifying their friends of when they were using or purchasing JUUL-related products [22]. Collectively, these interpersonal communications suggest that people bond around tobacco-related products on Twitter and that there may be co-use of tobacco among many people or social influences in which one person motivates another to use tobacco.

Hookah Use Behavior and *Polysubstance Use* were identified as topics of discussion and may represent a syndrome of risky behavior among select Twitter users. These findings are in line with earlier research on hookah posts to Tumblr [18] and Instagram [10] as well as survey-based research that demonstrated that those who use hookah were significantly more likely to use other substances including alcohol, cigarettes, marijuana, and cocaine compared with those who refrained from hookah use [25]. Individuals who combine the use of hookah with other substances may be at risk for substance misuse; for example, hookah use facilitates greater intake of alcohol and vice versa [26].

Posts in this study reflected Twitter users' interest in flavors, which is similar to earlier research on tobacco-related post to Twitter [22,27]. A recent study identified that flavors were a common reason for hookah use among a nationally representative sample of young adults (aged 18-24 years) [28]. Research has also documented that flavored tobacco products like hookah are perceived to be less harmful than cigarettes [29]. Restricting flavors, such as those identified in this study (Cinnamon, Watermelon, Blueberry, etc), to reduce the appeal of hookah may be a policy consideration to explore in the future.

Many posts found in this study reflected that Twitter users craved, enjoyed, or wanted hookah; this finding, when coupled with the finding that posts indicative of disliking hookah were rare, suggests that there is a current need for targeted interventions to discourage the appeal of hookah use. The common discussions about hookah's appeal may help normalize hookah use on Twitter, which may have consequences for offline behaviors [30].

Limitations

This study focused on posts to Twitter, and findings may not generalize to other social media platforms. The posts analyzed in this study were collected from a 12-month period and may not generalize to other time periods. While only one root word "hookah" (or "#hookah") was used in data collection, research has indicated that this is the common term to refer to waterpipe use on social media [10,13,18]. Data collection relied on Twitter's Streaming API, which prevented collection of tweets from private Twitter accounts. As a result, findings may not represent the attitudes and behaviors of individuals with private accounts.

Conclusion

Social events, appeal or abuse liability, flavors, and polysubstance use were common contexts and experiences associated with Twitter discussions about hookah in 2017-2018. Considered in concert with traditional data sources about hookah, these results suggest that social events, appeal or abuse liability, flavors, and polysubstance use warrant consideration as targets in future surveillance, public policy, and interventions addressing hookah. This study also highlights a clear benefit of using social media data in public health surveillance. Data from social media can serve as an ongoing system to inform public health researchers about tobacco products or ways in which these products are used by the public in near real time.

Acknowledgments

Research reported in this publication was supported by Grant #P50CA180905 from the National Cancer Institute and the Food and Drug Administration (FDA) Center for Tobacco Products. The National Institutes of Health (NIH) or FDA had no role in study design, collection, analysis, and interpretation of data; writing the report; and the decision to submit the report for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or FDA.

Authors' Contributions

JPA and LD conceived the study and analyzed the data. JPA drafted the initial manuscript. LD, AML, TBC, and JBU revised the manuscript for important intellectual content and approved the final manuscript. JBU and TBC received funding for the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Hookah word clouds.

[PDF File (Adobe PDF File), 2MB - jmir_v20i11e11669_app1.pdf]

References

1. Salloum RG, Asfar T, Maziak W. Toward a Regulatory Framework for the Waterpipe. Am J Public Health 2016 Dec;106(10):1773-1777. [doi: 10.2105/AJPH.2016.303322] [Medline: 27552262]
2. Allem JP, Unger JB. Emerging adulthood themes and hookah use among college students in Southern California. Addict Behav 2016 Dec;61:16-19 [FREE Full text] [doi: 10.1016/j.addbeh.2016.05.002] [Medline: 27208879]
3. El-Zaatari ZM, Chami HA, Zaatari GS. Health effects associated with waterpipe smoking. Tob Control 2015 Mar;24 Suppl 1:i31-i43 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051908] [Medline: 25661414]
4. Maziak W. The global epidemic of waterpipe smoking. Addict Behav 2011;36(1-2):1-5 [FREE Full text] [doi: 10.1016/j.addbeh.2010.08.030] [Medline: 20888700]
5. Mohammed KA, Geneus CJ, Yadgir S, Subramaniam DS, Burroughs TE. Correlates of Hookah Pipe Awareness and Perceived Harmfulness Among U.S. Adults. Am J Prev Med 2017 Apr;52(4):513-518. [doi: 10.1016/j.amepre.2016.10.032] [Medline: 27989450]
6. Primack BA, Hopkins M, Hallett C, Carroll MV, Zeller M, Dachille K, et al. US health policy related to hookah tobacco smoking. Am J Public Health 2012 Sep;102(9):e47-e51 [FREE Full text] [doi: 10.2105/AJPH.2012.300838] [Medline: 22827447]
7. Ayers JW, Leas EC, Allem JP, Benton A, Dredze M, Althouse BM, et al. Why do people use electronic nicotine delivery systems (electronic cigarettes)? A content analysis of Twitter, 2012-2015. PLoS One 2017;12(3):e0170702 [FREE Full text] [doi: 10.1371/journal.pone.0170702] [Medline: 28248987]
8. Allem JP, Escobedo P, Chu K, Boley CT, Unger JB. Images of Little Cigars and Cigarillos on Instagram Identified by the Hashtag #swisher: Thematic Analysis. J Med Internet Res 2017 Jul 14;19(7):e255 [FREE Full text] [doi: 10.2196/jmir.7634] [Medline: 28710057]
9. Allem JP, Escobedo P, Cruz TB, Unger JB. Vape pen product placement in popular music videos. Addict Behav 2017 Nov 03. [doi: 10.1016/j.addbeh.2017.10.027] [Medline: 29128149]
10. Allem JP, Chu K, Cruz TB, Unger JB. Waterpipe Promotion and Use on Instagram: #Hookah. Nicotine Tob Res 2017 Oct 01;19(10):1248-1252 [FREE Full text] [doi: 10.1093/ntr/ntw329] [Medline: 28077449]
11. Guidry J, Jin Y, Haddad L, Zhang Y, Smith J. How Health Risks Are Pinpointed (or Not) on Social Media: The Portrayal of Waterpipe Smoking on Pinterest. Health Commun 2016;31(6):659-667. [doi: 10.1080/10410236.2014.987468] [Medline: 26512916]
12. Ben Taleb Z, Laestadius LI, Asfar T, Primack BA, Maziak W. #Hookahlife: The Rise of Waterpipe Promotion on Instagram. Health Educ Behav 2018 Jun 01:1090198118779131. [doi: 10.1177/1090198118779131] [Medline: 29952224]
13. Allem JP, Ramanujam J, Lerman K, Chu K, Boley CT, Unger JB. Identifying Sentiment of Hookah-Related Posts on Twitter. JMIR Public Health Surveill 2017 Oct 18;3(4):e74 [FREE Full text] [doi: 10.2196/publichealth.8133] [Medline: 29046267]
14. Chen A, Zhu S, Conway M. Combining Text Mining and Data Visualization Techniques to UnderstandConsumer Experiences of Electronic Cigarettes and Hookah in OnlineForums. Online J Public Health Inform 2015;7(1):e117. [doi: 10.5210/ojphi.v7i1.5783]
15. 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]
16. Myslín M, Zhu S, Chapman W, Conway M. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. J Med Internet Res 2013 Aug 29;15(8):e174 [FREE Full text] [doi: 10.2196/jmir.2534] [Medline: 23989137]
17. Pew Research Center. Pew Research Center. 2017. Social media fact sheet URL: http://www.pewinternet.org/fact-sheet/social-media/ [accessed 2018-10-29] [WebCite Cache ID 73XHTTVnT]
18. Primack BA, Carroll MV, Shensa A, Davis W, Levine MD. Sex Differences in Hookah-Related Images Posted on Tumblr: A Content Analysis. J Health Commun 2016;21(3):366-375 [FREE Full text] [doi: 10.1080/10810730.2015.1095814] [Medline: 26890733]
19. Allem JP, Ferrara E. The Importance of Debiasing Social Media Data to Better Understand E-Cigarette-Related Attitudes and Behaviors. J Med Internet Res 2016 Dec 09;18(8):e219 [FREE Full text] [doi: 10.2196/jmir.6185] [Medline: 27507563]
20. Allem JP, Ferrara E. Could Social Bots Pose a Threat to Public Health? Am J Public Health 2018 Aug;108(8):1005-1006. [doi: 10.2105/AJPH.2018.304512] [Medline: 29995482]

21. Davis CA, Varol O, Ferrara E, Flammini A, Menczer F. Botornot: A system to evaluate social bots. 2016 Presented at: the 25th International Conference Companion on World Wide Web. International World Wide Web Conferences Steering Committee; 2016 Apr 11; Montreal, Canada p. 273-274.
22. Allem JP, Dharmapuri L, Unger JB, Cruz TB. Characterizing JUUL-related posts on Twitter. Drug Alcohol Depend 2018 Dec 01;190:1-5. [doi: 10.1016/j.drugalcdep.2018.05.018] [Medline: 29958115]
23. Allem JP, Ferrara E, Uppu SP, Cruz TB, Unger JB. E-Cigarette Surveillance With Social Media Data: Social Bots, Emerging Topics, and Trends. JMIR Public Health Surveill 2017 Dec 20;3(4):e98 [FREE Full text] [doi: 10.2196/publichealth.8641] [Medline: 29263018]
24. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems 2013:3111-3119.
25. Goodwin RD, Grinberg A, Shapiro J, Keith D, McNeil MP, Taha F, et al. Hookah use among college students: prevalence, drug use, and mental health. Drug Alcohol Depend 2014 Aug 01;141:16-20. [doi: 10.1016/j.drugalcdep.2014.04.024] [Medline: 24882367]
26. Jackson KM, Colby SM, Sher KJ. Daily patterns of conjoint smoking and drinking in college student smokers. Psychol Addict Behav 2010 Sep;24(3):424-435 [FREE Full text] [doi: 10.1037/a0019793] [Medline: 20853927]
27. Kavuluru R, Han S, Hahn EJ. On the popularity of the USB flash drive-shaped electronic cigarette Juul. Tob Control 2018 Apr 13:-. [doi: 10.1136/tobaccocontrol-2018-054259] [Medline: 29654121]
28. Silveira ML, Hilmi NN, Conway KP. Reasons for Young Adult Waterpipe Use in Wave 1 (2013-2014) of the Population Assessment of Tobacco and Health Study. Am J Prev Med 2018 Nov;55(5):650-655. [doi: 10.1016/j.amepre.2018.06.004] [Medline: 30219210]
29. Kowitt SD, Meernik C, Baker HM, Osman A, Huang L, Goldstein AO. Perceptions and Experiences with Flavored Non-Menthol Tobacco Products: A Systematic Review of Qualitative Studies. Int J Environ Res Public Health 2017 Dec 23;14(4):- [FREE Full text] [doi: 10.3390/ijerph14040338] [Medline: 28333107]
30. Unger JB, Urman R, Cruz TB, Majmundar A, Barrington-Trimis J, Pentz MA, et al. Talking about tobacco on Twitter is associated with tobacco product use. Prev Med 2018 Sep;114:54-56. [doi: 10.1016/j.ypmed.2018.06.006] [Medline: 29898418]

Abbreviations

API: application program interface

Edited by G Eysenbach; submitted 23.07.18; peer-reviewed by L Laestadius, J Colditz; comments to author 13.09.18; revised version received 05.10.18; accepted 08.10.18; published 19.11.18.

Please cite as:
Allem JP, Dharmapuri L, Leventhal AM, Unger JB, Boley Cruz T
Hookah-Related Posts to Twitter From 2017 to 2018: Thematic Analysis
J Med Internet Res 2018;20(11):e11669
URL: http://www.jmir.org/2018/11/e11669/
doi:10.2196/11669
PMID:30455162

Original Paper

Web-Based Signal Detection Using Medical Forums Data in France: Comparative Analysis

Marie-Laure Kürzinger[1], MSc; Stéphane Schück[2], MSc, MD; Nathalie Texier[2], PharmD; Redhouane Abdellaoui[3], MSc; Carole Faviez[3], MSc; Julie Pouget[4], MSc; Ling Zhang[5], MSc; Stéphanie Tcherny-Lessenot[1], MSc, MD; Stephen Lin[5], MD; Juhaeri Juhaeri[6], PhD

[1]Epidemiology and Benefit Risk Evaluation, Sanofi, Chilly-Mazarin, France
[2]Kappa Santé, Paris, France
[3]Kap Code, Paris, France
[4]Information Technology and Solutions, Sanofi, Lyon, France
[5]Global Pharmacovigilance, Sanofi, Bridgewater, NJ, United States
[6]Epidemiology and Benefit Risk Evaluation, Sanofi, Bridgewater, NJ, United States

Corresponding Author:
Marie-Laure Kürzinger, MSc
Epidemiology and Benefit Risk Evaluation
Sanofi
1 avenue Pierre Brossolette
Chilly-Mazarin, 91385
France
Phone: 33 1 69 74 59 42
Email: marie-laure.kurzinger@sanofi.com

Abstract

Background: While traditional signal detection methods in pharmacovigilance are based on spontaneous reports, the use of social media is emerging. The potential strength of Web-based data relies on their volume and real-time availability, allowing early detection of signals of disproportionate reporting (SDRs).

Objective: This study aimed (1) to assess the consistency of SDRs detected from patients' medical forums in France compared with those detected from the traditional reporting systems and (2) to assess the ability of SDRs in identifying earlier than the traditional reporting systems.

Methods: Messages posted on patients' forums between 2005 and 2015 were used. We retained 8 disproportionality definitions. Comparison of SDRs from the forums with SDRs detected in VigiBase was done by describing the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, receiver operating characteristics curve, and the area under the curve (AUC). The time difference in months between the detection dates of SDRs from the forums and VigiBase was provided.

Results: The comparison analysis showed that the sensitivity ranged from 29% to 50.6%, the specificity from 86.1% to 95.5%, the PPV from 51.2% to 75.4%, the NPV from 68.5% to 91.6%, and the accuracy from 68% to 87.7%. The AUC reached 0.85 when using the metric empirical Bayes geometric mean. Up to 38% (12/32) of the SDRs were detected earlier in the forums than that in VigiBase.

Conclusions: The specificity, PPV, and NPV were high. The overall performance was good, showing that data from medical forums may be a valuable source for signal detection. In total, up to 38% (12/32) of the SDRs could have been detected earlier, thus, ensuring the increased safety of patients. Further enhancements are needed to investigate the reliability and validation of patients' medical forums worldwide, the extension of this analysis to all possible drugs or at least to a wider selection of drugs, as well as to further assess performance against established signals.

(J Med Internet Res 2018;20(11):e10466) doi:10.2196/10466

KEYWORDS

adverse event; internet; medical forums; pharmacovigilance; signal detection; signals of disproportionate reporting; social media

Introduction

Adverse drug reactions (ADRs) are an important public health concern. Drug safety currently depends on postmarketing surveillance, which is conducted through spontaneous reporting systems based on voluntary reports. The reporting rate is low, and the time for getting access to reported data can be long. Thus, it is difficult to detect signals of disproportionate reporting (SDRs) in a timely manner, and some may not even be captured by those reporting systems. Alternative sources of data have already been used to detect drug-adverse event (AE) associations, including claims data [1], electronic medical records (EMRs) [2-4], and consumer search logs [5,6]. In addition, approaches have been developed to combine SDRs from data sources such as EMRs, claims, internet search logs with SDRs from the Food and Drug Administration Adverse Event Reporting System (FAERS) [5,7,8].

With the development and popularity of social media and medical forums, many internet users are exchanging health-related information, which may involve ADRs, and the question of consistency and usefulness of these new data sources for Web-based signal detection is under scrutiny. Recent projects have aimed at investigating the quality of social media data, as well as investigating the most performant method for the Web-based signal detection. The use of Web-based data (such as query logs and social media) is emerging among regulators (Food and Drug Administration and European Medicines Agency), industry, and academia [5,6,8-10]. As an example, a public-private partnership between the European Commission and European Federation of Pharmaceutical Industries and Associations, called WEB-RADR: Recognising Adverse Drug Reactions has been launched in 2014; this consortium made of organizations, including European medicines regulators, academics, and the pharmaceutical industry, aims to develop new ways of gathering information on suspected ADRs. One of the objectives is to investigate the potential for publicly available social media data for identifying potential drug safety issues. A recently published study [11] has focused on the performance evaluation of established statistical signal detection algorithms in Twitter or Facebook for a broad range of drugs and adverse events. Another example is related to the recent collaboration between Sanofi and Microsoft [8]. In this study, a Web-based search query method, called a query log reaction score, was developed to detect whether AEs associated with certain drugs could be found from search engine query data. The results were compared with reference signal detection algorithms commonly used with the FAERS.

The potential strength of Web-based data relies on their volume and real-time availability, allowing early SDR detection. This study aims to assess the consistency of SDRs detected from patients' forums in France and the ability to identify SDRs earlier than the traditional reporting systems. Three products were selected for this study (insulin glargine, teriflunomide, and zolpidem). SDRs detected from this Web-based source were compared with those detected from World Health Organization (WHO) AEs reporting system (VigiBase) using traditional SDR detection methods.

Methods

Data Sources

The sources of data used were (1) patients' medical forums in France and (2) VigiBase, the WHO individual case safety reports (ICSR) database; these sources of data are described later.

Medical Forum's Messages Database: Detec't

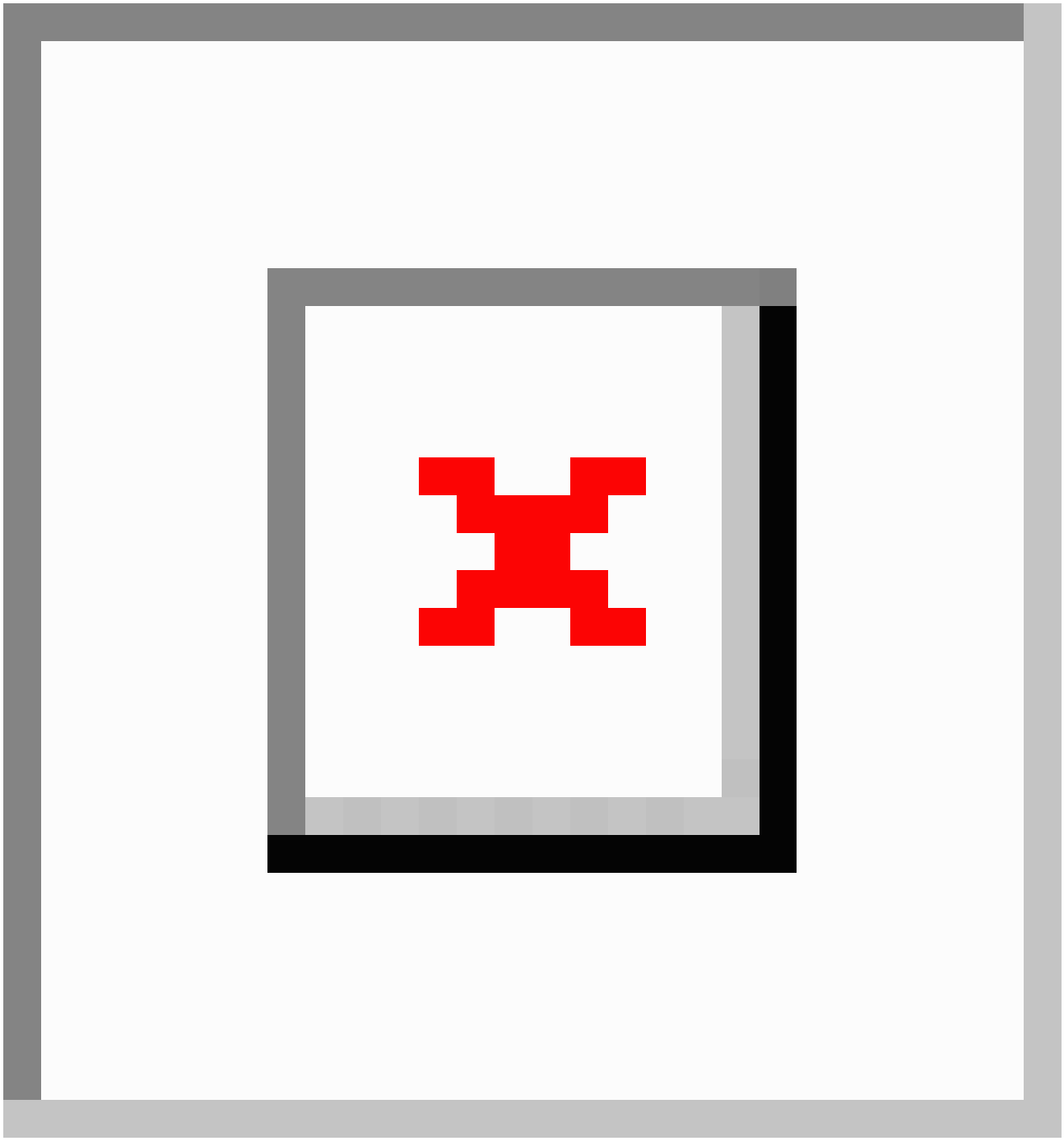
This was a retrospective study based on the secondary use of data from the Web—patients' medical forums in France. The Detec't database is a private database aggregating messages from social media, including safety information. We included 12 well-known medical forums in this study (see Multimedia Appendix 1). Every discussion publicly available at the time of the search and containing at least one message with the name of the drugs of interest (active substance or brand name) was extracted using a Web crawler (Detec't Extractor). This messages scrapping was done by targeting patients' messages using the HTML structure of each forum. This Web crawler was then adapted for each data source. Messages were extracted with all metadata related to the message (date, author of the post ID, URL of the discussion, and name of the forum). Finally, data were cleaned (deletion of ads, quotation of other Web users, and signature) using the HTML structure of the posts.

Retrieved messages went through several steps of processing so that the drug-event pairs necessary to perform the analysis were obtained. First, messages were deidentified. The data processing then automatically ensured that the messages contained the name of the drugs of interest and the concept of drug use or intake to constitute the corpora of messages related to the drugs under study. These corpora were formatted and screened to detect all references to medical events referring to potential AE (ie, not the indication of the drug, not an AE preceding the intake of the drug, not a question about adverse effects without having experienced it, etc).

The deidentification of messages was performed using an in-house algorithm based on the regular expression to automatically identify specific sequences of characters (like proper names, phone numbers, postal codes, mail addresses, etc). Messages containing the names of the drugs were identified by automatically detecting, among the set of extracted messages, references to active substances, and brand names of the drugs of interest; this step included the detection of common spelling mistakes. Next, the drug intake notion was identified with a specific algorithm (on this set of messages) based on the detection of regular expressions (identifying first-person personal pronouns, for instance). This procedure ensured that the person experiencing the drug could be identified (this person could be the author of the post or one of their relative). These "intake messages" represented the set of messages on which the following steps were applied.

As described earlier [12], the formatting of messages included the conversion of messages to lowercases, the removal of extra whitespaces, and the tokenization of messages. All words from messages were stemmed using Porter's algorithm to associate inflected and derived words together with their root form.

Figure 1. Extraction methodology and data preparation for analysis. AE: adverse events; AERS: Adverse Events Reporting System.



Medical concepts were detected using an extended version of Medical Dictionary for Regulatory Activities (MedDRA) version 15.0 adjusted and supplemented by vernacular vocabulary. In MedDRA, Preferred Terms (PT) correspond to unique descriptors of medical concepts. Lower-Level Terms (LLT) correspond to variants of PT and the lowest level of the terminology (each LLT is associated with a single parent PT and several LLTs can be associated with one PT). Medical concepts were detected at the LLT level and encoded as PT. The supplementation of vocabulary by vernacular terms was performed by manually reviewing a sample of messages from Web forums previously automatically annotated with MedDRA. The review was performed by 4 computer science experts who were familiar with the MedDRA dictionary. All medical concepts manually identified and not detected by MedDRA were saved and manually associated with a PT as if they were new LLTs. The evaluation of medical concepts detection, after supplementation on a sample of 157 messages, conducted to a recall of 71% and an accuracy of 93%. Regarding the adjustments provided to MedDRA, some PTs out of the study scope were removed (ie, “poverty” or “married”). For example, terms including the mention NOS (Not Otherwise Specified), that is, “Allergy NOS,” were cleaned by removing the mention. Eventually, all terms were stemmed. A manual cleaning was performed to deduplicate the terms obtained after stemming. The detection of medical concepts was performed by looking for exact matches between the stemmed versions of MedDRA and messages.

Potential AEs were identified by applying an algorithm [12] based on a Gaussian mixture model and then a support vector machine algorithm. This step identified drug-event pairs.

Extracted messages were stored in the database with (1) metadata associated with the messages; (2) results of the annotation process—drug intake, medical concepts, and their MedDRA code; and (3) the result of the AE detection algorithm.

The final step of the data preparation identified the number of messages for each drug-event pair (Figure 1).

The Comparison Database: VigiBase

The comparison database used as the gold standard was VigiBase, the WHO Global ICSR database. It consists of reports of adverse reactions received from member countries since 1968. VigiBase is updated with incoming ICSRs on a continuous basis. The VigiBase data resource is the largest and most comprehensive in the world, and it is developed and maintained by the Uppsala Monitoring Centre on behalf of the WHO. By May 2015, over 11 million reports were available in the database. The VigiBase database system includes linked databases containing medical and drug classifications—WHO Adverse Reactions Terminology or MedDRA, WHO International Classification of Diseases, and WHO Drug Dictionary; these classifications enable structured data entry, retrieval, and analysis at different levels of precision and aggregation.

Drugs

Three drugs (insulin glargine, teriflunomide, and zolpidem) were selected for the study to cover different therapeutic areas and different lengths of use since market authorization. Data corresponding to these selected drugs (active substance and brand name) were extracted from the Detec't database. The following drugs were therefore searched on the database: insulin glargine AND glargine AND Lantus; teriflunomide AND Aubagio; zolpidem AND Stilnox. Synonyms of the drug name and spelling mistakes were considered (ie, detail on misspelling). Misspellings were added by identifying most common errors and adding them as researched forms of the drug.

In order to carry out the analysis of disproportionate reporting, a background group of 327 drugs was randomly selected from the Detec't database. This later included drugs that were randomly chosen from an exhaustive list of French drugs. The messages corresponding to these 327 drugs were selected from the same set of forums and time period; they went through the same analysis and encoding steps.

Signals of Disproportionate Reporting Detection Metrics and Definition

The disproportionality analysis of spontaneous reports (comparing the number of observed cases with that of expected cases) was used. The quantitative method in signal detection relies on the principle of disproportionality [7,13,14]. We used the Proportional Reporting Ratio (PRR), Reporting Odds Ratio (ROR), Reporting Fisher's Exact Test (RFET), empirical Bayes geometric mean (EBGM), and the Information Component (IC). A total of 8 disproportionality definitions were considered for this study (Table 1).

Statistical Analysis

Descriptive Analysis

A description of messages for the overall period (cumulative from 2005 to 2015) and across time within the study period was provided—numbers of messages with the drug name, numbers of messages containing the concept of drug use or intake, medical concepts, and potential AEs.

Comparative Analysis

The comparison of SDRs detected from patients' forums in France to SDRs detected in VigiBase were described using sensitivity (true positive rate)=a/N1, specificity (true negative rate)=d/N2, positive predictive value (PPV)=a/M1, negative predictive value (NPV)=d/M2, and accuracy (a+d)/N (Tables 2 and 3). The receiver operating characteristics (ROC) curve and the area under the curve (AUC) were considered to measure the overall performance of the test to discriminate between positive and negative SDRs. The ROC curve represented the true positive rate (sensitivity) plotted in function of the false positive rate (100-specificity) for different thresholds of the metric.

Table 1. Definition of disproportionate signals.

Metric	Definition of disproportionate signal
Empirical Bayes Geometric Mean (EBGM)	EBGM≥2
Empirical Bayes Geometric Mean (EBGM)	EBGM≥4
Lower bound of the 95% CI of EBGM (EB05)	EB05≥2
Proportional Reporting Ratio (PRR)	PRR≥2, N ≥3, χ^2≥4
Lower bound of the 95% CI of PRR (PRR025)	PRR025≥1
Lower bound of the 95% CI of the Reporting Odds Ratio (ROR025)	ROR025≥1
Lower bound of the 95% CI of the Information Component (IC025)	IC025=0
Reporting Fisher's Exact Test (RFET) P	RFET P≤.05

Table 2. Signals: two-by-two contingency table for a combination of positive and negative signals from medical forums and VigiBase to measure performance.

Signals from medical forums	Signals from VigiBase		Total
	Positive	Negative	
Positive	a (true positive)	b (false positive)	M1
Negative	c (false negative)	d (true negative)	M2
Total	N1	N2	N

Table 3. Performance indicators.

Performance indicators	Value
Sensitivity (true positive rate)	*a*/N1
Specificity (true negative rate)	*d*/N2
Positive predictive value	*a*/M1
Negative predictive value	*d*/M2
Accuracy	(*a*+*d*)/*N*

Time Analysis

For SDRs identified in both data sources, we performed an analysis of time difference in months between the date of detection of SDRs from French patients' forums and the date of detection of SDRs from VigiBase.

Results

Descriptive Analysis

The data from 8 medical forums were considered for analysis and corresponded to messages published between January 1, 2005 and December 31, 2015, for insulin glargine and zolpidem and between April 1, 2014 and December 31 2015, for teriflunomide. For teriflunomide, this time restriction was because the product received its first marketing authorization application in September 2013 and was first launched in March 2014. Figure 2 shows the messages flowchart. The extraction was conducted for the identification of 102 messages for teriflunomide, 3326 messages for insulin glargine, and 4584 messages for zolpidem, which were relative to the drug intake. Among those, 61 messages for teriflunomide, 2335 messages for insulin glargine, and 3732 messages for zolpidem contained medical concepts. Among those, 41 messages for teriflunomide, 1799 messages for insulin glargine, and 2998 messages for zolpidem contained potential AEs. This resulted in 33 unique drug-event pairs for teriflunomide, 194 for insulin glargine, and 318 for zolpidem. The number of SDRs detected varied across the definition of disproportionate SDRs. For teriflunomide, insulin glargine and zolpidem, the number of SDRs varied from 4 to 12, 21 to 48, and 23 to 95, respectively. These SDRs were compared with the SDRs detected in VigiBase.

Comparative Analysis

Between 2005 and 2015, the medical forums data contained 545 drug-event pairs. Overall, 7618 pairs were identified in VigiBase, of which 422 drug-event pairs combinations overlapped with the forums data (Figure 3). The overlap considered an exact match with the event terminology. When restricting to pairs with at least 2 messages, the overlap was 275 drug-event pairs.

Among 545 drug-event pairs from the forums, only 123 were not identified in VigiBase; those 123 drug-event pairs corresponded each to only one message in the forums. The individual inspection showed that some PTs were not adequately specific. For example, some PTs identified in the forums were "weight," but it was not possible to match them with the PT, such as "underweight" or "overweight" or "abnormal weight gain" or "weight abnormal," which were identified in the drug-events pairs from VigiBase.

Among the overlap of 422 drug-event pairs, the specificity was high (87.5%-95.5%) depending on the SDR definition (Table 4). On the other hand, the sensitivity was low (29%-50.6%), indicating that an important proportion of SDRs from VigiBase was not identified in the forums. The PPV (51.2%-75.4%), NPV (68.5%-91.6%), and accuracy (68%-87.7%) were high.

Among 275 drug-event pairs (for which at least 2 reports or messages were considered; Table 5), the figures were slightly higher, improving the overall performance. The specificity varied between 86.9% and 93.4% and the sensitivity between 39.1% and 56.5%. The PPV (53.9%-76.6%), NPV (68.7%-91.5%), and accuracy (70.6%-86.2%) were high.

Whatever definition of disproportionate SDR used, the ROC curves and the AUC showed an overall good performance. AUC varied around 0.8. The highest AUC was shown with the EBGM metric (AUC=0.85; Figure 4).

Time Analysis

For SDRs detected both in VigiBase and patients' forums, we calculated the time difference in months between the date of detection of positive SDRs from the French patients' forums and the date of detection of these SDRs in VigiBase (Table 6).

Figure 2. Flowchart for the 3 drugs and the other 327 drugs (comparison group). AE: adverse events.

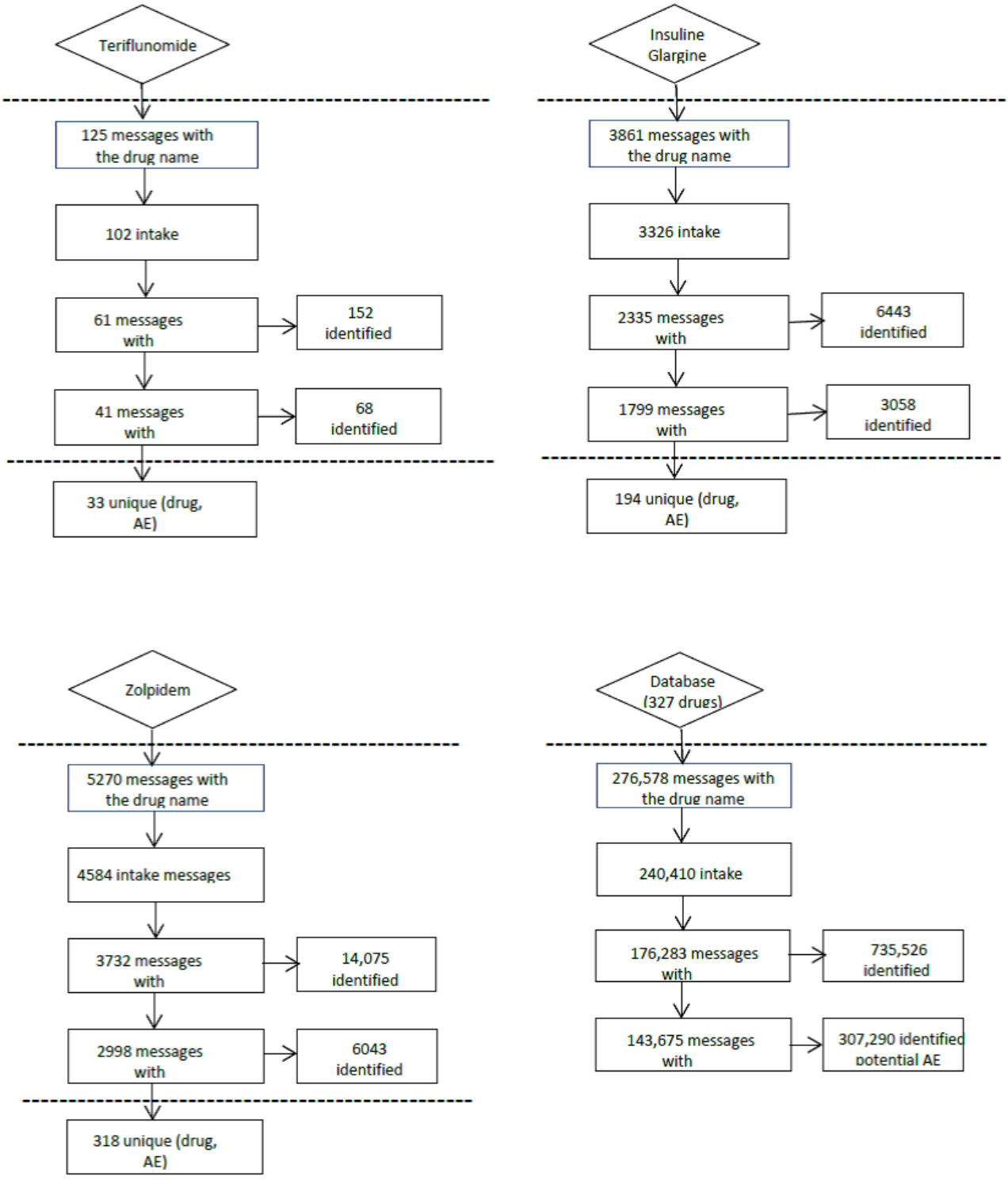


Figure 3. Time periods covered by VigiBase and the forums database and the number of drug-event pairs overlap, as well as pairs overlap with at least 2 messages (smallest circle).

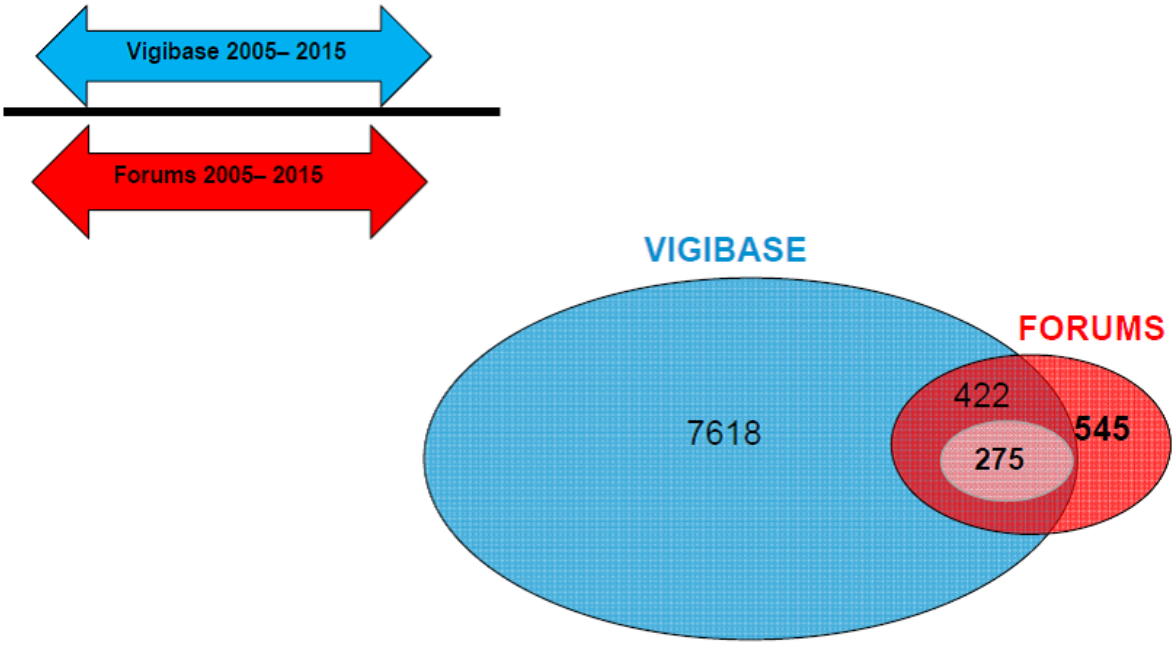


Table 4. The sensitivity, specificity, positive predictive value, negative predictive value, and accuracy among 422 drug-event pairs.

Definition	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)	Accuracy (%)
EB05[a]≥2	29.0	95.5	62.5	84.0	82.0
EBGM[b]≥2	48.2	89.3	62.5	82.3	78.2
EBGM≥4	39.6	94.6	51.2	91.6	87.7
PRR[c]≥2, N≥3, χ^2≥4	31.9	94.0	67.9	77.9	76.5
Lower 95% CI of PRR≥1	37.3	87.5	64.1	70.0	68.7
Lower 95% CI of ROR≥1	37.0	87.9	66.3	68.5	68.0
IC025[d]>0	33.3	94.2	75.4	72.5	73.0
RFET[e]: P≤.05	50.6	86.1	68.1	74.8	73.0

[a]EB05: Lower bound of the 90% CI of empirical Bayes geometric mean.

[b]EBGM: empirical Bayes geometric mean.

[c]PRR: Proportional Reporting Ratio.

[d]IC025: Lower bound of the 95% CI of the information component.

[e]RFET: Reporting Fisher's Exact Test.

Table 5. The sensitivity, specificity, positive predictive value, negative predictive value, and accuracy among 275 drug-event pairs.

Definition	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)	Accuracy (%)
EB05[a]≥2	39.1	93.4	64.1	83.5	80.7
EBGM[b]≥2	49.4	88.5	65.1	80.2	76.7
EBGM≥4	51.2	92.3	53.9	91.5	86.2
PRR[c]≥2, N≥3, χ^2≥4	44.2	91.5	70.4	78.3	76.7
Lower 95% CI of PRR≥1	48.3	88.7	75.7	70.2	71.6
Lower 95% CI of ROR≥1	47.1	88.5	75.7	68.7	70.6
IC025[d]>0	45.8	91.1	76.6	72.5	73.5
RFET[e]: P≤.05	56.5	86.9	75.6	73.6	74.2

[a]EB05: Lower bound of the 90% CI of empirical Bayes geometric mean.

[b]EBGM: empirical Bayes geometric mean.

[c]PRR: Proportional Reporting Ratio.

[d]IC025: Lower bound of the 95% CI of the information component.

[e]RFET: Reporting Fisher's Exact Test.

Figure 4. The receiver operating characteristics (ROC) curves and area under the curve applying empirical Bayes geometric mean (EBGM)≥4 in the VigiBase and EBGM in the forums.

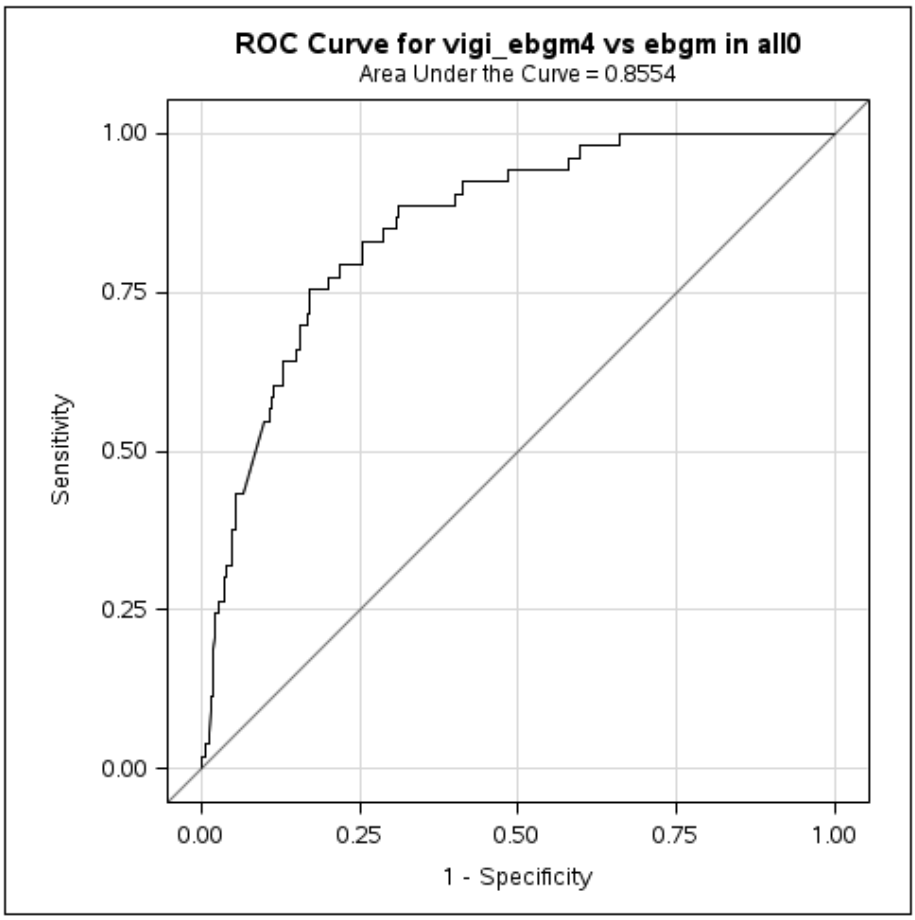


Table 6. The time difference in months of signals detection dates (Δtime) between patients' forums and VigiBase.

Definition	Δtime[a]<0, n (%)	Δtime[a]=0, n (%)	Δtime[a]>0, n (%)	Total number of pairs, n (%)
PRR[b]≥2, N≥3, χ^2≥4	15 (25.4)	3 (5.1)	41 (69.5)	59 (100)
EB05[c]≥2	10 (32.3)	3 (9.7)	18 (58.1)	31 (100)
EBGM[d]≥2	22 (26.5)	4 (4.8)	57 (68.7)	83 (100)
EBGM≥4	12 (37.5)	4 (12.5)	16 (50)	32 (100)
IC025[e]>0	13 (21.3)	3 (4.9)	45 (73.8)	61 (100)
Lower 95% CI of PRR≥1	29 (32.6)	5 (5.6)	55 (61.8)	89 (100)
Lower 95% CI of ROR≥1	29 (32.2)	5 (5.6)	56 (62.2)	90 (100)
RFET[f]: P≤.05	34 (30.6)	7 (6.3)	70 (63.1)	111 (100)

[a]Δtime: detection date in patients' forums−detection date in VigiBase.

[b]PRR: Proportional Reporting Ratio

[c]EB05: Lower bound of the 90% CI of empirical Bayes geometric mean.

[d]EBGM: empirical Bayes geometric mean.

[e]IC025: Lower bound of the 95% CI of the information component.

[f]RFET, Reporting Fisher's Exact Test.

Depending on the definition of SDRs, up to 38% (12/32) of common SDRs were detected earlier (up to 128 months earlier) in the forums than in VigiBase. In addition, up to 13% (4/32) were detected at the same date but were available earlier in the forums given the real-time availability of data on the Web. The qualitative exploration of SDRs detected earlier in the forums showed heterogeneity as some were related to serious medical events and other to patients-related symptoms (ie, stress and hunger).

Most signals that were detected earlier in VigiBase were linked to serious medical events, which probably led to medical consultation and, thus, to an AE reporting done through a health care professional. In addition, most of those events were related to the System Organ Class "nervous system disorders" and "psychiatric disorders."

Discussion

Principal Findings

This study aimed at assessing the consistency of SDRs detected from patients' forums in France over the last 11 years and the ability to identify SDRs earlier than that in VigiBase.

The potential strength of Web-based data relies on their volume and real-time availability, allowing early signal detection. This pilot study showed a good performance and earlier detection of SDRs in the French medical forums compared with SDRs detected in traditional sources. In addition, these pilot results indicate that using patients' medical forums may be considered as a complementary source of data to traditional sources, allowing SDRs to be detected earlier and, thus, facilitating the increased safety of patients.

We first compared SDRs (by considering several definitions of disproportionate SDRs) detected in the forums data and the WHO AEs reporting system (VigiBase). The comparison of positive and negative SDRs showed that whatever the definition of disproportionate SDR, the sensitivity was low and the specificity was very high. In addition, the PPV and NPV were high. The overall performance was good, showing that data from the medical forum may be a valuable source to be considered for signal detection. In another study [8] using query log data, results showed that the method had moderate sensitivity and low specificity in detecting signals in Web query data compared with reference signal detection algorithms in FAERS. In another study [11] using Twitter and Facebook, the authors suggested that broad-ranging statistical signal detection in Twitter and Facebook, using currently available methods for adverse event recognition, performs poorly and cannot be recommended at the expense of other pharmacovigilance activities; this indicates that results in terms of performance might vary according to the Web data source used and the metric used for SDRs.

Second, among SDRs from patients' forums and VigiBase, we calculated time differences in detection of SDRs to measure the ability of forums data to detect earlier SDRs compared with VigiBase. Up to 38% (12/32) of common SDRs could be detected earlier when using the forums data, which is an important finding. The qualitative exploratory analysis of the SDRs detected earlier showed that events were related to serious as well as patient-related symptoms. This finding is consistent with recent studies [15] that addressed the question of earlier detection of drug-related AEs in the social media compared with FAERS. The findings highlighted some of the promises of social media data sources for detecting early AE reports patterns compared with conventional pharmacovigilance sources and showed that social media AE reports helped predict the occurrence of FAERS reports several months later for one of the two drugs that were studied. In a study [16], the objective was to examine whether specific product-AE pairs were reported through social media before being reported to FAERS. In one of the positive cases, the first report occurred in social media prior to the SDR detection from FAERS. Authors concluded

that an efficient semiautomated approach to social media monitoring might provide earlier insights into certain AEs.

Strengths and Limitations

One of the strengths of this study was the quality of preprocessing and processing of the data extracted from the forums. Messages that were used for Web-signal detection in this study were not only containing the drug but also a medical event (cooccurrence) as this is done in other studies [17]. Several cleansing and validation steps were performed to ensure that the identified messages were related to an internet user who used the concept of use or intake of the drug and with potential AE.

This study has several limitations. First, the results only apply to 3 drugs and for the French medical forums. Thus, results are not generalizable to all drugs and at a worldwide scale. However, we do not have a strong hypothesis to believe that the use of Web-based medical forums and interactions of French internet users would be different in other developed countries. Thus, further studies focusing on worldwide patients' forums should be envisaged. Second, automatics algorithms have their limitations. The management of Web-based data needs continuous updates on modeling and data processing to ensure high quality and accuracy of the information retrieved. Although the data were processed, it is still possible that some drug names or medical concepts could be missed, that some AEs may be confused with drugs indications, or questions from Web users about AEs they have not experienced, or descriptions of symptoms that are not adverse reactions to drugs. Third, Web-based data rely on patients' perspective and declaration but not on a true medical diagnosis. Web-based data are sensitive to increase in the media coverage, resulting in increased searches or posts and are prone to changes in people's search or communication behavior. Finally, VigiBase is not a true gold standard, as it has its own limitations (such as lack of denominator and underreporting). VigiBase, however, has been used as a standard for signal detection by regulators and pharmaceutical companies, and our study showed that patients' forum could be used as a complementary data source to detect SDR earlier. Although the choice of the reference data remains challenging [18], further studies using a refined gold standard, such as drug-event pairs shown in the labels, should be considered.

Conclusions

This study shows a good performance and earlier detection of SDRs detected in patients' medical forums compared with SDRs detected in traditional sources. Those SDRs relate to serious medical events as well as subjective patients-related symptoms (eg, stress and hunger). These results indicate that using patients' medical forums may be considered as a complementary source of data to traditional sources, allowing SDRs to be detected earlier and, thus, ensuring the increased safety of patients. Further enhancements are needed to investigate the reliability and validation of patients' medical forums worldwide, the extension of this analysis to all possible drugs or at least to a wider selection of drugs, as well as to further assess performance against established signals.

Acknowledgments

This research was fully funded by Sanofi.

Conflicts of Interest

MLK, JP, STL, SL, and JJ are all employees of Sanofi, and this work was conducted as part of their employment. LZ also contributed to this paper as a former employee of Sanofi. SS and NT are employed by Kappa Santé and provided the data through funding from Sanofi. CF is employed by Kap Code, a Kappa Santé start-up that owns Detec't. RA also contributed to this paper as a former employee of Kap Code.

Multimedia Appendix 1

List of medical forums included in the study.

[PDF File (Adobe PDF File), 21KB - jmir_v20i11e10466_app1.pdf]

References

1. Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. Pharmacoepidemiol Drug Saf 2010 Aug;19(8):858-868 [FREE Full text] [doi: 10.1002/pds.1926] [Medline: 20681003]
2. Coloma PM, Trifirò G, Schuemie MJ, Gini R, Herings R, Hippisley-Cox J, EU-ADR Consortium. Electronic healthcare databases for active drug safety surveillance: is there enough leverage? Pharmacoepidemiol Drug Saf 2012 Jun;21(6):611-621. [doi: 10.1002/pds.3197] [Medline: 22315152]
3. LePendu P, Iyer SV, Bauer-Mehren A, Harpaz R, Mortensen JM, Podchiyska T, et al. Pharmacovigilance using clinical notes. Clin Pharmacol Ther 2013 Jun;93(6):547-555 [FREE Full text] [doi: 10.1038/clpt.2013.47] [Medline: 23571773]
4. Iyer SV, Harpaz R, LePendu P, Bauer-Mehren A, Shah NH. Mining clinical text for signals of adverse drug-drug interactions. J Am Med Inform Assoc 2014;21(2):353-362 [FREE Full text] [doi: 10.1136/amiajnl-2013-001612] [Medline: 24158091]

5. Yom-Tov E, Gabrilovich E. Postmarket drug surveillance without trial costs: discovery of adverse drug reactions through large-scale analysis of web search queries. J Med Internet Res 2013;15(6):e124 [FREE Full text] [doi: 10.2196/jmir.2614] [Medline: 23778053]
6. White RW, Harpaz R, Shah NH, DuMouchel W, Horvitz E. Toward enhanced pharmacovigilance using patient-generated data on the internet. Clin Pharmacol Ther 2014 Aug;96(2):239-246 [FREE Full text] [doi: 10.1038/clpt.2014.77] [Medline: 24713590]
7. Harpaz R, DuMouchel W, LePendu P, Bauer-Mehren A, Ryan P, Shah NH. Performance of pharmacovigilance signal-detection algorithms for the FDA adverse event reporting system. Clin Pharmacol Ther 2013 Jun;93(6):539-546 [FREE Full text] [doi: 10.1038/clpt.2013.24] [Medline: 23571771]
8. Colilla S, Tov EY, Zhang L, Kurzinger M, Tcherny-Lessenot S, Penfornis C, et al. Validation of New Signal Detection Methods for Web Query Log Data Compared to Signal Detection Algorithms Used With FAERS. Drug Saf 2017 May;40(5):399-408. [doi: 10.1007/s40264-017-0507-4] [Medline: 28155198]
9. Harpaz R, Callahan A, Tamang S, Low Y, Odgers D, Finlayson S, et al. Text mining for adverse drug events: the promise, challenges, and state of the art. Drug Saf 2014 Oct;37(10):777-790 [FREE Full text] [doi: 10.1007/s40264-014-0218-z] [Medline: 25151493]
10. Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance: A review. J Biomed Inform 2015 Apr;54:202-212 [FREE Full text] [doi: 10.1016/j.jbi.2015.02.004] [Medline: 25720841]
11. Caster O, Dietrich J, Kürzinger ML, Lerch M, Maskell S, Norén GN, et al. Assessment of the Utility of Social Media for Broad-Ranging Statistical Signal Detection in Pharmacovigilance: Results from the WEB-RADR Project. Drug Saf 2018(24) [FREE Full text] [doi: 10.1007/s40264-018-0699-2] [Medline: 30043385]
12. Abdellaoui R, Schück S, Texier N, Burgun A. Filtering Entities to Optimize Identification of Adverse Drug Reaction From Social Media: How Can the Number of Words Between Entities in the Messages Help? JMIR Public Health Surveill 2017 Jun 22;3(2):e36 [FREE Full text] [doi: 10.2196/publichealth.6577] [Medline: 28642212]
13. Hauben M, Zhou X. Quantitative methods in pharmacovigilance: focus on signal detection. Drug Saf 2003;26(3):159-186. [Medline: 12580646]
14. Berlin C, Blanch C, Lewis DJ, Maladorno DD, Michel C, Petrin M, et al. Are all quantitative postmarketing signal detection methods equal? Performance characteristics of logistic regression and Multi-item Gamma Poisson Shrinker. Pharmacoepidemiol Drug Saf 2012 Jun;21(6):622-630. [doi: 10.1002/pds.2247] [Medline: 21994119]
15. Duh MS, Cremieux P, Audenrode MV, Vekeman F, Karner P, Zhang H, et al. Can social media data lead to earlier detection of drug-related adverse events? Pharmacoepidemiol Drug Saf 2016 Dec;25(12):1425-1433 [FREE Full text] [doi: 10.1002/pds.4090] [Medline: 27601271]
16. Pierce CE, Bouri K, Pamer C, Proestel S, Rodriguez HW, Van Le H, et al. Evaluation of Facebook and Twitter Monitoring to Detect Safety Signals for Medical Products: An Analysis of Recent FDA Safety Alerts. Drug Saf 2017 Dec;40(4):317-331 [FREE Full text] [doi: 10.1007/s40264-016-0491-0] [Medline: 28044249]
17. Yang CC, Yang H, Jiang L. Postmarketing Drug Safety Surveillance Using Publicly Available Health-Consumer-Contributed Content in Social Media. ACM Trans. Manage. Inf. Syst 2014 Apr 01;5(1):1-21. [doi: 10.1145/2576233]
18. Hauben M, Aronson JK, Ferner RE. Evidence of Misclassification of Drug-Event Associations Classified as Gold Standard 'Negative Controls' by the Observational Medical Outcomes Partnership (OMOP). Drug Saf 2016 May;39(5):421-432. [doi: 10.1007/s40264-016-0392-2] [Medline: 26879560]

Abbreviations

AE: adverse event
AERS: Adverse Events Reporting System
AUC: area under the curve
EBGM: empirical Bayes geometric mean
FAERS: Food and Drug Administration Adverse Event Reporting System
IC: information component
ICSR: individual case safety reports
LLT: lower-level terms
NOS: Not Otherwise Specified
NPV: negative predictive value
PPV: positive predictive value
PRR: proportional reporting ratio
PT: preferred terms
RFET: Reporting Fisher's Exact Test
ROC: receiver operating characteristics
ROR: reporting odds ratio
SDR: signals of disproportionate reporting

WHO: World Health Organization
MedDRA: Medical Dictionary for Regulatory Activities

Edited by G Eysenbach; submitted 23.03.18; peer-reviewed by C Freifeld, M Gilbert; comments to author 25.04.18; revised version received 29.06.18; accepted 29.06.18; published 20.11.18.

Please cite as:
Kürzinger ML, Schück S, Texier N, Abdellaoui R, Faviez C, Pouget J, Zhang L, Tcherny-Lessenot S, Lin S, Juhaeri J
Web-Based Signal Detection Using Medical Forums Data in France: Comparative Analysis
J Med Internet Res 2018;20(11):e10466
URL: http://www.jmir.org/2018/11/e10466/
doi:10.2196/10466
PMID:30459145

Original Paper

Clinicians' Selection Criteria for Video Visits in Outpatient Care: Qualitative Study

Linda Sturesson[1*], MA; Kristina Groth[2,3*], MSc, PhD

[1]Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

[2]Centre for Innovation, Karolinska University Hospital, Stockholm, Sweden

[3]Department of Clinical Science, Intervention and Technology, Karolinska Institutet, Stockholm, Sweden

[*]all authors contributed equally

Corresponding Author:
Kristina Groth, MSc, PhD
Centre for Innovation
Karolinska University Hospital
C1:77 Huddinge
Stockholm,
Sweden
Phone: 46 700856664
Email: kristina.groth@sll.se

Abstract

Background: Video visits with patients were introduced into outpatient care at a hospital in Sweden. New behaviors and tasks emerged due to changes in roles, work processes, and responsibilities. This study investigates the effects of the digital transformation—in this case, how video visits in outpatient care change work processes and introduce new tasks—to further improve the concept of video visits. The overarching goal was to increase the value of these visits, with a focus on the value of conducting the treatment for the patient.

Objective: Through the real-time, social interactional features of preparing for and conducting video visits with patients with obesity, this study examines which patients the clinicians considered suitable for video visits and why. The aim was to identify the criteria used by clinicians when selecting patients for video visits to understand what criteria the clinicians used as the grounds for their selection.

Methods: Qualitative methods were used, including 13 observations of video visits at 2 different clinics and 14 follow-up interviews with clinicians. Transcripts of interviews and field notes were thematically analyzed, discussed, and synthesized into themes.

Results: From the interviews, 20 different arguments for selecting a specific patient for video visits were identified. Analyzing interviews and field notes also revealed unexpressed arguments that played a part in the selection process. The unexpressed arguments, as well as the implicit reasons, for why a patient was given the option of video visits can be understood as the selection criteria for helping clinicians in their decision about whether to offer video visits or not. The criteria identified in the collected data were divided into 3 themes: practicalities, patient ability, and meeting content.

Conclusions: Not all patients with obesity undergoing treatment programs should be offered video visits. Patients' new responsibilities could influence the content of the meeting and the progress of the treatment program. The selection criteria developed and used by the clinicians could be a tool for finding a balance between what the patient wants and what the clinician thinks the patient can manage and achieving good results in the treatment program. The criteria could also reduce the number and severity of disturbances and limitations during the meeting and could be used to communicate the requirements they represent to the patient. Some of the criteria are based on facts, whereas others are subjective. A method for how and when to involve the patient in the selection process is recommended as it may strengthen the patient's sense of responsibility and the relationship with the clinician.

(J Med Internet Res 2018;20(11):e288) doi:10.2196/jmir.9851

KEYWORDS

outpatient care; selection criteria; telemedicine; telehealth; ethnography

Introduction

Telemedicine solutions can be beneficial for different stakeholders at different levels and from different perspectives [1-3]. Through telemedicine, access to health care for all can be possible as the participants are not bound to a specific place, and care can be provided for patients who struggle to physically visit health care premises, for example, because they live far away or are not strong enough to travel. Hence, there is an expectation that digital communication will, to some extent, be more inclusive; for example, marginalized groups may gain benefits [4].

One telemedicine solution is video-mediated meetings or consultations (hereafter called video visits) between patients (including relatives) and clinicians, which introduce a new way of conducting such meetings. Video visits can be as effective as face-to-face interventions [3], and they appear to be safe and convenient in outpatient care, but there are complex challenges related to their adoption by clinicians [5]. Implementing telemedicine changes how work is organized in terms of roles, tasks, and processes [5], which in turn changes the power relationships between participants and their expectations of each other [6-9]. Video visits imply that technology is used for mediating the meeting, that there is a geographical separation between the clinician and the patient [10,11] and that a nonclinical space (the patient's) is added to the meeting [10,12]. During a video visit, the place chosen by the patient might influence the complex communication between the clinician and patient [12,13]. Introducing a nonclinical setting may affect those involved, the consultation, and the outcome of the meeting for example, because a health care environment manifests social orders [7] and facilitates the maintenance of professional and patient roles. In addition, limitations and disturbances may occur due to the technology, the patient's surroundings, or the procedures deployed when conducting video visits. These are all aspects that might affect the behaviors of the patient and clinician, how the video visit is conducted, and its outcome [12].

Aspects that may positively influence the outcome of video visits have been identified; for example, there is an established relationship and trust between the clinician and patient, there is a need for frequent contact, and those involved in the video visit need to have the necessary technical skills [5]. Other aspects that can influence whether or not a patient is considered suitable for telemedicine solutions are, for example, the patient's preferences and circumstances and system capacity [13], as well as how the technology may affect the patient's health condition [5]. It is recommended that patients living a long distance from the health care premises are offered telemedicine solutions, and physicians are recommended to begin with less complex cases [13]. The selection of patients for video visits may hence be determined by various criteria such as the complexity of the meeting, the patient's preferences and distance from the clinic, the clinician's experience with the technology, and so forth [2]. It is known that the physical clinical environment affects patient satisfaction, attitudes, and work performances [14], which implies that selecting patients for video visits needs careful consideration. Some patients may not be considered suitable for video visits [5,13], and this introduces the new task of choosing whether a patient is suitable when implementing video visits [5]. But how do clinicians choose patients for video visits? It has been reported that patients are more positive about *tele-homecare* (which, as well as video visits, also includes daily monitoring of data) than clinicians, and their perceptions of such care may differ from that of clinicians [15]. Even though *tele-homecare* includes daily monitoring of tasks, the perception of a video visit may also differ between patients and clinicians. The perception may also be influenced by the type of care and issues addressed during the video visit. In this study, we explored the specific task of selecting patients for video visits as part of a treatment program in outpatient care at a university hospital in Sweden. Video visits were not offered through spontaneous meetings initiated by the patients [5].

The aim of this study was to identify the criteria used by clinicians when selecting patients for video visits, by exploring which patients the clinicians considered suitable for video visits and why. The study was part of a broader investigation including how video visits in outpatient care change work processes and introduce new tasks, with the overarching goal of increasing the value of these visits by making further improvements to the concept. The value may differ between stakeholders, and our focus was on the treatment of the patient—not only on the value of benefits to the patient but also for the hospital and society. The value can be measured by the progress of the treatment.

Methods

Overview

The study was qualitative and exploratory in its approach. Interviews with clinicians and observations of video visits were conducted to generate data. The focus was on the situatedness in the use of video visits and the situated actions when clinicians conducted such visits [16]. In addition, both formal and informal settings in everyday work and ad hoc individual conversations related to video visits were observed and analyzed to understand the phenomenon of video visits and their role in a wider context.

Our analysis resulted in 2 overarching categories: the criteria used by clinicians when selecting patients for video visits and disturbances and limitations. The first category is presented in this paper, and the second has been addressed in another paper [12]. As the 2 categories result from the same research study, the same methods and materials were used.

Approach to the Research Area

Theoretical perspectives in symbolic interactionism provided a source of inspiration and a starting point, providing frameworks suitable for analyzing social reality and understanding human behavior and human feelings. Social interaction can be influenced by moods, weather, locations, and environments [17]. The individual defines the situation both consciously and unconsciously, and human behavior is seen in relation to the whole context [17]. Both diversity and commonalities are sought with an open mind, with attention being given "to what falls out of view or falls between the cracks" [18].

In our study, the video visits were part of a treatment program that included several consecutive meetings. The consultation is

a social interaction that involves a clinician, a patient and often a relative, in which at least 1 of them has a predetermined goal for the meeting. What happens between those involved can be understood as social acting and, more specifically, as an instrumental or planned action [16]. For example, a clinician may have the goal of learning about the patient's behavior since the last meeting, progress, side effects, etc. To achieve this goal, the clinician will prepare by reading the patient's medical record and making notes on what to address during the consultation. However, each consultation session is a link in a longer treatment chain—a path along which each situation affects the outcome of each session [16].

Clinicians develop skills based on physical consultations, and face-to-face visits become the norm for clinical meetings [9]. The clinicians' frame of reference is thus the traditional physical meeting or a follow-up by phone. When introducing video visits, clinicians are therefore likely to compare them with traditional clinical meetings. In our study, we explored video visits by gathering examples of the clinicians selecting patients for such visits, the reasons behind their choices, and how they shared their experiences and discussed their choices.

Ethical Approval and Consent

Ethical approval for the study was given by the Regional Ethical Review Board in Stockholm before data gathering (reference number: 2016/1027-31). The clinicians obtained written informed consent for participation and for publication (including information about participation, anonymity, the purpose and objectives of the study, and the responsible researcher) from patients, relatives, and guardians. Participants were offered video visits instead of physical meetings. The consent form was either sent by email or given by hand to the patient and, if applicable, to relatives. The clinicians signed a written consent for participation following review by the researcher.

Context

A total of 2 patient flows were involved in this study, named Clinic A and Clinic B. Both clinics treated patients with obesity. The clinics had congruent goals, agendas, and philosophies for their treatment. The content of care was mainly based on a humanistic perspective of health and disease, with a lesser focus on biomedical data such as weight and body composition. However, these variables were still used as treatment outcome assessments. Video visits at the clinics were part of a treatment program that included several consecutive meetings aimed at helping patients to implement lifestyle changes successfully. Clinicians supported patients in their efforts to achieve behavioral and lifestyle changes. Between visits, patients were asked to work actively on these changes by themselves. Both clinics shared the same view about using video visits as complements to face-to-face visits and for follow-ups. The staff consisted of doctors, nurses, psychologists, nutritionists, occupational therapists, and physiotherapists. Clinicians at Clinic B also had competence in cognitive behavioral therapy. Differences between the clinics are described in Table 1.

Technology and Devices

The concept used for video visits was developed for less complex meetings in outpatient care. The technology included an ordinary videoconferencing tool with encrypted communication, capable of producing adequate quality for seeing and hearing each other and for sharing documents. The technology could not be used to connect sensors used for monitoring parameters, and the quality of the video was not high enough to provide details of problems such as skin issues. A complex video visit, such as when a neurologist needs to see small detailed movements during care for patients with Parkinson disease or to demonstrate exercises to a patient [19], may require equipment of higher quality as well as additional space in front of the video camera for specific exercises.

The patient or relative typically used his or her own device such as a computer, mobile phone, or tablet with a webcam, speaker, internet connection, and Web browser or the videoconferencing app.

Respondents and Recruitment

In preparation for the study, 2 clinics were selected to participate. They were identified from the second author's work with introducing video visits in outpatient care settings at the hospital. One clinic was selected because it had successfully adopted the concept of video visits earlier in the year, and the other was selected because it had shown interest and carried out test video visits but had not yet started. Moreover, 2 clinicians from the first clinic, who were already conducting video visits, and 6 clinicians from the second clinic who wanted to start video visits, agreed to participate in the study.

At both clinics, the staff selected patients or relatives for video visits. Video visits were only offered to patients who were physically present at the clinic at the beginning of their treatment. The clinicians offered video visits to the selected patients either during a physical meeting or through a telephone contact. The patients had the opportunity to accept or decline video visits. During the study period, there were patients who declined. The clinicians who conducted the video visits had previously met face-to-face with the patients.

If patients accepted a video visit, the clinicians asked them if they wanted to participate in the research study. The question was asked to the patient and, if applicable, to the relative during a face-to-face meeting, phone call, or previous video visit (at Clinic B, where video visits were being used before the research study started). The staff, patients, and any guardian provided written informed consent to participate in the research.

Data Collection

The data collection, conducted by the first author, consisted of a total of 13 observations and 14 interviews; see Table 2 for more details. In all, 6 clinicians conducted 2 video visits each and were, therefore, observed and interviewed twice. However, 1 of the interviews was conducted without an observation (see below), resulting in a total of 13 observations and 14 interviews.

Each observation started before the actual video visit and included the time for the clinician's immediate preparation. The researcher was located in the same room as the clinician and was visually and verbally presented to the patient and relative at the beginning of the video visit, giving each patient a chance to withdraw his or her consent. During the video visit, the

researcher observed the meeting from a position out of sight of the webcam, that is, the patient and relative could not see the researcher. The observations were partly exploratory and partly structured. Some aspects such as start and end time, patient's location, and number of participants were predetermined and noted in the observation protocol. These were combined with field diaries that contained the exploratory observation notes. The observations were not recorded, photographed, or filmed.

The interviews were in-depth, contextual, and semistructured and were conducted with the clinicians after, and in addition to, each video visit. Of the 14 interviews, 13 were conducted face-to-face and one through phone. Furthermore, 1 of the interviews occurred without an observation, as the patient withdrew consent to participate in the study as the observation was about to start. The interview was still conducted after the video visit. The interviews were recorded and transcribed verbatim.

In addition, the first author attended formal encounters (eg, treatment conferences with clinicians) as a passive observer and participated in informal gatherings (eg, lunches and other breaks), taking field notes to capture the clinical discourse and clinicians' perceptions and thoughts about video visits, without interfering in the discussions taking place. All data were gathered during a contiguous period of 3 months during 2016.

Analysis

The analysis process follows a qualitative approach [20] in which the transcripts of interviews and field notes were read through several times to achieve a familiarization with the content. During the reading, themes were identified and noted in a blank document. Corresponding transcripts and field notes were read iteratively to gain a full picture of the collected data and then a conceptual framework was created. After this initial process, the transcripts of interviews and field notes were analyzed thematically [21]. The data were then read through again and coded to match the themes in the developed conceptual framework.

Spreadsheets were used to organize and sort the data. To find and keep track of patterns in the material, themes were separated into different rows in the spreadsheet, and each interview and the corresponding observations were sorted into different columns. Pieces of the text were sorted to the appropriate cells. The principle of spreadsheets was also used to analyze and find patterns in the quantitative data. The data, themes, and sorting were continuously discussed throughout the analysis.

The themes were synthesized into 2 overarching categories: "Selecting patients for video visits" and "Disturbances and limitations." From the analysis, it became clear that selecting patients had added a new task for clinicians, and video meetings had introduced disturbances and limitations related to both the technology and the surroundings. This paper focuses on the first category "Selecting patients for video visits" and the criteria the clinicians used in the selection process. The themes sorted under the category "Selecting patients for video visits" were issues of the patient's ability, practical matters, and the meeting content. Each theme represents a number of criteria.

In the Results section, quotes are used to illustrate situations in which selection considerations were made. The quotes chosen represent situations that occurred once or several times, illustrating an effect of something that may occur in other situations. When illustrating a situation related to the criteria with excerpts from the data, we use the notation Clinic X, Int_Y, or Obs_Y, where Int stands for interview and Obs for observation. The interview that followed an observation of a video visit was given the same number as the observation, that is, the number of the video visit.

Table 1. Comparison of patient population, implementation stages, and settings of Clinic A and Clinic B.

Aspect of the setting	Clinic A	Clinic B
Patient population	Children and adolescents with obesity (aged 2-18 years).	Adults with obesity (aged >18 years).
Responsible for and involved in the treatment	Relatives were responsible for treatment and provided an important role in its implementation. Relatives of young children visited the clinic together with the child. Follow-ups and reconciliations were made by phone with relatives of young children and not with the child. Teenage patients were assessed by the clinicians to decide whether they were mature enough to take responsibility for their own treatment. If so, the relative usually did not participate in follow-ups.	Patients were responsible for their own treatment. Relatives were not present during meetings.
Stages of implementation of video visits	Video visits began when the research study started.	6-month history of carrying out video visits.
Setting for video visits	One room was used for video visits. The room was equipped with a computer, camera, and headset. Clinicians booked the room before the video visits.	The clinicians each used their own room, with their computer equipped with camera and headset.

Table 2. Number of observations and interviews conducted at Clinic A and Clinic B with clinicians, patient, relatives, or both patient and relatives.

Method	Total, n	Clinic A, n					Clinic B, n		
		Total	Clinician	Patient	Relative	Both	Total	Clinician	Patient
Observation	13	9	6	5	3	1	4	2	4
Interview	14	10	6	—[a]	—	—	4	2	—

[a]Dashes indicate patients or relatives were not interviewed.

Results

Overview

When clinics introduced video visits, we found that a new work task emerged while preparing for these visits, because the clinicians selected patients for video visits instead of offering them to everyone. The clinicians said in the interviews that they used different arguments for including or excluding patients for video visits. In addition, implicit reasons for inclusion or exclusion were identified in the observations. These reasons were based on the clinicians' thoughts and assumptions regarding a patient's condition and needs and on the content the clinician had planned or expected for the meeting. These arguments and implicit reasons can be understood as selection criteria for helping clinicians in their decision about whether or not to offer video visits to a patient. The selection was usually conducted individually by the clinicians, but sometimes they discussed different criteria and specific patients in advance with colleagues. Such discussions were both formal (during treatment conferences) and informal (chatting during breaks).

We identified 20 selection criteria, summarized in Table 3, that were either expressed in interviews, identified during observations, or both. Often several criteria interacted with each other. We divided the identified criteria into 3 themes: practicalities, patient ability, and meeting content. Each criterion is described, without any preferred order, under its respective theme. Due to the small number of respondents, we have not been able to evaluate the degree to which each criterion is relevant. This may also be dependent on the individual patient, on the diagnosis, or on the type of care.

Practicalities

The criteria regarding practicalities are about the essential conditions that need to be fulfilled to achieve video visits. They are also related to clinicians' desire to provide a good service for their patients. For example, for patients who find it cumbersome traveling to the hospital, video visits can facilitate the start or continuation of a treatment program. The criteria regarding practicalities were used by the clinicians to make an assessment from 2 perspectives: conditions and facilitating the treatment.

From the conditions perspective (1), 3 criteria needed to be fulfilled:

a. the patient is positive about video visits
b. the patient has access to the necessary technology
c. the patient has had previous face-to-face meetings at the clinic

All 3 criteria are fairly easy to assess, as there can only be 2 answers: yes or no (eg, the patient either has the required technology or not). In 1 case, the family did not have access to a camera (1b), but they were so positive about video visits (1a) that they borrowed a tablet from a relative (Clinic A, Int_5). Furthermore, 1 clinician expressed this by saying:

> *They must be positive and have the right conditions for using the technology.* [Clinic B, Int_11]

It is important to bear in mind that having access to the technology does not imply being able to handle it (see 4d below). Regarding the technology, the clinicians need to have access to the requirements and be updated on any future changes. The third criterion (1c) is seen as a condition for being able to establish a relationship with the patient (5a), as shown below.

From the perspective of facilitating the treatment (2), 5 criteria were used to include patients for video visits. These criteria were used for patients:

a. living far away from the clinic
b. with economic issues that made it difficult to fulfill the care
c. with other illnesses or disabilities that could be affected by physical visits
d. with family, work, or school-related issues that complicated physical visits
e. with lack of time due to school, work, or medical status

These 5 criteria could be answered with either "yes" or "no" and were thus easy to assess. However, what was considered a long distance to the clinic was a subjective assessment made by the clinician. The most commonly expressed reason for why a patient (or relative) was offered a video visit was related to having a long distance to travel to the hospital (2a). The clinicians discussed how long a time it would take for the patient and relative to get to the clinic, or where the patient lived. An hour or more of traveling time 1 way for the patient was not uncommon. The distance was sometimes also connected to lack of time (2e), for example, patients not wanting to be away from school. A long distance to the hospital (2a) could also affect how positive the patient is to video visits (1a):

> *...for this family, with a long distance [to the hospital] and several other contacts with health care, the mother has been very positive about meeting in this way instead [through video visits].* [Clinic A, Int_1]

Some clinicians also mentioned stigma and how the patient might suffer from leaving school for obesity treatment (2d). Leaving a lesson for a video visit was described as being less stigmatizing than leaving school for half a day to travel to the clinic.

One patient at Clinic B was considered to "really benefit from the video visits," because he had physical difficulties getting to the clinic and had a weak immune system (2c), thus making him:

> *...very susceptible to infection [...] I think it's a good service for such a patient, to save some of his energy and force.* [Clinic B, Int_13]

Clinicians hence offered video visits if there were practical issues making it difficult for the patient to travel to the hospital but also because the clinician wanted to protect the patient from uncomfortable situations. These criteria can be seen as fairly easy to assess. The patient's fulfillment of some criteria in this theme may change over time. For example, patients might change their minds about video visits, get the right equipment, or move.

Patient Ability

The criteria regarding the patient's ability were used by the clinicians to make an assessment from 3 perspectives: the patient's well-being, mindset, and relationship with the clinician.

From the well-being perspective (3), the following criteria were used:

a. the patient is stable in weight
b. the patient is mentally stable
c. the patient does not have multiple diagnoses

Table 3. A summary of the 20 criteria divided into 3 themes.

Theme, perspective, and code	Criterion
Practicalities	
Conditions	
1a	The patient is positive about video visits
1b	The patient has access to the necessary technology
1c	The patient has had previous face-to-face meetings at the clinic
Facilitating treatment	
2a	For patients living far away from the clinic
2b	For patients with economic issues making it difficult to fulfill the care
2c	For patients with other illnesses or disabilities that could be affected by physical visits
2d	For patients with family, work, or school-related issues that complicated physical visits
2e	For patients with a lack of time due to school, work, or medical status
Patient ability	
Well-being	
3a	The patient is stable in weight
3b	The patient is mentally well
3c	The patient does not have multiple diagnoses
Mindset	
4a	The patient can take responsibility
4b	The patient has an understanding of the disease
4c	The patient can understand instructions about how to use the technology
4d	The patient is able to handle the technology
Relationship	
5a	The clinician has an established relationship with the patient
Meeting content	
Enablers	
6a	An expressed need, from patient or relative, for more frequent contact or encouragement
6b	An identified need, from clinicians, for more frequent contact, more encouragement or feedback to achieve a better treatment outcome
6c	Video communication adds something in comparison to other communication media
6d	Video communication does not include sensitive issues

Criteria 3a and 3c are easy to assess as there can only be 2 answers, "yes" or "no." Criterion 3b is more difficult to determine because if not diagnosed, it requires a subjective assessment by the clinician. For example, 1 reason given for not offering video visits to a patient who is mentally unstable was the uncertainty about what to do if the patient began to cry. The clinician wanted to have the opportunity to comfort the patient if they cried by giving them a handkerchief, a glass of water, or a pat on the shoulder.

Patients who were considered less suited for video visits were those with multiple diagnoses (3c):

> *In those cases where I feel no, it might not be relevant [with video visits] sometimes it depends on multiple diagnoses, how it will work out with the communication and instructions, and so on.* [Clinic A, Int_5]

This example also addresses criterion 4c below.

From the mindset perspective (4), the following criteria were used:

a. the patient can take responsibility
b. the patient has an understanding of the disease
c. the patient can understand instructions about how to use the technology
d. the patient is able to handle the technology

These criteria might be difficult to assess as they are subjective and depend on how well the clinician knows the patient. The assessment is therefore, to a large extent, grounded in the established relationship between patient and clinician and in the effects and progress of the treatment program. If the patient is fairly new to the clinician, and if there are other criteria that make the patient eligible for video visits, the trust in the patient's ability to take responsibility (4a) needs to be established through other means, for example, first impressions when talking to the patient (1c).

To be included for video visits, the patients needed to be able to take responsibility for choosing a suitable location for the video visit and having their weight taken in advance. In other words, new tasks were introduced for the patients. Weight is an important parameter for the clinician to assess the treatment of the patient. For patients at Clinic A, 1 condition for being offered video visits was that the school nurse could take the weight and send the figures to the clinic, where it could be written into the medical journal. In 1 of our observations, the clinician asked and explained to the patient:

> *Did you take your weight with the school nurse [...]? It's a requirement for video visits, that you go there, but the nurse cannot chase you.* [Clinic A, Obs_2]

For this specific patient, the video visits in combination with physical meetings were important to maintain the treatment; see criterion 6b below. In another example, from Clinic B, 1 of the clinicians reasoned around the patients' ability to take responsibility (4a) based on personality, and the patients' understanding of the disease (4b):

> *The patient is one of those who you cannot only have the video visit with [...], because I don't know how the treatment would turn out. However, there are of course some others [...] who have the opposite personality, who are super organized and super skilled in that way, if [a specific patient] had had a somewhat fairly reliable scale at home, [the patient] would not have had to come here to be weighed.* [Clinic B, Int_12]

Uncertainty related to criterion 4a about responsibility was expressed during a lunch break when clinicians at Clinic A discussed what to do if a patient suddenly turned off the video and left the video meeting due to being upset.

From the relationship perspective (5), 1 criterion was used:

a. the clinician has an established relationship with the patient

This criterion is subjective and depends on how well the clinician knows the patient (1c). Knowing about the patient's ability to take responsibility (4a) implies that the clinician must have a well-functioning collaboration and an established relationship (5a) with the patient:

> *Those who I have been thinking about to offer video visits, are those I know well and who have been here for quite some time. I do not think a video visit is just as useful for new visits, when you may have only met the patient a few times. I think you must have developed some kind of treatment alliance and relationship and have explained the growth curves. It's a feeling I get, and those I've chosen, these I know and feel that we have a relationship. We have some kind of treatment alliance. I think that it's difficult to build a treatment alliance via video visit, and the treatment alliance is the foundation of our treatment.* [Clinic A, Int_1]

Using video visits with new patients was seen as less suitable because the relationship between the patient and clinician had not yet been established. In all cases but 1, the clinician had met the patient (or relative) face-to-face before the video visit (1c). In the 1 case where the clinician had not met the patient, the patient had previously met with another clinician at the clinic. It was stated that the relationship between a clinician and patient is established (5a) during physical meetings. Maintenance of a well-established relationship, however, was not considered to need close physical proximity, and video visits were therefore considered an option only when the clinician and the patient had already established a relationship.

Criteria about the patient's ability can be summarized as follows: the patient having a personality suited for video visits, he or she is able to take responsibility, and there is an established relationship between the clinician and patient. Most of these criteria may change over time, for example, criterion 5a, as the relationship between clinician and patient can develop over time. Thus, this indicates that the assessment of who is considered suitable for video visits (or not) is dynamic over time.

Meeting Content

Criteria about the content of the meeting or the treatment plan concern aspects that can affect the outcome of the meeting or

the treatment in the longer run. They were used by the clinicians to make an assessment from 1 perspective: being an enabler for the treatment program.

From the enabler perspective (6), the following criteria were used:

a. an expressed need, from patient or relative, for more frequent contact or encouragement
b. an identified need, from clinicians, for more frequent contact, more encouragement or feedback to achieve a better treatment outcome
c. video communication adds something in comparison to other communication media
d. video communication does not include sensitive issues

These criteria could be seen as driving forces for using video visits or enablers of the meeting content. Criteria 6a and 6b can be identified by the patient, relative, or clinician, by needs that may grow over time, during conversations and as the relationship develops. In the following example, the clinician identified a need for more frequent contact (6b) and suggested video visits to the patient, who then expressed the same need (6a):

> *I suggested [video visits] pretty quickly and directly they were very positive about this and they were also quite clear that, yes, we would like to have a little more frequent contact and pep talk.* [Clinic A, Int_5]

The possibility of offering more frequent contact, for example, by dividing a 1-hour meeting into 4 15-min meetings, is only practically possible through long-distance communication.

In contrast to the first 2 criteria, criterion 6c may rather be something that is triggered by a need while planning for the meeting. In the following example from our data, a planned face-to-face meeting was considered to benefit from being rescheduled into a video visit (6c):

> *...one of the clinicians first thought of just calling the relative, but instead chose to offer a video visit when she realized that she could show the growth curve in orderto calm the relative, showing that the patient was doing fine.* [Clinic A, Int_10]

Another example is when a clinician expressed the difference between using video and a phone (6c):

> *When the patients are younger it feels important to engage the parents, and video visits provide the ability to involve more people in the conversation compared to phone conversations.* [Clinic A, Int_5]

A third example is when the patient and the relative lacked time to attend a meeting (2e), but the clinician still wanted to see them to grasp their interplay (6c; Clinic A, Int_6).

Sensitive issues (6d) may be needs triggered while planning for the meeting and more difficult to assess; perhaps not in terms of what may be sensitive or not, but whether a video visit is suitable for that specific issue or not. The presumption of the technical limitations of using video visits, therefore, generated selection criteria regarding what kind of consultation the clinician thought would be useful; for example, the fact that sensitivities could not be included in the treatment session using video visits. However, there seemed to be a difference in these presumptions between experienced and inexperienced clinicians. For the clinicians at Clinic B, who had conducted video visits several times during a period of 1 year, this did not seem to be an issue. One clinician said that when implementing video visits, they worried about whether the technology would work or not and, therefore, did not have as much focus on the patient. However, now while being used to the technology and trusting it to work, the clinician felt that they were able to catch the patients' mood even if they were not located in the same place (Clinic B, Int_14). Being able to include sensitive issues in the video visit could interact with criterion 4a: the patient's ability to take responsibility regarding their choice of environment. A schoolyard, for example, was not seen as a proper place for the patient when discussing certain issues.

Usually, the clinicians stated several different reasons for offering video visits, reasons that could be connected to several of the criteria: "The mother had just had a baby" (2d), "they live far away" (2a), "and we need fairly close contact. Every second or third month doesn't work here" (6b; Clinic A, Int_2). In some cases, video visits may be necessary to meet criteria 6a, 6b, and 6c; that is, if more frequent contact is needed, then it may only be possible through video visits. Whether or not to suggest video visits to a patient also depended on the content of the meeting. Some of the criteria may be identified while planning for the meeting.

Summary of Findings

Table 3 summarizes the criteria identified and described above, divided into 3 themes. There is no specific order or strength in the criteria, that is, no one is more important than the others. However, in specific situations, when the clinician is assessing the criteria, some may be seen as more valid or more important than others.

Discussion

Principal Findings

The clinicians did not consider all patients suitable for video visits, which added the new work task of selecting patients [5]. When doing so, the clinicians relied on their understanding and knowledge about the patient and his or her progress during treatment. While assessing whether a patient was suitable for video visits, the clinicians developed and used different criteria. We identified 20 criteria, which we themed under practicalities, patient ability, and meeting content. The criteria can be seen as requirements involving the patients' external and internal circumstances. External circumstances, that is, most criteria under practicalities, relate to the patients' surroundings and environment and appear easy to assess (eg, access to technical equipment and distance to the clinic). Internal circumstances, that is, most criteria under patient ability, relate to the patient's needs and abilities (eg, the patient has to be mentally well and able to manage the technology). External and internal circumstances can change over time, with the effect that the assessment may change and result in a different decision.

Assessment of the Criteria

Video visits, which involve geographical distance between clinician and patient, provide good opportunities for qualitative care [1-3,5,10-12]. We found that the criterion about the patient's distance from the clinic was a strong incentive for the patient to use video visits instead of face-to-face meetings, thereby becoming the most common inclusion criterion used by the clinicians. We also found other criteria concerning different practical conditions, the patient's abilities, and the meeting content that, to a greater or lesser extent, influenced the possibility to conduct video visits and their suitability. Some criteria could easily be assessed by a yes or no, for example, whether the patient had the required technology or not. Other criteria required a more qualitative and subjective assessment of the patient's needs and abilities, for example, a need for more frequent contact. Some assessments were based on interpretations, for example, whether the patient could understand instructions. From the clinician's perspective, knowing about the patient's needs and abilities required an established relationship with the patient.

Usually, the clinicians stated several reasons for offering video visits. Several criteria thus interacted in the assessment and during the selection process. Some criteria could be reinforced by others, and some were considered more important. For example, if the patient was not stable in weight (a strong indication for not offering video visits), lived far away from the clinic, and the clinician had identified a need for more frequent contact (2 strong indications for offering video visits), the patient might still be selected for video visits. Similarly, some criteria could exclude video visits even though other important criteria were fulfilled. For example, if the patient did not have the technical equipment required, it would not be possible to conduct video visits even if the patient wanted to and was mentally well.

The assessment was made by the clinicians, often on their own and sometimes in discussion with a colleague based on their knowledge about and trust in the patient. Both knowledge and trust are built through the relationship between clinician and patient. Trust is also related to the patient's abilities—something the clinician learns as the relationship develops. Research from other studies indicates that the perception of video visits may differ between the clinician and patient [15]—something that may coincide with the relationship.

Consequences of the Selection Process

The clinicians at both clinics chose to take responsibility for selecting patients for video visits. Our interpretation is that they did this simply because they wanted to achieve the best possible outcome from the treatment [5]. What would happen if the patient could make the choice of having video visits by themselves—not spontaneous meetings [5] but meetings as part of the treatment program? It seemed that the clinicians did not see this as an option, mainly because it could have a negative effect on treatment. However, it may be difficult for clinicians to fully understand "the patients' lives" [5] and thereby make a proper assessment. Selecting patients for video visits implies that some patients will be excluded and thus not given the same opportunity due to differing circumstances [4]. The question is how the patient's needs and desires can be taken into account without having a negative effect on the treatment. The complexity of the meetings within the treatment program makes it reasonable to think that clinicians should have a large say in deciding whether a meeting should take place face-to-face or as a video visit [5].

Some of the criteria used for deciding whether a patient is suitable for a video visit or not have to do with distance or whether it is cumbersome for the patient to travel in any other way. How do the clinicians balance between such aspects, which can be highly relevant for the patient, and other criteria that, for example, could have an effect on the content of the meeting? How do they balance between what the patient wants and what they think the patient can manage and still achieve good results in the treatment program? The clinicians need time to find this level of balance [5] to understand the transformation of roles and responsibilities [12] and how these issues affect the content of the treatment program. The selection criteria developed and used by the clinicians, together with influence from or even in collaboration with the patient, can be a tool for finding this balance, something that we expand upon below.

Practicalities

The criteria within the theme practicalities include both conditions that are necessary for conducting video visits and aspects that can be seen as enablers for the patient to fulfill or undergo the treatment at all. To a large extent, the enablers may influence the clinicians in their decision, especially if the patient lives far away from the clinic, if they cannot get away from work or school, or if there are other family-related issues that may complicate face-to-face meetings. If the incentive for the patient is strong regarding these criteria, then the clinician may need to overlook other criteria related to the patient's abilities, just to be able to proceed with the treatment.

Patient Ability

The criteria within the theme patient ability include aspects regarding the patient's well-being and mindset and the relationship between patient and clinician. Several of these criteria are closely related to trust and judgment. When video visits were introduced, responsibilities were transferred from the clinician to the patient, for example, taking their weight. The clinicians were dependent on knowing the patient's weight to judge how well the treatment was proceeding. Thus, this new responsibility for the patient influenced the clinician's ability to follow the patient's progress in the treatment program, which in turn affected the clinician's assessment of the selection criteria. The new responsibility for the patient implies that the clinician has to trust the patient's ability to fulfill this responsibility.

Using video visits, the patient's nonclinical place and space is added to the meeting [10-12]. It was the patient's responsibility to choose his or her location for the video visits, a location often unknown to the clinicians beforehand. The patient's location, including its surroundings, sometimes resulted in disturbances and limitations that in turn could affect the meeting [12]. Sometimes the clinicians had to adjust the content of a meeting to allow for disturbances and limitations situated in the patient's

location. For example, if a planned session includes sensitive issues to be discussed, these can instead be brought up in a later meeting that is conducted either face-to-face or as a video visit with the condition that the patient chooses a more suitable location. The clinicians based their selection on their knowledge of the patient's ability to choose a location with as few disturbances or limitations as possible. Hence, the patient's ability to take responsibility for their place and space influenced the content of the meeting, which in turn affected the assessment of the selection criteria.

Assessing the patient's ability to take responsibility for selecting a suitable location or taking their weight might not be easy. If the patient has an understanding of the disease, his or her ability to take on the responsibilities that come with the new role may increase. Issues related to responsibility can, however, be difficult for the clinician to communicate and for the patient to understand. Still, such aspects need to be considered by the clinician when selecting a patient for video visits. Knowing the patient is 1 condition for understanding their ability to take responsibility; an understanding that is subjective and related to several criteria. An established relationship with the patient is therefore necessary and might also be a condition for the consultation to work better [5]. Face-to-face meetings with the patient were considered necessary to establish a relationship. This might grow over time, which means that the clinician's assessment of the patient's suitability for video visits may also change over time.

Meeting Content

The criteria within the theme meeting content include needs related to more frequent contact, the possibility to add content to the meeting that cannot be added using other communication media, and sensitive issues. These criteria are more complex and are affected by issues of a more subjective character, such as trust. Either the patient or clinician could require more frequent contact or encouragement; for example, a 1-hour meeting could be divided into 4 15-min meetings. In addition, being in different locations could add new content to the meeting, for example, providing the clinician with a view into the patient's home and thereby open up the patient's private sphere [12]. This gives the clinician an opportunity to meet the patient in his or her own context, adding new details to the treatment program and to the clinician-patient relationship. Video visits could thus generate content in the meeting that would not be practically possible in face-to-face meetings. If sensitive issues were to be discussed, this could affect the clinician either in the selection process, for example, scheduling that specific meeting as a face-to-face meeting, or in the planning process, for example, bringing up the sensitive issues in the next meeting.

Using the Criteria as a Tool in the Selection Process

Introducing video visits changes ordinary work practices [5], and this in turn can affect the conduct of the meeting with a risk of introducing disturbances and limitations [5,12]. Greater requirements on the patient's abilities could reduce this risk. For example, if the patient can handle the technology, then the risk of the clinician becoming first technical support is reduced [5,12]. Similarly, if the patient can take responsibility, the risk of ending up in a meeting with a large number of unaccounted-for disturbances is reduced [5,12]. Hence, the criteria used in the selection process can reduce the number and severity of disturbances and limitations in the meeting, but the criteria can also be used to communicate the requirements they represent to the patient.

The criteria developed for and used in the selection process can, to some extent, be used to guide the clinician in how to think during this selection process. However, the reality is never easy, which means the selection process is intertwined with pros and cons when deciding whether the patient is suitable for video visits or not. Our interpretation is that the selection process is based to a large extent on the clinician's thoughts, knowledge, and assumptions regarding the patient and his or her abilities [5]. We cannot give a clear-cut answer to what would happen if the patient could decide on video visits or at least be an active part in the selection process. Greater patient involvement in the selection process may strengthen the patient's sense of responsibility and perhaps also reduce the difference in perception of the video visits [15]. Several of the criteria used in the selection process involve the patient's opinion, but it is still the clinician who makes the assessment and decides. Perhaps the patient's ability to take responsibility could be strengthened if the patient was more involved in the selection process. This could increase the patient's understanding of the new responsibilities. A first bonding, during which the clinician and patient establish a good relationship, appears to be necessary, however. Methods for involving the patient in the assessment of the criteria may also need to be developed.

Conclusions

We conclude that not all patients, adults, or children with obesity undergoing treatment programs should be offered video visits. The patient's new responsibilities of choosing a suitable place and taking their own weight could influence the content of the meeting and the progress of the treatment program. The selection criteria developed and used by the clinicians could be used as a tool for finding a balance between what the patient wants and what the clinician thinks the patient can manage and achieving good results in the treatment program. The criteria could also reduce the number and severity of disturbances and limitations in the meeting and be used to communicate the requirements they represent to the patient. Some of the criteria are based on facts; for example, if the patient does not have access to the required technology, then a video visit is not an option. Other criteria are subjective; that is, they are more formal and used rather as a checklist to avoid individual interpretations among the clinicians. A method for how and when to involve the patient in the selection process is recommended since it may strengthen the patient's sense of responsibility and relationship with the clinician.

Implications for Further Research

Further research is required to understand the full effects of selecting patients for video visits. Our study provides 1 piece of the puzzle and can guide other researchers in studying the selection task. We have focused on a small path of patient flows, a chronic disease for which the patients undergo a treatment program with several consecutive meetings. There are no data

that need to be monitored daily; only a little data collected by the patient are used to follow the progress of the treatment. The selection criteria are likely to differ in other patient flows and in other types of care programs. Due to the low number of respondents in the study, we can neither prioritize the criteria nor generalize the results. Instead, our results can be seen as something upon which further research can be built. We have not measured the specific benefits of video visits compared with face-to-face visits regarding the progress of the treatment, and we know that the values may differ between diagnosis areas. A health economics study of the benefits, for different stakeholders, would be most interesting.

Acknowledgments

The authors would like to thank the respondents at Clinic A and Clinic B for their time and for fruitful discussions. The project has been cofunded by the county council in Stockholm and Vinnova.

Conflicts of Interest

None declared.

References

1. Greenhalgh T, Vijayaraghavan S, Wherton J, Shaw S, Byrne E, Campbell-Richards D, et al. Virtual online consultations: advantages and limitations (VOCAL) study. Br Med J Open 2016 Jan 29;6(1):e009388. [doi: 10.1136/bmjopen-2015-009388]
2. Powell RE, Stone D, Hollander JE. Patient and health system experience with implementation of an enterprise-wide telehealth scheduled video visit program: mixed-methods study. JMIR Med Inform 2018 Feb 13;6(1):e10 [FREE Full text] [doi: 10.2196/medinform.8479] [Medline: 29439947]
3. Speyer R, Denman D, Wilkes-Gillan S, Chen YW, Bogaardt H, Kim JH, et al. Effects of telehealth by allied health professionals and nurses in rural and remote areas: a systematic review and meta-analysis. J Rehabil Med 2018 Feb 28;50(3):225-235 [FREE Full text] [doi: 10.2340/16501977-2297] [Medline: 29257195]
4. Huxley CJ, Atherton H, Watkins JA, Griffiths F. Digital communication between clinician and patient and the impact on marginalised groups: a realist review in general practice. Br J Gen Pract 2015 Dec;65(641):e813-e821 [FREE Full text] [doi: 10.3399/bjgp15X687853] [Medline: 26622034]
5. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-world implementation of video outpatient consultations at macro, meso, and micro levels: mixed-method study. J Med Internet Res 2018 Dec 17;20(4):e150 [FREE Full text] [doi: 10.2196/jmir.9897] [Medline: 29625956]
6. Groth K, Olin K, Gran O, Permert J. The role of technology in video-mediated consensus meetings. J Telemed Telecare 2008 Oct;14(7):349-353. [doi: 10.1258/jtt.2008.007006] [Medline: 18852315]
7. Halford S, Lotherington AT, Obstfelder A, Dyb K. Getting the whole picture? New information and communication technologies in healthcare work and organization. Inf Commun Soc 2010 Apr;13(3):442-465. [doi: 10.1080/13691180903095856]
8. Nicolini D. Stretching out and expanding work practices in time and space: the case of telemedicine. Hum Relat 2016 Apr 22;60(6):889-920. [doi: 10.1177/0018726707080080]
9. Pappas Y, Seale C. The opening phase of telemedicine consultations: an analysis of interaction. Soc Sci Med 2009 Apr;68(7):1229-1237. [doi: 10.1016/j.socscimed.2009.01.011] [Medline: 19201514]
10. Petersson J. Technospatialities and telehealthcare: unfolding new spaces of visibility. Inf Commun Soc 2015 Jul 17;19(6):824-842. [doi: 10.1080/1369118X.2015.1061575]
11. Oudshoorn N. How places matter: telecare technologies and the changing spatial dimensions of healthcare. Soc Stud Sci 2012 Feb;42(1):121-142. [doi: 10.1177/0306312711431817] [Medline: 22530385]
12. Sturesson L, Groth K. Effects of the digital transformation: qualitative study on the disturbances and limitations of using video visits in outpatient care. J Med Internet Res 2018 Jun 27;20(6):e221 [FREE Full text] [doi: 10.2196/jmir.9866] [Medline: 29950290]
13. Sabesan S, Allen D, Loh PK, Caldwell P, Mozer R, Komesaroff PA, Royal Australasian College of Physicians Telehealth Working Group. Practical aspects of telehealth: are my patients suited to telehealth? Intern Med J 2013 May;43(5):581-584. [doi: 10.1111/imj.12119] [Medline: 23668268]
14. LeRouge C, Garfield MJ, Collins RW. Telemedicine: technology mediated service relationship, encounter, or something else? Int J Med Inform 2012 Sep;81(9):622-636. [doi: 10.1016/j.ijmedinf.2012.04.001] [Medline: 22579395]
15. Radhakrishnan K, Xie B, Berkley A, Kim M. Barriers and facilitators for sustainability of tele-homecare programs: a systematic review. Health Serv Res 2016 Feb;51(1):48-75 [FREE Full text] [doi: 10.1111/1475-6773.12327] [Medline: 26119048]
16. Suchman L. Plans and Situated Actions: The Problem of Human-Machine Communication. New York, NY: Cambridge University Press; 1987.
17. Trost J, Levin I. [Att förstå vardagen: Med ett symboliskt interaktionistiskt perspektiv]. Lund: Studentlitteratur; 2010.

18. Pigg SL. On sitting and doing: ethnography as action in global health. Soc Sci Med 2013 Dec;99:127-134. [doi: 10.1016/j.socscimed.2013.07.018] [Medline: 23953135]
19. Willows T, Dizdar N, Nyholm D, Widner H, Grenholm P, Schmiauke U, et al. Initiation of levodopa-carbidopa intestinal gel infusion using telemedicine (video communication system) facilitates efficient and well-accepted home titration in patients with advanced Parkinson's disease. J Parkinsons Dis 2017 Nov;7(4):719-728 [FREE Full text] [doi: 10.3233/JPD-161048] [Medline: 28984615]
20. Ritchie J, Spencer L, O'Connor W. Carrying out qualitative analysis. In: Ritchie J, Lewis J, editors. Qualitative Research Practices: A Guide for Social Science Students Researchers. London: SAGE Publications; 2007:219-262.
21. Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? Int J Qual Stud Health Well-being 2014;9:26152 [FREE Full text] [doi: 10.3402/qhw.v9.26152] [Medline: 25326092]

Edited by G Eysenbach; submitted 17.01.18; peer-reviewed by J Moll, RK B; comments to author 05.04.18; revised version received 04.07.18; accepted 26.08.18; published 05.11.18.

Please cite as:
Sturesson L, Groth K
Clinicians' Selection Criteria for Video Visits in Outpatient Care: Qualitative Study
J Med Internet Res 2018;20(11):e288
URL: https://www.jmir.org/2018/11/e288/
doi:10.2196/jmir.9851
PMID:30401661

Original Paper

Using Video Technology to Increase Treatment Completion for Patients With Latent Tuberculosis Infection on 3-Month Isoniazid and Rifapentine: An Implementation Study

Chee Kin Lam[1,2], MPH; Kara McGinnis Pilote[1,2], MPH, MA; Ashraful Haque[2], MCH; Joseph Burzynski[2], MD, MPH; Christine Chuck[2], MPA; Michelle Macaraig[2], DrPH

[1]Centers for Disease Control and Prevention, Atlanta, GA, United States

[2]Bureau of Tuberculosis Control, New York City Department of Health and Mental Hygiene, Long Island City, NY, United States

Corresponding Author:
Chee Kin Lam, MPH
Bureau of Tuberculosis Control
New York City Department of Health and Mental Hygiene
42-09 28th Street
CN72b, WS 24-017
Long Island City, NY, 11101
United States
Phone: 1 718 310 2538
Email: clam4@health.nyc.gov

Abstract

Background: Since January 2013, the New York City (NYC) Health Department Tuberculosis (TB) Program has offered persons diagnosed with latent TB infection (LTBI) the 3-month, once-weekly isoniazid and rifapentine (3HP) treatment regimen. Patients on this treatment are monitored in-person under directly observed therapy (DOT). To address patient and provider barriers to in-person DOT, we piloted the use of a videoconferencing software app to remotely conduct synchronous DOT (video directly observed therapy; VDOT) for patients on 3HP.

Objective: The objective of our study was to evaluate the implementation of VDOT for patients on 3HP and to assess whether treatment completion for these patients increased when they were monitored using VDOT compared with that using the standard in-person DOT.

Methods: Between February and October 2015, patients diagnosed with LTBI at any of the four NYC Health Department TB clinics who met eligibility criteria for treatment with 3HP under VDOT (V3HP) were followed until 16 weeks after treatment initiation, with treatment completion defined as ingestion of 11 doses within 16 weeks. Treatment completion of patients on V3HP was compared with that of patients on 3HP under clinic-based, in-person DOT who were part of a prior study in 2013. Furthermore, outcomes of video sessions with V3HP patients were collected and analyzed.

Results: During the study period, 70% (50/71) of eligible patients were placed on V3HP. Treatment completion among V3HP patients was 88% (44/50) compared with 64.9% (196/302) among 3HP patients on clinic DOT (P<.001). A total of 360 video sessions were conducted for V3HP patients with a median of 8 (range: 1-11) sessions per patient and a median time of 4 (range: 1-59) minutes per session. Adherence issues (eg, >15 minutes late) during video sessions occurred 104 times. No major side effects were reported by V3HP patients.

Conclusions: The NYC TB program observed higher treatment completion with VDOT than that previously seen with clinic DOT among patients on 3HP. Expanding the use of VDOT may improve treatment completion and corresponding outcomes for patients with LTBI.

(J Med Internet Res 2018;20(11):e287) doi:10.2196/jmir.9825

KEYWORDS

computer-assisted therapy; directly observed therapy; mobile phone; telemedicine; videoconferencing

Introduction

In 2015, the World Health Organization, a leading public health organization, published an agenda that outlines the strategic direction to promote the integration of digital health concepts into tuberculosis (TB) prevention and care activities [1]. One digital health product identified that supports their strategy is the use of electronic observation of treatment [1]. In the United States, the use of video to remotely monitor patient treatment for active TB is rapidly growing [2]; however, the use of technology to monitor adherence to preventive treatment for latent TB infection (LTBI) has not been widely documented [3].

Nearly a quarter of the world's population is infected with TB and, left untreated, many are at risk of progressing to active TB disease [4]. An important component of the US Centers for Disease Control and Prevention's (CDC) TB elimination strategy is to expand efforts to treat individuals diagnosed with LTBI using shorter treatment regimens [5]. In 2011, CDC began recommending the use of a shorter treatment regimen, a 3-month, once-weekly regimen of isoniazid and rifapentine (3HP) under directly observed therapy (DOT), for treatment of LTBI in otherwise healthy individuals aged ≥12 years and in HIV-infected patients not taking antiretroviral medications [6,7].

In 2013, the New York City (NYC) Department of Health and Mental Hygiene (DOHMH) began offering 3HP at its TB clinics and found that treatment completion increased from a baseline of 34% with 9 months of isoniazid (9H) to 65% with 3HP [8], but it was still lower than the 82% treatment completion observed in the 3HP clinical trial [9]. Stennis et al attributed the lower than expected treatment completion to the inconvenience associated with the DOT requirement [8]. Patients in this study were treated with 3HP under in-person clinic DOT. Furthermore, among patients who chose a non-3HP treatment, 96% reported the clinic DOT requirement and 77% reported concerns about taking time away from work, child care, or other responsibilities for clinic visits as reasons they did not choose the 3HP regimen [8].

In the United States, DOT is the standard of care for monitoring patients on treatment for active TB disease, particularly those who are infectious, to ensure adherence to medication [10]. DOT requires substantial public health resources and generally is not the standard of care for patients on treatment for LTBI, a noninfectious form of TB. DOT involves trained workers observing patients ingest each dose of medication throughout the duration of treatment. DOT requires patients to either go to a clinic or have DOT workers visit patients' homes or other locations to observe medication ingestion [11]; this can be inconvenient and disruptive for patients [12,13]. Several TB programs have explored the use of videoconferencing to remotely monitor patients on treatment for active TB, known as video DOT (VDOT). These programs have reported better or equal rates of treatment completion compared with those with in-person DOT while providing a more convenient and flexible option for patients [2,12,14-16]. VDOT uses videoconferencing software to allow patients and staff to communicate remotely via smartphones, tablets, or desktop computers. An NYC study found VDOT to be a feasible alternative to in-person DOT while improving the treatment adherence and maximizing health department resources [12]. However, to date, only one published instance known to the authors has reported using VDOT to monitor patients on treatment for LTBI [3].

To improve treatment completion for patients on 3HP, the NYC DOHMH piloted the use of live-videoconferencing technology to conduct weekly DOT observations for patients on 3HP (V3HP). The intent of the V3HP pilot was to alleviate barriers to DOT to improve treatment completion among patients started on the 3HP regimen. The objectives of this evaluation were as follows: (1) to determine the feasibility of using VDOT on patients prescribed 3HP and assess resources required to implement; (2) to compare treatment completion of patients in the V3HP pilot with previously measured 3HP treatment data; and (3) to describe challenges encountered during the pilot implementation.

Methods

Integration of Video Directly Observed Therapy for Treatment of Latent Tuberculosis Infection

For the V3HP implementation study, NYC DOHMH adapted the existing videoconferencing software, educational and enrollment materials, and protocols used in the previous NYC 3HP and VDOT pilot experiences [8,12]. Clinic staff received in-service training and job aids for assessing patient eligibility and referring patients to the V3HP pilot. Three nonclinical staff were trained to perform observations for V3HP, even though one performed nearly all of the observations. In addition, staff were trained in the installation and operation of the software and basic troubleshooting. Furthermore, staff were trained in documentation procedures for monitoring patients in the implementation study.

Study Population

Eligible patients treated for LTBI with 3HP between February and October 2015 at any of the four NYC Health Department TB clinics and who met NYC DOHMH eligibility requirements for VDOT [12] were offered participation in V3HP. The diagnosis of LTBI and the prescription of 3HP were left to the discretion of providers. The eligibility for V3HP included the possession of a smartphone, tablet, or computer with videoconferencing capability; patients' willingness to use their personal devices for VDOT sessions; access to a reliable internet connection; and agreement to a VDOT schedule. Participants were followed through the completion of treatment or up to 16 weeks from treatment initiation, whichever came first. Eligible minors were enrolled at the provider's discretion if parental consent was obtained. Patients and guardians of minors signed a DOT agreement, which included the use of videoconferencing for observation sessions and acknowledgment of personal responsibility for costs incurred due to the use of personal devices and data service. Patients ineligible for or refused V3HP were still able to be treated with 3HP with in-person clinic DOT at any of the Health Department TB clinics but were excluded from the implementation study.

V3HP patients were prescribed medication, as per CDC guidelines [7], monthly at one of the four Health Department TB clinics. Patients returned to the clinic for monthly follow-up evaluation and medication refills. During these monthly visits, patients had the option of taking their medications in-person in lieu of their weekly VDOT sessions.

Process Conducting Video Sessions

3HP patients were assigned to a VDOT worker who contacted the patients to verify enrollment eligibility, schedule weekly video observation sessions, remotely assist the patients in installing the Health Department-approved videoconferencing software, and test the stability of the internet connection similar to the process in a prior NYC study [12]. During each observation session, the VDOT worker logged into the videoconference at the scheduled time using a conference identifier unique to each patient and waited for the patient to log in. Observation sessions were conducted using NYC's standard VDOT practice [12], which includes a VDOT worker asking patients at the beginning of each session if they experienced any side effects since their previous dose, and if no side effects were reported, patients were observed ingesting all prescribed medications. Patients reporting or experiencing any side effects during VDOT sessions were asked to return to the clinic or were contacted by a provider to determine the course of action. Patients were observed through the completion of therapy. No additional follow-up was performed after treatment completion or 16 weeks after treatment initiation.

Patients who failed to log in within 5 minutes of their scheduled appointment were contacted by the VDOT worker via telephone. If patients could not be reached within 30 minutes of the appointment time, a voicemail or short message service (SMS) text message was left requesting the patient to call the worker to reschedule. In addition, SMS text messaging was used to remind patients of their appointment but was used only after obtaining patient approval in accordance with the NYC DOHMH policy. The VDOT worker would attempt to call patients the following day if they had not returned the original phone call. Treatment outcomes, issues with completing VDOT observations, and other evaluation variables were documented in a V3HP database by the VDOT worker following all successful and failed VDOT sessions.

Data Collection

Patient demographics, treatment outcomes (ie, treatment completion), and information on monthly clinic visits and clinic DOT were obtained from the TB clinic's electronic medical record system. Treatment outcomes were categorized as follows: treatment completion (ie, completion of treatment using 3HP on VDOT), lost (ie, unable to locate after treatment initiation), refused treatment, switched treatment types, discontinued due to side effects per physician advice, and other (eg, moved). Duration of the VDOT observation sessions, outcomes of the sessions, issues encountered during sessions, and other comments pertinent to therapy sessions or failed attempts to contact patients were obtained from the V3HP database. Issues encountered during the sessions were captured as predefined codes and free text by the VDOT staff.

Definitions

Patients were considered to have completed treatment successfully if they received at least 11 doses of 3HP within 16 weeks of treatment initiation. Issues were categorized into adherence, medical, and technical. Adherence issues consisted of 4 subcategories as follows: patient lateness (defined as >15 minutes late to a scheduled session); patient lateness for more than a day; missing or lost medications; and unapproved self-administered doses. Medical issues were included if patients reported side effects to a VDOT worker or other clinical staff or if they were documented in the electronic medical record. Technical issues were subdivided into the following 3 categories: DOHMH error, including health department computer or phone connection errors, videoconference software crashes, and audio or visual hardware malfunctions; patient equipment error, including connection difficulties, software errors, and hardware errors; and patient knowledge, including inability to operate phone or software and misunderstanding of observation requirements for VDOT.

Analysis

The characteristics and treatment outcomes of patients in the V3HP implementation study were compared with those of 3HP patients on in-person clinic DOT who were part of an earlier NYC study implemented from January to November 2013 [8]. Participants for both studies were enrolled at NYC Health Department clinics and were included if they met the following criteria: patients being treated for LTBI; those aged ≥12 years; males or nonpregnant, nonnursing females; HIV-uninfected or -infected individuals who were not on highly active antiretroviral medications, and patients who could be contacted via telephone in case of a missed DOT visit. The significant differences in demographics and treatment outcomes between patients on 3HP with VDOT and those on clinic DOT were calculated using Pearson's chi-square or Fisher's exact test for categorical variables and Wilcoxon rank-sum test for continuous variables.

This implementation study was considered a public health program evaluation activity, not research, and, therefore, it did not meet the criteria to undergo review by the NYC DOHMH Institutional Review Board. Furthermore, this project was reviewed and approved at the CDC as program evaluation activity.

Results

Treatment Outcomes

From February to October 2015, 70% (50/71) of patients who initially agreed to V3HP were placed on VDOT. Among the V3HP patients, 88% (44/50) completed their treatment on 3HP under VDOT (Figure 1); 6% (3/50) of the additional patients completed treatment after switching to a non-3HP treatment regimen following 1-2 VDOT sessions. Of 3 patients who did not complete treatment, 2 patients opted to discontinue the treatment after experiencing headache and dizziness, respectively; 1 patient moved out of the jurisdiction after completing a single VDOT session and was referred for follow-up in the other jurisdiction. Furthermore, 21 patients

who initially agreed to V3HP subsequently did not start on V3HP for various reasons (Figure 1).

There were few differences in patient demographics between V3HP patients and patients in the prior NYC 3HP study who were monitored under clinic DOT, although a higher proportion of V3HP patients were recently exposed to an infectious TB patient (Table 1). Treatment completion for V3HP patients was higher than that for 3HP patients on clinic DOT (44/50, 88%, vs 196/302, 64.9%; *P*<.001) [8].

Video Directly Observed Therapy Sessions

Of 549 3HP treatment doses ingested by 50 V3HP patients, 65.6% (360/549) were observed under VDOT, 30.4% (167/549) doses were observed in the clinic by staff, and 4.0% (22/549) were self-administered. Patients had a median of 8 VDOT (range: 1-11) sessions. In addition, 42 patients completed 3HP treatment with 12 doses of medication and 2 patients received physician approval to discontinue therapy after 11 doses. Session times were captured for 95.8% (345/360) of VDOT sessions. The median session time was 5 (range: 1-59) minutes.

Figure 1. Outcomes of patients on 3-month, once-weekly treatment with isoniazid and rifapentine referred for video directly observed therapy (DOT; V3HP). MD: medical doctor.

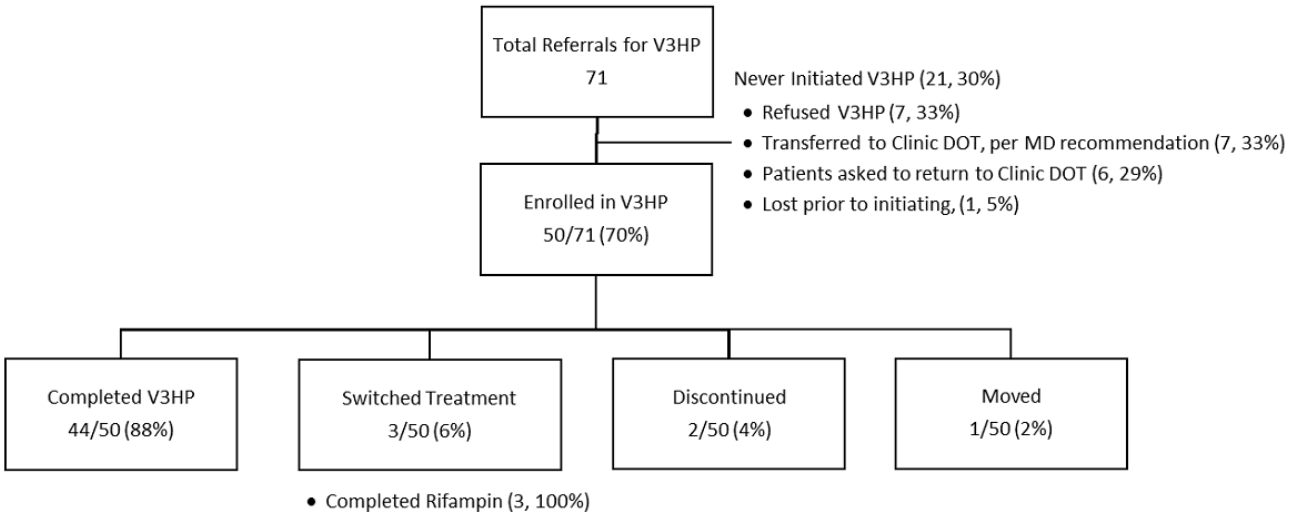


Table 1. Characteristics and treatment outcomes of patients on 3-month, once-weekly treatment with isoniazid and rifapentine (3HP) on clinic directly observed therapy (DOT; January to November 2013) versus video directly observed therapy (VDOT; February to October 2015) in New York City.

Characteristic	3HP Clinic DOT (n=302)	3HP VDOT (n=50)	*P* value[a]
Male sex, n (%)	154 (51.0)	25 (50.0)	.90
Age (years), median (interquartile range)	33 (22-45)	33.5 (25-46)	.51
US born, n (%)	70 (23.2)	7 (14.0)	.15
HIV status, n (%)			.33
Positive	1 (0.3)	0 (0.0)	
Negative	161 (53.3)	32 (64.0)	
Unknown	140 (46.4)	18 (36.0)	
Tuberculosis risk category, n (%)			.02[b]
Population Risk	187 (61.9)	27 (54.0)	
Medical Risk	56 (18.5)	8 (16.0)	
Contact to an active TB case	42 (13.9)	15 (30.0)	
Other	17 (5.6)	0 (0.0)	
Treatment outcomes, n (%)			.001[b]
Completed Treatment	196 (64.9)	44 (88.0)	
Did Not Complete	106 (35.1)	6 (12.0)	

[a]*P* value calculated using the Pearson's chi-square or Fisher's exact test for categorical variables and Wilcoxon rank-sum test for continuous variables.
[b]Significance at *P*<0.05.

Table 2. Issues encountered during the implementation study (n=205).

Issue type	Value, n (%)
Adherence	104 (50.7)
Unapproved self-administer	12 (11.5)
Patient misplaced or forgot meds	5 (4.8)
Patient late >1 day	15 (14.4)
Patient late >15 minutes	72 (69.2)
Technical	75 (36.6)
Health department related	29 (38.7)
Patient equipment	43 (57.3)
Patient knowledge	3 (4.0)
Medical	26 (12.7)

Textbox 1. Justifications for self-administered doses of 3-month, once-weekly isoniazid and rifapentine (3HP) regimen (n=22).

Unapproved self-administered justifications (n=12)

- Unable to reach patient. Self-administered prior to the callback: 6.
- The patient was to be observed in the clinic. The patient self-administered instead: 2.
- Technical issue. The patient self-administered rather than awaiting troubleshooting: 2.
- Administered prior to initial contact by the pilot staff: 1.
- The patient went on vacation without prior notice and did not have a video-enabled device: 1.

Approved self-administered justification (n=10)

- Administrative or holiday closure of office. Unable to schedule alternate time with the patient: 3.
- Technical issue. The patient thought observation was underway, but the video was nonfunctional: 3.
- Physician excused absence
  - Patient overseas and unable to connect: 3
  - Patient away on a meditation: 1

A total of 205 issues were encountered during the V3HP pilot among 47 patients (Table 2); 76.6% (157/205) occurred during 149 unique VDOT sessions. The remaining issues were side effects reported in-person during monthly clinic follow-ups, problems resulting in clinic DOT visits, and instances where patients self-administered medication was not under observation. Of all the issues, 50.7% (104/205) were related to patient adherence, including 12 instances where patients self-administered treatment without prior physician or program staff approval (Textbox 1). There were 26 medical issues, most of which were reported within the first 6 doses of medication (n=20). Of 37.1% (76/205) technical issues identified, a majority resulted from patient equipment errors (n=43). Health Department-related equipment errors (n=29) typically occurred in the beginning of the pilot and earlier in patients' treatment course.

Discussion

Principal Considerations

This implementation study examined over 300 VDOT sessions among 50 patients on 3HP. Our analysis found that the 3HP treatment completion for patients in the implementation study increased compared with that in a prior NYC study that offered 3HP with clinic DOT (196/302, 64.9%, vs 44/50, 88%). Our evaluation supports the inclusion of VDOT to improve the completion of therapy with 3HP. Although patient nonadherence was prominent during the pilot period, with nearly half of the scheduled VDOT sessions having some form of adherence issue, the implementation study still demonstrated that staffing needs were minimal to account for the variable rescheduling time for monitoring nonadherent patients and providing reminders calls and SMS text messages when patients were late. In this implementation study, a single VDOT worker managed all observation sessions for 50 patients. Furthermore, technical issues did not prohibit the continuation of the observation sessions and the completion of treatment. This suggests that VDOT can successfully monitor patients on 3HP using minimal health department resources while offering an effective alternative for treating LTBI that removes some of the barriers to treatment completion.

In spite of these circumstances, the occurrence of patients self-administering doses remained low. A variety of causes

resulted in self-administered doses, including nearly half that were approved absence by a physician or NYC staff (Textbox 1). The occurrence of self-administered doses is anticipated, and the minimal unexcused absence adds to its acceptability as an option for patient-centered care.

A recent clinical trial by Belknap et al found that in the United States, 3HP under self-administration was noninferior to 3HP under DOT [17]; however, further evaluation is needed under program settings. Therefore, the wider use of VDOT for monitoring patients on 3HP may contribute toward efforts to more rapidly reduce TB in the United States by increasing treatment completion and preventing disease.

Strengths and Limitations

This V3HP implementation study was successfully implemented in NYC by integrating two existing programs—VDOT for monitoring patients on treatment for active TB and the 3HP short-course treatment regimen, recommended to be administered with DOT [8,12]. Staff experienced with the two initiatives were consulted to inform the implementation plan, and the few technological issues encountered were easily resolved because staff could quickly identify and address problems. The V3HP pilot required one staff person working part-time to conduct the VDOT sessions for all 50 patients enrolled during the 8-month pilot period. Furthermore, we found that patients were willing to use their own phones for VDOT sessions, a potential cost saving for health departments. However, additional evaluations including cost-effectiveness analyses comparing VDOT with in-person DOT or self-administered treatment would help quantify the value to TB programs.

This pilot also had a number of limitations. One limitation was the use of data from a prior study as the comparison group. While there were few differences between the study population for the V3HP pilot and the previous 3HP study, the V3HP pilot did enroll a greater proportion of high-risk patients with recent contact to someone with active TB disease; these high-risk individuals may have been more motivated to adhere to treatment and could have impacted treatment completion. In addition, the 2-year time difference between the previous study and the current V3HP pilot may have given clinicians an increased level of comfort in offering 3HP and less likely to discontinue the treatment because of mild or anticipated side effects. No other programmatic changes were identified between the two time periods that could have impacted the clinic population and influenced treatment completion among V3HP patients.

Enrollment in the program also presented several limitations. First, patients had to be available for VDOT sessions during limited business hours (ie, between 8:00 AM and 4:30 PM), so patients who preferred to take their medications in the early morning, late evening, or weekend could not participate. Patients opting for clinic DOT had a wider window of time to receive their weekly dose because several NYC Health Department TB clinics offer weekend and evening hours. Use of asynchronous video technology (technology that can record and timestamp video) could alleviate the scheduling constraint of live-video DOT [18]. Second, not only did patients have to use their own videoconferencing-enabled devices, but their devices also had to be compatible with the videoconferencing software approved by the DOHMH, and they were required to demonstrate the ability to use the software. Patients having a device may represent a population more inclined to accept treatment and complete through the use of video monitoring. Providing additional software options or loaner devices may make the supervision of treatment more convenient for patients and reach a wider population. Finally, patients still had to travel to the clinic for monthly follow-up visits to receive their medication. Aside from being a deterrent to patients who may not want to have to return to the clinic to receive medication, this may also have had a positive effect on treatment completion. However, as participants were able to receive medications in-person during those visits, this decreases the number of opportunities for VDOT observations. It likely accounted for the median of 8 DOT sessions per participant; the remaining observations were likely performed in-person. Further evaluation would be necessary to assess treatment completion when the number of clinic visits is reduced.

Additionally, as eligibility for and offering of 3HP is not collected as part of the program practice, the data were not available for analysis; this limits our ability to quantify the acceptance of 3HP and whether VDOT potentially increased its acceptance among patients. While there was a high proportion (21/71, 30%) of patients who did not start on V3HP, it is uncertain how these missing data affect the treatment outcome. Thus, additional studies are needed to assess DOT monitoring preferences for LTBI.

Concurrent to the pilot, a policy change in the NYC TB clinics preferentially offered another short-course treatment regimen, 4 months of daily rifampin, as an alternative to 3HP for the treatment of LTBI, which could have impacted the enrollment into the pilot. Furthermore, a shortage of rifapentine interrupted providers from offering 3HP for approximately 2 months. The overall impact of this shortage is difficult to determine because the shortage was initially not reported, but clinicians may have been aware and restricted their offering of the regimen. These events prevented the analysis to determine whether the treatment initiation with 3HP increased with VDOT. However, in taking a patient-centered approach to care, several treatment options may give patients alternatives to achieve better outcomes. Additional analysis is needed to determine whether preferentially offering of multiple short-course treatment regimens increases treatment initiation for LTBI as well as treatment completion.

Conclusions

This evaluation shows that the use of VDOT with 3HP for the treatment of LTBI is feasible and could be integrated into the current NYC LTBI treatment practice with minimal disruption to staff time and training. Treatment completion of patients on 3HP for LTBI increased with the use of VDOT. VDOT addressed some of the barriers to in-person DOT for patients with LTBI. Programs looking to implement 3HP for the treatment of LTBI should consider evaluating the use of VDOT. Further research is necessary to assess the use of VDOT for patients on treatment with 3HP compared with self-administration in a programmatic setting. Additionally, it

may be worth exploring the expansion of the use of asynchronous videoconferencing technology, thereby further reducing the intrusiveness of DOT.

Acknowledgments

The authors acknowledge the contributions of the V Vazquez-Stewart, N Mitropoulos, and G Henry. The authors would also like to thank the physicians, nurses, public health advisors, and other staff at the NYC DOHMH for their efforts in managing the patients in this pilot.

Conflicts of Interest

None declared.

References

1. Falzon D, Timimi H, Kurosinski P, Migliori GB, Van Gemert W, Denkinger C, et al. Digital health for the End TB Strategy: developing priority products and making them work. Eur Respir J 2016 Jul;48(1):29-45 [FREE Full text] [doi: 10.1183/13993003.00424-2016] [Medline: 27230443]
2. Macaraig M, Lobato MN, McGinnis Pilote K, Wegener D. A National Survey on the Use of Electronic Directly Observed Therapy for Treatment of Tuberculosis. J Public Health Manag Pract 2017 Jul 07;24(6):567-570. [doi: 10.1097/PHH.0000000000000627] [Medline: 28692611]
3. Holzschuh EL, Province S, Johnson K, Walls C, Shemwell C, Martin G, et al. Use of Video Directly Observed Therapy for Treatment of Latent Tuberculosis Infection - Johnson County, Kansas, 2015. MMWR Morb Mortal Wkly Rep 2017 Apr 14;66(14):387-389 [FREE Full text] [doi: 10.15585/mmwr.mm6614a3] [Medline: 28406884]
4. Houben RMGJ, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. PLoS Med 2016 Oct;13(10):e1002152 [FREE Full text] [doi: 10.1371/journal.pmed.1002152] [Medline: 27780211]
5. Centers for Disease Control and Prevention. US Department of Health and Human Services. Division of Tuberculosis Elimination Strategic Plan 2016-2020 URL: https://www.cdc.gov/tb/about/strategicplan.htm [WebCite Cache ID 6tTkqbTI5]
6. Centers for Disease Control and Prevention, American Thoracic Society. Targeted tuberculin testing and treatment of latent tuberculosis infection. MMWR Recomm Rep 2000 Jun 09;49(RR-6):1-51 [FREE Full text] [Medline: 10881762]
7. Centers for Disease Control and Prevention. Recommendations for use of an isoniazid-rifapentine regimen with direct observation to treat latent Mycobacterium tuberculosis infection. MMWR Morb Mortal Wkly Rep 2011 Dec 09;60(48):1650-1653 [FREE Full text] [Medline: 22157884]
8. Stennis NL, Burzynski JN, Herbert C, Nilsen D, Macaraig M. Treatment for Tuberculosis Infection With 3 Months of Isoniazid and Rifapentine in New York City Health Department Clinics. Clin Infect Dis 2016 Jan 01;62(1):53-59. [doi: 10.1093/cid/civ766] [Medline: 26338781]
9. Sterling TR, Villarino ME, Borisov AS, Shang N, Gordin F, Bliven-Sizemore E, TB Trials Consortium PREVENT TB Study Team. Three months of rifapentine and isoniazid for latent tuberculosis infection. N Engl J Med 2011 Dec 08;365(23):2155-2166. [doi: 10.1056/NEJMoa1104875] [Medline: 22150035]
10. Centers for Disease Control and Prevention. National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Division of Tuberculosis Elimination. 2013. Core curriculum on tuberculosis: what the clinician should know URL: https://www.cdc.gov/tb/education/corecurr/pdf/corecurr_all.pdf [WebCite Cache ID 73GeORpNK]
11. Centers for Disease Control and Prevention. Essential components of a tuberculosis prevention and control program. Recommendations of the Advisory Council for the Elimination of Tuberculosis. MMWR Recomm Rep 1995 Sep 08;44(RR-11):1-16 [FREE Full text] [Medline: 7565539]
12. Chuck C, Robinson E, Macaraig M, Alexander M, Burzynski J. Enhancing management of tuberculosis treatment with video directly observed therapy in New York City. Int J Tuberc Lung Dis 2016 May;20(5):588-593. [doi: 10.5588/ijtld.15.0738] [Medline: 27084810]
13. Yellappa V, Lefèvre P, Battaglioli T, Narayanan D, Van der Stuyft P. Coping with tuberculosis and directly observed treatment: a qualitative study among patients from South India. BMC Health Serv Res 2016 Dec 19;16:283 [FREE Full text] [doi: 10.1186/s12913-016-1545-9] [Medline: 27430557]
14. Gassanov MA, Feldman LJ, Sebastian A, Kraguljac MJ, Rea E, Yaffe B. The use of videophone for directly observed therapy for the treatment of tuberculosis. Can J Public Health 2013 May 14;104(3):e272. [Medline: 23823897]
15. Wade VA, Karnon J, Eliott JA, Hiller JE. Home videophones improve direct observation in tuberculosis treatment: a mixed methods evaluation. PLoS One 2012;7(11):e50155 [FREE Full text] [doi: 10.1371/journal.pone.0050155] [Medline: 23226243]
16. DeMaio J, Schwartz L, Cooley P, Tice A. The application of telemedicine technology to a directly observed therapy program for tuberculosis: a pilot project. Clin Infect Dis 2001 Dec 15;33(12):2082-2084. [doi: 10.1086/324506] [Medline: 11698993]

17. Belknap R, Holland D, Feng P, Millet J, Caylà JA, Martinson NA, TB Trials Consortium iAdhere Study Team. Self-administered Versus Directly Observed Once-Weekly Isoniazid and Rifapentine Treatment of Latent Tuberculosis Infection: A Randomized Trial. Ann Intern Med 2017 Nov 21;167(10):689-697 [FREE Full text] [doi: 10.7326/M17-1150] [Medline: 29114781]
18. Garfein RS, Collins K, Muñoz F, Moser K, Cerecer-Callu P, Raab F, et al. Feasibility of tuberculosis treatment monitoring by video directly observed therapy: a binational pilot study. Int J Tuberc Lung Dis 2015 Sep;19(9):1057-1064 [FREE Full text] [doi: 10.5588/ijtld.14.0923] [Medline: 26260824]

Abbreviations

3HP: 3-month, once-weekly isoniazid and rifapentine regimen
9H: 9-month daily isoniazid regimen
CDC: United States Centers for Disease Control and Prevention
DOHMH: Department of Health and Mental Hygiene
DOT: directly observed therapy
LTBI: latent tuberculosis infection
NYC: New York City
SMS: short message service
TB: tuberculosis
VDOT: video directly observed therapy
V3HP: video directly observed therapy for patients on a 3-month, once-weekly isoniazid and rifapentine regimen

Edited by G Eysenbach; submitted 16.01.18; peer-reviewed by R Moro, D Falzon, A DeLuca; comments to author 05.04.18; revised version received 31.05.18; accepted 28.06.18; published 20.11.18.

Please cite as:
Lam CK, McGinnis Pilote K, Haque A, Burzynski J, Chuck C, Macaraig M
Using Video Technology to Increase Treatment Completion for Patients With Latent Tuberculosis Infection on 3-Month Isoniazid and Rifapentine: An Implementation Study
J Med Internet Res 2018;20(11):e287
URL: https://www.jmir.org/2018/11/e287/
doi:10.2196/jmir.9825
PMID:30459146

Original Paper

Feasibility, Acceptability, and Adoption of Digital Fingerprinting During Contact Investigation for Tuberculosis in Kampala, Uganda: A Parallel-Convergent Mixed-Methods Analysis

Elizabeth B White[1], AB; Amanda J Meyer[1,2], MPH; Joseph M Ggita[2], BSc; Diana Babirye[2], MS; David Mark[2], BA; Irene Ayakaka[2], MBChB, MIPH; Jessica E Haberer[3,4], MS, MD; Achilles Katamba[2,5], MBChB, PhD; Mari Armstrong-Hough[1,2*], MPH, PhD; John Lucian Davis[1,2,6*], MD

[1]Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT, United States

[2]Uganda Tuberculosis Implementation Research Consortium, Makerere University, Kampala, Uganda

[3]Massachusetts General Hospital Center for Global Health, Boston, MA, United States

[4]Harvard Medical School, Boston, MA, United States

[5]Clinical Epidemiology Unit, Department of Medicine, College of Health Sciences, Makerere University, Kampala, Uganda

[6]Pulmonary, Critical Care, and Sleep Medicine Section, Yale School of Medicine, New Haven, CT, United States

[*]these authors contributed equally

Corresponding Author:
John Lucian Davis, MD
Department of Epidemiology of Microbial Diseases
Yale School of Public Health
Laboratory of Epidemiology and Public Health
60 College Street
New Haven, CT, 06510
United States
Phone: 1 (203) 785 2867
Email: lucian.davis@yale.edu

Abstract

Background: In resource-constrained settings, challenges with unique patient identification may limit continuity of care, monitoring and evaluation, and data integrity. Biometrics offers an appealing but understudied potential solution.

Objective: The objective of this mixed-methods study was to understand the feasibility, acceptability, and adoption of digital fingerprinting for patient identification in a study of household tuberculosis contact investigation in Kampala, Uganda.

Methods: Digital fingerprinting was performed using multispectral fingerprint scanners. We tested associations between demographic, clinical, and temporal characteristics and failure to capture a digital fingerprint. We used generalized estimating equations and a robust covariance estimator to account for clustering. In addition, we evaluated the clustering of outcomes by household and community health workers (CHWs) by calculating intraclass correlation coefficients (ICCs). To understand the determinants of intended and actual use of fingerprinting technology, we conducted 15 in-depth interviews with CHWs and applied a widely used conceptual framework, the Technology Acceptance Model 2 (TAM2).

Results: Digital fingerprints were captured for 75.5% (694/919) of participants, with extensive clustering by household (ICC=.99) arising from software (108/179, 60.3%) and hardware (65/179, 36.3%) failures. Clinical and demographic characteristics were not markedly associated with fingerprint capture. CHWs successfully fingerprinted all contacts in 70.1% (213/304) of households, with modest clustering of outcomes by CHWs (ICC=.18). The proportion of households in which all members were successfully fingerprinted declined over time (ρ=.30, P<.001). In interviews, CHWs reported that fingerprinting failures lowered their perceptions of the quality of the technology, threatened their social image as competent health workers, and made the technology more difficult to use.

Conclusions: We found that digital fingerprinting was feasible and acceptable for individual identification, but problems implementing the hardware and software lead to a high failure rate. Although CHWs found fingerprinting to be acceptable in principle, their intention to use the technology was tempered by perceptions that it was inconsistent and of questionable value. TAM2 provided a valuable framework for understanding the motivations behind CHWs' intentions to use the technology. We

emphasize the need for routine process evaluation of biometrics and other digital technologies in resource-constrained settings to assess implementation effectiveness and guide improvement of delivery.

(J Med Internet Res 2018;20(11):e11541) doi:10.2196/11541

KEYWORDS

biometrics; mHealth; mobile phone; tuberculosis

Introduction

The ability to uniquely identify individuals in health care settings is important for patient care, health system monitoring, and health research. For patients, unique identifiers may facilitate continuity of care, linking of encounters into a longitudinal health record, and prevention of errors during treatment. For health systems, these linkages provide richer evidence for monitoring and evaluation than aggregated data [1]. In clinical and public health research, unique identification helps preserve the integrity of data and protects against misclassification [2]. In resource-constrained settings, however, there are many barriers to unique patient identification: lack of national identification systems, inconsistent spelling of names, uncertainty about date of birth, continually changing phone numbers, a lack of street addresses, and intentional avoidance of identification procedures to escape stigma. A reliable identification method that circumvents these barriers could improve data accuracy and patient retention in care in resource-constrained settings.

Biometric identification techniques offer a novel and appealing solution to these challenges in settings where other identification methods are not feasible or acceptable. Biometric methods rely on an individual's physical characteristics, such as fingerprints, facial structure, iris geometry, or actions, including handwriting or gait pattern [3]. A number of biometric identifiers, including fingerprint and ocular characteristics, have demonstrated technical feasibility in various studies [4]. However, fingerprint scanning has become the most widely used because of the development and widespread availability of portable, low-cost technologies for digital capture [2] and its high sensitivity and specificity for verification [5]. Others have reported that fingerprinting is feasible [2,5-9] and acceptable [9,10]. However, few published reports exist regarding the actual use of fingerprinting technologies in resource-constrained settings. Therefore, we sought to perform a detailed process evaluation of digital fingerprint scanning by community health workers (CHWs) in urban Uganda to understand the feasibility, acceptability, and adoption of this technology for patient identification [11]. Additionally, we sought to better understand the determinants of CHWs' intended and actual use of fingerprint scanning technology by applying a widely used conceptual framework, the Technology Acceptance Model 2 (TAM2) [12].

Methods

Study Design, Objectives, Setting, and Population

We conducted a parallel-convergent, mixed-methods study of digital fingerprinting in the context of a household-randomized trial of enhanced tuberculosis (TB) contact investigation. Specifically, the trial (called the parent study) sought to evaluate the effects of home sputum collection and short message service text messaging on completion of evaluation for TB among household contacts living with index TB patients. This substudy sought to determine the *feasibility* of digital fingerprinting as measured by the proportion of participants and households successfully identified via fingerprints at baseline and follow-up; to describe the reasons for not capturing fingerprints; and to ascertain the technology's *acceptability* in principle and *adoption* in practice among CHWs with experience using it.

The parent study was conducted in Kampala, Uganda, from July 2016 to July 2017. In the parent study, we used digital fingerprinting to avoid duplicate registrations of index patients and contacts and to verify follow-up visits at clinics for those needing additional evaluation. Those referred for follow-up evaluation at the clinic included contacts who were persons living with HIV; those who had tuberculosis symptoms but did not produce a sputum sample at the household visit; and those who had an inconclusive diagnostic result for sputum collected during the home visit. All others were not referred for a follow-up visit. In this substudy, we analyzed quantitative data from participants enrolled in the parent study and qualitative data from interviews with CHWs who carried out digital fingerprinting and other study procedures. Children aged <5 years were not eligible for scanning because digital fingerprints are difficult to capture and less accurate in young children [13,14].

Study Procedures

Prior to implementation, all CHWs completed a course introducing the rationale for the use of fingerprints as biometric identifiers, describing different fingerprint patterns, and training them to capture high-quality fingerprints using a digital scanner. CHWs participated in hands-on training, including "role-play" sessions that allowed them to practice acquiring good-quality fingerprints and troubleshooting commonly encountered problems with fingerprint scanning. All CHWs were trained in infection control practices prior to initiating their work and provided with disposable personal protective equipment to protect them during patient encounters. CHWs performed digital fingerprinting and collected individual age, sex, and self-reported HIV status from household members during contact investigation visits. Fingerprinting was performed using multispectral fingerprint scanners (Lumidigm M301, HID Global, Austin, TX, USA) linked to embedded matching software (Biometrac, Louisville, KY, USA). Matching was available offline and fully integrated as an application programming interface within a customized survey app (CommCare, Dimagi, Boston, MA, USA). The app logged each health worker and time-stamped each encounter. Data were

uploaded to a cloud-based server (CommCareHQ, Dimagi). Fingerprint images were not stored but instead recorded as a series of unique characters decipherable only using a secured, proprietary algorithm.

Quantitative Analysis

For individual contacts, the outcome of interest was the failure to record a complete fingerprint scan in the database, categorized as a binary outcome. A complete scan required successful imaging of the fingerprint with sufficient clarity and resolution to allow adequate feature extraction; scans failing to meet quality criteria (eg, because of degraded ridges, dirt, or fingerpad placement excluding the fingerprint core) were immediately rejected. A complete scan required capture of right and left thumbprints, followed by right and left index fingerprints; any scan that failed to capture all four fingerprints was deemed unsuccessful. Although fingerprinting is an individual procedure, it is frequently offered to multiple household members on a single hardware device during a household visit for contact investigation. To reflect these conditions, we also defined failure at the level of the household encounter; any encounter that did not capture fingerprints from all present household contacts was deemed unsuccessful. If a household required multiple visits to enroll all contacts, we included only the first household encounter in our analyses. Two investigators (EBW and DB) independently reviewed free text explanations from CHWs for fingerprinting failures and classified each as a hardware problem, a software problem, or as another unclassified problem.

We described the population characteristics of individual study participants, including age, sex, and HIV status, as well as characteristics of households, including which CHW captured fingerprints and the time period of enrollment. We examined differences in success by age, using the standard categories employed by the World Health Organization Stop TB Department (5-14 years and ≥15 years); sex; and HIV status. We examined the trend in fingerprinting success over time by the quarter of study enrollment by calculating Spearman rho. In addition, we examined differences in household-level fingerprinting success by CHWs using chi-square test. To test associations between individual characteristics and fingerprinting success, we fit bivariate logistic regression models using generalized estimating equations and a robust covariance estimator to account for clustering by household. We report *P* values based on cluster-robust standard errors (SEs). To estimate the extent of clustering of outcomes by household and CHW, we calculated intraclass correlation coefficients (ICCs).

Qualitative Interview Procedures

During the last 2 months of the study, we carried out parallel in-depth interviews with each of the 15 CHWs who conducted study procedures using a semistructured interview guide. We developed the interview guide to elicit responses related to 3 overarching topics as follows: the CHWs' first interactions with digital fingerprinting; their experiences using digital fingerprinting during the study; and their opinions regarding the usability of digital fingerprinting. The guide was developed in English and is reported in Multimedia Appendix 1. One English-speaking investigator (EBW) interviewed all 15 CHWs who conducted study procedures. All but one reported feeling comfortable completing the interview in English; a native Luganda-speaking investigator (JMG) reinterviewed this CHW in Luganda to give the respondent the opportunity to elaborate on experiences and opinions in his or her native language. During the interview, each CHW was also asked to mock-fingerprint the interviewer as a means of eliciting the user's experiences and interactions with digital fingerprinting. All interviews were recorded, transcribed, and uploaded to a secure Web-based server for qualitative data analysis (Dedoose, Manhattan Beach, CA, USA). In addition, interviewers used a structured debriefing form (Multimedia Appendix 2) to organize emergent themes immediately following each interview. Additional details were added iteratively after reviewing interview recordings and transcripts.

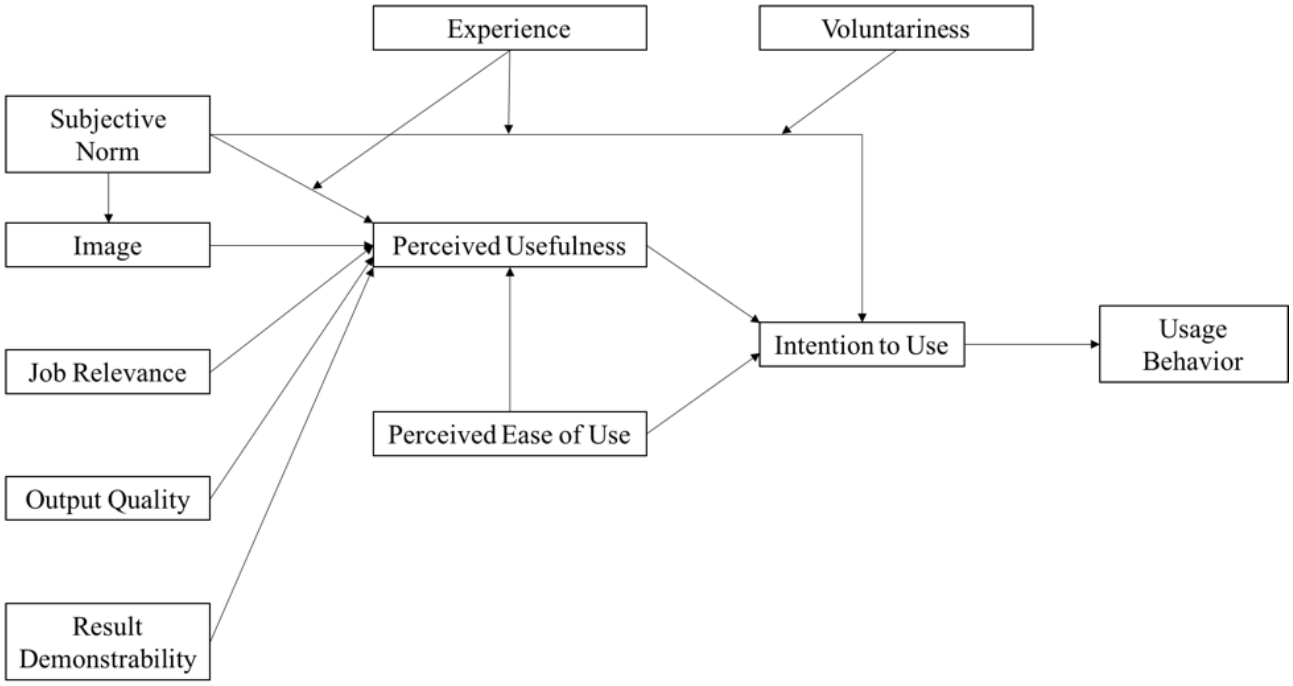
Qualitative Analysis

We carried out the qualitative analysis using the debriefing forms to identify key themes [15]. Using the TAM2 framework, one investigator (EBW) categorized themes into prespecified antecedents of "behavioral intention" to use fingerprinting technology (Figure 1). TAM2 theorizes that behavioral intention precedes and predicts actual use. Behavioral intentions are influenced by perceptions of the technology's usefulness and ease of use. Five domains independently contribute to the perceived usefulness of a technology: the perception that important others expect one to use the technology (*subjective norm*); the perception that social status is enhanced through its use (*image*); the perception that the technology supports an important job function (*job relevance*); the performance of the technology (*output quality*); and tangible results of its use (*result demonstrability*) [12,16-19].

Human Subjects Considerations

Each participant or the parent or guardian of minors provided written informed consent as part of the parent study. Furthermore, participants aged 8-17 years provided written assent. For this substudy, CHWs provided verbal consent prior to the interview. Institutional review boards at the Makerere College of Health Sciences, the Uganda National Council for Science and Technology, and Yale University approved the study protocol.

Figure 1. Technology Acceptance Model 2; adapted from Venkatesh and Davis.



Results

Study Population and Results of Quantitative Analysis

Of all household contacts eligible for the parent study, 75.5% (694/919) of individuals aged ≥5 years were eligible for digital fingerprinting (Figure 2). Of those eligible, 74.2% (515/694) had a successful fingerprint scan during the household visit. Of the contacts without successful fingerprint scans during the household visit, 60.3% (108/179) of fingerprint scan failures were classified as software problems, 36.3% (65/179) as hardware problems, and 3.4% (6/179) as unclassified problems; none were classified as refusals. We found similar baseline fingerprinting success rates and failure reasons among index patients; because these were individual data collected separately and in a clinic setting, we have reported them separately in Multimedia Appendix 3. Only 3% (1/32) of the contacts fingerprinted at the household visit and referred to the clinic for evaluation were identified via fingerprint at the follow-up visit. Among individual contacts, clustering of unsuccessful scans by household was extensive (ICC=.99). Household contacts who were not successfully fingerprinted did not differ significantly with respect to sex, age, or HIV status from those who were successfully fingerprinted (Table 1).

CHWs successfully fingerprinted all consenting contacts in 70.0% (213/304) of households. Among households, clustering of fingerprint scan outcomes by CHW was modest (ICC=.18). The frequency of successfully fingerprinting all contacts in a household by CHW ranged from 45% to 97%, with a median of 71% (P<.001). The proportion of households where all contacts were successfully fingerprinted decreased over time—87% (69/79) in quarter 1, 77% (48/62) in quarter 2, 68% (52/76) in quarter 3, and 51% (44/87) in quarter 4 (ρ=.30, P<.001).

Qualitative Interviews

The CHWs involved in the parent study were recruited on the basis of their high level of previous work experience and their ability to speak both English and Luganda. All 15 CHWs who carried out fingerprint scans were interviewed. The median interview length was 37 minutes (interquartile range: 33.5-42 minutes). CHWs ranged in age from 24 to 54 years with a median of 33 years, and 80% (12/15) of CHWs were females. Most (13/15, 87%) had completed ordinary secondary education (O-Level) or higher and a few (3/15, 20%) had completed university-level education. Most of the CHWs had prior experience using information technology, including smartphones (14/15, 93%), and fewer had previously used computers (8/15, 53%), or tablets (5/15, 33%). All 15 CHWs had worked in a lay health worker role prior to joining the study.

In the interviews, CHWs emphasized how specific experiences with the fingerprinting technology affected their sense of identity, their interactions with household contacts, and their ability to carry out their work. These experiences informed CHWs' perceptions of the fingerprinting technology's ease of use and usefulness, two key determinants of intention to use, or acceptance, in the TAM2 model.

Figure 2. Flow diagram showing enrollment of household contacts.

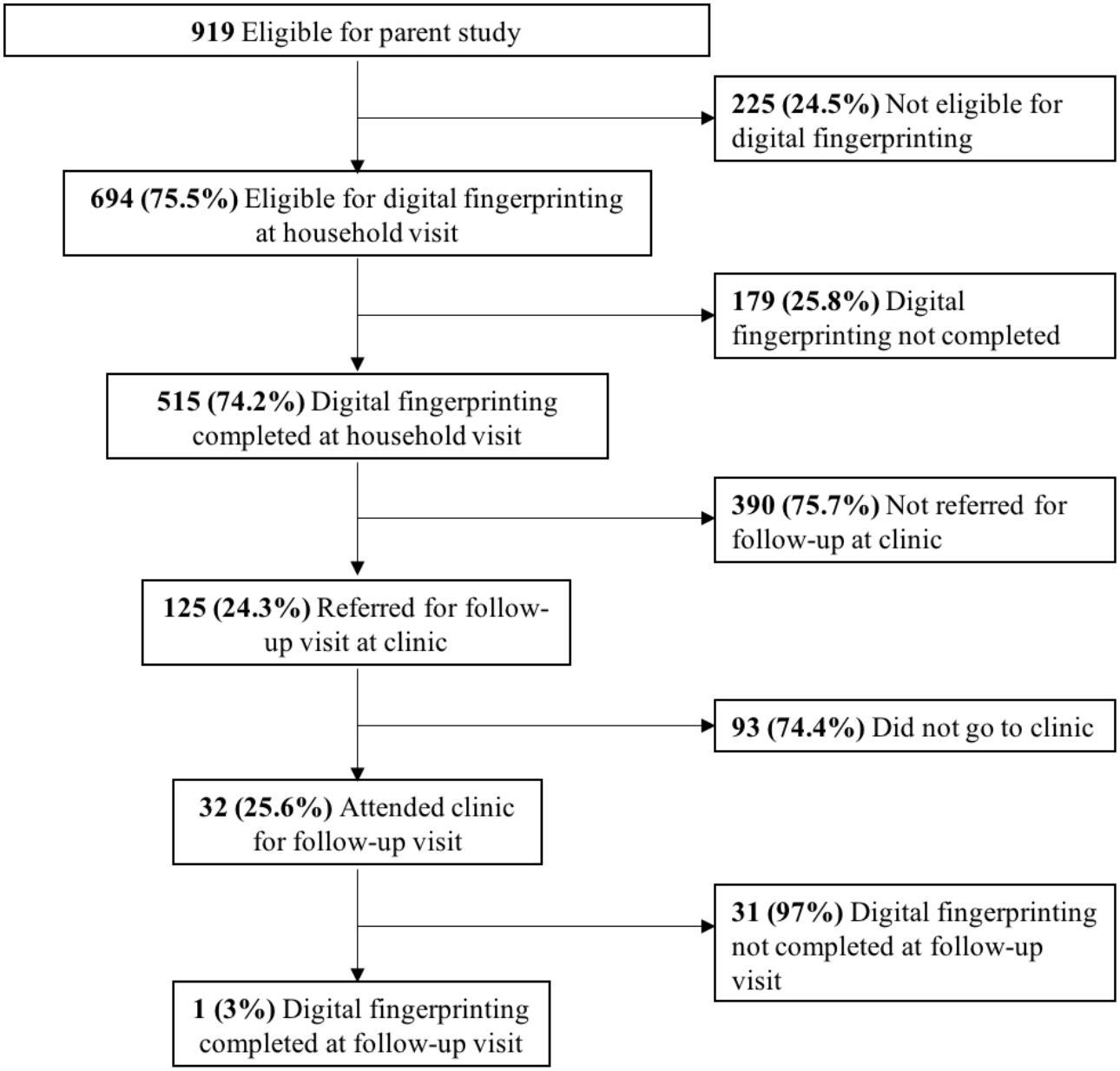


Table 1. Characteristics of study participants (n=694).

Characteristic	Fingerprint scan (n=515)	No fingerprint scan (n=179)	*P* value[a]
Age, n (%)			.56
Children 5-14 years	162 (31.5)	59 (33.0)	
Adults ≥15 years	353 (68.5)	120 (67.0)	
Sex, n (%)			.83
Female	336 (65.2)	108 (60.3)	
Male	179 (34.8)	71 (39.7)	
Proportion living with HIV, n (%)			.87
Positive	41 (8.0)	17 (9.5)	
Negative or unknown	474 (92.0)	162 (90.5)	

[a]Corrected for clustering of fingerprint scan outcomes by household with robust SEs.

Idealized Views of Fingerprinting

CHWs described the usefulness of fingerprinting in an idealized way, reflecting many of the potential applications of fingerprinting that were introduced during training. CHWs consistently said that they believed that fingerprinting would prevent duplicate enrollment and help identify patients who came for follow-up, even if they visited a different study facility.

> *It's useful. I get to know exactly I am with the right patient. And if he has ever, for example you have so many facilities, maybe that patient has ever been to [a different health center], and they have ever scanned, so the scanner will refuse or it will tell me already the patient is in the system.* [Female, CHW6]

Even while acknowledging that the technology did not work perfectly, many CHWs said that they believed that fingerprinting could be useful and should continue.

> *Me I just wish [the use of fingerprinting] would continue and it could be stable, it could not stop, you know you go to the field and it stops, and you have to do restart, do things, it takes a lot of time.…So me, I just wish in case [fingerprinting] continues, let us do those challenges so we can remove those.* [Female, CHW8]

By expressing a desire for fingerprinting to continue, despite substantial challenges with the technology, CHWs revealed how much their perceptions of its potential usefulness were driven by their optimism to make it work.

Positive and Negative Consequences of Digital Fingerprinting for the Self-Image of Community Health Workers

CHWs described their role in the community with pride. They said that they felt that they were providing important services to their patients, whom they often referred to as "clients." However, fingerprint scanning had complicated implications for CHWs' self-image. CHWs explained that the technology could both elevate and threaten their social status. On one hand, fingerprint scanning represented an additional service they could offer to their clients, which elevated the capabilities they projected as CHWs. They perceived digital fingerprinting to be an important technology because it is associated with registering for a national identification card and for identification at commercial banks. The excitement of getting to use this important technology in their work helped motivate CHWs to learn and implement fingerprint scanning.

> *So I was so excited, and I even asked myself, "Who am I, to be in this?" So, I put on my brains in there to really understand what is going to be done. And it took me only two days to get everything in the tablet because I was so attached to it, I wanted it so much.* [Female, CHW12]

On the other hand, when CHWs struggled to use the fingerprinting technology in front of clients, they felt that their credibility was diminished.

> *When you're printing someone and it fails? They just look at you like you don't know what you're doing.* [Female, CHW13]

CHWs placed high importance on their competence in carrying out contact investigation, and a failed fingerprinting attempt could damage one's credibility. Thus, CHWs perceived that the technology enhanced their social and professional status when it worked smoothly but threatened their status when it failed in the presence of a client.

Variable Views on the Need and Appropriateness of Digital Fingerprinting

While CHWs generally acknowledged the need for some way to identify patients and contacts to carry out contact investigation, views were mixed regarding whether fingerprinting was necessary. These mixed opinions arose from different perceptions of the job relevance of fingerprinting, or the belief that fingerprinting is important to contact investigation. Some CHWs thought that fingerprinting could be the best way to uniquely identify people:

> *Even if you give three names, someone might come with, another person might come with three names which are the same. Yet here the fingerprints identify the very person you want.* [Female, CHW3]

However, others suggested that the name, health center, patient identification number, signatures, photos, or voice recordings would suffice as alternatives. In practice, most CHWs described using some combination of name and other identifiers to identify contacts at follow-up, rather than using the fingerprint. One CHW distinguished between the usefulness of fingerprinting for identifying contacts versus index patients. He said that it was more useful for contacts who are numerous and who come to the clinic months after the CHW meets them. Because index patients are fewer in number, sicker when the CHW meets them, and come back to the clinic often, CHWs felt that they were more memorable and that there was no need to rely on a fingerprint to identify them.

Impact of Failures to Capture Fingerprints Digitally

Even before interviewers asked about technology failures that prevented the successful capture of fingerprints, CHWs repeatedly turned the discussion toward their experiences with technology failure. CHWs linked the output quality, or how well the technology performed, to their perceptions of its usefulness. A small number of CHWs who reported never having issues with the technology described fingerprinting as being useful. Most CHWs, however, described an increase in technology failures over time, preventing them from capturing fingerprints and adding unnecessary time to the study procedures. When asked whether fingerprinting was useful and should continue in the future, almost all of these CHWs still responded yes, but only if it worked consistently and did not take too much time.

> *It would be good, like I've told you, but the technical issues around it can make the work difficult.* [Female, CHW9]

Thus, the perceived usefulness of digital fingerprinting depended on it being reliable, fast, and free from technology problems.

Voluntary Abandonment of Digital Fingerprinting

Most CHWs described instances when they chose to "bypass" the fingerprint scan during contact investigation enrollment; this option was built into the software to allow them to continue with the encounter even when fingerprint scanning failed. They did not indicate any negative impacts of failing to capture a fingerprint on contact investigation procedures. These descriptions suggest that result demonstrability was low and the effect of capturing a fingerprint was not tangible to CHWs.

> *When it has refused. That's when I decide to go back and I bypass the fingerprint scanner, and I continue with my patients. I jump it and go to the next question.* [Female, CHW5]

In addition, CHWs described troubleshooting measures that they used when the fingerprint scanner failed: disconnecting and reconnecting the cable linking the scanner to the tablet, powering the tablet off and back on again, and asking a colleague for help. However, most CHWs said that they only attempted to troubleshoot one to three times—or sometimes not at all—before bypassing the fingerprint scan altogether. In the view of CHWs, whether a fingerprint was successfully captured did not seem to change the contact investigation procedures.

Variable Confidence in Using the Technology

CHWs differed in their perceptions of the ease of use of the technology, including the scanner itself and the tablet that they used to control the scanner. Some said that it was consistently easy to navigate through the app on the tablet and obtain a fingerprint using the scanner. Others described relying on colleagues or study staff for support when they had problems, which were frequent and which they came to anticipate.

> *I'm expecting I will go and then I will call [the technology support officer] that this thing has blacked out. So it's expected…I don't think I'm the only one complaining about the scanner. They disturb us a lot.* [Female, CHW9]

This range of comfort with the technology was also reflected during the interview prompt exercise in which CHWs demonstrated the fingerprinting process. Some worked quickly, while others were hesitant when navigating through the app; some were able to describe the process in their own words, while others read directly from the text on the screen. Individuals' confidence using the tablet and scanner varied greatly.

Personal Risks to Health Workers

CHWs described two forms of risk that they associated with digital fingerprinting and that influenced their perceptions of its ease of use. First, some CHWs worried about the risk of infection through close contact with patients during the fingerprinting procedure, exacerbated by lack of adequate space and ventilation while performing fingerprinting.

> *When you're doing this and this [demonstrating placing fingers on the scanner], you're kind of getting closer to the patient who is HIV—I mean TB positive, so somehow you are risking. Just try to demonstrate, just try to put your finger here [on the scanner]. So as I'm a community health worker and you have to get closer to me, I'm also breathing in.* [Female, CHW9]

Second, CHWs said they worried about personal security when carrying the tablet and scanner to household visits.

> *When we move, some of our places are not in…they are not easy to go there alone. Because you have slums, very dangerous to go with the gadget…And TB is mostly in those places.* [Female, CHW7]

The risk of infection and lack of personal security introduced psychological and logistical challenges that CHWs had to overcome to carry out fingerprinting.

Discussion

The inability to uniquely and accurately identify individuals in resource-constrained settings remains a major barrier to improving the quality of health information management and public health research. We found that digital fingerprint scanning was feasible but not reliable—failing to capture fingerprints in about one-quarter of cases—during household contact investigation for TB. Importantly, we found evidence that failures were tightly clustered by household, that they increased substantially over the course of the study, and that there were no systematic differences by clinical or demographic factors. The low rate of fingerprinting at follow-up suggests that CHWs saw little value in the digital fingerprinting system's usefulness as a verification tool. A systematic qualitative analysis indicated that CHWs continued to find digital fingerprinting acceptable in principle despite the technology's inconsistent reliability and an accumulating experience with technology failures that decreased their confidence in its usefulness in this setting.

The patterns of fingerprinting failures during the household visit pointed toward problems with the implementation of both software and hardware. Fingerprinting outcomes were almost completely clustered at the household level, suggesting that rather than being driven by sporadic, individual-level failures or refusals, the fingerprinting technology either worked or did not work on a given visit to a household. We identified no individual patient characteristics associated with failure, including age and sex, which argues against degraded individual fingerprints as a cause of failure, as might be expected among adult manual laborers. Furthermore, the predominance of software and hardware problems as explanations for failure and the modest clustering by CHW imply that technology failures were responsible rather than the skills of individual health workers. Finally, the significantly increasing proportion of fingerprinting failures over time reflects the declining functionality of the technology, whether due to health worker disengagement from the technology, software issues, hardware issues or, perhaps, all three.

Previous studies have shown that CHWs without prior experience with digital fingerprinting describe the technology as acceptable in principle [10]. However, we observed that

CHWs' assessments of fingerprint scanning could change as they gain experience with the technology. We found that the TAM2 domains of image, job relevance, and output quality were especially relevant to CHWs' perceptions of the usefulness of digital fingerprinting in the study. Technology failures lowered CHWs' perception of the quality of the system, threatened CHWs' social image, and made the technology more difficult for CHWs to use. Although the technology worked as intended in the majority of interactions, workarounds and a lack of a tangible benefit of fingerprinting ultimately limited its job relevance and perceived usefulness among CHWs. After regular use, CHWs continued to express enthusiasm for fingerprint scanning in principle, but their intention to use the technology was tempered by perceptions that it was inconsistent and of questionable value, ultimately undermining their intention and usage behaviors.

Our findings add to a relatively limited literature on the use of digital fingerprinting for public health applications in sub-Saharan Africa. Our findings differ from a study of the same technology among female sex workers in Zambia, where digital fingerprinting was feasible for and acceptable to clients in the clinic setting, but not acceptable to clients in the field [5]. Perhaps, because participants were at greater risk for stigma or arrest and prosecution, the most common reasons for refusal related to clients' concerns about a potential loss of confidentiality or privacy. In contrast, we found that a majority of community members underwent fingerprinting during study registration without differences by demographic or clinical characteristics or documented refusals. Similar to a previous study of a mobile health tool for reporting adverse effects of treatments for drug-resistant TB in South Africa, we found that reported enthusiasm for technology—fingerprinting, in this case—did not translate into usage [20]. The acceptance of fingerprinting technology among CHWs serving these clients may decline if they experience technology failures during their work and may be more impactful in terms of its use than the perceived acceptability by community members. There may be a role for communities of practice—learning and peer support networks established to facilitate continuous quality improvement—as patient identification technology is being introduced to help address these challenges [21-23].

Finally, the almost universal failure of lay health workers in this study to use digital fingerprinting at follow-up contrasts with the findings of a study of a biometric identification system for monitoring TB treatment in rural Uganda, which found that fingerprinting improved follow-up among patients engaged in daily directly observed therapy at the clinic [24]. A low background rate of clinic follow-up in our study limited opportunities for digital fingerprinting in this context and, perhaps, therefore, its utility. In settings where digital fingerprinting has been shown to be feasible and acceptable, researchers should conduct larger, well-controlled studies to assess whether fingerprinting is an effective tool for monitoring and improving adherence to follow-up visits in combination with feedback communications. Finally, a limitation of digital fingerprinting is that it is unable to reliably capture fingerprints of children aged <5 years [14], resulting in their exclusion from the analysis. Further studies should evaluate whether newer technologies can accurately capture fingerprints for children aged <5 years. Future studies could also include interviews with household contacts to gain a more comprehensive understanding of the acceptability and challenges of fingerprinting from the perspective of contacts.

This study has a few limitations. First, we had limited data on the technical reasons for each fingerprinting failure. While we were able to categorize failures broadly as related to hardware or software problems, these groupings are not specific enough to guide improvement strategies. Detailed logs itemizing the circumstances of each fingerprinting failure should be included in future evaluations. Second, incomplete data for household-level covariates, such as income, limited our ability to identify predictors of failure to capture fingerprints digitally, although the lack of reported refusals and the very small number of unattributable explanations for failure make patient factors an unlikely explanation.

This study also has several strengths. First, the mixed-methods design enabled complementary analyses of the use of fingerprint scanning during household contact investigation for TB. The quantitative analysis revealed evidence of extensive clustering of failures within household encounters, while the qualitative analysis showed the influence of these failures on CHWs' perceptions of the technology's usefulness. Second, we organized key themes offered by CHWs into TAM2 subdomains such as image, job relevance, and output quality, showing how these perceptions shape CHWs' evolving understanding of the usefulness of fingerprinting technology. Third, we were able to interview the entire CHW population involved in the study rather than relying on a sample. Finally, we evaluated a multispectral fingerprinting technology integrated with and offered as a standard commercial product by a leading global health software platform, increasing the generalizability.

The ability to accurately collect and link individual data to preserve privacy and enhance the generation of quality measures for patients moving through complex care pathways should be a major global health priority [25]. Despite the feasibility and acceptability of biometric identification methods as a means of bringing unique patient identification to resource-constrained settings, the technology we evaluated was not widely adopted by health professionals tasked with using it. As biometric technologies are increasingly introduced in resource-constrained health contexts, our findings point to the importance of theory-informed, mixed-methods evaluation of adoption of these technologies. Mixed-methods data may guide iterative improvements to hardware, software, and the user interface to ensure that the technology aligns with tasks that users find useful and important and engages health workers so that they voluntarily apply the technology to improve the experience of patients. Furthermore, future studies should consider whether detailed process evaluation using mixed methods can be applied to other biometric technologies.

Acknowledgments

We would like to thank the study participants, including the index patients and their household contacts who enrolled in the parent study, and the community health workers who carried out digital fingerprinting and participated in the in-depth interviews.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Interview guide used for in-depth interviews with community health workers.

[PDF File (Adobe PDF File), 66KB - jmir_v20i11e11541_app1.pdf]

Multimedia Appendix 2

Debriefing form completed by interviewer immediately following each interview, from which emerging themes were identified and analyzed.

[PDF File (Adobe PDF File), 42KB - jmir_v20i11e11541_app2.pdf]

Multimedia Appendix 3

Quantitative results for index patients corresponding to those for household contacts that are presented in the Results section.

[PDF File (Adobe PDF File), 53KB - jmir_v20i11e11541_app3.pdf]

References

1. Beck EJ, Shields JM, Tanna G, Henning G, de Vega I, Andrews G, et al. Developing and implementing national health identifiers in resource limited countries: why, what, who, when and how? Glob Health Action 2018;11(1):1440782 [FREE Full text] [doi: 10.1080/16549716.2018.1440782] [Medline: 29502484]
2. SonLa Study Group. Using a fingerprint recognition system in a vaccine trial to avoid misclassification. Bulletin of the World Health Organization 2007 Jan;85(1):64-67 [FREE Full text] [Medline: 17242760]
3. Jain AK. Technology: biometric recognition. Nature 2007 Sep 06;449(7158):38-40. [doi: 10.1038/449038a] [Medline: 17805286]
4. Unar J, Seng WC, Abbasi A. A review of biometric technology along with trends and prospects. Pattern Recognition 2014 Aug;47(8):2673-2688. [doi: 10.1016/j.patcog.2014.01.016]
5. Wall KM, Kilembe W, Inambao M, Chen YN, Mchoongo M, Kimaru L, et al. Implementation of an electronic fingerprint-linked data collection system: a feasibility and acceptability study among Zambian female sex workers. Global Health 2015 Jun 27;11:27 [FREE Full text] [doi: 10.1186/s12992-015-0114-z] [Medline: 26115656]
6. Weibel D, Schelling E, Bonfoh B, Utzinger J, Hattendorf J, Abdoulaye M, et al. Demographic and health surveillance of mobile pastoralists in Chad: integration of biometric fingerprint identification into a geographical information system. Geospat Health 2008 Nov;3(1):113-124. [doi: 10.4081/gh.2008.237] [Medline: 19021114]
7. Odei-Lartey EO, Boateng D, Danso S, Kwarteng A, Abokyi L, Amenga-Etego S, et al. The application of a biometric identification technique for linking community and hospital data in rural Ghana. Global Health Action 2016;9:29854 [FREE Full text] [Medline: 26993473]
8. Harichund C, Haripersad K, Ramjee R. Participant verification: prevention of co-enrolment in clinical trials in South Africa. South African Medical Journal 2013 May 15;103(7):491-493. [Medline: 23802218]
9. Serwaa-Bonsu A, Herbst AJ, Reniers G, Ijaa W, Clark B, Kabudula C, et al. First experiences in the implementation of biometric technology to link data from Health and Demographic Surveillance Systems with health facility data. Global Health Action 2010 Feb 24;3:1-8 [FREE Full text] [doi: 10.3402/gha.v3i0.2120] [Medline: 20200659]
10. van Heerden A, Harris DM, van Rooyen H, Barnabas RV, Ramanathan N, Ngcobo N, et al. Perceived mHealth barriers and benefits for home-based HIV testing and counseling and other care: Qualitative findings from health officials, community health workers, and persons living with HIV in South Africa. Soc Sci Med 2017 Dec;183:97-105 [FREE Full text] [doi: 10.1016/j.socscimed.2017.04.046] [Medline: 28475904]
11. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ 2015 Mar 19;350:h1258 [FREE Full text] [Medline: 25791983]
12. Venkatesh V, Davis FD. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science 2000;46(2):169-332 [FREE Full text]
13. Schumacher G. Fingerprint Recognition for Children. JRC Technical Reports 2013:1-102 [FREE Full text] [doi: 10.2788/3086]

14. Jain AK, Arora SS, Cao K, Best-Rowden L, Bhatnagar A. Fingerprint Recognition of Young Children. IEEE Transactions on Information Forensics and Security 2017;12(7):1501-1514 [FREE Full text]
15. Simoni J, Beima-Sofie K, Amico K, Hosek S, Johnson M, Mensch B. Can Debrief Reports Expedite Qualitative Studies? 2017 Presented at: 12th International Conference on HIV Treatment and Prevention Adherence; 2017; Miami, FL URL: https://www.hanc.info/Documents/4_Adherence%202017%20Qual%20Debrief%20poster_27%20May%202017%20kra%20JMS-KBS%20SH%20JMS_submitted.pdf
16. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319-350. [doi: 10.2307/249008]
17. Moore GC, Benbasat I. Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation. Information Systems Research 1991 Sep;2(3):192-222. [doi: 10.1287/isre.2.3.192]
18. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. Journal of Biomedical Informatics 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
19. Fishbein M, Ajzen I. Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research. Reading, MA: Addison-Wesley; 1975.
20. Chaiyachati KH, Loveday M, Lorenz S, Lesh N, Larkan L, Cinti S, et al. A pilot study of an mHealth application for healthcare workers: poor uptake despite high reported acceptability at a rural South African community-based MDR-TB treatment program. PLoS One 2013;8(5):e64662 [FREE Full text] [doi: 10.1371/journal.pone.0064662] [Medline: 23724075]
21. Wenger E. Communities of practice: learning, meaning, identity. Cambridge, U.K: Cambridge University Press; 1998.
22. Li LC, Grimshaw JM, Nielsen C, Judd M, Coyte PC, Graham ID. Use of communities of practice in business and health care sectors: a systematic review. Implement Sci 2009 May 17;4(27):1-9 [FREE Full text] [doi: 10.1186/1748-5908-4-27] [Medline: 19445723]
23. Meessen B, Kouanda S, Musango L, Richard F, Ridde V, Soucat A. Communities of practice: the missing link for knowledge management on implementation issues in low-income countries? Trop Med Int Health 2011 Aug;16(8):1007-1014 [FREE Full text] [doi: 10.1111/j.1365-3156.2011.02794.x] [Medline: 21564426]
24. Snidal SJ, Barnard G, Atuhairwe E, Ben Amor Y. Use of eCompliance, an innovative biometric system for monitoring of tuberculosis treatment in rural Uganda. Am J Trop Med Hyg 2015 Jun;92(6):1271-1279 [FREE Full text] [doi: 10.4269/ajtmh.14-0413] [Medline: 25846297]
25. Davis J, Katamba A, Vasquez J, Crawford E, Sserwanga A, Kakeeto S, et al. Evaluating tuberculosis case detection via real-time monitoring of tuberculosis diagnostic services. Am J Respir Crit Care Med 2011 Aug 01;184(3):362-367 [FREE Full text] [doi: 10.1164/rccm.201012-1984OC] [Medline: 21471088]

Abbreviations

CHW: community health worker
ICC: intraclass correlation coefficient
TAM2: Technology Acceptance Model 2
TB: tuberculosis

Edited by G Eysenbach; submitted 11.07.18; peer-reviewed by C Muzoora, L Li; comments to author 06.09.18; revised version received 10.09.18; accepted 10.09.18; published 15.11.18.

Please cite as:
White EB, Meyer AJ, Ggita JM, Babirye D, Mark D, Ayakaka I, Haberer JE, Katamba A, Armstrong-Hough M, Davis JL
Feasibility, Acceptability, and Adoption of Digital Fingerprinting During Contact Investigation for Tuberculosis in Kampala, Uganda: A Parallel-Convergent Mixed-Methods Analysis
J Med Internet Res 2018;20(11):e11541
URL: http://www.jmir.org/2018/11/e11541/
doi:10.2196/11541
PMID:30442637

Original Paper

Automated Identification of Hookahs (Waterpipes) on Instagram: An Application in Feature Extraction Using Convolutional Neural Network and Support Vector Machine Classification

Youshan Zhang[1], MS; Jon-Patrick Allem[2], MA, PhD; Jennifer Beth Unger[2], PhD; Tess Boley Cruz[2], PhD

[1]Department of Computer Science, Lehigh University, Bethlehem, PA, United States

[2]Keck School of Medicine of USC, Los Angeles, CA, United States

Corresponding Author:
Jon-Patrick Allem, MA, PhD
Keck School of Medicine of USC
2001 N Soto Street, 3rd Floor Mail
Los Angeles, CA, 90032
United States
Phone: 1 8586030812
Email: allem@usc.edu

Abstract

Background: Instagram, with millions of posts per day, can be used to inform public health surveillance targets and policies. However, current research relying on image-based data often relies on hand coding of images, which is time-consuming and costly, ultimately limiting the scope of the study. Current best practices in automated image classification (eg, support vector machine (SVM), backpropagation neural network, and artificial neural network) are limited in their capacity to accurately distinguish between objects within images.

Objective: This study aimed to demonstrate how a convolutional neural network (CNN) can be used to extract unique features within an image and how SVM can then be used to classify the image.

Methods: Images of waterpipes or hookah (an emerging tobacco product possessing similar harms to that of cigarettes) were collected from Instagram and used in the analyses (N=840). A CNN was used to extract unique features from images identified to contain waterpipes. An SVM classifier was built to distinguish between images with and without waterpipes. Methods for image classification were then compared to show how a CNN+SVM classifier could improve accuracy.

Results: As the number of validated training images increased, the total number of extracted features increased. In addition, as the number of features learned by the SVM classifier increased, the average level of accuracy increased. Overall, 99.5% (418/420) of images classified were correctly identified as either hookah or nonhookah images. This level of accuracy was an improvement over earlier methods that used SVM, CNN, or bag-of-features alone.

Conclusions: A CNN extracts more features of images, allowing an SVM classifier to be better informed, resulting in higher accuracy compared with methods that extract fewer features. Future research can use this method to grow the scope of image-based studies. The methods presented here might help detect increases in the popularity of certain tobacco products over time on social media. By taking images of waterpipes from Instagram, we place our methods in a context that can be utilized to inform health researchers analyzing social media to understand user experience with emerging tobacco products and inform public health surveillance targets and policies.

(J Med Internet Res 2018;20(11):e10513) doi:10.2196/10513

KEYWORDS

convolutional neural network; feature extraction; image classification; Instagram; social media; support vector machine

Introduction

Instagram, with millions of posts per day [1], can be used to inform public health surveillance targets and policies. However, this research relying on image-based data often relies on hand coding of images [2,3], which ultimately limits the scope of the study. Images from social media may be more useful than findings from text-based platforms alone (eg, Twitter and Reddit) when attempting to understand health behaviors, for

example, user experiences with emerging tobacco products [4]. While automated image classification is useful for large-scale image classification (eg, processing and assigning labels to millions of images), current best practices in automated image classification are limited in their capacity to accurately distinguish between objects within images [5-7]. Automated image classification has been used in supervised, unsupervised, and hybrid approaches in classifying data [8-10]. Compared with unsupervised methods, supervised methods can be divided into stages of training and testing. The training stage consists of training a classifier by images and its labels, for example, describing image content, such as a person, dog, elephant, etc. The testing stage predicts the labels of the test images (in a new set of images) by a trained classifier.

Prior research has focused on ways to overcome the methodological challenges of automated image classification such as low accuracy. For example, Perronnin et al improved the Fisher Kernel approach to extend the bag-of-visual-words, also called bag-of-features (BOF), for large-scale image classification using internet images from ImageNet and Flickr, which increased precision from 47.9% to 58.3% but did not improve accuracy [5]. Verma et al used the backpropagation neural network approach to classify large images with good accuracy (97.02%), but this approach could not identify multiple categories of an image [6]. To reduce the time and spatial complexity of images, Simonyan et al proposed 2 visualization techniques using deep convolutional networks (ConvNets) to classify artificial images [7]. They combined understandable visualizations of ConvNets, maximizing the scores of images within different classes with gradient-based ConvNets visualization generating the saliency map (also called features map, which can represent the influence of pixels in image on image classification results) of every image (corresponding to one class) to use a deconvolution (also called transpose of convolution, which performs upsampling tasks instead of downsample in convolutional layer) network to segment objects in the images [7].

These earlier approaches have moved automated image classification forward; however, there are still a number of significant limitations to overcome [11-13]. For example, the large number of images that need to be extracted to train a model requires great computational power. In addition, the BOF method cannot localize the objects within an image and cannot use visual word positions (eg, if a cup was in an image, BOF could not find its position) [14,15]. Support vector machine (SVM) have a limitation in showing the transparency of results, as the final model is difficult to visualize. Moreover, it is a challenge to choose a suitable kernel in kernel SVM [16-18]. A convolutional neural network (CNN), on the other hand, can improve the generalization of the algorithm and can solve nonlinear problems. While a CNN has high accuracy, to get better results, the parameters should be fine-tuned (eg, input image size, patch size, and the number of convolutional layers), and network performance is hard to optimize [19,20].

The purpose of this study is to determine whether combining CNN and SVM can achieve higher accuracy in image classification compared with CNN or SVM alone. To this end, data from Instagram containing images of waterpipes, also known as hookah (an emerging tobacco product possessing similar harms to that of cigarettes), were examined. By taking data from Instagram, we place our methods in a context that can be utilized to later inform researchers in the health domain who wish to analyze social media to understand user experience with emerging tobacco products and inform public health surveillance targets and public policies [2-4,21-25].

Methods

Data Acquisition

Data used in this study comprised posts on Instagram between February 19, 2016 and May 19, 2016, in the United States that included the hashtag #hookah. A total of 820 images was used in this study. The ground truth was manually labeled (hookah and nonhookah images). To balance the data and classes, the training images included 420 images (210 hookah and 210 nonhookah images), and test images also included 420 images (210 hookah and 210 nonhookah images). Further details on data collection are described elsewhere [24]. MATLAB was used to classify images into 2 categories: images containing a waterpipe (hookah) and those not containing.

Convolutional Neural Network

Image features comprising 25 layers were extracted using AlexNet [26-28] (a well-trained CNN software). Figure 1 shows the architecture of AlexNet. Among these 25 layers, there are input and output layers, 7 rectified linear units (ReLU) layers, 2 normalization layers, 3 pooling layers, 2 dropout layers (drop), 1 softmax layer (prob), and 8 learnable weights layers, which contain 5 convolutional layers (conv) and 3 fully connected layers (fc) [26]. The input layer comprised 227×227-pixel images. The ReLU layer reduces the number of epochs to achieve the training error rate higher than traditional tanh units. The normalization layer (norm) increases the generalization and reduces the error rate. The pooling layers summarize the outputs of adjacent pooling units [29]. The dropout layer efficiently decreases the test errors [30], and both dropout layer and the softmax layer reduce the overfitting phenomenon, while the output layer is the categories of images. To extract the features, we fine-tuned the network by removing the last 2 layers of the original 25 layers, as all layers are not suitable for extracting the features. As the layers at the beginning of the network can only detect the edges of the images, we used the results of the fully connected layers to extract features.

Figure 1. The architecture of AlexNet, which comprised 25 layers.

Support Vector Machine

SVM, a supervised learning model with algorithms that analyze data for classification, has been used to predict the categories of objects in images [18,31]. Our proposed method goes beyond earlier research as the input (feature vectors) was based on the results of the CNN, which can boost accuracy and save time. AlexNet was used to extract features, and those features were used to then train the SVM classifier, requiring only minutes to train all images, thereby saving time [11]. Once the SVM classifier was trained using the feature vectors, the categories of images were predicted.

Analytical Approach

First, we classified images into 2 categories—hookah and nonhookah images—and labeled accordingly. Figure 2 shows the classification scheme, for example, the input image dimension is 227×227×3 pixels, and the output of the CNN is the 4096×1×1 feature maps, which are used to train the SVM classifier; then, the classifier is used to predict the categories (hookah vs nonhookah) of test images. The hookah images contain a waterpipe, and the nonhookah images do not contain a waterpipe (Figure 3). Next, we divided image sets into training and test images; the training images were used to extract and learn the features (n=420, randomly selected), while the test images were used to calculate the accuracy of the method (n=420, randomly selected). To extract features of the images, the dimension of the input images was made uniform, for example, the image size was 227×227, as the image dimensions of 227×227 are the default of AlexNet. If an image is larger or smaller, we resized the dimensions of the input image to 227×227. We loaded the pretrained CNN by utilizing AlexNet [26], which has been trained by >1 million images. As discussed above, AlexNet was fine-tuned in our method, for example, we removed the last 2 layers of the AlexNet and used the data of the final fully connected layer. Based on the data of the last fully connected layer, we computed the features of the training and test images based on the CNN. Then, the class labels were extracted from the training and test image sets.

To optimize the SVM classifier, we automatically optimized hyperparameters (such as learning rate, the number of layers in the CNN, and mini-batch size) of the waterpipe features vector, and based on the optimized results, we arrived at an optimized SVM classifier (Figure 2) [18,32,33]. The input images dimension is 227×227×3. The output of the CNN was 4096×1×1 features maps of 2 image classes. These features were trained by the SVM classifier, and the trained classifier was later used to predict the categories of test images. We then assessed the performance of the SVM classifier by using the test images and increased the number of images to improve accuracy (the number of images increased from 42 to 420; Figures 3 and 4). Features of the waterpipes in the yellow box were extracted to train the SVM classifier. Based on the trained classifier, we predicted the classes of new images.

Figure 2. The scheme of our method. SVM: support vector machine; CNN: convolutional neural network.

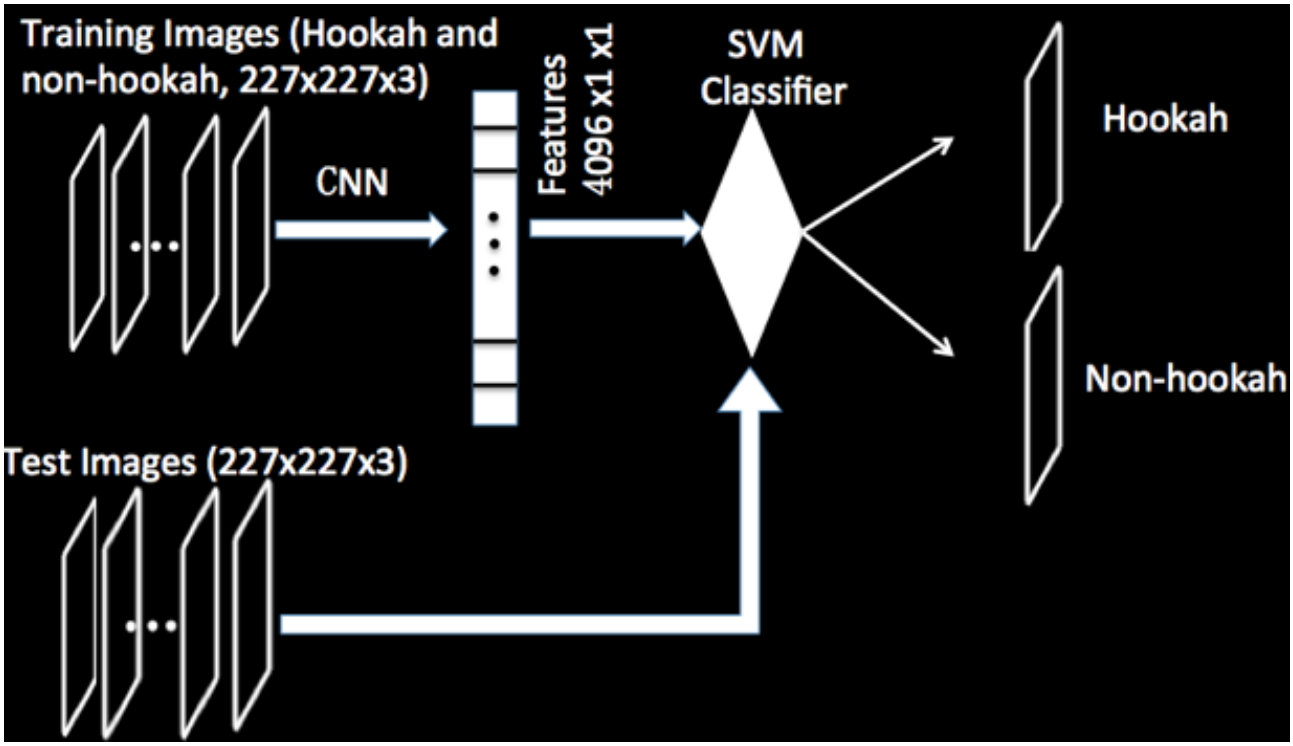


Figure 3. Examples of images with waterpipes (left) and without waterpipes (right).

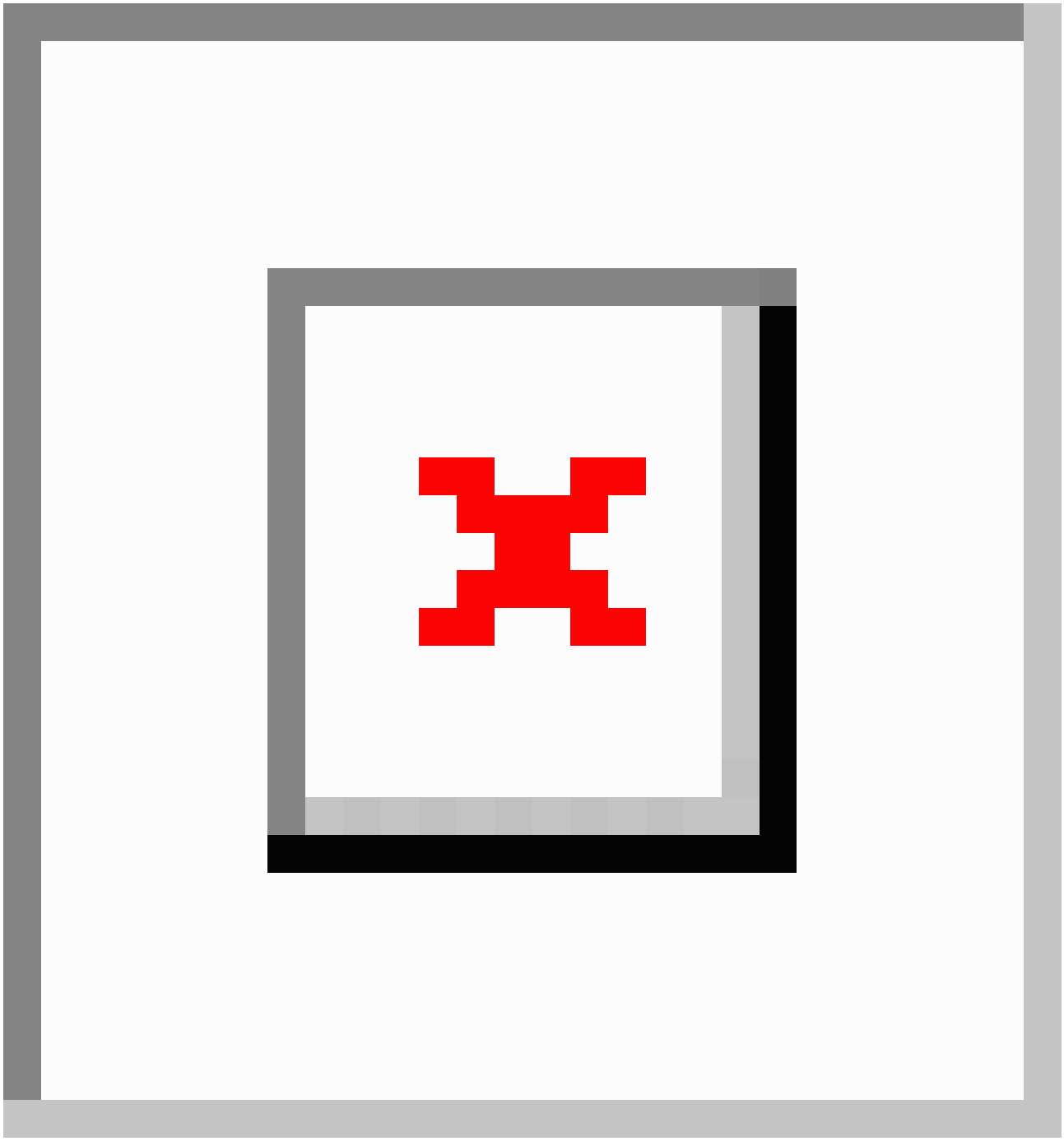


Figure 4. The localization of waterpipes in one image.

Results

Feature Extraction

Results demonstrated that hookah features could be extracted by the CNN, with image categories classified by the SVM, maintaining a high level of accuracy: highest, 99.5% (418/420). Figure 5 shows the features that were extracted from the first convolutional layer; this layer can only detect the edges and blobs, while more features were extracted from the remaining convolutional layers. The original hookah image is on the left. The feature images (right) contains a montage of 96 images, which can reflect the processing of extracting features. Figure 6 shows the feature vectors of the 420 training images, with range –20 to 20; the majority of feature vectors are located between −10 and 10. The x-axis is the image features vector with 4096 total feature vectors. The y-axis is the range of the features with the range between –20 and 20. Figure 7 presents the histogram of the feature vectors. The maximum number of features was between –2 and 2; this interval reflected the most important features of the hookah images. Figure 8 shows the relationship between the function evaluations and the minimum objective. When the function evaluation was 25, the error between the minimum objective and the estimated minimum objective was the highest. Function evaluations demonstrated how many times to evaluate the optimized output. The minimum objective was the minimum observed value of the objective function; it is the smallest overall observation point if there are coupled constraints or evaluation errors.

Figure 5. The extracted features of the first layer using the convolutional neural network.

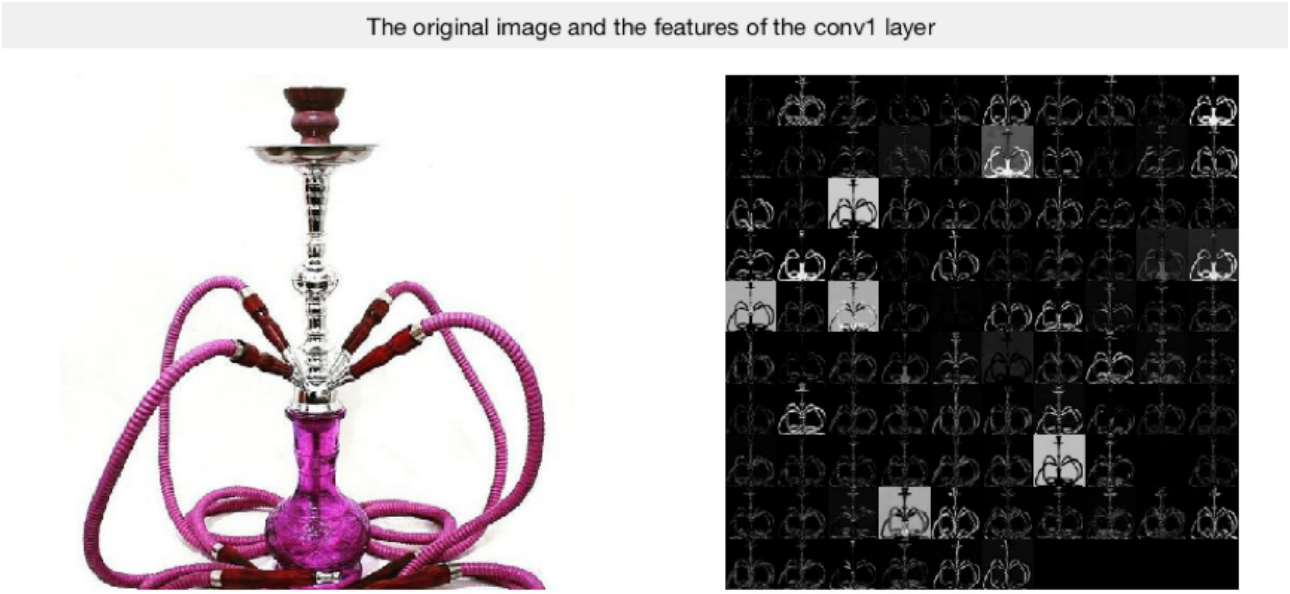


Figure 6. The features of the total image sets (420 images).

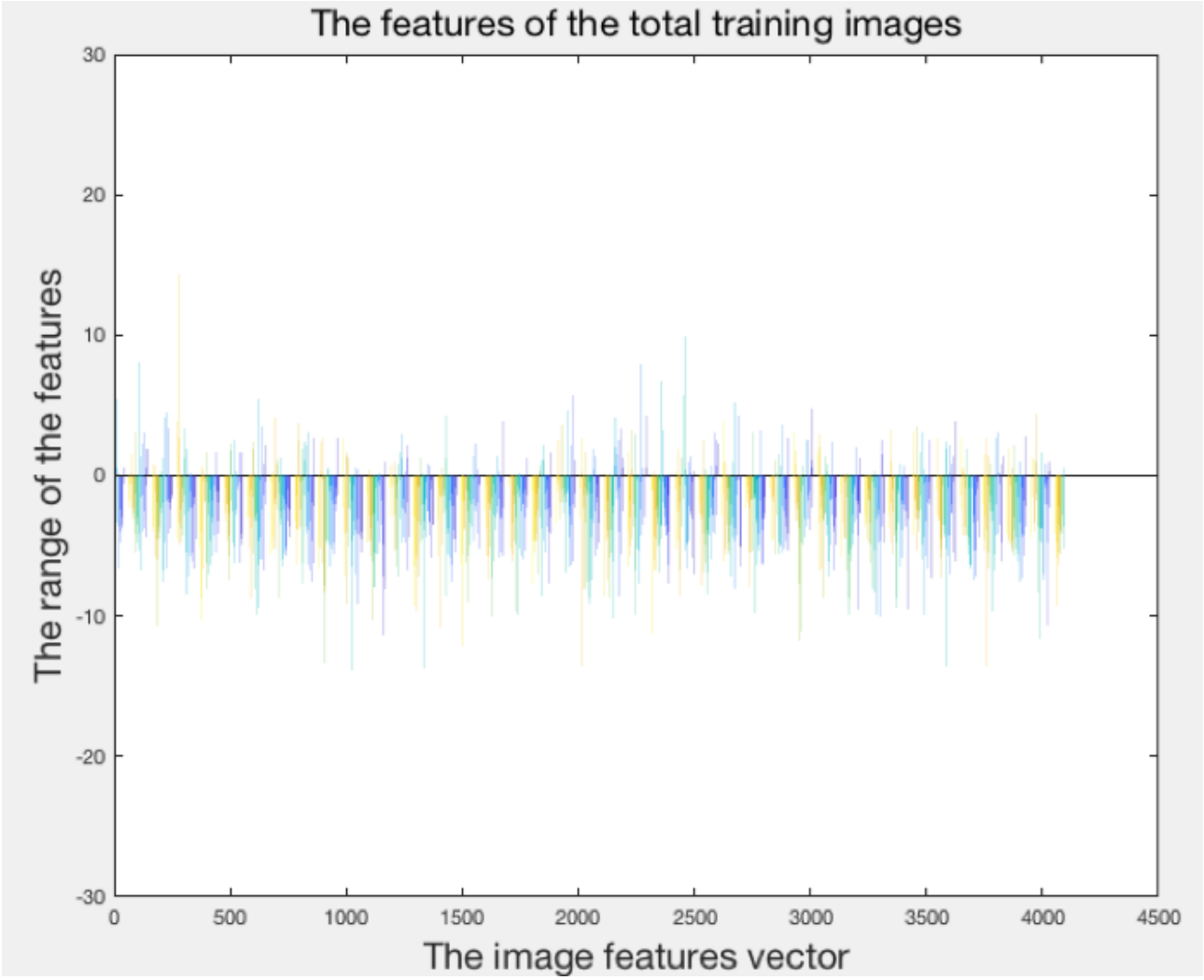


Figure 7. The histogram of the features. The interval of –2 and 2 contains the maximum number of features.

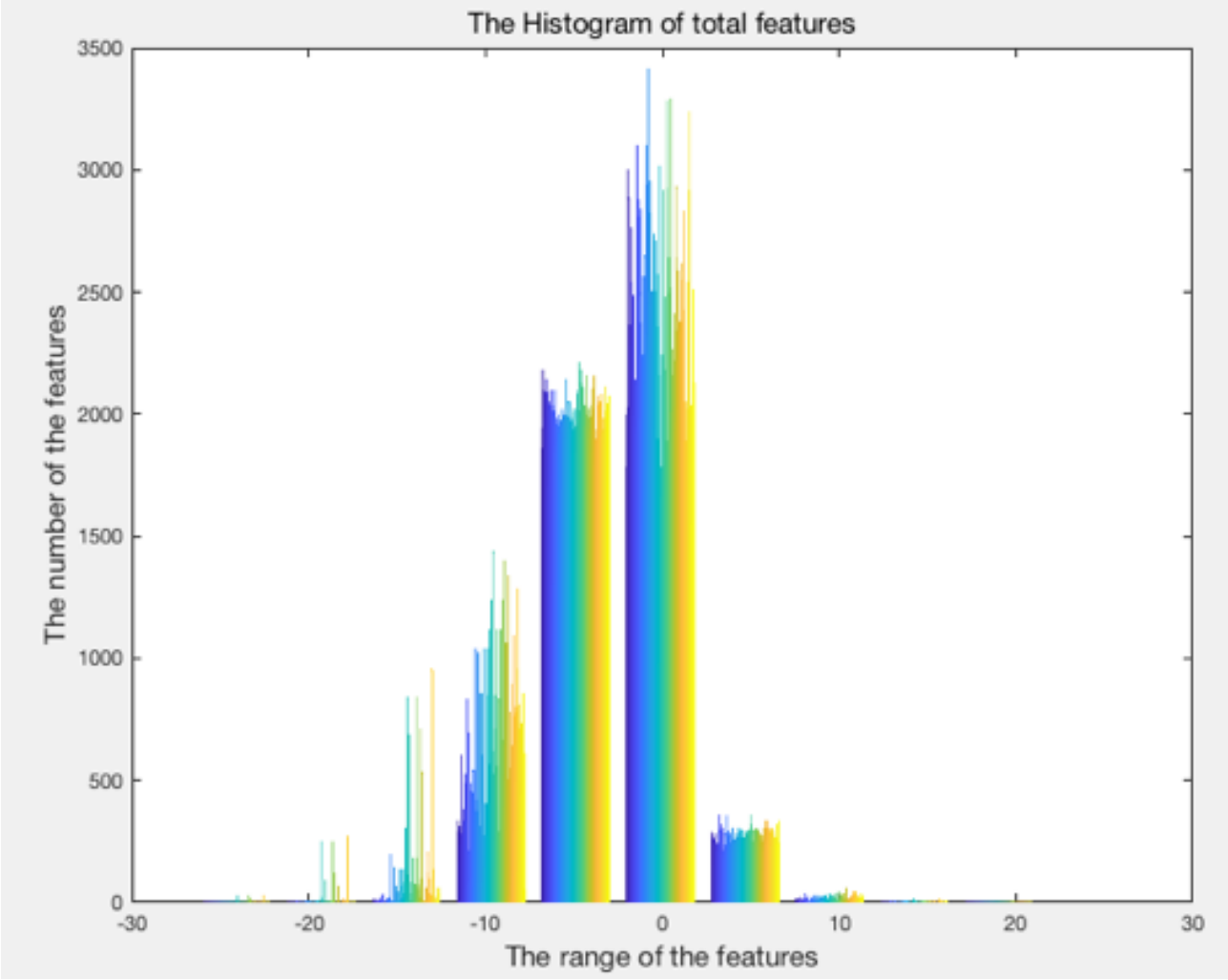
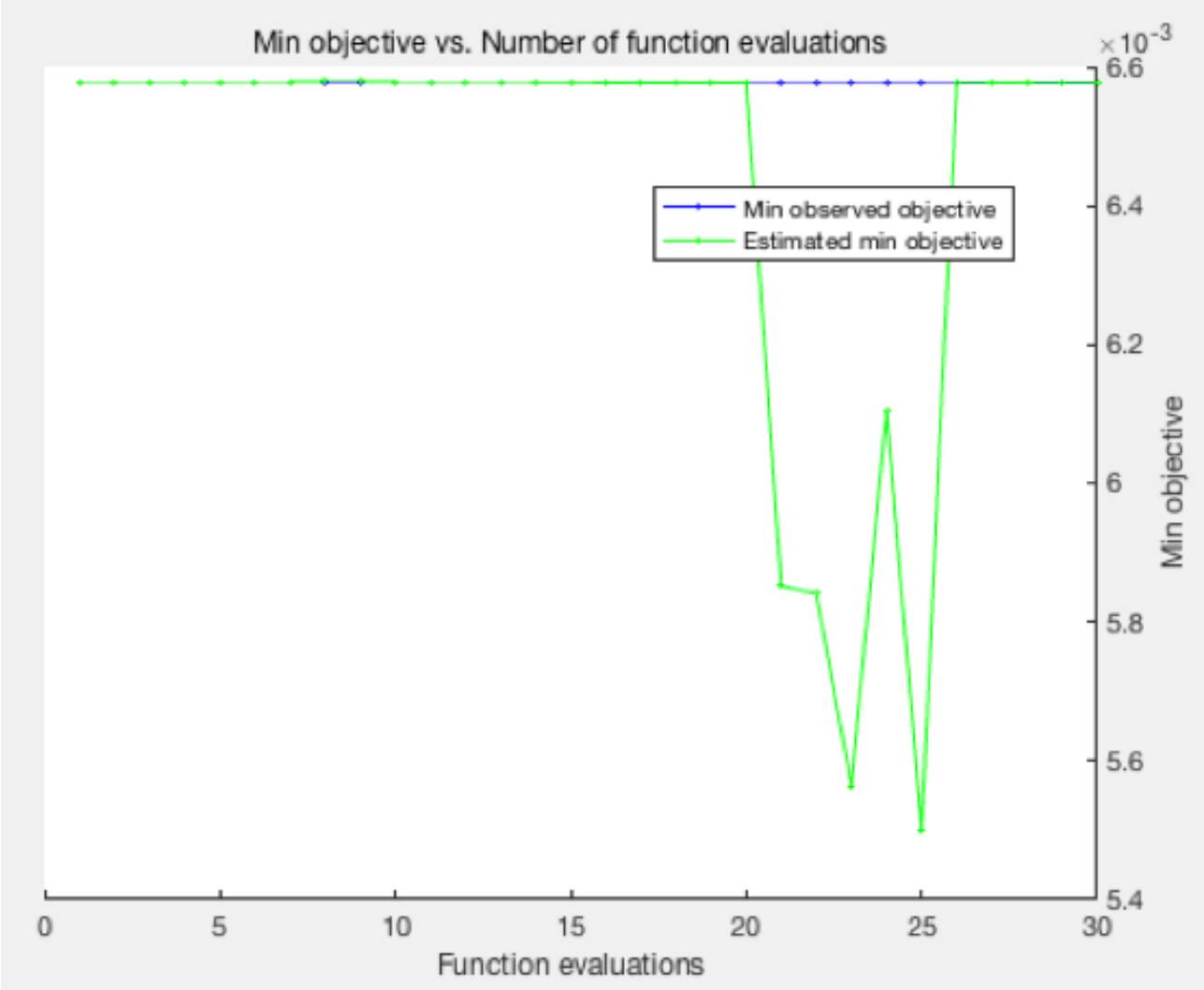


Figure 8. The relationship between the function evaluations and the minimum objective.



The estimation of minimum objective functions can show the difference between the estimated (optimized) minimum objective and real minimum objective. The minimum objective and the estimated minimum objective are similar; however, there are differences across certain function evaluations. The maximum proportion of error is <.01, which is acceptable [19]. Based on the optimized SVM classifier, we evaluated the performance of our method by the test images.

Test Image Classification Results

Figure 9 presents the learning curve showing the relationship between the percentage of validated images (eg, the training images, excluding the test images) and the average level of accuracy of the method. From the chart, with increase in the percentage of validated images, some of the accuracies boosted significantly. For example, from 80% (336/420), the accuracies increase faster than previous percentages, demonstrating that more training images are beneficial in predicting results. As the number of validated training images increased, the total number of extracted features increased: For 1 image, we can extract 4096 features; therefore, with the number of the validated training images (n) increased, the total number of extracted features can increase into n ×4096. In addition, as the number of features learned by the SVM classifier increased, the average level of accuracy increased. The number of validated images was equal to the percentage × the number of the training set; for example, if 10%, then the validated images=10%×420=42.

Overall, 99.5% (418/420) of images classified were correctly identified as either hookah or nonhookah images (Figure 10). The first 2 green squares show the number of the test images and the percentage of the correct image classifications. For example, there were 208 images correctly classified as hookah, and this number accounted for 49.5% (208/420) of all test images. Similarly, 210 images were properly classified as nonhookah, and this accounted for 50% (210/420) of all test images. In the first row, all nonhookah images were correctly classified as such. In the second row, there were 2 hookah images incorrectly classified as nonhookah images, representing 0.5% (2/420) of all data. In the first row, 100% (208/208) of hookah images were correctly classified. In the second row, 99.1% (210/212) were correctly classified as hookah images. In the first column, 99% (208/210) were correctly classified as hookah images, and 0.1% (2/210) were correctly classified as nonhookah images. In the second column, out of 210 nonhookah images, 100% (210/210) were correctly classified as nonhookah images, and all images were correctly classified as hookah images.

Figure 9. The learning curve showing the line graph of the accuracy of the classifier with a different number of validated images.

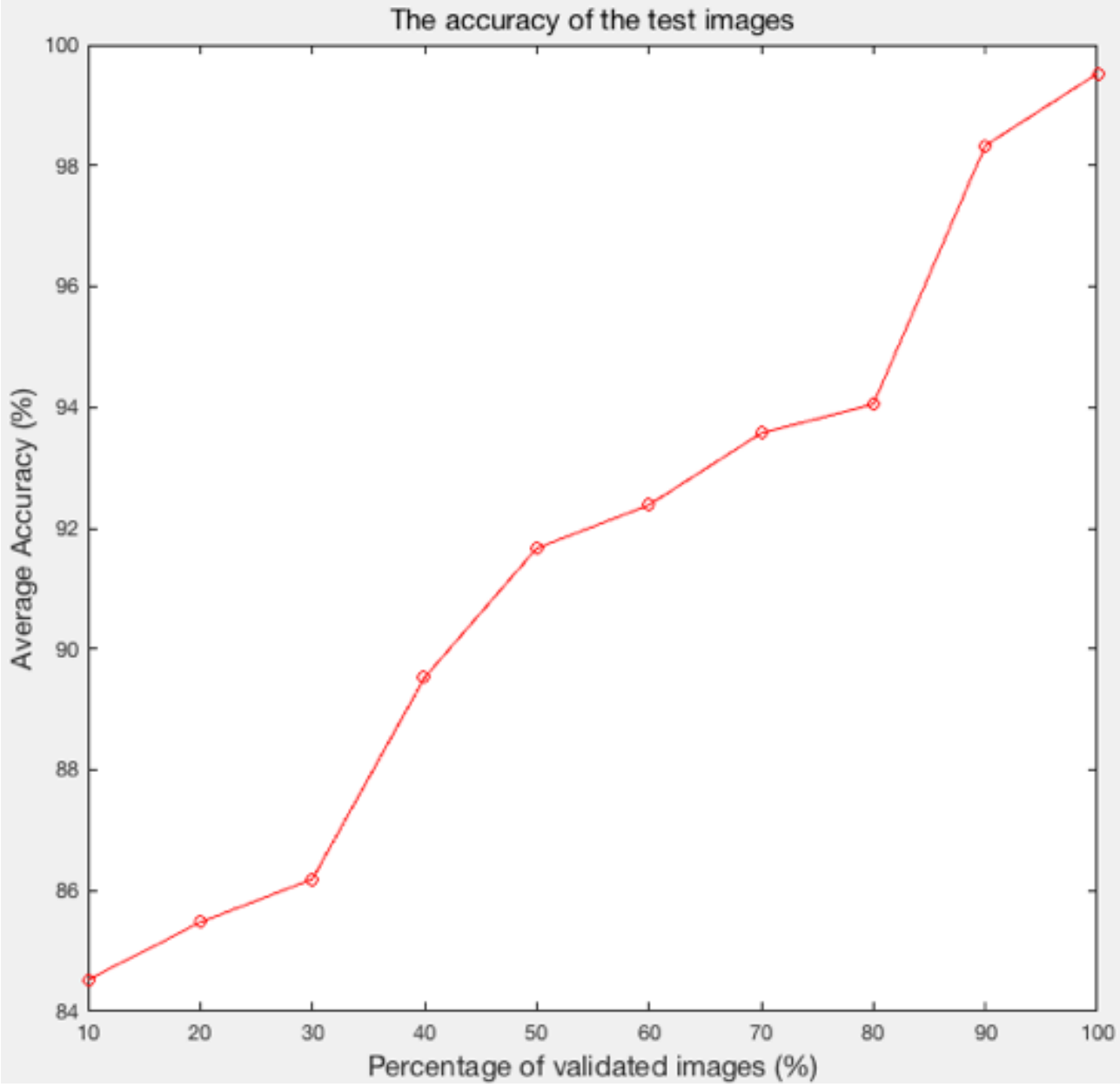
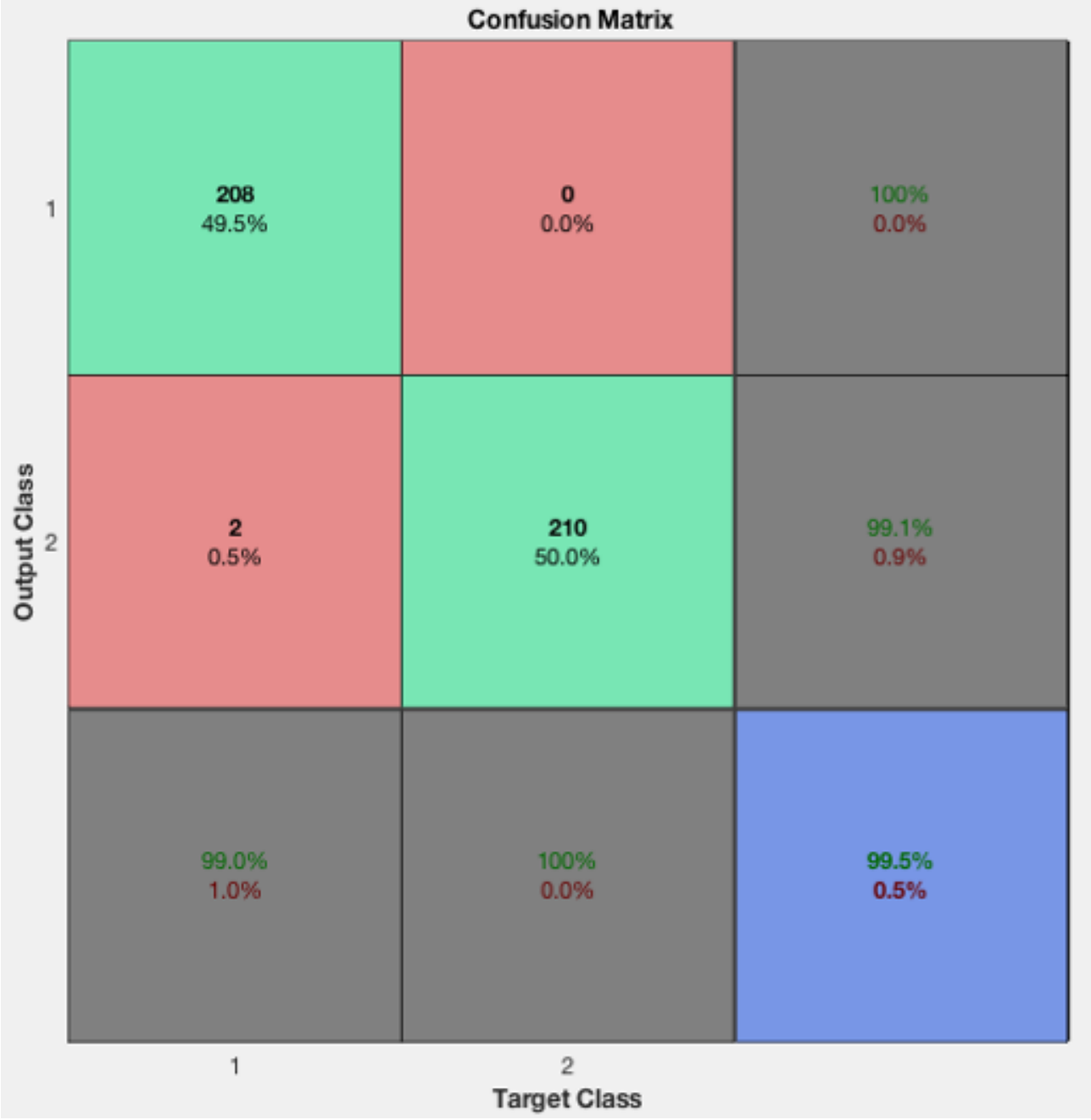


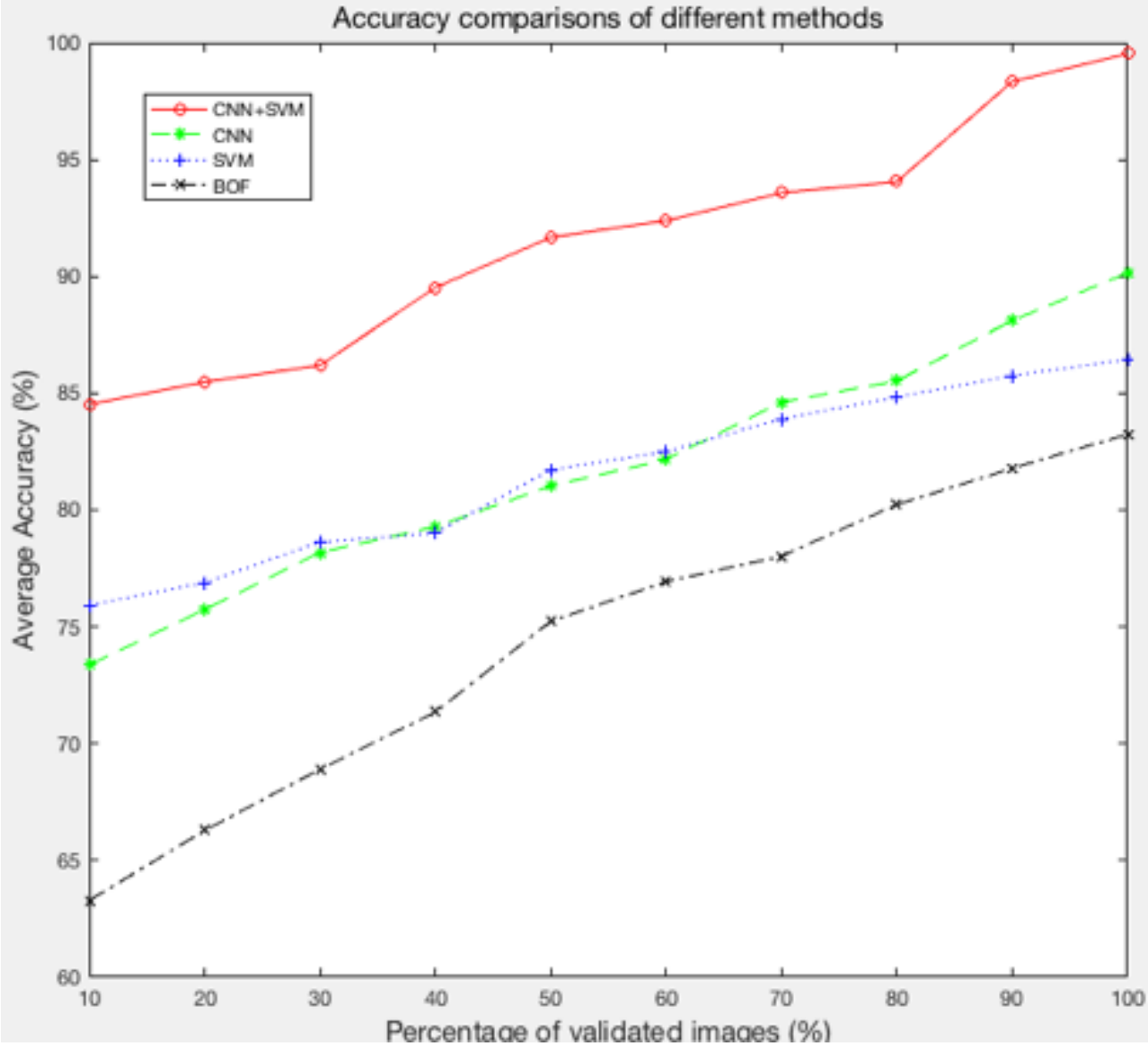
Figure 10. The confusion matrix of the test images (columns 1 and 2 are the hookah and nonhookah categories, respectively, column 3 is the accuracy of classified results).



Comparison With Other Methods

We compared our method with CNN, SVM, and BOF [12,34]. For SVM and BOF, the input was the original image (raw pixel values). Figure 11 shows how the accuracy of various models can be improved as a function of the size of the training data. Our method (CNN+SVM) had the highest accuracy, 99.5% (418/420), compared with other models (CNN, SVM, and BOF).

Figure 11. The prediction accuracy of different methods with different percentages of validated images. CNN: convolutional neural network; SVM: support vector machine; BOF: bag-of-features.



Discussion

Principal Findings

This study showed that the use of a CNN to extract features and SVM to classify images results in higher accuracy in automated image classification compared with using CNN or SVM alone. One crucial advantage of our pipelined approach is that we extracted sufficient features (4096 features from each image representing the details of each image) from a pretrained CNN model (AlexNet), taking advantage of SVM to train the features, saving time. Compared with earlier work using CNN, SVM, and BOF, our method improves accuracy when the number of training images is increased with accuracy reaching 99.5% (418/420), illustrating that our method is suitable for distinct images-like waterpipes.

The methods presented here could help detect increases in the popularity of certain tobacco products over time on social media. By identifying waterpipes in images from Instagram, we can identify Instagram users who may need tobacco-related education to curb hookah use. Instagram may be used to bolster the reach and delivery of health information that communicates the risk of hookah use [35-38]. Earlier research used Instagram images to capture and describe the context in which individuals use and are marketed tobacco products [3,24,25]. For example, the analysis of Instagram data on electronic cigarettes demonstrated that a majority of images were either individuals showing their favorite combinations of products (eg, type of electronic cigarette device and flavored juice) or people performing tricks with the products (eg, blowing a large aerosol cloud in competition with others) [25], demonstrating how and why people use this tobacco product. Previous analyses of hookah-related posts to social media websites provide information about hookah-related contexts, including the importance of stylized waterpipes, use of hookah in social settings, copromotion with alcohol [24], and primarily positive user experiences [39-41].

Earlier studies using image-based data provided timely information from a novel data source; however, their methods

relied on hand coding of images—a process requiring time, expertise, and sample sizes small enough to reasonably code by hand, ultimately limiting the scope of the work. The findings from this study showed how automated image classification could be used to overcome such limitations. In addition, the methods from this study can help researchers in tobacco control identify what proportion of viewers on a social media site are interested in certain products; such methods may be crucial to document the every changing tobacco landscape.

Limitations

The findings from this study should be considered with several limitations in mind, including the fact that our task was a simple binary classification (hookah vs nonhookah), which may result in high accuracy. To eliminate the problem of overfitting, we used ReLU, softmax, dropout layers in a CNN, and utilized several different training datasets (the number of datasets is different, which increased from 42 to 420; Figure 11). The methods developed in this study were only applied in the context of images from Instagram that focused on waterpipes and should be applied in more categories and other contexts in the future. While we had high accuracy in classification, accuracy could be improved with better input features from the CNN model. In the future, researchers should try to enlarge the sets of training images to extract specific features of an image, which may achieve higher accuracy with less computation power.

Conclusions

Findings demonstrated that by combining CNN and SVM to classify images resulted in 99.5% (418/420) accuracy in image classification, which is an improvement over earlier method using SVM, CNN, or BOF alone. A CNN extracts more features of the images, allowing the SVM classifier to be better informed, which results in higher accuracy compared with methods that extract fewer features. Future research can use our method to reduce computational time in identifying objects in images.

Acknowledgments

Research reported in this publication was supported by Grant # P50CA180905 from the National Cancer Institute and the Food and Drug Administration (FDA) Center for Tobacco Products. The National Institutes of Health or FDA had no role in study design, collection, analysis, and interpretation of data, writing the report, and the decision to submit the report for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or FDA.

Conflicts of Interest

None declared.

Authors' Contributions

YZ and JPA conceived of the study and analyzed the data. YZ and JPA drafted the initial manuscript. JBU and TBC received funding for the study. JBU and TBC revised the manuscript for important intellectual content.

References

1. Instagram Business Team. Instagram.com. 2016 Sep 22. Hitting 500,000 Advertisers URL: https://business.instagram.com/blog/500000-advertisers/ [accessed 2018-09-26] [WebCite Cache ID 72j5U9uDE]
2. Allem JP, Cruz TB, Unger JB, Toruno R, Herrera J, Kirkpatrick MG. Return of cartoon to market e-cigarette-related products. Tob Control 2018 Jul 26. [doi: 10.1136/tobaccocontrol-2018-054437] [Medline: 30049688]
3. Allem JP, Escobedo P, Chu KH, Boley TC, Unger JB. Images of Little Cigars and Cigarillos on Instagram Identified by the Hashtag #swisher: Thematic Analysis. J Med Internet Res 2017 Jul 14;19(7):e255 [FREE Full text] [doi: 10.2196/jmir.7634] [Medline: 28710057]
4. Allem JP, Dharmapuri L, Unger JB, Cruz TB. Characterizing JUUL-related posts on Twitter. Drug Alcohol Depend 2018 Dec 01;190:1-5. [doi: 10.1016/j.drugalcdep.2018.05.018] [Medline: 29958115]
5. Perronnin F, Sánchez J, Mensink T. Improving the fisher kernel for large-scale image classification. In: Computer Vision-ECCV. 2010 Presented at: European conference on computer vision; September 5, 2010; Springer, Berlin, Heidelberg p. 143-156.
6. Verma K, Verma LK, Tripathi P. Image Classification using Backpropagation Algorithm. Journal of Computer Science 2014;1:7-15.
7. Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: visualising image classification models and saliency maps.: arXiv:1312.6034v2; 2014. URL: https://arxiv.org/pdf/1312.6034.pdf [accessed 2018-10-30] [WebCite Cache ID 73YkRWIDb]
8. Orlov N, Shamir L, Macura T, Johnston J, Eckley DM, Goldberg IG. WND-CHARM: Multi-purpose image classification using compound image transforms. Pattern Recognit Lett 2008 Jan;29(11):1684-1693 [FREE Full text] [Medline: 18958301]
9. Dehkordy HI. Automated Image Classification Via Unsupervised Feature Learning by K-Means. Indiana University-Purdue University Indianapolis 2015:1-80 [FREE Full text]

10. Sun J, Yang J, Zhang C, Yun W, Qu J. Automatic remotely sensed image classification in a grid environment based on the maximum likelihood method. Mathematical and Computer Modelling 2013 Aug;58(3-4):573-581. [doi: 10.1016/j.mcm.2011.10.063]
11. Chapelle O, Haffner P, Vapnik VN. Support vector machines for histogram-based image classification. IEEE Trans Neural Netw 1999;10(5):1055-1064. [doi: 10.1109/72.788646] [Medline: 18252608]
12. Nowak E, Jurie F, Triggs B. Sampling strategies for bag-of-features image classification. In: European Conference on Computer Vision. 2006 Presented at: European conference on computer vision; May 7, 2006; Springer, Berlin, Heidelberg p. 490-503.
13. Cireşan D, Meier U, Masci J, Schmidhuber J. Multi-column deep neural network for traffic sign classification. Neural Netw 2012 Aug;32:333-338. [doi: 10.1016/j.neunet.2012.02.023] [Medline: 22386783]
14. Jiang YG, Ngo CW CW, Yang J. Towards optimal bag-of-features for object categorizationsemantic video retrieval. 2007 Presented at: Proceedings of the 6th ACM international conference on Image and video retrieval; July 9, 2007; Amsterdam, Netherlands p. 494-501.
15. Marszaek M, Schmid C. Spatial weighting for bag-of-features. 2006 Presented at: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; June 17-22, 2006; New York, New York, United States p. 2118-2125.
16. Auria L, Moro R. Support Vector Machines (SVM) as a Technique for Solvency Analysis. 2007 Presented at: DIW Diskussionspapiere - DIW Berlin Publikationen von Forscherinnen und Forschern des DIW Berlin; August 2008; Berlin, Germany.
17. Burges CJ. A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery 1998;2(2):121-167.
18. Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics 2000 Oct;16(10):906-914. [Medline: 11120680]
19. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 2015:1-14.
20. Vedaldi A, Lenc K. Matconvnet: Convolutional neural networks for matlab. 2015 Presented at: Proceedings of the 23rd ACM international conference on Multimedia; October 26 - 30, 2015; Brisbane, Australia p. 689-692.
21. Ayers JW, Leas EC, Allem JP, Benton A, Dredze M, Althouse BM, et al. Why do people use electronic nicotine delivery systems (electronic cigarettes)? A content analysis of Twitter, 2012-2015. PLoS One 2017;12(3):e0170702 [FREE Full text] [doi: 10.1371/journal.pone.0170702] [Medline: 28248987]
22. Allem JP, Ferrara E, Uppu SP, Cruz TB, Unger JB. E-Cigarette Surveillance With Social Media Data: Social Bots, Emerging Topics, and Trends. JMIR Public Health Surveill 2017 Dec 20;3(4):e98 [FREE Full text] [doi: 10.2196/publichealth.8641] [Medline: 29263018]
23. Allem JP, Ramanujam J, Lerman K, Chu KH, Boley Cruz T, Unger JB. Identifying Sentiment of Hookah-Related Posts on Twitter. JMIR Public Health Surveill 2017 Oct 18;3(4):e74 [FREE Full text] [doi: 10.2196/publichealth.8133] [Medline: 29046267]
24. Allem JP, Chu KH, Cruz TB, Unger JB. Waterpipe Promotion and Use on Instagram: #Hookah. Nicotine Tob Res 2017 Oct 01;19(10):1248-1252. [doi: 10.1093/ntr/ntw329] [Medline: 28077449]
25. Chu KH, Allem JP, Cruz TB, Unger JB. Vaping on Instagram: cloud chasing, hand checks and product placement. Tob Control 2016 Dec;26(5):575-578 [FREE Full text] [doi: 10.1136/tobaccocontrol-2016-053052] [Medline: 27660111]
26. Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. 2010 Presented at: Proceedings of the 27th international conference on machine learning (ICML-10); June 21-24, 2010; Haifa, Israel p. 807-814.
27. Ballester P, de Araújo RM. On the Performance of GoogLeNet and AlexNet Applied to Sketches. 2016 Presented at: Thirtieth AAAI Conference on Artificial Intelligence; February 12 - 17, 2016; Phoenix, Arizona.
28. Yu W, Yank K, Bai Y, Xiao T, Yao H, Rui Y. Visualizing and comparing AlexNet and VGG using deconvolutional layers. 2016 Presented at: ICML'16 Proceedings of the 33rd International Conference on International Conference on Machine Learning; June 19 - 24, 2016; New York, NY, United States.
29. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Communications of the ACM 2017;60(6):84-90.
30. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR. Improving neural networks by preventing co-adaptation of feature detectors. arXiv:1207.0580 2012:1-18.
31. Hua S, Sun Z. Support vector machine approach for protein subcellular localization prediction. Bioinformatics 2001 Aug;17(8):721-728. [Medline: 11524373]
32. Escalera S, Pujol O, Radeva P. On the decoding process in ternary error-correcting output codes. IEEE Trans Pattern Anal Mach Intell 2010 Jan;32(1):120-134. [doi: 10.1109/TPAMI.2008.266] [Medline: 19926903]
33. Escalera S, Pujol O, Radeva P. Separability of ternary codes for sparse designs of error-correcting output codes. Pattern Recognition Letters 2009;30(3):285-297.
34. Jágou H, Douze M, Schmid C. Improving Bag-of-Features for Large Scale Image Search. Int J Comput Vis 2009 Aug 11;87(3):316-336. [doi: 10.1007/s11263-009-0285-2]

35. Pechmann C, Pan L, Delucchi K, Lakon CM, Prochaska JJ. Development of a Twitter-based intervention for smoking cessation that encourages high-quality social media interactions via automessages. J Med Internet Res 2015;17(2):e50 [FREE Full text] [doi: 10.2196/jmir.3772] [Medline: 25707037]
36. Pechmann C, Delucchi K, Lakon CM, Prochaska JJ. Randomised controlled trial evaluation of Tweet2Quit: a social network quit-smoking intervention. Tob Control 2017 Dec;26(2):188-194 [FREE Full text] [doi: 10.1136/tobaccocontrol-2015-052768] [Medline: 26928205]
37. Allem JP, Escobedo P, Chu KH, Soto DW, Cruz TB, Unger JB. Campaigns and counter campaigns: reactions on Twitter to e-cigarette education. Tob Control 2017 Dec;26(2):226-229 [FREE Full text] [doi: 10.1136/tobaccocontrol-2015-052757] [Medline: 26956467]
38. Naslund JA, Kim SJ, Aschbrenner KA, McCulloch LJ, Brunette MF, Dallery J, et al. Systematic review of social media interventions for smoking cessation. Addict Behav 2017 Oct;73:81-93. [doi: 10.1016/j.addbeh.2017.05.002] [Medline: 28499259]
39. Chen AT, Zhu SH, Conway M. Combining Text Mining and Data Visualization Techniques to Understand Consumer Experiences of Electronic Cigarettes and Hookah in Online Forums. Online J Public Health Inform 2015;7(1):e117.
40. Krauss M, 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]
41. Myslín M, Zhu S, Chapman W, Conway M. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. J Med Internet Res 2013 Aug 29;15(8):e174 [FREE Full text] [doi: 10.2196/jmir.2534] [Medline: 23989137]

Abbreviations

BOF: bag-of-features
CNN: convolutional neural network
ConvNets: convolutional networks
FDA: Food and Drug Administration
ReLU: rectified linear units
SVM: support vector machine

Edited by G Eysenbach; submitted 27.03.18; peer-reviewed by M Paul, A Benis; comments to author 05.06.18; revised version received 30.07.18; accepted 07.08.18; published 21.11.18.

Please cite as:
Zhang Y, Allem JP, Unger JB, Boley Cruz T
Automated Identification of Hookahs (Waterpipes) on Instagram: An Application in Feature Extraction Using Convolutional Neural Network and Support Vector Machine Classification
J Med Internet Res 2018;20(11):e10513
URL: http://www.jmir.org/2018/11/e10513/
doi:10.2196/10513
PMID:30452385

Original Paper

Patients' Experiences of Accessing Their Electronic Health Records: National Patient Survey in Sweden

Jonas Moll[1], PhD; Hanife Rexhepi[2], PhD; Åsa Cajander[1], PhD; Christiane Grünloh[3,4], MSc, Tech Lic; Isto Huvila[5], PhD; Maria Hägglund[6], PhD; Gunilla Myreteg[1,7], PhD; Isabella Scandurra[8], PhD; Rose-Mharie Åhlfeldt[2], PhD

[1]Department of Information Technology, Uppsala University, Uppsala, Sweden

[2]School of Informatics, University of Skövde, Skövde, Sweden

[3]School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden

[4]Institute of Informatics, TH Köln University of Applied Sciences, Gummersbach, Germany

[5]Department of ALM, Uppsala University, Uppsala, Sweden

[6]Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

[7]Department of Business Studies, Uppsala University, Uppsala, Sweden

[8]Centre for Empirical Research on Information Systems, Örebro University School of Business, Örebro, Sweden

Corresponding Author:
Jonas Moll, PhD
Department of Information Technology
Uppsala University
Lägerhyddsvägen 2, House 2
Uppsala, SE-751 05
Sweden
Phone: 46 703174383
Email: jonas.moll@it.uu.se

Abstract

Background: Internationally, there is a movement toward providing patients a Web-based access to their electronic health records (EHRs). In Sweden, Region Uppsala was the first to introduce patient-accessible EHRs (PAEHRs) in 2012. By the summer of 2016, 17 of 21 county councils had given citizens Web-based access to their medical information. Studies on the effect of PAEHRs on the work environment of health care professionals have been conducted, but up until now, few extensive studies have been conducted regarding patients' experiences of using PAEHRs in Sweden or Europe, more generally.

Objective: The objective of our study was to investigate patients' experiences of accessing their EHRs through the Swedish national patient portal. In this study, we have focused on describing user characteristics, usage, and attitudes toward the system.

Methods: A national patient survey was designed, based on previous interview and survey studies with patients and health care professionals. Data were collected during a 5-month period in 2016. The survey was made available through the PAEHR system, called *Journalen*, in Sweden. The total number of patients that logged in and could access the survey during the study period was 423,141. In addition to descriptive statistics reporting response frequencies on Likert scale questions, Mann-Whitney tests, Kruskal-Wallis tests, and chi-square tests were used to compare answers between different county councils as well as between respondents working in health care and all other respondents.

Results: Overall, 2587 users completed the survey with a response rate of 0.61% (2587/423,141). Two participants were excluded from the analysis because they had only received care in a county council that did not yet show any information in *Journalen*. The results showed that 62.97% (1629/2587) of respondents were women and 39.81% (1030/2587) were working or had been working in health care. In addition, 72.08% (1794/2489) of respondents used *Journalen* about once a month, and the main reason for use was to gain an overview of one's health status. Furthermore, respondents reported that lab results were the most important information for them to access; 68.41% (1737/2539) of respondents wanted access to new information within a day, and 96.58% (2454/2541) of users reported that they are positive toward *Journalen*.

Conclusions: In this study, respondents provided several important reasons for why they use *Journalen* and why it is important for them to be able to access information in this way—several related to patient empowerment, involvement, and security. Considering the overall positive attitude, PAEHRs seem to fill important needs for patients.

(J Med Internet Res 2018;20(11):e278) doi:10.2196/jmir.9492

KEYWORDS

eHealth; medical records; national survey; patients; patient-accessible electronic health records; patient portal; personal health records

Introduction

Internationally, there is movement toward providing patients with Web-based access to their electronic health records (EHRs); this parallels a global shift toward increased patient empowerment and patient participation. In the United States, for instance, the *OpenNotes* initiative for providing patients access to their EHR began as a pilot and evaluation project that included 105 volunteer primary care physicians and their 19,000 patients [1,2]. The project started in 2010 and has since spread throughout the United States [3]. *Blue Button* is a similar initiative by the United States Department of Veteran Affairs, which enables Veteran Affairs patients to access data from their EHR through Web, such as clinical notes, Veteran Affairs appointments, test results, etc [4,5]. Similar schemes have been initiated in Australia [6], Finland [7], Canada [8], Denmark [9], Estonia [10], the United Kingdom [11], and Sweden [12]. However, different strategies and approaches have affected the uptake and impact. The implementation progress has in several countries been slow because of legal constraints [13,14] and concerns about, for example, security and privacy among medical professionals [8,12,15].

In Sweden, citizens are provided with the service *Journalen* for patient-accessible EHR (PAEHR), accessible through Web via the national patient portal. The PAEHR service accesses the EHR information through a national health information exchange platform. Hence, patients have one access point to all their health record information regardless of (1) how many health care providers they have visited, and (2) which EHR system their health care providers use [16]. However, there are limitations and exceptions to patient access. Whether patients have access to their EHR depends on whether they receive care from a public or private health care provider. For example, it is possible that although the county has implemented *Journalen*, specific private providers do not give access to their notes. At the time of the study, each health authority could choose whether they give patients immediate access to signed (ie, confirmed) notes, unsigned notes, or whether a delay of 14 days is implemented to either or both of them [17,18]. There seems to be no standard practice for physicians to sign notes [19] and some notes are never signed [20]. Following a court decision that deemed health care notes in the records as "public documents" and, thus, patients having the right to read them, the implementation in Region Uppsala was changed to let patients decide what kind of notes they want to read [20].

Figure 1. The patient-accessible electronic health record Journalen after log-in, showing the functions and information available (partially translated). Licensed under fair use. Source: https://e-tjanster.1177.se/. Service produced by Inera AB under the auspices of Swedish county councils and regions.

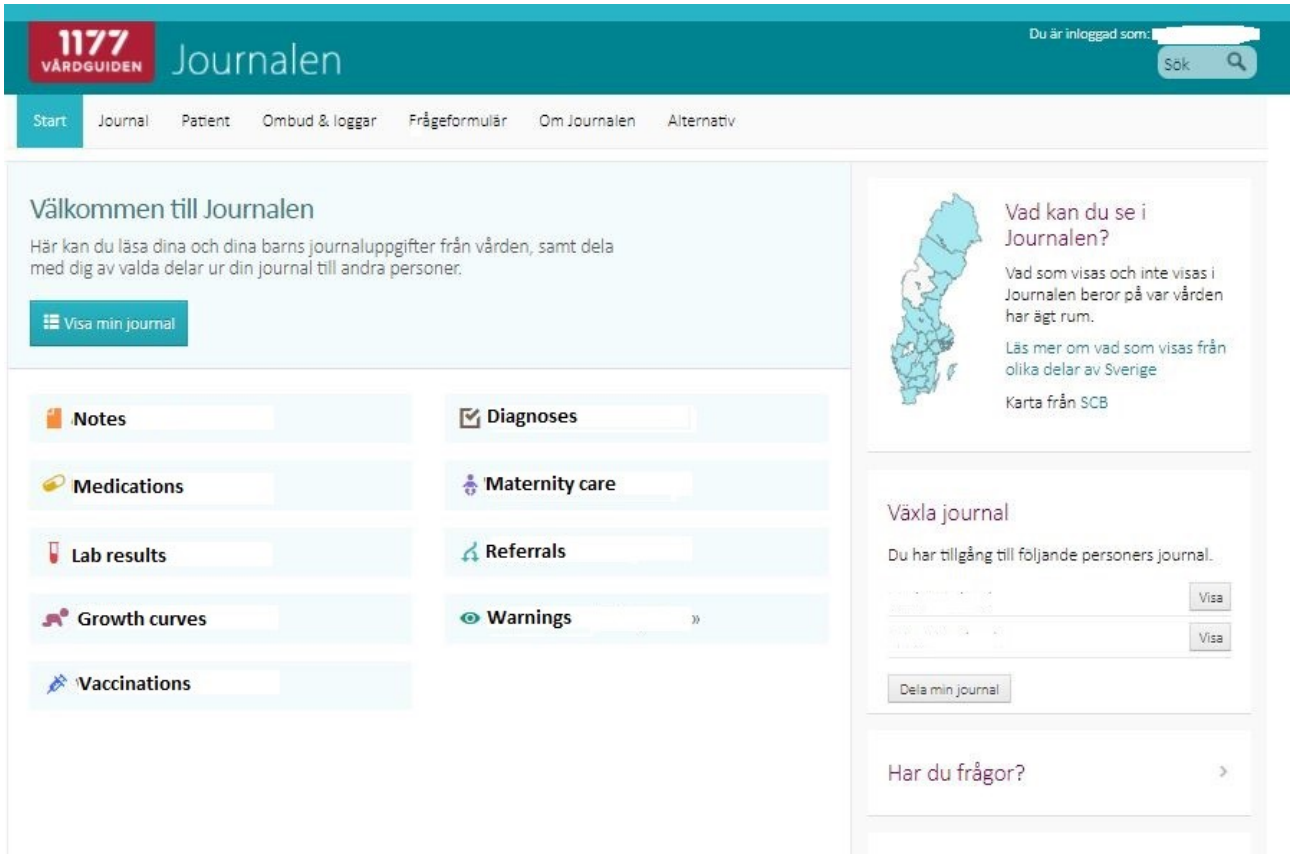


Figure 2. Information shown in Journalen depending on county council or health care provider (blue squares) during the time of the survey.

	Medical Notes	Diag-noses	Visits (time/date / provider)	Lab Results	Medi-cations	Immuni-zations	Referrals	Medical alerts	Access Log Lists	Psychi-atry	Health Declaration Forms	Blocking of record	TOTAL
Blekinge	■	■	■	■	■			■					6
Dalarna	■		■				■						3
Gotland													0
Gävleborg													0
Halland	■	■	■	■		■							5
Jämtland/ Härjedalen													0
Jönköping	■	■				■							3
Kalmar	■	■			■	■							4
Kronoberg	■	■								■			3
Norrbotten	■		■		■	■							4
Skåne	■	■	■							■			4
Stockholm	■	■	■	■									4
Södermanland	■	■	■					■					4
Uppsala	■	■		■	■		■	■	■		■	■	9
Värmland	■	■		■		■	■		■				6
Västerbotten	■	■	■		■								4
Västernorrland													0
Västmanland	■						■		■				3
VGR	■	■	■					■					4
Örebro	■	■	■	■	■	■							6
Östergötland	■	■	■	■	■	■	■						7
Capio (private care provider)	■	■	■	■									4
TOTAL of regions (of 20)	18	15	12	8	7	7	5	4	3	2	1	1	

Currently, when patients access the PAEHR, they find varying clinical content, such as medical notes from the EHRs (from all health care professions and all connected health care providers who have agreed to give access, both public and private), a list of prescribed medications, lab results, warnings, diagnosis, maternity care records, referrals, and vaccinations (Figure 1). Although the PAEHR interface is identical for all users, there are significant differences in how much information to which each health care provider gives access. Figure 2 gives an overview of what types of clinical content the health care providers have chosen to give access to. The *access log list* (also called "consult audit trail") presents a list of everyone who has accessed the record [13], including their name, role, and the date they accessed the record. The list includes health care professionals as well as anyone patients have chosen to share their record with, and patients can make use of the list to check whether there has been unexpected access.

Although patients' use of and attitudes toward PAEHRs have been studied to some degree, research on PAEHRs has up until today mostly focused on the health care professionals' perspective [8,12,19,21,23], and most studies originate from the US context [23]. For example, health care professionals have had several concerns such as the negative impact on workload, privacy risks, and fear of increased anxiety in patients [23]. However, in contrast to the fears of many health care professionals, a Swedish interview study with cancer patients showed that Web-based access did not generate substantial anxiety, concerns, or increase in the number of telephone contacts to care units [22]. The same study showed that the patients were generally positive toward the system and the possibilities it gives to read and follow their medical treatment. In addition, an extensive meta-study by Mold et al [24] on patients' attitudes showed a similar generally positive attitude. Improved communication with clinicians, as well as improved satisfaction and self-care, were among the most important findings. Similar conclusions were drawn in a meta-study by de Lusignan et al [23], indicating that Web-based access to own EHR and related services offers increased convenience and satisfaction to patients. A review published in 2015 called for more empirical testing regarding the effect of PAEHR on outcomes for both patients and health care providers [25]; this is especially true for follow-up studies, investigating the effects of long-term use of PAEHR systems. In addition, the meta-study by Mold et al [24] concluded that most studies on PAEHR focusing on patients' usage and attitudes had been conducted in relation to the *OpenNotes* movement in the United States. To date, few follow-up studies have been conducted in Europe and none in Sweden. The Swedish interview study with patients

performed by Rexhepi et al [22] was only concerned with patients with cancer in Region Uppsala and was conducted shortly after *Journalen* was introduced. Hence, it did not answer questions about which types of patients used the Swedish PAEHR system or attitudes toward the system in the wider population representing patients with other conditions and from different county councils with different implementations. The large Swedish national survey study about PAEHR use and attitudes presented in this paper aims to fill the abovementioned research gaps by answering the following research questions several years after *Journalen* was introduced:

- Why do patients in Sweden use *Journalen*? And how often do they use it?
- What types of information are most valued by patients?
- What are the general attitudes by patients toward *Journalen*?
- What differences can be identified with regards to attitudes between different county councils in Sweden?

After the survey study has been introduced in section 2, results regarding demographics, usage, and attitudes toward *Journalen* will be presented in section 3. Section 4 includes discussions of key results and their relations to earlier studies with patients as well as health care professionals. The paper ends with conclusions and a short discussion about the need for further studies in this area.

Methods

A survey study was conducted to elicit opinions and experiences of patients using *Journalen*. The study was conducted from June to October 2016, after ethical approval of the research was granted by the Regional Ethical Review Board in Uppsala, Sweden (EPN 2016/129). Participants were recruited through the national electronic health (eHealth) service *Journalen*. When patients logged into the service *Journalen*, they received a request for voluntary survey participation together with information about the study. Thus, only active users of *Journalen* received a request for participation. Patients were automatically presented with standard consent on Web prior to completing the survey. Participants accessed the survey, and the possibility to give consent, by following a link beneath the information about the study.

An anonymous self-completion questionnaire was designed covering six different topic areas with a total of 24 questions (see questionnaire in Multimedia Appendix 1):

- General questions related to the eHealth service *Journalen*
- Questions targeting experiences from accessing and using the content of *Journalen*
- Information security
- General questions about information needs, behavior, and information-seeking styles
- Personal health-related questions
- Demographics

The questionnaire included questions with various response options (5-point Likert scale, multiple choice, and free text form). The selection of topic areas and the design of the questionnaire was informed by previous studies, including Huvila et al [26], Rexhepi et al [22], Grünloh et al [19], Huvila et al [27], Ålander & Scandurra [28], and Scandurra et al [21]. The usability and technical functionality of the electronic questionnaire had not been tested before fielding the questionnaire. However, participants received information about whom to contact in case of technical issues. Overall, 2587 patients from 21 county councils completed the survey. The number of respondents for each county council or region varied. Notably, it was not possible to statistically verify whether the number of respondents is adequate to provide other than some tentative county and group-wise comparisons.

The collected data were managed by the eHealth service provider Inera in accordance with the security requirements presented in the ethical application and approved from the Regional Ethical Review Board. The survey data were stored in the same database system as the PAEHR, meaning that the collected data, including patient identification (ID), had the same security protection as all patient information handled in the PAEHR. A patient ID was stored during the collection period to ensure that patients have not left duplicate responses. When the collection period was completed, the patient ID was removed and all stored information was anonymized. After being collected, the anonymized dataset was exported to researchers for analysis. In order to provide a clear focus on the research questions on which this study is based, a selection of themes and questions were made. Questions regarding demography, general usage data, and attitudes were selected for analysis. Before the analysis began, data from patients only treated in Region Gävleborg were excluded (2 patients) because *Journalen* had not yet been introduced in that county council at the time the survey was open. It was, however, possible to access the e-service and the survey without having access to any EHR data, and as these respondents had only received care in that county council, their experiences were considered not relevant for this study.

The results presented in this paper are based on the descriptive analysis showing general trends related to the themes mentioned above. The reported percentages are based on those who answered each specific question—the total number of respondents on each question are indicated either in the main text or Multimedia Appendix 2. Only completed questionnaires have been analyzed, as the answers were stored in the database only when the respondent chose to submit on the last page. The Mann-Whitney U test was used for Likert scale items after a transformation (strongly agree=5; strongly disagree=1) for pairwise comparisons between the group of respondents who worked or had worked in health care and the group consisting of all other respondents. In addition, chi-square test was used for questions with nominal scales. The Kruskal-Wallis test was used to detect possible differences in ratings between county councils. The significance level was set to 95% for all tests. The SPSS 25 software was used for all analyses.

Results

Result Presentation

Results regarding the users of *Journalen*, as well as their usage of the system and attitudes toward it, are presented below.

References to specific questions in the appended questionnaire will be given as *{qX}*, where *X* denotes the question number.

Demographic Information

During the survey period, 423,141 users logged into *Journalen*, of which 2587 patients completed the survey (unique users that logged in; response rate, 0.61%, 2587/423,141). Although not everyone answered all the questions, >90% (2338/2587) responded to each question in the survey. Of all respondents, 62.97% (1629/2587) identified as women and 30.85% (798/2587) as men; 0.39% (10/2587) respondents chose "other" and 150 did not answer this question. According to the statistics about *Journalen* from the company providing the service, Inera [29], this reflects the gender distribution of the users in general (in 2016: 60% women and 40% men). Of all respondents, 39.81% (1030/2587) stated that they were working or had been working within health care and 54.54% (1411/2587) stated that they had no professional relation to health care (146 respondents did not answer this question). Independent of the county council or region a person lives, it is possible to receive care in a different county as well. In addition, 1674 respondents indicated that they, indeed, had received care from a county council other than their home county council. Table 1 shows from which county council or region respondents come from, as well as the number of respondents that received care in the respective council or region. Of note, county councils, which had not yet introduced *Journalen* at the time when the survey was closed are not included in the third column of Table 1. Table 2 shows the educational levels represented among respondents. Respondents with at least 3 years of higher education background are in the majority.

To sum up, the survey results regarding user characteristics on a national level indicate that a majority of respondents were women and that the majority had studied at least 3 years on the higher education level. In addition, results indicate that many users of *Journalen* had experiences both of being patients and working as medical professionals.

Table 1. The number of participating patients from each county council or region and the total number of respondents who have received care in the respective county council or region.

County council or region	Respondents from respective county council or region, n	Respondents who received care in the respective county council or region, n
Region Skåne	520	692
Region Uppsala	333	520
Region Östergötland	241	364
Region Västra Götaland	179	328
Region Jönköping	154	218
Värmland county council	143	180
Västmanland county council	103	163
Region Örebro	102	185
Sörmland county council	94	149
Region Kronoberg	94	133
Dalarna county council	93	160
Västerbotten county council	93	144
Kalmar county council	78	133
Norrbotten county council	57	98
Region Halland	54	106
Blekinge county council	51	101
Stockholm county council	41	299
Region Gävleborg	7	N/A[a]
Västernorrland county council	6	N/A
Region Gotland	1	N/A
Region Jämtland Härjedalen	1	N/A
Not specified	142	N/A

[a]N/A: not applicable.

Table 2. Educational level of persons who answered the survey (N=2455).

Educational level	Value, n (%)
Research education	75 (3.05)
Higher education ≥3 years	945 (38.49)
Higher education <3 years	467 (19.02)
High school ≥3 years	410 (16.70)
High school <3 years	248 (10.10)
Less than high school	159 (6.48)
No formal education	66 (2.69)
Other	85 (3.46)

Usage

Regarding the frequency of use *{q2}*, 72.08% (1794/2489) of the patients answering the survey logged into *Journalen* about once a month, whereas others logged in about once a week (240/2489, 9.64%), a few times a week (393/2489, 15.79%), and several times a day (62/2489, 2.49%). Thus, the majority of patients answering the survey were infrequent users. The chi-square test showed no statistically significant association between respondents who were working or had been working in health care and those who had not (χ^2_4=1.5, *P*=.83).

Table 3 shows a mapping between the frequency of use data and self-reported demographic data about health conditions *{q23, q24}*. Persons who considered their health as "very good" were among the least frequent users, whereas those with cancer, diabetes, or other chronic conditions were among the most frequent users. Many health care providers chose not to give access to psychiatric records, yet respondents who stated that they have a psychiatric condition appeared to access the record similarly to the aforementioned conditions. They may, of course, have other health issues that they are interested in but based on these results, there appears to be no major difference in how patients who identify as belonging to psychiatry access their records.

One part of the patient survey focused on why they were using *Journalen {q4}*. The three most common reasons were gaining an overview, following up on visits, and becoming more involved, and the least common reason was that they suspected inaccuracies (Figure 3). The most common reason for using *Journalen* was to obtain an overview of one's medical history and treatment. Multimedia Appendix 2 provides more detailed results from all items in *{q4}*.

Mann-Whitney *U* test showed no statistically significant difference between the group of current or former health care professionals and the group of all other respondents regarding the reasons *overview* (*U*=717,488, *P*=.13), *follow-up* (*U*=699,543, *P*=.83), *become more involved* (*U*=655,657, *P*=.16), and *suspect error* (*U*=653,566, *P*=.50). Significant differences between these groups were only found for *not being sure about treatment* (*U*=704,309, *P*=.001) and *prepare for a visit* (*U*=714,054, *P*=.003). In both these cases, current and former health care professionals gave significantly lower ratings. See detailed results in Multimedia Appendix 2.

When asked about how long respondents were willing to wait until information was available after a visit *{q6}*, 68.41% (1737/2539) wanted access to new information same day or after a day (in other words, within 24 hours). Furthermore, 19.22% (488/2539) respondents wanted access to new information within 2 weeks, 1.42% (36/2539) within 1 month, and 10.95% (278/2539) chose "other." Respondents were informed that the alternatives "Same day" and "After a day" would mean that the physician may not yet have signed the notes and probably not have seen, for instance, new test results. The chi-square test showed no statistically significant association between respondents who were working or had been working in health care and those who had not (χ^2_5=7.8, *P*=.17).

To sum up, the survey results showed that most respondents were infrequent users of *Journalen*, which was especially true for users who considered themselves to be in good health. Patients with chronic conditions were among the most frequent users. The main reasons for using *Journalen* were (1) to receive an overview of one's own medical history and treatment; (2) to follow up on doctor's visits; and (3) to become more involved in one's care.

User Attitudes and Perceptions

Overall, patients who answered the survey were positive toward *Journalen*, as indicated in Table 4. Of all respondents, >96% (2455/2528 and 2454/2541 for the respective questions presented in Table 4) showed a positive attitude (strongly agree or agree) toward *Journalen {q3}*. As can be seen in Table 5, there are no substantial differences between county councils regarding attitudes toward the reform. In both Tables 4 and 5 the responses "Strongly agree" and "Agree" have been summed up as "Positive" and "Strongly disagree" and "Disagree" have been summed up as "Negative." A Kruskal-Wallis test showed no statistical differences in attitude ratings between the different county councils (χ^2_{15}=10.7, *P*=.77), showing no support for an effect of the county council. The most positive respondents (98.04%, 150/153, positive and 0.65%, 1/153, negative) belonged to the Dalarna county council, and the least positive (93.62%, 88/94, positive) respondents belonged to Blekinge county council. More detailed results from *{q3}*, for the respective county council, can be found in Multimedia Appendix 2.

Statistically significant differences between the group of current and former health care professionals and the group of all other respondents were found both regarding *Journalen* as a reform (*U*=784,071, *P*<.001) and *Journalen* as good for them (*U*=728,196, *P*=.005). In both these cases, current or former health care professionals were significantly more negative in their answers. See detailed results for these two groups in Multimedia Appendix 2.

Owing to large differences between county councils regarding what functions were available in *Journalen*, it is important to consider the county council when evaluating user attitudes toward and perceived importance of *{q17}* different functions. In Table 6, answers about the perceived importance from respondents who belonged to county councils or regions where a certain type of information was shown are compared with the answers from respondents from county councils or regions where the information was not shown. Overall, access to test results is perceived to be the most important category and log list the least important. More detailed results from the chosen items in *{q17}*, for the respective county council, can be found in Multimedia Appendix 2.

When comparing current and former health care professionals against the group of all other respondents for the information types presented in Table 6, significant differences were found for immunizations (*U*=640,633, *P*<.001), health declaration forms (*U*=615,435, *P*<.001), and log list (*U*=640,253, *P*=.004). In all these cases, current or former health care professionals gave significantly higher ratings. See detailed results in Multimedia Appendix 2. Statistically significant differences in ratings between county councils or regions that present a certain type of information and those who do not were only found for test results (*U*=710,736, *P*=.049) and visit history (*U*=632,152, *P*=.005). Thus, whether or not the information is accessible in a particular county council does not appear to have a significant impact on the rating of importance of the most of the information types.

Table 3. Frequencies of usage of *Journalen* by respondents belonging to different condition categories.

Condition	Several times a day, n (%)	Few times a week, n (%)	Once a week, n (%)	About once a month, n (%)
Cancer (n=335)	18 (5.4)	70 (20.9)	40 (11.9)	207 (61.8)
Diabetes (n=260)	6 (2.3)	50 (19.2)	35 (13.5)	169 (65.0)
High blood pressure (n=589)	17 (2.9)	84 (14.3)	60 (10.2)	428 (72.7)
Psychiatry (n=487)	11 (2.3)	98 (20.1)	48 (9.9)	330 (67.8)
In good health (n=1096)	22 (2.0)	114 (10.4)	86 (7.8)	874 (79.7)

Figure 3. Percentage of respondents who strongly agree, agree, are neutral, disagree, or strongly disagree with the statement that they use Journalen to get an overview of their medical history and treatment, to follow up on what has been said during visits, to become more involved in their own care, and because they suspect inaccuracies, respectively {q4}.

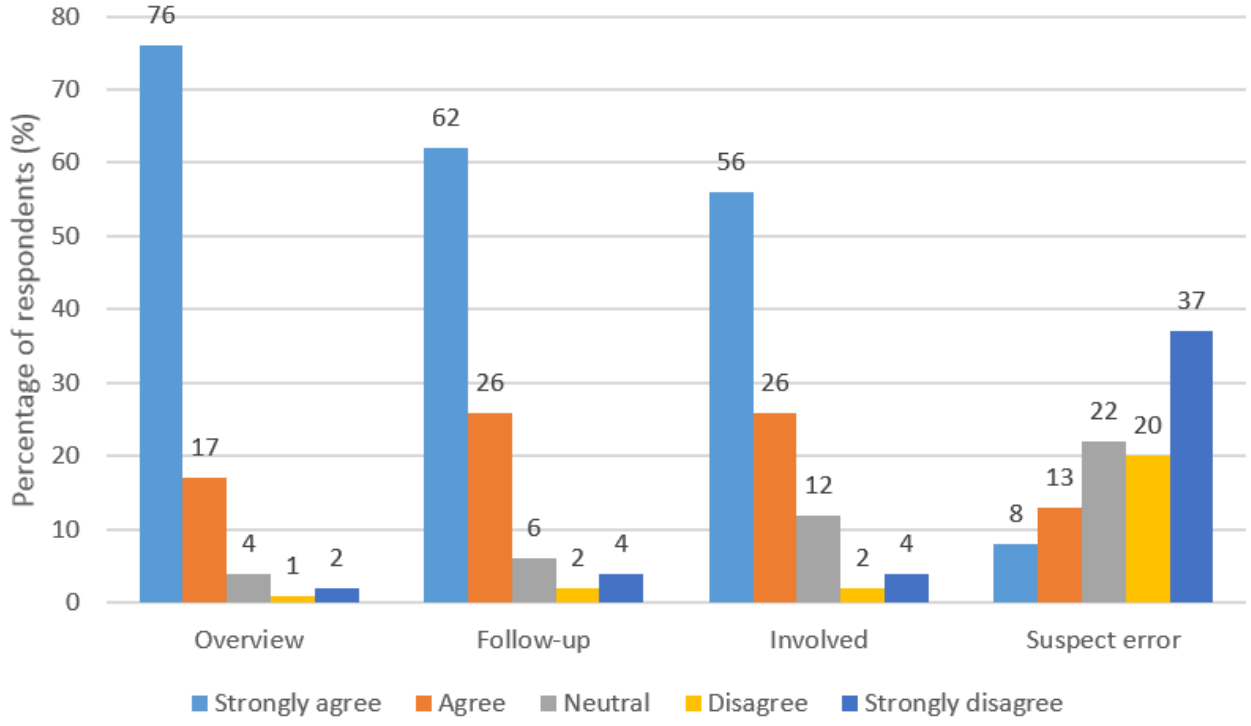


Table 4. Number of respondents who considered access to *Journalen* as good for them (N=2528) and as a good reform in general (N=2541) *{q3}*.

Question	Positive, n (%)	Neutral, n (%)	Negative, n (%)
"Getting access to *Journalen* is good for me"	2455 (97.11)	39 (1.54)	34 (1.34)
"Allowing access to *Journalen* is generally a good reform"	2454 (96.58)	38 (1.50)	49 (1.93)

Table 5. Number of participants from different county councils who were positive, neutral, and negative toward *Journalen* as a reform within health care.

County council	Positive, n (%)	Neutral, n (%)	Negative, n (%)
Region Skåne (n=653)	636 (97.3)	6 (1.0)	11 (1.7)
Region Uppsala (n=491)	477 (97.1)	6 (1.2)	8 (1.6)
Region Östergötland (n=341)	327 (95.9)	6 (1.8)	8 (2.3)
Region Västra Götaland (n=304)	286 (94.1)	8 (2.6)	10 (3.3)
Stockholm county council (n=291)	278 (95.5)	5 (1.7)	8 (2.8)
Region Jönköping (n=203)	199 (98.0)	2 (1.0)	2 (1.0)
Region Örebro (n=169)	162 (95.9)	6 (3.5)	1 (0.6)
Värmland county council (n=170)	163 (95.9)	4 (2.3)	3 (1.8)
Västmanland county council (n=153)	148 (96.7)	3 (2.0)	2 (1.3)
Dalarna county council (n=153)	150 (98.0)	2 (1.3)	1 (0.7)
Sörmland county council (n=138)	134 (97.1)	2 (1.4)	2 (1.4)
Västerbotten county council (n=139)	132 (95.0)	2 (1.4)	5 (3.6)
Region Kronoberg (n=128)	123 (96.1)	2 (1.6)	3 (2.3)
Kalmar county council (n=128)	123 (96.1)	2 (1.6)	3 (2.3)
Region Halland (n=101)	96 (95.0)	3 (3.0)	2 (2.0)
Blekinge county council (n=94)	88 (93.6)	3 (3.2)	3 (3.2)
Norrbotten county council (n=94)	92 (97.9)	0 (0.0)	2 (2.1)

When asked about the importance of being able to access patient-related data, >93% (2348/2506) of respondents strongly agreed or agreed with the statement that it made them feel more informed (Figure 4). The diagram shows the four highest rated reasons to why respondents believed the access to patient-related information to be of importance to them *{q5}*. These are as follows: (1) it makes them feel informed; (2) it improves communication between medical staff and them; (3) it improves the understanding of their condition; and (4) it makes them feel safe. As a comparison, the results for the alternative "Not important" are also provided in Figure 4. It is clear that the majority of respondents found value in being able to access information about their health. More detailed results from all items in *{q5}* can be found in Multimedia Appendix 2.

Several significant differences were found when using the Mann-Whitney U test to compare ratings from current and former health care professionals with all other respondents. Significant differences were found regarding the items feel informed (U=725,164, P=.02), improve communication (U=763,062, P<.001), better understanding (U=746,053, P<.001), and not important (U=585,143, P=.02). In all these cases, those who did not belong to the group of current or former health care professionals gave significantly more positive ratings. More details can be found in Multimedia Appendix 2.

Finally, 26.07% (655/2512) of respondents stated that they had felt worried about something they had read in *Journalen {q12}*. Thus, the majority of respondents have not been worried by the contents of *Journalen*. The chi-square test did not show any significant association between the group of current or former health care professionals and the group of all other respondents (χ^2_2=2.9, P=.24).

When asked about what actions they took in cases when they felt worried after reading in *Journalen*, the most common answer of the respondents was that they had searched for information on the internet (339/655), followed by calling the hospital (237/655) and asking during the next doctor's visit (235/655). The least common action to take was to contact a patient association (36/655).

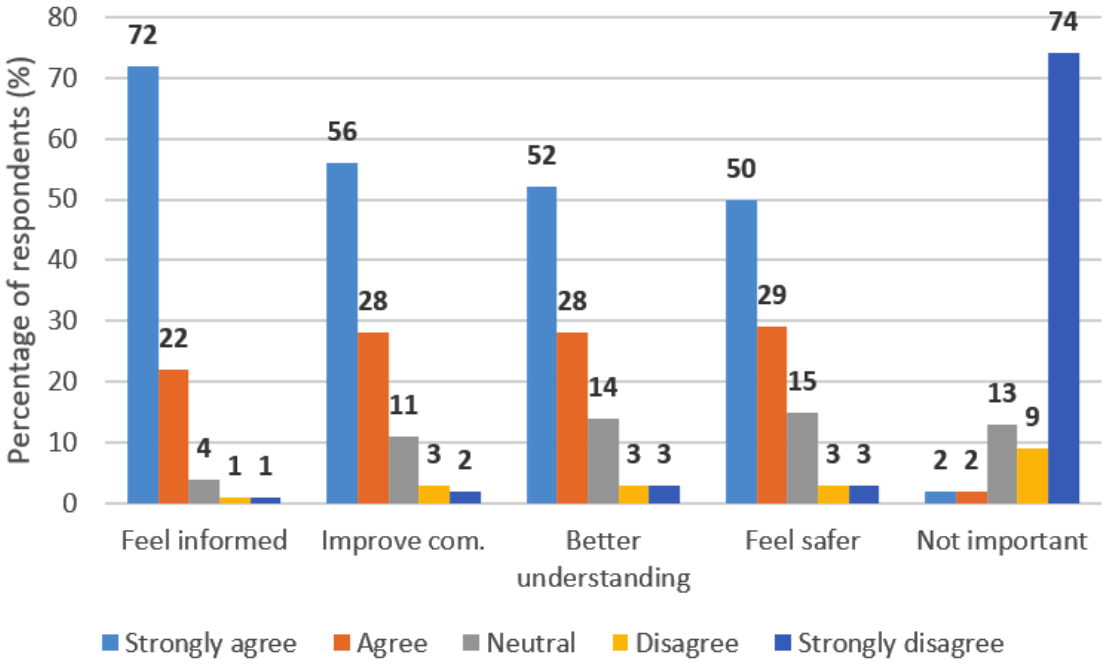
To sum up, the survey results revealed a strong positive attitude toward *Journalen* as a reform and a resource. The top three reasons why patients believed that *Journalen* is important were as follows: (1) it makes them feel more informed; (2) it improves their communication with medical staff; and (3) it results in a better understanding of one's health status. The most important resource, according to the survey, was test results.

Table 6. Comparisons between answers from county councils that present a particular type of information (Yes) and those who do not (No).

Information types	Value, n (%)	Mean (SD)	*P* value
Referral			.62
Yes	894 (37.33)	4.60 (0.75)	—[a]
No	1501 (62.67)	4.59 (0.79)	—
Medlist			.53
Yes	933 (38.88)	4.61 (0.79)	—
No	1467 (61.12)	4.61 (0.83)	—
Immunization			.88
Yes	809 (33.96)	4.39 (0.94)	—
No	1573 (66.04)	4.37 (1.02)	—
Labresult			.049
Yes	948 (39.45)	4.75 (0.61)	—
No	1455 (60.55)	4.77 (0.66)	—
Visithistory			.005
Yes	1489 (62.43)	4.64 (0.75)	—
No	896 (37.57)	4.57 (0.80)	—
Healthforms			.37
Yes	560 (23.18)	4.14 (1.21)	—
No	1856 (76.82)	4.09 (1.24)	—
Loglist			.26
Yes	320 (13.22)	4.08 (1.04)	—
No	2101 (86.78)	4.12 (1.07)	—

[a]Not applicable.

Figure 4. Percentage of respondents who strongly agree, agree, are neutral, disagree, or strongly disagree with that accessing patient information is important for them because it makes them feel informed, improves communication between medical staff and them, improves the understanding of their condition, or makes them feel safe; the “Not important” option is shown as a comparison {q5}.



Discussion

Principal Results

Regarding attitudes, the main finding is that, similarly to earlier research, respondents are generally positive toward the service; this is congruent with findings related to other PAEHR studies, such as *Blue Button* [5] or *OpenNotes* [30]. Furthermore, attitudes do not differ greatly between patients from different county councils in Sweden; this shows that large differences between county councils, regarding which type of information is accessible, do not have a major effect on the overall acceptance of the service. The same tendency can be observed when considering the importance of having access to the different types of information (eg, test results) or the availability of a function (eg, ability to provide a health declaration for the next visit). Whether this function or information is currently available in that county or not does not to a large extent affect the rating of its importance. Statistical differences between county councils that provide a particular type of information and those who do not were only found for test results and health declaration forms; this result was expected, as the question regarded the information types as such and not their possible implementations in *Journalen*. In addition, it indicates that the implementations of the information sources in *Journalen* do not have a negative impact in this respect. According to Table 6, test results are perceived to be the most important information source. This is interesting because currently most county councils in Sweden do not show this information. These results give indications about which information types to prioritize in future development iterations. Regarding the perceived effects of using *Journalen*, the alternatives "Feeling informed" and "Improved communication with medical staff" are the most highly rated selected by respondents, which is supported by earlier findings from interviews with patients conducted right after launch in Region Uppsala [22].

When it comes to the usage, it is clear that the majority of respondents are not frequent users of *Journalen* (use frequency about once a month); this may be explained by that the majority of respondents indicated that they are in good health. Logging into *Journalen* frequently would not be relevant if an individual is currently not having an ongoing health issue with an active health care contact. Table 3 also shows that nearly 80% (874/1096) of those respondents answered that they are not frequent users. We do, however, know little about the patterns of use in relation to contact with health care and other health-related events; it is a topic to be explored further. In addition, Table 3 reveals that respondents with certain conditions, such as cancer or a psychiatric condition, belong to the more frequent users. The highest rated reasons for using *Journalen* among respondents (getting an overview, follow up on visits, and becoming more involved in the care process) correspond with results from earlier interviews with patients in Sweden [22], as well as studies from other countries [11,30,31]. It is clear that searching for errors in the record is not the main reason to use the service and that most patients who answered the survey have not been worried of something they have read in *Journalen*; this is also in line with earlier findings [22]. However, it is in contrast to the results from earlier interviews with physicians [11,19,28], who expressed concerns that patients will be worried when they read their records, or that they will start looking for errors and, hence, believed that PAEHR is not useful for patients. The fact that most respondents want to have information in *Journalen* within a day (thus potentially unsigned information) is also interesting to relate to the concern of physicians that patients will have access to results before the professionals do. Only around 26% (655/2512) of respondents have felt worried because of something they have read in *Journalen* and, thus, the results indicate that early access to, for instance, test results may not be a big issue in this respect. Further analysis of the relationship between immediate access and worry is, however, needed.

Most county councils where respondents have received care are represented by, at least, 250 respondents and only one of these councils is represented by <100 respondents (Norrbotten county council, where only 98 respondents had received care). This is important, as it shows that there are adequate data that make it meaningful to study each county council separately. Moreover, respondents of this survey seem to be quite mobile in their care seeking, and a total of 1674 positive responses were given to the question whether health care had been received outside respondents' home county. The result of this mobility can, for instance, be that a patient can see results from lab tests performed in one council but not another. This mobility highlights the importance of having one national PAEHR service giving patients access to information from many different health care providers in one aggregated view. It does, however, also highlight the need for a more unified national regulatory framework, ensuring a streamlined information provision across health care providers in all county councils [17].

Limitations

There are some limitations to this study. First, the link to the survey was available on *Journalen* 's log-in page. Thus, only persons who have logged in or, at least, visited the log-in page to *Journalen* during the time that the survey was open are among respondents; this creates a positive bias, as previous users who no longer use *Journalen* for whatever reason have not answered the survey. This could, in part, explain the overwhelmingly positive overall attitudes toward the system. Another explanation for the overall positive attitudes could be that patients really are positive. There has been a lot of negative coverage on *Journalen* in Swedish media. Health care professionals, especially physicians, were critical to the service. Patients who are aware of this debate might have been reluctant to criticize the service out of fear of losing access altogether and rather express their support to ensure that the service continues. Either way, the results can only be interpreted as strong support of the service from respondents.

The results regarding the frequency of use show that most users log in a few times a month; this may be because most of the respondents do not have regular contact with health care. It is important to keep in mind that this question does not capture irregularities of use—users of *Journalen* probably log in more frequently when they are treated for an illness. In addition, it is important to consider that 1030 of 2441 respondents stated that they work or have worked within health care. It has been shown

in earlier studies that, for example, physicians who had used *Journalen* themselves were more positive toward the service than those who had not [20,28]. The answers given by current and former health care professionals and all other respondents, respectively, have, however, been separated in Multimedia Appendix 2 and were similar for most questions used in the paper. Additionally, the Mann-Whitney tests did, for most items, not show any support for significant differences between these groups. The exception being the items in question 5 about why it is important to be able to access patient information.

Like in most surveys, respondents form a small sample of all possible users. A lot more users than those who answered the survey logged into *Journalen* during the 5 months the survey was open. Technically, 2587 respondents should be considered a good sample size; nevertheless, we do not know whether, for example, the demographic distribution is representative. Respondents have a higher education level than the general population. Among our respondents, 61% had higher education, whereas 42% of the general population does. Whether this is because users of *Journalen* are well-educated or whether this is a subgroup of users who are more inclined to answer a survey we cannot tell. Therefore, further studies of users of *Journalen* are needed.

Conclusions and Future Work

Up until now, no study has investigated the long-term effects of PAEHRs in Sweden, and few follow-up studies on PAEHRs have been conducted in other European countries. In this paper, the much-needed follow-up work begins with a focus on patients and their usage preferences and attitudes toward the system. The results show that the majority of respondents are women, which is in line with the overall statistics on *Journalen*. Use frequency varies among patients with different conditions—the results indicate that patients with chronic conditions are among the most frequent users. In addition, results show that there is an overall strong positive attitude toward *Journalen* and no statistical differences in attitude ratings could be found between county councils. Thus, the difference in the information shown in *Journalen* between county councils does not have a strong effect in this respect. The highest rated reasons for using *Journalen* are to acquire an overview of one's health status and follow up on visits (eg, memory aid). Furthermore, results show that respondents view test results as the most important type of information, underlining the need to implement support for that information on a national level.

Further research is needed as there are many aspects of the patient survey which are outside the scope of this overview paper, like privacy and security, means of sharing information, usability, etc. In addition, within the recently started research project patient-centered assessment of patients' online access to EHRs, the current implementation and use of PAEHR in Sweden will be evaluated through in-depth qualitative case studies in different regions to achieve a better understanding of how roles, relationships, and organizational structures are affected on micro, meso, and macro levels [16]. Furthermore, it is of importance to study the long-term effects that *Journalen* has had on the work environment for health care professionals, as this is also an area that is highly under investigated, especially in Europe.

Acknowledgments

This research was partly funded through the patient-centered assessment of patients' online access to electronic health records (PACESS) project (2016-00623) supported by FORTE—the Swedish Research Council for Health, Working Life and Welfare. The funding agency FORTE has not played any part in the study. PACESS is hosted by the DOME consortium which studies the Development of Online Medical Records and eHealth Services. IH's work was partially funded by the Academy of Finland grant for the project "Taking health information behaviour into account." We would also like to thank Inera AB (www.inera.se) for providing the data on the usage of the national e-services and managing the survey and data collection through *Journalen*.

Authors' Contributions

JM and HR are the guarantors of this work. JM is responsible for the study and led the analysis and the writing process. HR led the design of the study and contributed to analysis and writing. ÅC, CG, IH, IS, and RMÅ worked with the study design, took part in discussions about analysis and data interpretations, and contributed to all sections in the study. MH and GM took part in discussions about analysis and data interpretation and contributed to all sections in the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The questionnaire including information about the study.
[PDF File (Adobe PDF File), 114 KB - jmir_v20i11e278_app1.pdf]

Multimedia Appendix 2
Detailed results from survey questions 3, 4, 5, and 17.
[PDF File (Adobe PDF File), 136 KB - jmir_v20i11e278_app2.pdf]

References

1. Walker J, Meltsner M, Delbanco T. US experience with doctors and patients sharing clinical notes. BMJ 2015 Feb 10;350:g7785. [Medline: 25670182]
2. 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 2;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]
3. Bell SK, Delbanco T, Walker J. NEJM Catalyst. 2017 Oct 12. OpenNotes: How the Power of Knowing Can Change Health Care URL: https://catalyst.nejm.org/opennotes-knowing-change-health-care/ [accessed 2018-09-19] [WebCite Cache ID 72XrEhkl0]
4. Klein DM, Fix GM, Hogan TP, Simon SR, Nazi KM, Turvey CL. Use of the Blue Button Online Tool for Sharing Health Information: Qualitative Interviews With Patients and Providers. J Med Internet Res 2015 Aug 18;17(8):e199 [FREE Full text] [doi: 10.2196/jmir.4595] [Medline: 26286139]
5. 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 [FREE Full text] [doi: 10.1136/amiajnl-2014-003144] [Medline: 25352570]
6. Pearce C, Bainbridge M. A personally controlled electronic health record for Australia. J Am Med Inform Assoc 2014;21(4):707-713 [FREE Full text] [doi: 10.1136/amiajnl-2013-002068] [Medline: 24650635]
7. Kaipio J, Lääveri T, Hyppönen H, Vainiomäki S, Reponen J, Kushniruk A, et al. Usability problems do not heal by themselves: National survey on physicians' experiences with EHRs in Finland. Int J Med Inform 2017 Jan;97:266-281 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.10.010] [Medline: 27919385]
8. Wiljer D, Urowitz S, Apatu E, DeLenardo C, Eysenbach G, Harth T, et al. Patient accessible electronic health records: exploring recommendations for successful implementation strategies. J Med Internet Res 2008;10(4):e34 [FREE Full text] [doi: 10.2196/jmir.1061] [Medline: 18974036]
9. Nørgaard JR. E-record - access to all Danish public health records. Stud Health Technol Inform 2013;192:1121. [Medline: 23920895]
10. Tiik M, Ross P. Patient opportunities in the Estonian Electronic Health Record System. Stud Health Technol Inform 2010;156:171-177. [Medline: 20543352]
11. Mold F, de Lusignan S. Patients' Online Access to Their Primary Care Electronic Health Records and Linked Online Services: Implications for Research and Practice. J Pers Med 2015 Dec 04;5(4):452-469 [FREE Full text] [doi: 10.3390/jpm5040452] [Medline: 26690225]
12. Erlingsdóttir G, Lindholm C. When patient empowerment encounters professional autonomy: The conflict and negotiation process of inscribing an eHealth service. Scandinavian Journal of Public Administration 2015;19(2):27-48 [FREE Full text]
13. Lyttkens L. Support users to access information and services - final pilot evaluation Uppsala county council. 2015. URL: http://www.sustainsproject.eu/sustainsproject/attachment/d63v12.pdf [accessed 2016-02-22] [WebCite Cache ID 6fUJue03s]
14. Tiik M. Rules and access rights of the Estonian integrated e-Health system. Stud Health Technol Inform 2010;156:245-256. [Medline: 20543358]
15. Mandl KD, Szolovits P, Kohane IS. Public standards and patients' control: how to keep electronic medical records accessible but private. BMJ 2001 Feb 03;322(7281):283-287 [FREE Full text] [Medline: 11157533]
16. Hägglund M, Scandurra I. Patients' Online Access to Electronic Health Records: Current Status and Experiences from the Implementation in Sweden. Stud Health Technol Inform 2017;245:723-727 [FREE Full text] [Medline: 29295193]
17. Scandurra I, Lyttkens L, Eklund B. Implications of Swedish National Regulatory Framework of the Patient Accessible Electronic Health Record. Stud Health Technol Inform 2016;228:695-699 [FREE Full text] [Medline: 27577474]
18. Hägglund M, Moll J, Åhlfeldt RM, Scandurra I. Timing It Right - Patients' Online Access to Their Record Notes in Sweden. Stud Health Technol Inform 2018;247:336-340 [FREE Full text] [Medline: 29677978]
19. Grünloh C, Cajander Å, Myreteg G. "The Record is Our Work Tool!"-Physicians' Framing of a Patient Portal in Sweden. J Med Internet Res 2016 Jun 27;18(6):e167 [FREE Full text] [doi: 10.2196/jmir.5705] [Medline: 27349531]
20. Scandurra I, Jansson A, Forsberg-Fransson ML, Ålander T. Patient Accessible EHR is Controversial: Lack of knowledge and diverse perceptions among care professions. International Journal of Reliable and Quality E-Healthcare 2017;6(1):29-45. [doi: 10.4018/IJRQEH.2017010103]
21. Scandurra I, Jansson A, Forsberg-Fransson ML, Ålander T. Is 'Patient's Online Access to Health Records' a Good Reform? — Opinions from Swedish Healthcare Professionals Differ. Procedia Computer Science 2015;64:964-968 [FREE Full text] [doi: 10.1016/j.procs.2015.08.614]
22. Rexhepi H, Åhlfeldt RM, Cajander Å, Huvila I. Cancer patients' attitudes and experiences of online access to their electronic medical records: A qualitative study. Health Informatics J 2018 Jun;24(2):115-124 [FREE Full text] [doi: 10.1177/1460458216658778] [Medline: 27440056]
23. de Lusignan S, Mold F, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic interpretative review. BMJ Open 2014;4(9):e006021 [FREE Full text] [doi: 10.1136/bmjopen-2014-006021] [Medline: 25200561]

24. Mold F, de Lusignan S, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic review in primary care. Br J Gen Pract 2015 Mar;65(632):e141-e151 [FREE Full text] [doi: 10.3399/bjgp15X683941] [Medline: 25733435]
25. Jilka SR, Callahan R, Sevdalis N, Mayer EK, Darzi A. "Nothing About Me Without Me": An Interpretative Review of Patient Accessible Electronic Health Records. J Med Internet Res 2015 Jun 29;17(6):e161 [FREE Full text] [doi: 10.2196/jmir.4446] [Medline: 26123476]
26. Huvila I, Cajander Å, Daniels M, Åhlfeldt RM. Patients' perceptions of their medical records from different subject positions. J Assn Inf Sci Tec 2015 Apr 02;66(12):2456-2470. [doi: 10.1002/asi.23343]
27. Huvila I, Daniels M, Cajander Å, Åhlfeldt RM. Patients Reading their Medical Records: Differences in Experiences and Attitudes between Regular and Inexperienced Readers. Information Research 2016 Mar;21(1):n1.
28. Ålander T, Scandurra I. Experiences of Healthcare Professionals to the Introduction in Sweden of a Public eHealth Service: Patients' Online Access to their Electronic Health Records. Stud Health Technol Inform 2015;216:153-157 [FREE Full text] [Medline: 26262029]
29. Inera AB. Statistik för Journalen [Statistics on Journalen]. 2018. URL: https://www.inera.se/aktuellt/statistik/journalen_statistik/ [accessed 2018-07-03] [WebCite Cache ID 70dapQwNg]
30. Esch T, Mejilla R, Anselmo M, Podtschaske B, Delbanco T, Walker J. Engaging patients through open notes: an evaluation using mixed methods. BMJ Open 2016 Jan 29;6(1):e010034 [FREE Full text] [doi: 10.1136/bmjopen-2015-010034] [Medline: 26826154]
31. Shah SGS, Fitton R, Hannan A, Fisher B, Young T, Barnett J. Accessing personal medical records online: a means to what ends? Int J Med Inform 2015 Feb;84(2):111-118 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.10.005] [Medline: 25453275]

Abbreviations

eHealth: electronic health
EHR: electronic health record
ID: identification
PAEHR: patient-accessible electronic health record

Edited by G Eysenbach; submitted 29.11.17; peer-reviewed by B Fisher, H Yu, J Lalor, S Shah, G Smith, R Pankomera; comments to author 15.03.18; revised version received 08.06.18; accepted 19.07.18; published 01.11.18.

Please cite as:
Moll J, Rexhepi H, Cajander Å, Grünloh C, Huvila I, Hägglund M, Myreteg G, Scandurra I, Åhlfeldt RM
Patients' Experiences of Accessing Their Electronic Health Records: National Patient Survey in Sweden
J Med Internet Res 2018;20(11):e278
URL: https://www.jmir.org/2018/11/e278/
doi:10.2196/jmir.9492
PMID:30389647

Original Paper

Insights Into Older Adult Patient Concerns Around the Caregiver Proxy Portal Use: Qualitative Interview Study

Celine Latulipe[1], PhD; Sara A Quandt[2], PhD; Kathryn Altizer Melius[2], MS; Alain Bertoni[3], MD, MPH; David P Miller Jr[4], MD, MS; Douglas Smith[5], MBA; Thomas A Arcury[6], PhD

[1]Department of Software and Information Systems, University of North Carolina at Charlotte, Charlotte, NC, United States

[2]Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, United States

[3]Department of Epidemiology and Prevention, Division of Public Health Sciences, Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, United States

[4]Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, United States

[5]Greene County Health Care, Inc, Snow Hill, NC, United States

[6]Department of Family and Community Medicine, Wake Forest School of Medicine, Winston-Salem, NC, United States

Corresponding Author:
Celine Latulipe, PhD
Department of Software and Information Systems
University of North Carolina at Charlotte
9201 University City Avenue
Charlotte, NC, 28233
United States
Phone: 1 7046878195
Email: clatulip@uncc.edu

Abstract

Background: Electronic patient portals have become common and offer many potential benefits for patients' self-management of health care. These benefits could be especially important for older adult patients dealing with significant chronic illness, many of whom have caregivers, such as a spouse, adult child, or other family member or friend, who help with health care management. Patient portals commonly contain large amounts of personal information, including diagnoses, health histories, medications, specialist appointments, lab results, and billing and insurance information. Some health care systems provide proxy accounts for caregivers to access a portal on behalf of a patient. It is not well known how much and in what way caregivers are using patient portals on behalf of patients and whether patients see any information disclosure risks associated with such access.

Objective: The objective of this study was to examine how older adult patients perceive the benefits and risks of proxy patient portal access by their caregivers.

Methods: We conducted semistructured interviews with 10 older adult patients with chronic illness. We asked them about their relationship with their caregivers, their use of their patient portal, their caregiver's use of the portal, and their perceptions about the benefits and risks of their caregiver's use of the portals. We also asked them about their comfort level with caregivers having access to information about a hypothetical diagnosis of a stigmatized condition. Two investigators conducted a thematic analysis of the qualitative data.

Results: All patients identified caregivers. Some had given caregivers access to their portals, in all cases by sharing log-in credentials, rather than by setting up an official proxy account. Patients generally saw benefits in their caregivers having access to the information and functions provided by the portal. Patients generally reported that they would be uncomfortable with caregivers learning of stigmatized conditions and also with caregivers (except spouses) accessing financial billing information.

Conclusions: Patients share their electronic patient portal credentials with caregivers to receive the benefits of those caregivers having access to important medical information but are unaware of all the information those caregivers can access. Better portal design could alleviate these unwanted information disclosures.

(J Med Internet Res 2018;20(11):e10524) doi:10.2196/10524

KEYWORDS

caregivers; patient portals; proxy; proxy portal access; proxy portal accounts

Introduction

Background

The introduction of electronic patient portals over the past decade has the potential to offer many benefits to patients, such as faster and more direct access to health information and test results and the ability to easily renew prescriptions, make appointments, and communicate with health care providers. However, these benefits may not accrue uniformly across patient populations, and research has shown that patient portal adoption is lower among older adults, who typically have more chronic health problems and increased needs for health services [1].

Many older adults have limited ability and motivation to use electronic patient portals [2]. However, their caregivers (such as adult children) may be able to leverage the benefits of such portals on their behalf. This usage of patient portal accounts by a caregiver is referred to as *proxy portal use*. Some portal systems allow patients to provide a caregiver with access to their portal by setting up a proxy account. While health care systems that offer proxy access may encourage patients to set up proxy accounts for their caregivers, patients may simply share their portal credentials (username and password) with their caregivers.

The objective of our research was to gain insight into how older adult patients with chronic illness think about caregiver access to information available on patient portals. We have presented a qualitative interview study with lower-income, older adult patients in an urban area in the United States. We interviewed patients about their caregiver relationship, their own use of their patient portal, their caregiver's use of their patient portal, and their comfort levels in sharing all the information and functionality that a standard patient portal provides. We discuss our findings and provide a set of preliminary design guidelines for patient portal designers who wish to support this proxy portal use.

Barriers to Adoption of Patient Portals

Electronic patient portals can be viewed as an information and communication technology that supports "aging in place" because it allows older adults to access tools and information to manage their chronic illness from their own homes. Connelly et al have found that the evaluation of information and communication technologies for aging in place requires multiple methods because of the wide variety of nuanced contexts in which older adults live [3].

Lyles et al showed that lack of technical support and the fear of losing the doctor-patient relationship were barriers to the adoption of patient portals for African American and Latino patients [4,5]. Similarly, Ancker et al found that the odds of being given an access code for a patient portal were higher for patients who were young, English speakers, white, insured, and female [6]. Patients with lower socioeconomic status were less likely to make use of patient portals [7].

Health literacy may be another significant factor impacting the adoption of patient portals, as it is correlated with higher perceived ease of use and usefulness of health information technologies [8]. Functional impairment may negatively impact the ability of older adults to use patient portals [9]. Given that functional impairment tends to be lower with younger populations [9] and that health literacy tends to be higher among younger populations [10], younger caregivers may be better equipped to use patient portals than some older patients.

Caregivers and Information Sharing

Caregivers (often family members) play an important role in assisting older adults living with chronic medical conditions. According to a 2013 Pew Research report, 36% of US adults acted as caregiver to an adult relative or friend [11]. Some recent research suggests expanding the concept of "personal health informatics" to "family health informatics" [12]. Pang et al have noted the need to design for patient privacy while sharing information with relatives [13]. The needs of informal caregivers, such as their need for social support, recognition as caregivers, and communication with other stakeholders, are complex [14-18]. Bosch and Kanis noted the importance of making caregivers better-informed through access to systems such as patient portals [19].

Recent work looking into chronic illness patients with spousal caregivers demonstrates that it is important to support situations in which the values between patient and caregiver are in tune, as well as situations in which these values conflict [20]. While that work targeted spousal caregivers, similar patterns of conflicting values may emerge with other types of caregivers such as adult children, siblings or close friends, and neighbors. Designing to handle conflicting values around sharing and privacy between patient and caregiver becomes a critical component when looking at proxy access to patient portals.

Proxy Portal Accounts

Most electronic patient portal systems have some mechanism for the provision of proxy accounts. This is common for parents of young children, where the parents get automatic proxy access to their children's portals. However, only some patient portal systems allow proxy accounts to be established for adult patients. Health care systems that do not offer proxy accounts for adult patients may recommend that adults wishing to share their patient portal information share their log-in credentials with their caregiver. Sarkar and Bates argued that current health care systems do not adequately engage caregivers and noted, "…although the Office of the National Coordinator acknowledges the importance of caregivers and family, broadly adopted standards for caregiver access to patient portals are not available" [21].

Caine et al have investigated patient attitudes toward the sharing of their electronic health records within medical systems [22]. They found that patients were unaware of how much information was in their electronic health records and wanted much more fine-grained control over who could access this information. In addition, they wanted to know when their information was accessed and by whom. While their study did not specifically address issues of patient portals and caregiver access, the issues they discussed are relevant to the patient portal context.

In a survey of patients across all age groups, Wolff et al found that patients share portal access with caregivers not only for information sharing purposes and emergency reasons but also

because they need technical help [23]. In a focus group study of adults aged >75 years, participants raised concerns about autonomy and control, and the authors reported that it would be difficult to create a single model of access control for proxy accounts that would be suitable for all patient-caregiver dyads [24]. In addition, Wolff et al conducted a scan of 20 large health systems in the United States to understand how many offer proxy portal accounts for adult caregivers and noted that while many do offer such accounts, only some health systems advertise the availability of proxy accounts on their websites [25].

Health Information Security

How health information is shared within systems has been a subject of study, particularly when patient charts include information about stigmatized conditions such as mental illness, substance abuse, or sexual health [26]. In addition, Caine and Hanania and Schwartz et al showed that patients want granular control over which health care stakeholders see what parts of their medical data [27,28]. We are unaware of any studies specifically examining how patients deal with such information when sharing patient portal access with a caregiver, who may have a proxy account or who may have the patients' portal password.

There are 2 interrelated issues with password sharing. Individuals frequently share their passwords with friends, coworkers, and relatives, allowing them to access a system account using the account holder's credentials [29,30]. In addition, individuals often engage in cross-system password sharing, in which they use the same password (or a similar password) across many systems [31,32]. The sharing of passwords with others creates a significant information disclosure risk if the individual owning the credentials has also used the same password across multiple systems, such as a patient portal, Web-based banking, and social networking.

One approach to improving privacy in the sharing of health information is the "break-glass" access control protocol [33-35]. This idea is based on the metaphor of the need to break the glass around a fire alarm. This can be applied to health care information because many people want to keep their health information private, but in the case of a life-threatening medical emergency would want caregivers to have access to that health information. When this approach is applied to health information systems, any emergency that satisfies the "break the glass" criteria not only allows specified people to access the information but also creates an audit trail so that patients and providers can see that the information has been shared, when it was shared, and with whom. To the best of our knowledge, this approach has not been used for managing access to patient portals for patient caregivers.

Methods

Research Project

This research fits within a larger longitudinal research project investigating older adult patients and caregiver usage of electronic patient portals [2,36]. In that project, a set of patients and caregivers have been interviewed periodically over a period of 2 years to understand the patient portal adoption and usage. The patient portal at the academic medical center that is the home for our research team does offer proxy accounts for caregivers of adult patients. Creating this account requires completing and submitting paper forms.

For this particular study into patient perceptions around caregiver access, we began by first doing a walkthrough of the patient portal system. We analyzed the functionality available in this patient portal and compared that with the functionality in other portal systems to generate a set of common features. We then developed an interview study to more deeply understand how older adults perceive caregiver's proxy portal access within the context of this standard feature set.

Participants

The recruitment objective was to engage patients who were racially diverse and representative of the population of older, low-income adults in the area surrounding the academic medical center. Participants were selected from among individuals who had participated in the larger longitudinal survey interview study of patient portal utilization [2,36]. Low-income older adults were the focus of the larger study because this population has relatively low use of patient portals and digital technology, in general, and a goal was to determine leverage points for improving the patient portal utilization in this segment of the population.

For the larger study, 120 participants were recruited from patients receiving care from a clinic in the academic medical center that serves predominantly Medicaid and noninsured patients. Participants were not asked to meet low-income criteria as individuals. These participants were community-dwelling adults aged ≥55 years, who were being treated for a chronic disease (diabetes, hypertension, dyslipidemia, or cardiovascular disease), spoke English or Spanish, and were in sufficiently good health to give informed consent and complete a lengthy fixed response survey interview. Recruitment for the larger study included a 3-step process: (1) clinic staff and physicians generated a list of patients who met the inclusion criteria; (2) we sent recruitment letters to a randomly generated list of these patients; and (3) we made follow-up phone calls to describe the study and schedule interviews with those who were sent the letters. The overall study had a refusal rate of 49.2%.

Of 70 academic medical center patients who had a caregiver when they completed the survey interview component and who remained in sufficiently good health to continue participation, we recruited 10 to complete the semistructured interviews (age range, 60-71 years; 5 male, 5 females; 7 African American, and 3 white individuals; see Table 1). Participants were contacted individually until 10 agreed to complete the interview.

Table 1. The summary of participants, portal use, caregiver relations, caregiver portal use, and the presence of stigmatized health condition.

Participant identifier	Sex	Race	Uses portal	Caregiver relation	Caregiver uses portal	2nd caregiver relation	2nd caregiver uses portal	Stigmatized health condition[a]
P1	Female	African American	Yes	Child	Yes	N/A[b]	N/A	N/A
P2	Male	African American	Yes	Friend	No	N/A	N/A	HIV
P3	Female	African American	Yes	Children	Yes	Home health assistant	Yes	MH[c]
P4	Male	White	No	Spouse	Yes	N/A	N/A	N/A
P5	Male	African American	Yes	Brother	No	Home health assistant	No	N/A
P6	Male	African American	Yes	Spouse	No	N/A	N/A	MH
P7	Male	White	Yes	Sister	Yes	Neighbor	No	N/A
P8	Female	White	Yes	Spouse	No	N/A	N/A	MH
P9	Female	African American	No	Niece	No	N/A	N/A	MH
P10	Female	African American	No	Child	No	N/A	N/A	N/A

[a]All patients suffered from at least one chronic condition such as high blood pressure, arthritis, or heart disease.

[b]N/A: not applicable.

[c]MH: mental health.

Interview Structure

Data collection was completed between June and August 2017 by 2 trained interviewers. Interviewers met participants in their homes. The interviews typically lasted between 30 and 60 minutes and were audiorecorded. The interviewers explained the study and obtained signed consent. Participants received a US $20 incentive for completing the interview. The interview study was approved by the academic medical center's Institutional Review Board.

The semistructured interview covered patient's background, caregiver relationship, patient's portal use, and caregiver's use of the portal (see Multimedia Appendix 1). We defined caregiver as "someone with whom you share your health information, and who helps you with your health care." We asked all patients about how secure they considered their health information on the portal, and whether they considered it more or less secure than social networking sites and Web-based banking. The interview then specifically investigated the patients' comfort level with caregiver access to the portal in the hypothetical situation of a stigmatized illness, as well as the comfort level with the caregiver accessing information such as past medical records and billing or insurance data. This part of the interview included laminated screenshots of the patient portal with notional data (such as screens showing provider messaging, prescription renewal, test results, visit summaries, appointment scheduling, and billing), which were used as visual prompts to remind participants of the various features being discussed and the kinds of information available on various screens of the portal. We specifically asked participants about whether they would be concerned with their caregivers seeing financial billing information on the portal. We asked this to ascertain whether participants were aware of this information being on the portal and how they felt about sharing such information.

Data Analysis

Transcripts from the audiorecorded interviews were completed by a professional service and then edited for accuracy. Data analysis was conducted throughout the process, with the team performing ongoing reflection on interview transcripts as they became available. As the researchers reviewed the transcript narratives, they met periodically to discuss the themes, patterns, and issues they found in the data [37]. The research team created an initial coding dictionary based on these discussions. The codes included in the dictionary reflect *a priori* themes reflected in the interview guide and themes that arose from the review of the data. Two of the interviews were coded and reviewed by 4 researchers who then refined the coding dictionary. All interview transcripts were then coded in a 2-step process. Each transcript was coded by one research team member and then checked for coding accuracy by a second team member. All of the investigators discussed the interpretation of the themes and patterns, and agreed on the final presentation.

Results

Principal Results

In total, 7 patients had used the patient portal provided by the academic medical center, and 4 of them had allowed one or more of their caregivers to access their patient portal (see Table 1). Given the small sample size, these numbers are not meant to suggest typical levels of portal usage by patients or caregivers but suggest that there is some proportion of older adult patients

that use the portal and some portion that do allow their caregivers to access their portals on their behalf.

Caregiver Relations and Tasks

All participants identified at least one caregiver. Most participants identified a close relative, such as a spouse, sibling, or adult child, as their primary caregiver. Caregivers helped with both health-related and household tasks such as cleaning and cooking. The health care–related tasks that caregivers engaged in included getting patients to appointments, reminding them of medications, helping with diet and exercise, and communicating with doctors.

Eight participants had given their caregiver Health Insurance Portability and Accountability Act of 1996 (HIPAA) authorization by signing an authorization document. Two participants were unfamiliar with HIPAA and were unsure whether or not they had granted anyone such authorization. Three participants identified multiple caregivers, and in these cases, there was a split of responsibilities with a local friend or paid assistant helping with day-to-day care, and a family member acting as a caregiver from a distance.

Patient Portal Use

Overall, 7 of 10 participants used their patient portal, though their perceptions of it and usage levels varied. P3 reported using her portal 4 times per week and mentioned using the portal for appointment tracking. P1 reported using the portal but explained that she did not like it at all, mainly because her health care providers had not responded to messages through the portal. Other participants had logged into the portal occasionally, and 2 of the participants had forgotten their portal passwords and had not been able to log in recently. Participants reported using the portal to monitor appointment schedules, message with doctors, and look at test results. None of the participants reported using the portals to look at billing or make payments.

Security Concerns

All participants remarked that they expected the information to be seen by health care staff only and that information on the portal was kept more private than information on Facebook. In comparing the portal to Web-based banking, a number of participants felt they could not comment because they did not do Web-based banking, while others felt the portal was equally secure or more secure than Web-based banking. Most participants thought it was very unlikely that someone could hack into their portal and steal their health information. Participant responses demonstrated that they do not consider their medical information to be of high value or interest and are therefore not very concerned about information security issues when using the portal.

Caregiver Portal Use

Of 10 participants, 4 had given their passwords to one or more caregivers so that the caregiver could access the portal. P1 talked about how she had set her portal password to something easy to remember and shared it with her daughter, who mostly checks it for appointments. In asking about who has access to her portal, P3 noted that multiple children and her home health care assistant all have access, but only use it on an as-needed basis. P5 has 2 caregivers—a daily home health care assistant and a brother with whom he shares his health care issues. His brother, who has HIPAA authorization and power of attorney, can log into his patient portal but has not done so to his knowledge. P5 has not given his home health care assistant access to his patient portal, despite noting that they are “good friends.” In addition, P7 has 2 caregivers, a local neighbor and a sister who lives in another state. His sister has access to the portal and used it to follow appointments as P7 went through cancer treatment. However, P7 has not given his neighbor caregiver access. The neighbor takes him to the hospital in medical emergencies and helps him when he is not well, and P7 says if it became important or useful, he would consider giving her access to the portal.

Thematic Analysis

A number of themes emerged around comfort level with caregiver portal access. Most participants, even those who were not on the portal or who had not granted caregivers access to the portal, saw benefit in their caregiver being able to access the portal. Although 2 of 10 patients interviewed (P6 and P8) used their portal themselves, they had caregivers who were not internet users and would be unable to use the patient portal.

Health Literacy Assistance

One of the main reasons participants felt caregiver access was beneficial was in the caregiver’s ability to help them understand the information on the portal, such as doctor’s messages and test results. For example, P3 described how her caregivers help her understand the information that is there.

> *I don’t understand all this, all these abbreviations. So she [health assistant] opened it up, she and my daughter-in-law, and they were reading and letting me know that my red blood cell count was very low and my white blood cells were my hemoglobin was out of whack, my TSH levels were out of whack.*

Similarly, P8 mentioned a close friend who works in the medical field and could act in the role of caregiver (though does not currently). She noted how helpful it could be to have this friend access her portal, saying “she may understand some of the stuff better than I do.”

Caregivers as Communication Gatekeepers

In multiple instances, caregivers helped to keep the rest of the family informed or helped to explain difficult medical situations to family members. For example, P3 who gives her daughter-in-law portal access, noted the following:

> *My son is one of those people that he can’t take what’s happening, so I explain it to her [daughter-in-law]. She’s in the medical field as well, and she’s a CNA [Certified Nursing Assistant] at [clinic] and she and her best friend [name] is a registered nurse, they get together and explain to him what’s going on with me.*

This quote demonstrates both the benefit of caregivers with medical knowledge having portal access and those caregivers using that information to communicate the situation to other

family members. P7 described the benefits of his sister having portal access.

> *She, sort of keeps the other family members informed, so it serves a purpose there, too.*

Stigmatized Health Issues

The portal in use at the academic medical center is full featured and includes full medical records and past diagnoses. When asked about the security of information on the portal, P1 who is a regular portal user and was familiar with the amount of information available noted the following:

> *Now, probably if I were HIV+ and was trying to get a job, I might be a little more sensitive about things like that. But I don't know. I'm not real worried about it.*

This comment demonstrated that the concerns about the privacy of information on the patient portal may be moderated by the presence or absence of stigmatized health issues (both from the perspective of the portal being hacked and the perspective of the portal being accessed by caregivers). P2 had an HIV+ diagnosis, and he expressed reticence about sharing his patient portal with his caregiver friend. This participant noted that he had not shared his HIV+ status with his caregiver, although he admitted that he would be okay with her finding out about his status through the portal if there was an emergency and she needed access to help him.

P5, when asked about caregivers accessing information about a hypothetical HIV diagnosis, responded as follows:

> *You don't want everybody to know, but then the ones that are close to you that are actually gonna have to be the ones there for you, you would have to let them know.*

This quote highlights that there is a tension between needing care help and feeling embarrassed about such a stigmatized condition.

P7 raised a number of concerns about caregiver portal access. One concern was that his neighbor caregiver, who does not drink alcohol, would see that he drinks alcohol if she had access to the portal. This demonstrates that having caregivers who are slightly more distant in relationship can cause tensions in considering sharing access to a patient portal because lifestyle information is often captured and recorded as part of routine health checkups. In addition, P7 did not like the idea of his caregivers learning about a hypothetical mental health diagnosis. He also expressed a concern about a hypothetical diagnosis of HIV.

> *I don't think I'd want them to know. No, I don't think so. I think that would be considered sort of private, … I mean I hope that never happens, but I wouldn't feel real comfortable.*

Emergency Access

Several patients discussed the benefits of caregivers being able to access the portal in case of an emergency. While patients who had granted portal access to their caregivers saw benefits of this access in cases of emergency, this was also true for participants who had *not* granted portal access to their caregivers. This was clear in the case of P2, the participant who had an HIV+ status, but had not shared that status with his caregiver. He admitted that he would be okay with his friend caregiver accessing his portal in case of emergency, but his comment expresses a high level of reticence, even about emergency access: "…if it had to be, so be it. If it came down to it." The finality of this comment suggests that giving her access would definitely be a last resort, only if he really needed help from her. P7 has a sister living out of town who serves as a distant caregiver and a neighbor who serves as a close-by caregiver. This was the participant mentioned previously who felt that he would not want his neighbor to have access to the portal because she might be able to see that he drinks alcohol. But he expressed that in an emergency, he could see the benefits of her being able to access the portal and communicate with his out-of-town family.

Billing Privacy

Most participants, even those who were regular users of the portal, were unaware that billing information was available on the portal. Some participants who had close family members as caregivers were not concerned with those caregivers seeing billing information. P4 explained her openness in this area:

> *That would be fine too because, if something happens to us, we're older, they're gonna be responsible for that.*

P9 noted that she would not want her niece caregiver to see the balance in her bank account, but she would be okay with her niece seeing the medical bills on the patient portal.

Some participants had definite concerns about their caregivers accessing their portal and seeing billing information, and this tended to vary with the relationship between the patient and the caregiver. For example, P5 responded that he would be comfortable with his brother seeing billing information but not his home health care assistant. P8, who has a husband caregiver who does not use the portal at all, spoke about possibly giving her best friend portal access and described how she would trust her friend with health information, making appointments, and renewing prescriptions, but, "Well, I'd rather keep the billing stuff private… I've just always been, my finances are my business."

P7 was not comfortable with anyone having access to his billing information. He did not realize that there was billing information on the portal, and in thinking about his sister and neighbor caregivers, noted "I would rather they not see that…an invasion of my private life, I guess." These comments show that feelings about caregivers seeing medical billing information are quite varied, with some participants feeling quite uncomfortable about such information being disclosed to caregivers through the portal.

Discussion

Principal Findings

Participants in this study typically shared their current health information with their caregivers, whether through the portal

or other means of communication. The only exception to this rule was when there was a stigmatized condition, such as in the example of the participant who is HIV+ and had not shared that status with his caregiver, but would be okay if she found out the status if she needed to, in case of emergency. This situation is a perfect example of where the "break-glass" paradigm of access control [33] could be applied to good effect. That 8 of 10 participants had given their caregivers formal HIPAA authorization reflects their willingness to share current health information. The other 2 participants indicated they were not familiar with HIPAA and were unsure if they had granted anyone authorization. With HIPAA authorization, caregivers can receive participant health information directly from health care providers.

We found that older adult patients see benefits in having their caregivers access the patient portals, though this was more likely when the caregiver was a family member. In this study, 4 of the 10 participants gave caregivers access by sharing their log-in credentials. The fact that none of these participants set up official proxy accounts for their caregivers suggests that they either did not know those accounts were available or they considered the process to get those accounts too burdensome. Regardless of the reason for password sharing, the practice is concerning, especially given research that shows people use the same or similar passwords across multiple systems [30,31]. Older patients who share their portal password with a caregiver may also be inadvertently giving that caregiver access to their bank account or email. However, this practice is an easy way for older patients to share their information with their caregivers and is likely to continue regardless of how easy proxy account setup becomes.

We observed that some patients have concerns about sharing information about stigmatized medical issues (such as mental health conditions or infections that could have been transmitted through sexual activity), though it tends to be hard for patients to consider these hypotheticals. The hypotheticals that caused the most privacy concerns were related to mental illness and sexually transmitted infections.

Furthermore, we noted significant concern from some of our participants regarding the billing information that is available in the portal being accessible to caregivers. While people tend to be quite private about their personal finances, this issue did not pose a concern when the caregiver was a spouse.

Design Considerations

While patient portal systems allow for the setting up of proxy accounts, those accounts provide proxy users with access to everything that the patient can access. The following considerations are based on the idea that existing proxy accounts could be modified to allow the benefits of caregiver access, while addressing the common privacy concerns noted by our interview participants. These considerations should be taken as starting points for further research and discussion, given that our findings are based on a small sample of 10 patient interviews.

- Promote the use of proxy accounts by allowing easy setting up of proxies online, using simple, clear language.
- Provide a simple checklist of access controls, with screenshots, to help patients decide what information or functionality to grant the caregiver.
- Provide a default proxy account configuration that includes access to most information and functions, but requires an opt-in for the complete medical record, billing, and insurance information.

These guidelines will only be useful if patients actually set up official proxy accounts for their caregivers. Some patients will likely continue to share their portal credentials with caregivers, and some health care systems do not offer proxy accounts for caregivers, which means password sharing is the only way for caregivers to access a patient's portal. Hence, we offer the following design guidelines for systems to help mitigate issues when passwords are shared with caregivers:

- Remind users when creating or changing passwords on the portal that they should choose unique passwords that are different from passwords on other important systems such as email and Web-based banking.
- Provide a "break-glass" mechanism that allows patients to specify who can be given access to the portal in case of emergency. Then, ensure that the system logs that access and provides clear alerts and log-in history on the portal so that the patient is made aware when someone has used the emergency "break-glass" mechanism.
- Ask users to identify themselves when logging into a portal. For example, after logging into a portal, the system could prompt the user to define themselves as either the "patient" or a "caregiver." If the user chooses caregiver, the system could ask their name and relationship. This could then be added to all of the logs and communications inside the portal so that the patient and caregiver would be separately identified. This could help the patient be able to see when their caregiver logged in and monitor for abuse of the system. Moreover, this would help health care providers know who they are communicating with through the portal.

One issue with the last guideline of having an identification step after log-in is that it could actually be seen as condoning or encouraging the sharing of portal credentials. One way to ensure that patients and caregivers do not view this as condoning the sharing of passwords is to respond when a caregiver self-identifies by asking them to talk to their patient about setting up proxy access and by sending a message to the patient that encourages them to set up a proxy account for that caregiver. In this way, the practice is allowed, but the system also nudges users toward a more secure mode of interaction with the system. This only makes sense if the patient portal system provides adult proxy accounts, and it is currently unclear how many portal systems in the United States actually provide this functionality.

Limitations

This qualitative study has a small number of participants, which is a limitation to the generalizability of the results. Further studies with larger populations are needed to understand the prevalence of these caregiver portal access issues. Similarly, the elderly, low-income nature of our population limits the generalizability of our results. Younger and wealthier patients may have different concerns about caregiver portal access.

Conclusions

We have presented the results of a qualitative study with 10 older, low-income adults who receive outpatient primary care through a university medical center in a small city in the United States. We have investigated how these older patients share health information with their caregivers, and how these patients feel about sharing electronic patient portal access with caregivers. While 2 previous studies have investigated patient attitudes about proxy portal use, ours is the first study to frame these attitudes around the full set of standard portal features and consider both the security and privacy concerns that may come into play. Our results suggest that patients typically share their log-in credentials with caregivers rather than setting up official proxy accounts. Regardless of the access mechanism, this proxy portal access provides no granularity of control over the information shared through the portal, and patients express some discomfort with the sharing of data around stigmatized illnesses and financial obligations. We suggest some guidelines to improve both official proxy portal accounts and standard portal accounts, to allow all stakeholders to reap the benefits of caregiver proxy portal use, without incurring inadvertent information disclosure risks or other security breaches.

Acknowledgments

This research was supported by grant R01 HS021679 from the Agency for Healthcare Research and Quality.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Interview guide.

[PDF File (Adobe PDF File), 53KB - jmir_v20i11e10524_app1.pdf]

References

1. 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]
2. Latulipe C, Gatto A, Nguyen HT, Miller DP, Quandt SA, Bertoni AG, et al. Design Considerations for Patient Portal Adoption by Low-Income, Older Adults. In: Proc SIGCHI Conf Hum Factor Comput Syst. New York, US: ACM; 2015 Apr Presented at: 33rd annual ACM Conference on Human Factors in Computing Systems; 2015 Apr 18; Seoul p. 3859-3868 URL: http://europepmc.org/abstract/MED/27077140 [doi: 10.1145/2702123.2702392]
3. Connelly K, ur Rehman Laghari K, Mokhtari M, Falk TH. Approaches to understanding the impact of technologies for aging in place: a mini-review. Gerontology 2014;60(3):282-288 [FREE Full text] [doi: 10.1159/000355644] [Medline: 24457288]
4. Goel MS, Brown TL, Williams A, Cooper AJ, Hasnain-Wynia R, Baker DW. Patient reported barriers to enrolling in a patient portal. Journal of the American Medical Informatics Association 2011;18(Supplement_1):8-12.
5. Lyles C, Allen J, Poole D, Tieu L, Kanter M, Garrido T. "I Want to Keep the Personal Relationship With My Doctor": Understanding Barriers to Portal Use among African Americans and Latinos. Journal of Medical Internet Research 2016 Dec 03;18(10):e263 [FREE Full text] [doi: 10.2196/jmir.5910] [Medline: 27697748]
6. 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]
7. Perzynski AT, Roach MJ, Shick S, Callahan B, Gunzler D, Cebul R, et al. Patient portals and broadband internet inequality. J Am Med Inform Assoc 2017 Sep 01;24(5):927-932. [doi: 10.1093/jamia/ocx020] [Medline: 28371853]
8. Mackert M, Mabry-Flynn A, Champlin S, Donovan E, Pounders K. Health Literacy and Health Information Technology Adoption: The Potential for a New Digital Divide. J Med Internet Res 2016 Dec 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
9. Greysen SR, Chin GC, Sudore RL, Cenzer IS, Covinsky KE. Functional impairment and Internet use among older adults: implications for meaningful use of patient portals. JAMA Intern Med 2014 Jul;174(7):1188-1190 [FREE Full text] [doi: 10.1001/jamainternmed.2014.1864] [Medline: 24839165]
10. Kutner M, Greenburg E, Jin Y, Paulsen C. The Health Literacy of America's Adults: Results from the 2003 National Assessment of Adult Literacy. National Center for Education Statistics 2006;NCES:2006-2483.
11. Fox S, Duggan M, Purcell K. Pew Research Center's Internet & American Life Project. Washington, DC: Pew Research Center; 2013 Jun 20. Family caregivers are wired for health URL: http://www.pewinternet.org/2013/06/20/family-caregivers-are-wired-for-health/ [WebCite Cache ID 72zZNc4dC]

12. Pina LR, Sien SW, Ward T, Yip JC, Munson SA, Fogarty J, et al. From Personal Informatics to Family Informatics: Understanding Family Practices around Health Monitoring. New York, NY, USA: ACM; 2017 Presented at: 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing; February 25 - March 1, 2017; Portland p. 2300-2315. [doi: 10.1145/2998181.2998362]
13. Pang CE, Neustaedter C, Riecke BE, Oduor E, Hillman S. Technology preferences and routines for sharing health information during the treatment of a chronic illness. In: CHI '13 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.: ACM; 2013 Presented at: ACM SIGCHI Conference on Human Factors in Computing Systems; April 27 - May 1, 2013; Paris, France p. 1759-1768.
14. Chen Y, Ngo V, Park SY. Caring for caregivers: designing for integrality. In: CSCW '13 Proceedings of the 2013 conference on Computer supported cooperative work. New York, NY, USA: ACM; 2013 Presented at: ACM 2013 Conference on Computer Supported Cooperative Work; February 23 - 27, 2013; San Antonio p. 91-102.
15. Hensely-Schinkinger S, de Carvalho AFP, Glanznig M, Tellioglu H. The definition and use of personas in the design of technologies for informal caregivers. 2015 Presented at: International Conference on Human-Computer Interaction; 2015 Aug 2; Los Angeles p. 202-213.
16. Schorch M, Wan L, Randall DW, Wulf V. Designing for Those who are Overlooked: Insider Perspectives on Care PracticesCooperative Work of Elderly Informal Caregivers. In: CSCW '16 Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. New York, NY, USA: ACM; 2016 Presented at: 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing; 2016 Feb 27; San Francisco p. 787-799.
17. Tixier M, Lewkowicz M. "Counting on the group": reconciling online and offline social support among older informal caregivers. New York, NY, USA: ACM; 2016 Presented at: 2016 ACM Conference on Human Factors in Computing Systems; 2016 May 7; San Jose p. 3545-3558.
18. Long K, Bakewell LL, McNaney RC, Vasileiou K, Atkinson M, Barreto M, et al. Connecting Those That Care: Designing for Transitioning, Talking, Belonging and Escaping. New York, NY, USA: ACM; 2017 Presented at: 2017 ACM Conference on Human Factors in Computing Systems; 2017 May 2; Denver p. 1339-1351.
19. Bosch LB, Kanis M. Design opportunities for supporting informal caregivers. New York, NY, USA: ACM; 2016 Presented at: ACM CHI 2016 Conference Extended Abstracts on Human Factors in Computing Systems; 2016 May 7; San Jose p. 2790-2797.
20. Berry AB, Lim C, Hartzler AL, Hirsch T, Wagner EH, Ludman E, et al. How values shape collaboration between patients with multiple chronic conditions and spousal caregivers. New York, NY, USA: ACM; 2017 Presented at: 2017 ACM Conference on Human Factors in Computing Systems; 2017 May 2; Denver p. 5257-5270.
21. Sarkar U, Bates DW. Care partners and online patient portals. JAMA 2014;311(4):357-358. [doi: 10.1001/jama.2013.285825] [Medline: 24394945]
22. Caine K, Kohn S, Lawrence C, Hanania R, Meslin EM, Tierney WM. Designing a patient-centered user interface for access decisions about EHR data: implications from patient interviews. J Gen Intern Med 2015 Jan;30 Suppl 1:S7-16 [FREE Full text] [doi: 10.1007/s11606-014-3049-9] [Medline: 25480719]
23. Wolff JL, Berger A, Clarke D, Green JA, Stametz R, Yule C, et al. Patients, care partners, and shared access to the patient portal: online practices at an integrated health system. J Am Med Inform Assoc 2016 Dec;23(6):1150-1158. [doi: 10.1093/jamia/ocw025] [Medline: 27026614]
24. Crotty BH, Walker J, Dierks M, Lipsitz L, O'Brien J, Fischer S, et al. Information Sharing Preferences of Older Patients and Their Families. JAMA Intern Med 2015 Sep;175(9):1492-1497. [doi: 10.1001/jamainternmed.2015.2903] [Medline: 26147401]
25. Wolff J, Kim V, Mintz S, Stametz R, Griffin J. An environmental scan of shared access to patient portals. J Am Med Inform Assoc 2018 Apr 01;25(4):408-412. [doi: 10.1093/jamia/ocx088] [Medline: 29025034]
26. Shenoy A, Appel JM. Safeguarding Confidentiality in Electronic Health Records. Camb Q Healthc Ethics 2017 Apr;26(2):337-341. [doi: 10.1017/S0963180116000931] [Medline: 28361730]
27. Schwartz P, Caine K, Alpert S, Meslin E, Carroll A, Tierney W. Patient preferences in controlling access to their electronic health records: a prospective cohort study in primary care. J Gen Intern Med 2015 Jan;30 Suppl 1:S25-S30 [FREE Full text] [doi: 10.1007/s11606-014-3054-z] [Medline: 25480721]
28. Caine K, Hanania R. Patients want granular privacy control over health information in electronic medical records. J Am Med Inform Assoc 2013 Jan 01;20(1):7-15 [FREE Full text] [doi: 10.1136/amiajnl-2012-001023] [Medline: 23184192]
29. Kaye J. Self-reported password sharing strategies. : ACM; 2011 Presented at: ACM 2011 Conference on Human Factors in Computing Systems; 2011 May 7; Vancouver p. 2619-2622.
30. Singh S, Cabraal A, Demosthenous C, Astbrink G, Furlong M. Password sharing: implications for security design based on social practice. New York, NY, USA: ACM; 2007 Presented at: ACM Conference on Human Factors in Computing Systems; 2007 Apr 29; San Jose p. 895-904.
31. Das A, Bonneau J, Caesar M, Borisov N, Wang X. The Tangled Web of Password Reuse. NDSS 2014 Feb 27;14:23-26.
32. Florencio D, Herley C. A large-scale study of web password habits. New York, NY, USA: ACM; 2007 Presented at: 16th International Conference on World Wide Web; 2007 May 8; Banff p. 657-666.

33. Marinovic S, Craven R, Ma J, Dulay N. Rumpole: a flexible break-glass access control model. : ACM; 2011 Presented at: 16th ACM Symposium on Access Control Models and Technologies; 2011 Jun 15; Innsbruck p. 73-82.
34. Brucker A, Petritsch H. Extending access control models with break-glass. New York, NY, USA: ACM; 2009 Presented at: 14th ACM Symposium on Access Control Models and Technologies; 2009 Jun 3; Stresa p. 197-206.
35. Kotz D, Fu K, Gunter C, Rubin A. Security for mobile cloud frontiers in healthcare. Communications of the ACM 2015 Jul 23;58(8):21-23.
36. Arcury T, Quandt S, Sandberg J, Miller JD, Latulipe C, Leng X, et al. Patient Portal Utilization Among Ethnically Diverse Low Income Older Adults: Observational Study. JMIR Med Inform 2017 Nov 14;5(4):e47 [FREE Full text] [doi: 10.2196/medinform.8026] [Medline: 29138129]
37. Luborsky M. The identification and analysis of themes and patterns. Qualitative methods in aging research 1994;168:189-210.

Abbreviations

HIPAA: Health Insurance Portability and Accountability Act of 1996

Edited by G Eysenbach; submitted 30.03.18; peer-reviewed by G Strudwick, M Honary, D Klein; comments to author 10.05.18; revised version received 04.07.18; accepted 17.07.18; published 02.11.18.

Please cite as:
Latulipe C, Quandt SA, Melius KA, Bertoni A, Miller Jr DP, Smith D, Arcury TA
Insights Into Older Adult Patient Concerns Around the Caregiver Proxy Portal Use: Qualitative Interview Study
J Med Internet Res 2018;20(11):e10524
URL: https://www.jmir.org/2018/11/e10524/
doi:10.2196/10524
PMID:30389654

Original Paper

Dual Use of a Patient Portal and Clinical Video Telehealth by Veterans with Mental Health Diagnoses: Retrospective, Cross-Sectional Analysis

Erica A Abel[1,2,3], PhD; Stephanie L Shimada[4,5,6], PhD; Karen Wang[1,2,7], MD; Christine Ramsey[1,2], PhD; Melissa Skanderson[1], MSW; Joseph Erdos[1,2,3], MD, PhD; Linda Godleski[3,8], MD; Thomas K Houston[4,6], MD, MPH; Cynthia A Brandt[1,2,9], MD, MPH

[1]Pain Research, Informatics, Multimorbidities and Education Center, VA Connecticut Healthcare System, West Haven, CT, United States

[2]Yale Center for Medical Informatics, Yale School of Medicine, New Haven, CT, United States

[3]Department of Psychiatry, Yale School of Medicine, New Haven, CT, United States

[4]Center for Healthcare Organization and Implementation Research, Edith Nourse Rogers Memorial Veterans Hospital, Bedford, MA, United States

[5]Department of Health Law, Policy, and Management, Boston University School of Public Health, Boston, MA, United States

[6]Division of Health Informatics and Implementation Science, Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, United States

[7]Department of Internal Medicine, Yale School of Medicine, New Haven, CT, United States

[8]National Telemental Health Center, VA Connecticut Healthcare System, West Haven, CT, United States

[9]Department of Emergency Medicine, Yale School of Medicine, New Haven, CT, United States

Corresponding Author:
Erica A Abel, PhD
Pain Research, Informatics, Multimorbidities and Education Center
VA Connecticut Healthcare System
950 Campbell Avenue, 11-ACSLG
West Haven, CT,
United States
Phone: 1 203 932 5711 ext 5323
Fax: 1 203 479 8126
Email: erica.abel@va.gov

Abstract

Background: Access to mental health care is challenging. The Veterans Health Administration (VHA) has been addressing these challenges through technological innovations including the implementation of Clinical Video Telehealth, two-way interactive and synchronous videoconferencing between a provider and a patient, and an electronic patient portal and personal health record, My HealtheVet.

Objective: This study aimed to describe early adoption and use of My HealtheVet and Clinical Video Telehealth among VHA users with mental health diagnoses.

Methods: We conducted a retrospective, cross-sectional analysis of early My HealtheVet adoption and Clinical Video Telehealth engagement among veterans with one or more mental health diagnoses who were VHA users from 2007 to 2012. We categorized veterans into four electronic health (eHealth) technology use groups: My HealtheVet only, Clinical Video Telehealth only, dual users who used both, and nonusers of either. We examined demographic characteristics and mental health diagnoses by group. We explored My HealtheVet feature use among My HealtheVet adopters. We then explored predictors of My HealtheVet adoption, Clinical Video Telehealth engagement, and dual use using multivariate logistic regression.

Results: Among 2.17 million veterans with one or more mental health diagnoses, 1.51% (32,723/2,171,325) were dual users, and 71.72% (1,557,218/2,171,325) were nonusers of both My HealtheVet and Clinical Video Telehealth. African American and Latino patients were significantly less likely to engage in Clinical Video Telehealth or use My HealtheVet compared with white patients. Low-income patients who met the criteria for free care were significantly less likely to be My HealtheVet or dual users than those who did not. The odds of Clinical Video Telehealth engagement and dual use decreased with increasing age. Women were more likely than men to be My HealtheVet or dual users but less likely than men to be Clinical Video Telehealth users.

Patients with schizophrenia or schizoaffective disorder were significantly less likely to be My HealtheVet or dual users than those with other mental health diagnoses (odds ratio, OR 0.50, CI 0.47-0.53 and OR 0.75, CI 0.69-0.80, respectively). Dual users were younger (53.08 years, SD 13.7, vs 60.11 years, SD 15.83), more likely to be white, and less likely to be low-income than the overall cohort. Although rural patients had 17% lower odds of My HealtheVet adoption compared with urban patients (OR 0.83, 95% CI 0.80-0.87), they were substantially more likely than their urban counterparts to engage in Clinical Video Telehealth and dual use (OR 2.45, 95% CI 1.95-3.09 for Clinical Video Telehealth and OR 2.11, 95% CI 1.81-2.47 for dual use).

Conclusions: During this study (2007-2012), use of these technologies was low, leaving much potential for growth. There were sociodemographic disparities in access to My HealtheVet and Clinical Video Telehealth and in dual use of these technologies. There was also variation based on types of mental health diagnosis. More research is needed to ensure that these and other patient-facing eHealth technologies are accessible and effectively used by all vulnerable patients.

(J Med Internet Res 2018;20(11):e11350) doi:10.2196/11350

KEYWORDS

mental health; patient portals; telemedicine; telehealth; eHealth; United States Department of Veterans Affairs

Introduction

Background

Veteran access to mental health care within the US Department of Veterans Affairs Health Administration (VHA), especially for rural veterans, has been a challenge [1]. About 2.9 million, 56% of all rural veterans, are enrolled in VHA care [2]. As compared with urban veterans, rural veterans with mental health diagnoses are sicker and face challenges accessing health care, including stigma, increased distance needed to travel to care, lack of access to transportation, and lack of specialists or providers in rural areas [3,4]. Although the VHA offers mental health services nationally through a combination of regional medical centers and community-based outpatient clinics, some locations may not have the specialty services or the staff to meet the demand for needed services [5].

To address access barriers and to enhance patient care, the VHA has been transforming the provision of clinical care, in part, through technological innovation. The VHA has been a leader nationally through its deployment of several health information technologies [6], including an integrated electronic medical record, Home Telehealth (eg, in-home, messaging, and peripheral devices such as blood pressure and heart rate monitors), mobile health apps, Clinical Video Telehealth (CVT), and an integrated Web-based personal health record and patient portal, My HealtheVet (MHV). These patient-facing technologies are consistent with recommendations by the Institute of Medicine to support continuous healing relationships through use of the internet and technologies that provide patients with access to care outside of face-to-face visits, and access to their medical information, when and where they need it most [7]. They also offer unique ways for patients and providers to communicate in addition to or in lieu of traditional face-to-face encounters [8]. The concept of *complementary use* of health care technology —the use of features and functionality of two different technologies in conjunction with one another—is a way of enhancing access to care and supporting patient-centered care by supporting communication, information sharing, and increasing patient involvement in their care [8]. The integration of technologies in a complementary way has the potential to increase veterans' access to care and improve the quality of delivered care. Dual use of technologies—veteran adoption and use of more than one technology at any point—may be a precursor to *complementary use*. For example, veterans who have adopted the MHV portal and have engaged in CVT might be willing or inclined to use both tools in a complementary way.

History of My HealtheVet and Clinical Video Telehealth

MHV and CVT are two established virtual care technologies in the VHA. Over a decade ago, the VHA launched its Web-based personal health record and patient portal, My HealtheVet (MHV), to complement traditional health care services, to improve comanaged care, and to promote active engagement of patients and their families in the patient's health care [9]. The MHV portal allows users to create and maintain a comprehensive personal health record by using a range of MHV features, including secure messaging, Web-based prescription refills, access to information in their VHA health record (eg, laboratory results, clinical progress notes, discharge summaries, and medication lists), and tracking of personal and self-reported health information using a variety of tools (eg, food, activity and allergy journals, family health history, and other data). Access to features depends on the type of MHV enrollment and account type. MHV registration creates a basic account that provides access to the MHV self-report features (ie, self-entered information or journals). An advanced account is limited to veterans and/or VHA patients and gives these users the ability to refill prescriptions and to view some of their information in their health records. A premium account, also only for veterans or VHA patients, is the highest level of MHV access and requires users to verify their identity either in-person or on the Web, a process known as authentication. In addition to using MHV to view many parts of their VHA health record and Department of Defense Military Service Information, premium users can send secure messages to communicate with their health care providers and health care teams [10]. As of July 2018, of the 4.45 million who registered (since November 2004), about 3.9 million indicated they were veterans. Of the 4.45 million registrants, about 2.74 million have authenticated since January 1, 2007 [11].

The VHA has also expanded access to health care through a range of telehealth services, including CVT which is two-way interactive and synchronous videoconferencing between a

provider and a patient at a distance in settings such as a VHA medical center and a community clinic or home. CVT is available across a range of medical and mental health specialties. Unlike MHV, a Web app, access to CVT is less ubiquitous for two primary reasons: (1) clinicians refer patients and place consults to CVT programs (patients may not be able to self-refer) and (2) CVT programs that do exist are not available to every veteran, everywhere. Some programs are regional—they emanate from a regional medical center or hub and provide remote specialty services to clinics that are closer to a veteran's home, thus reducing travel time and time away from work or family. A few programs provide services across state lines, although to specific sites (the provider and veteran must be located in a federal facility), and some programs provide CVT into the home.

Devices used in CVT within VHA are often attached to computers (eg, video cameras and microphones) or are stand-alone (eg, videoconferencing equipment). For visits conducted at a VHA medical center or community-based clinic, a VHA employee at the patient site (known as a telehealth clinical technician [TCT]) seats the patient in the CVT-equipped space and coordinates initiation of the CVT call with the provider. The TCT then leaves the room and is available as needed to provide technical or administrative support throughout the clinical encounter. If the patient is at home, the process is different as a TCT is not involved. The provider and the patient connect with each other using preplanned processes. Whether CVT is home- or clinic-based, the providers have well-defined, documented, and tested emergency procedures available.

There is evidence that veterans with mental health conditions are interested in, and have used, both MHV and CVT. Among veterans who reported enrollment and mental health use data in the 2010 National Survey of Veterans, 25% of those who indicated they used VHA mental health services endorsed that they used MHV to obtain information about their personal VHA health care [12]. In a study of veterans receiving care in VHA, those with trauma-related conditions and common mental health conditions (eg, depression, bipolar disorder, or posttraumatic stress disorder [PTSD]) were among the highest early adopters of MHV. The adoption of MHV among veterans with depression, anxiety, and PTSD was high compared with patients with other diagnoses [13].

In addition, VHA CVT services for individuals with mental health conditions (known also as telemental health) have grown tremendously. In October 2002, VHA began coding telehealth activity distinctly to enable its measurement. Between October 2002 and August 2018, there were more than 3.75 million CVT telemental health patient encounters. In 2017, VHA delivered more than 470,000 CVT telemental health patient encounters to over 150,000 unique patients [14]. During this study period (2007-2012), there were over 320,000 unique CVT users.

Recent research has explored whether veterans with mental health diagnoses are willing to use different electronic health (eHealth) technologies, including CVT. A survey of veterans in Hawaii (VHA and non-VHA users) found that 32% to 57% of those surveyed were receptive to using different technologies in their mental health care (ie, telephone calls, CVT into the home or clinic, Web-based computer-based interventions, personalized messages to computer, short message service (SMS) text messages to cell phone, or social networking with a peer group). Veterans' willingness varied and depended on the technology and their PTSD screen status [15]. Veterans with probable PTSD were significantly less likely than those with no PTSD to report willingness for CVT in clinic (20.4% vs 45.6%) or CVT in the home (25.5% vs 52.7%). The survey did not include questions about use of specific MHV features.

Research also has examined veterans' preferences on their use of technologies in managing their health care [16]. Whealin et al conducted a survey of a random sample of veterans who received VHA care and who had registered to use MHV [16]. Among those with a PTSD diagnosis and at least one chronic medical condition, 44.6% used health-related technology 1 to 3 times per month and 21.4% used it less than once per month. Most common uses of the technologies included searching for health information (78.9%), communicating with providers (71.1%), and tracking medications (64.9%). Respondents reported they were most experienced and comfortable with using computers, the Web, email (99%-100%) and had less experience and comfort with other modes, including social media (73.0%), mobile apps (79.6%), and Clinical Video Telehealth (67.3%).

Despite these early studies, there has been little work examining the extent to which veterans might be using and benefitting from the use of multiple eHealth technologies. This information is increasingly important as the VHA strives to improve access for all veterans, especially those with mental health conditions. Information on whether and how veterans use these two technologies provides information that could be leveraged to identify additional opportunities for mental health–related treatments and interventions, improve access to and patient engagement in mental health treatment, and improve the overall quality of mental health care. To our knowledge, research has not yet examined veterans' use of multiple eHealth technologies on a large scale using administrative data. This study examines (1) the early adoption and feature use of MHV, engagement in CVT, and the dual use of these two technologies and (2) the sociodemographic and mental health characteristic associated with the use of these technologies. The use of two technologies may lay the groundwork for and lead to the complementary use of technologies.

Methods

Study Design

We conducted a cross-sectional analysis of veterans' use of two VHA technologies: (1) the MHV personal health record and patient portal and (2) Clinical Video Telehealth.

Study Population

The sample for this study (N=2,171,325) is drawn from a retrospective cohort study evaluating technology adoption in VHA users [13]. This sample includes all veterans aged between 18 and 100 years, who received inpatient or outpatient care at VHA during our study period of October 1, 2007, through March 31, 2012, and had one or more common or high-priority mental health diagnoses by the time of MHV registration or first CVT

visit (see variable: mental health conditions). We chose to study this early period (ending March 2012), which coincided with the initial rollout of secure messaging to providers and patients. The study period overlaps with both the MHV pilot period (October 2007 and October 2009) and the 2010 to 2012 early national rollout of MHV and secure messaging. The Human Research Protection Program at the Veterans Affairs (VA) Connecticut Healthcare System and the Yale School of Medicine and Institutional Review Board at the Edith Nourse Rogers Memorial Veterans Hospital in Bedford, Massachusetts, approved this study.

Data Sources

Administrative, clinical, CVT, and MHV data for October 1, 2007, through March 31, 2012, the study period, were pulled from the VHA Corporate Data Warehouse. Variables extracted include patient demographics, medical and mental health diagnoses, MHV enrollment and authentication statuses, use of certain MHV features (ie, secure messaging and prescription refill), and CVT engagement.

Variables

Dependent Variables: Electronic Health Technology Use Groups—Dual Use, My HealtheVet Adoption, Clinical Video Telehealth Engagement

We created indicators for MHV adoption, CVT engagement, and dual use as dependent variables in our multivariable logistic regression models. To establish MHV adoption, we used MHV data on registration, authentication, and feature use. The available data included flags for adoption such as MHV *registration* (ie, the process of creating a personal profile, log-in, and access account to gain access to MHV), *authentication* (ie, the verification of identity before granting access to personal health information), and MHV *feature use*, including secure messaging use (ie, ever sent or ever read a secure message) and Web-based prescription refill (ie, ever refilled prescriptions on the Web). To establish CVT engagement, we identified whether the patient had ever had a CVT visit during the study period. A patient was determined to be a dual user if he or she had adopted MHV and engaged in CVT during the study period; the MHV adoption and CVT engagement did not have to be concurrent.

For our other analyses, we grouped veterans into four mutually exclusive eHealth technology use groups based on their MHV adoption and their CVT engagement (*CVT-MHV groups*): (1) dual users—veterans who had adopted MHV and had engaged in CVT at any point during the study period; (2) MHV only—veterans who had adopted MHV and did not have a CVT visit; (3) CVT only—veterans who had a CVT visit for mental health during the study period but had no MHV adoption; and (4) neither—veterans who had neither adopted MHV nor engaged in CVT during the study period.

Independent Variables: Mental Health Conditions

We focused on common or high priority mental health conditions in the veteran population, including bipolar disorder, major depression, other depression (ie, depressive disorders not meeting criteria for major depressive disorder such as adjustment disorder, depression not otherwise specified), anxiety, PTSD, schizophrenia or schizoaffective disorder, and other psychotic disorders. Veterans in the full study cohort may have multiple mental health diagnoses. We used previously validated diagnostic code groupings [17] and ascertained from the administrative data if veterans had one or more of the mental health diagnoses documented during the study period. We counted mental health conditions coded at least once for an inpatient stay or at least twice for an outpatient visit during the study period. Prior research has demonstrated that this approach improves the accuracy of the identification of disorders in administrative data [18,19] because outpatient codes are assigned by health care providers and may be less accurate than inpatient codes, which are assigned by professional coders in the VHA. Diagnoses were classified according to *International Classification of Diseases, Ninth Revision, Clinical Modification.*

Covariates: Demographic Characteristics

Demographic variables included gender, age, race or ethnicity, rural residence, and economic need. Rural residence was determined based on zip code of residence using VHA Office of Rural Health definitions based on the Rural-Urban Commuting Areas system (ie, urban, rural, and highly rural) [20]. As a proxy for socioeconomic status to capture economic need, we created a flag to indicate patients who qualified for free VHA health care based on a financial assessment.

Analyses

We used descriptive statistics to examine sociodemographic and mental health characteristics for the full cohort and within each of the four eHealth technology groups. We compared MHV feature use between the MHV only and dual users groups using chi-square statistics. In addition, among MHV adopters (MHV only and dual users), we examined the proportion of veterans in the full cohort and within each of the diagnostic code groupings who used any features of MHV (any MHV) and who used two specific features of MHV: prescription refills and secure messaging.

We used separate multivariable logistic regression models to examine predictors of MHV adoption versus nonadoption, CVT engagement versus nonengagement, and dual use versus nondual use. We examined associations between mental health diagnoses and eHealth technology use, adjusting for patient sociodemographic characteristics previously shown to be significantly associated with MHV use [13], and accounted for clustering of veterans within VHA health care regions (ie, Veterans Integrated Service Networks [VISNs]) by including VISN as a random effect in the models. SAS 9.4 was used to run all analyses (SAS institute, Cary, NC).

Results

Patient Demographics and Electronic Health Technology Adoption

From a cohort of over 6 million active users of the VHA, 2,171,325 patients aged between 18 and 100 years had one or more of the mental health conditions. As shown in Table 1, the majority were male, white and resided in urban areas. Mean age was 60.11 (SD 15.83), and over a quarter qualified for free care

in VHA based on economic need. Among these 2.17 million veterans, 1.51% (32,723/2,171,325) were dual users of MHV and CVT. Most patients (1.56 million) had neither adopted or used MHV features or engaged in CVT (71.72%, 1,557,218/2,171,325); 23.00% (499,445/2,171,325) of the patients were MHV only users and 3.77% (81,939/2,171,325) had engaged in CVT only. Compared with the overall cohort, dual users included a larger proportion of women veterans (8.49%, 184,331/2,171,173 vs 12.8%, 4212/32,721), white patients (77.70%, 1,521,828/1,958,625), and patients from rural areas (25.73%, 545,755/2,121,406 vs 44.27%, 14,256/32,199). Compared to the overall cohort, dual users had a lower mean age 53.08 (SD 13.74) years versus 60.11 (SD 15.83) years and a lower percentage of individuals with high economic need (26.15%, 567,728/2,170,948 vs 17.56%, 5746/32,723).

Over a third of the entire cohort had a diagnosis of PTSD (36.47%, 791,839/2,171,325), 31.47% (683,268/2,171,325) had a diagnosis of anxiety disorder, 23.95% (520,088/2,171,325) had a diagnosis of major depression, 17.61% (382,438/2,171,325) had other psychotic disorder diagnoses, and 12.68% (275,331/2,171,325) had a diagnosis of bipolar disorder (Table 2). Nearly two-thirds of individuals in the entire cohort (62.41%, 1,355,039/2,171,325) had other depression diagnoses. CVT engagement and MHV use varied by mental health diagnosis. Compared with veterans with other psychotic or schizophrenia or schizoaffective disorders, those with a diagnosis of depression, PTSD, anxiety, or bipolar disorder were higher users of both MHV and CVT.

The unadjusted comparisons of MHV feature use between MHV only and dual users show some differences (Table 3). Significantly more dual users authenticated or ever filled a prescription on the Web than MHV only users. There were no differences between dual and MHV users in secure messaging use.

As shown in Figure 1, there were variations in use of MHV and its primary features across mental health diagnoses. The percentages shown in the figure reflect the overall proportion of patients with that diagnosis who engage in each type of use. Individuals with major depression, PTSD, or bipolar diagnoses had overall higher levels of any MHV feature use and used the prescription refill feature more often than individuals with other diagnoses. Individuals diagnosed with schizophrenia or schizoaffective disorders and other psychotic disorders had consistently lower levels of any MHV feature use, prescription refill, and secure message use.

Table 1. Demographics by electronic health technology use groups.

Demographics[a]	Full cohort (N=2,171,325)	Dual users (n=32,723)	MHV[b] only (n=499,445)	CVT[c] only (n=81,939)	Neither (n=1,557,218)
Gender (male), n (%)	1,986,842 (91.51)	28,509 (87.13)	435,459 (87.19)	75,733 (92.44)	1,447,141 (92.94)
Residence, n (%)					
Urban	1,575,651 (74.27)	27,954 (55.73)	382,066 (78.13)	40,298 (49.97)	1,135,344 (74.71)
Rural	545,755 (25.73)	14,256 (44.27)	106,920 (21.87)	40,342 (50.03)	384,237 (25.29)
Race or ethnicity, n (%)					
White	1,521,828 (77.70)	26,075 (86.20)	373,057 (81.93)	61,208 (82.13)	1,061,488 (75.90)
African American	364,107 (18.59)	3116 (10.30)	65,704 (14.43)	10,206 (13.69)	285,081 (20.38)
Latino	12,813 (0.65)	114 (0.38)	2048 (0.45)	445 (0.60)	10,206 (0.73)
Other[d]	59,877 (3.06)	943 (3.12)	14,533 (3.19)	2669 (3.58)	41,732 (2.98)
High economic need, n (%)	567,728 (26.15)	5746 (17.56)	98,779 (19.78)	20,545 (25.07)	442,658 (28.43)
Age (years), n (%)					
<40	277,140 (12.76)	6712 (20.51)	81,624 (16.34)	12,632 (15.42)	176,172 (11.31)
40 to 59	653,918 (30.12)	12,770 (39.02)	179,169 (35.87)	27,136 (33.12)	434,843 (27.92)
60 to 79	985,240 (45.38)	12,696 (38.80)	213,008 (42.65)	37,585 (45.87)	721,951 (46.36)
>80	255,027 (11.75)	545 (1.67)	25,644 (5.13)	4586 (5.60)	224,252 (14.40)
Age (years), mean (SD)	60.11(15.83)	53.08 (13.74)	56.18 (14.68)	57.02 (14.70)	61.67 (16.00)

[a]Numbers may not sum because of missing data, and percentages may not sum to 100% because of rounding. The listed column percentages exclude missing data.

[b]MHV: My HealtheVet.

[c]CVT: Clinical Video Telehealth.

[d]Other category includes American Indian, Asian, and Native Hawaiian.

Table 2. Mental health conditions by electronic health technology use groups.

Mental health conditions	Full cohort, n[a] (%)	Dual users, n (%)	MHV[b] only, n (%)	CVT[c] only, n (%)	Neither, n (%)
Other depression	1,355,039 (62.41)	23, 679 (1.75)	329,706 (24.33)	56,414 (4.16)	945,240 (69.76)
Posttraumatic stress disorder	791,839 (36.47)	18,611 (2.35)	206,644 (26.10)	41,036 (5.18)	525,548 (66.37)
Anxiety	683,268 (31.47)	13,209 (1.93)	167,405 (24.50)	30,799 (4.51)	471,855 (69.06)
Major depression	520,088 (23.95)	12,868 (2.47)	141,864 (27.28)	26,937 (5.18)	338,419 (65.07)
Bipolar disorder	275,331 (12.68)	6666 (2.42)	69,764 (25.34)	15,436 (5.61)	183,465 (66.63)
Other psychotic disorders	382,438 (17.61)	5073 (1.33)	72,563 (18.97)	14,690 (3.84)	290,112 (75.86)
Schizophrenia or schizoaffective	124,879 (5.75)	1229 (0.98)	16,506 (13.22)	6261 (5.01)	100,833 (80.78)

[a]Veterans may have multiple mental health diagnoses; therefore, percentages do not sum to 100%.

[b]MHV: My HealtheVet.

[c]CVT: Clinical Video Telehealth.

Table 3. My HealtheVet adoption and feature use with and without dual use (N=2,171,325).

My HealtheVet feature use	Dual users (n=32,723), n (%)[a]	MHV[b] only (n=449,445), n (%)[a]	*P*[c] value
Authenticated	23,573 (72.04)	347,610 (69.60)	<.001
Ever filled a prescription on the Web	20,589 (63.88)	293,409 (59.59)	<.001
Ever used secure messaging[d]	3561 (10.88)	57,126 (11.44)	.002

[a]The total number of Veterans in each group and the percent of the overall study population.

[b]MHV: My HealtheVet.

[c]*P* value for chi-square.

[d]Only authenticated users who opt-in can secure message; however, the denominator for the percentage calculation is based on the total number of dual users or MHV only users to show the overall penetration of secure messaging activity among the entire population in each column.

Figure 1. My HealtheVet use by mental health diagnosis. MHV: My HealtheVet; RX: Prescription.

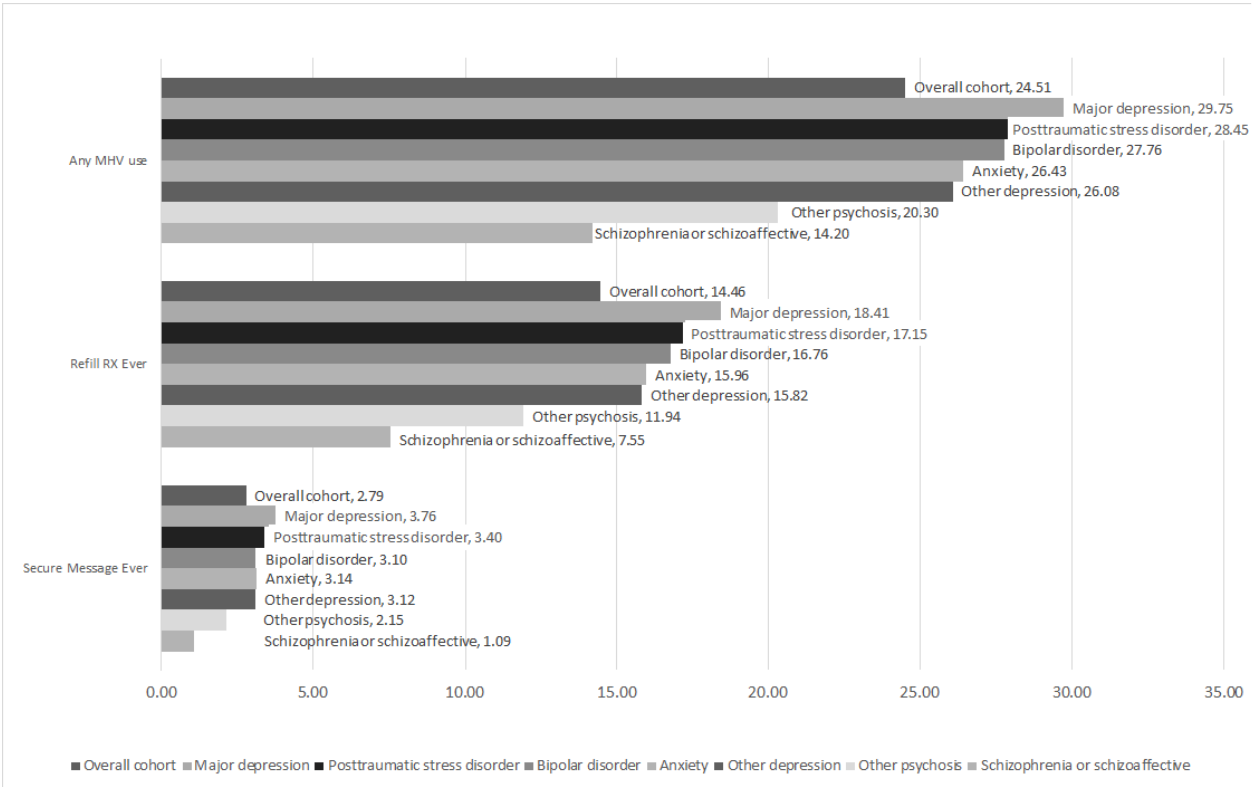


Table 4. Adjusted odds ratios of My HealtheVet, Clinical Video Telehealth, and dual use based on demographic characteristics (N=1,911,085).

Demographic characteristics	Model[a] predicting MHV[b] adoption, OR[c] (95% CI)	Model[a] predicting CVT[d] engagement, OR (95% CI)	Model[a] predicting dual use of both MHV and CVT, OR (95% CI)
Age in years			
<40	Reference	Reference	Reference
40 to 59	1.04 (1.01-1.07)	0.91 (0.87-0.95)	0.91 (0.88-0.94)
60 to 79	0.70 (0.65-0.75)	0.71 (0.65-0.77)	0.56 (0.51-0.61)
>80	0.27 (0.24-0.30)	0.40 (0.31-0.51)	0.17 (0.13-0.21)
Gender			
Female	1.57 (1.51-1.62)	0.92 (0.89-0.96)	1.16 (1.11-1.20)
Male	Reference	Reference	Reference
Race or ethnicity			
African American	0.51 (0.48-0.54)	0.72 (0.62-0.85)	0.51 (0.46-0.57)
Latino	0.53 (0.46-.062)	0.88 (0.79-0.98)	0.58 (0.41-0.82)
Other	0.83 (0.78-0.89)	0.98 (0.90-1.06)	0.82 (0.76-0.89)
White	Reference	Reference	Reference
Residence			
Rural	0.83 (0.80-0.87)	2.45 (1.95-3.09)	2.11 (1.81-2.47)
Urban	Reference	Reference	Reference
Income			
High economic need	0.64 (0.63-0.66)	0.99 (0.96-1.02)	0.75 (0.71-0.79)
Other	Reference	Reference	Reference
Bipolar disorder			
Yes	1.09 (1.07-1.12)	1.43 (1.35-1.51)	1.45 (1.37-1.53)
No	Reference	Reference	Reference
Major depression			
Yes	1.25 (1.22-1.28)	1.38 (1.26-1.52)	1.56 (1.45-1.68)
No	Reference	Reference	Reference
Posttraumatic stress disorder			
Yes	1.19 (1.16-1.22)	1.74 (1.58-1.91)	1.86 (1.77-1.96)
No	Reference	Reference	Reference
Schizophrenia or schizoaffective disorder			
Yes	0.50 (0.47-0.53)	1.25 (1.17-1.33)	0.75 (0.69-0.80)
No	Reference	Reference	Reference
Other psychosis			
Yes	1.04 (1.01-1.07)	1.14 (1.10-1.19)	1.13 (1.08-1.19)
No	Reference	Reference	Reference
Other depression			
Yes	1.20 (1.18-1.22)	1.32 (1.24-1.40)	1.42 (1.34-1.51)
No	Reference	Reference	Reference
Anxiety			
Yes	1.09 (1.07-1.11)	1.25 (1.17-1.33)	1.27 (1.20-1.35)
No	Reference	Reference	Reference

[a]Models accounted for clustering of veterans within VHA health care regions (known as Veterans Integrated Service Networks [VISN]) by including

VISN as a random effect to adjust for VISN-level differences.

[b]MHV: My HealtheVet.

[c]OR: odds ratio.

[d]CVT: Clinical Video Telehealth.

Models Predicting Adoption of My HealtheVet, Clinical Video Telehealth, and Dual Use

Sociodemographic Characteristics Associated With My HealtheVet, Clinical Video Telehealth, and Dual Use

Our logistic regression models showed that holding all other demographic characteristics and diagnoses constant, age is a strong predictor of any type of technology use (Table 4). Veterans with mental health diagnoses who were over the age of 60 years had significantly lower odds of MHV adoption, CVT engagement, or dual use than those under the age of 40 years. For example, the odds of a veteran being a dual user were 44% lower for those aged 60 to 79 years (OR 0.56, 95% CI 0.51-0.61) and 83% lower for those aged over 80 years (OR 0.17, 95% CI 0.13-0.21) compared with those aged under 40 years. The odds of veterans aged 40 to 59 years adopting MHV were 4% higher than the odds for veterans under 40 years (OR 1.04, 95% CI 1.01-1.07); however, those aged 40 to 59 years were less likely to engage in CVT than those aged under 40 years (OR 0.91, 95% CI 0.87-0.95).

There were also differences in technology use for other demographic characteristics. Women, as compared with men, had higher odds of both MHV adoption and dual use (OR 1.57, 95% CI 1.51-1.62 for MHV and OR 1.16, 95% CI 1.11-1.20 for dual use) but slightly lower odds of CVT engagement (OR 0.92, 95% CI 0.89-0.96). African American and Latino veterans had significantly lower odds of MHV adoption, CVT engagement, or dual use as compared with white veterans using the same technologies. The odds of an African American veteran adopting MHV, engaging in CVT, or being a dual user were 49%, 28%, and 49% lower, respectively, whereas those of Latino veteran were 47%, 12%, and 42% lower, respectively. Although being low income did not predict CVT engagement, low-income veterans eligible for free VHA care based on income had odds of MHV adoption that were 36% lower (OR 0.64, 95% CI 0.63-0.66) and odds of dual use that were 25% lower (OR 0.75, 95% CI 0.71-0.79) than patients who were not eligible. Rural patients had 17% lower odds of MHV adoption compared with urban patients (OR 0.83, 95% CI 0.80-0.87) but substantially higher odds of CVT engagement and dual use (OR 2.45, 95% CI 1.95-3.09 for CVT and OR 2.11, 95% CI 1.81-2.47 for dual use).

Diagnoses Associated With My HealtheVet, Clinical Video Telehealth, and Dual Use

There were differences in MHV adoption, CVT engagement, and dual users across mental health diagnoses, holding all sociodemographic variables and comorbid mental health diagnoses constant. Patients diagnosed with major depression were more likely to be a MHV adopter than the veterans with other diagnoses. The odds of a veteran diagnosed with major depression adopting MHV were 25% higher than patients not diagnosed with major depressive disorders (OR 1.25, 95% CI 1.22-1.28). Patients diagnosed with PTSD had substantially higher odds of both CVT engagement (OR 1.74, 95% CI 1.58-1.91) and being a dual user (OR 1.86, 95% CI 1.77-1.96) than the odds for patients not diagnosed with PTSD. Finally, patients diagnosed with schizophrenia or schizoaffective disorder were significantly less likely than veterans with other mental health diagnoses to be a MHV adopter (OR 0.50, 95% CI 0.47-0.53) or dual user (OR 0.74, 95% CI 0.69-0.80).

Discussion

Principal Findings

This study is one of the first examinations of multiple (*dual*) health information technology use among veterans with mental health diagnoses. We explored an early period (2007-2012) of MHV use and CVT engagement among veterans with mental health diagnoses. Our findings suggest that during the study period, overall use or engagement with these health information technology tools was low and that dual use was exceedingly low. Among these veterans diagnosed with mental health conditions, differences in use by patient characteristics were like the differences previously reported in the overall veteran populations, with some exceptions [13]. To place this work into context, we have noted these exceptions below.

The unadjusted results suggest different patterns of early use of CVT and MHV across residential areas, gender, economic need, and race or ethnicity. Patients living in rural areas had higher odds of being CVT only or dual users, whereas those in urban areas were more likely to be MHV only users. Prior work has found higher use of MHV in urban veterans but has not explored dual use in rural veterans. This finding may be because of the emphasis on implementing CVT programs in rural areas and, thus, increased availability where they are more important for providing veterans with access to mental health services. It may also reflect differences in access to technology or the internet at home (used in accessing MHV) with greater availability in urban versus rural settings.

Although women represent less than 10% of veterans receiving VHA health care, the number of veterans who are women continues to grow. Female veterans had higher odds of being MHV adopters or dual users compared with male veterans. This finding is consistent with other literature showing that women VHA users are more likely to adopt patient portals than male VHA users [13]. Male veterans had higher odds of CVT engagement compared with female veterans. Although these findings describe a period of early use, recent research has highlighted the potential benefits of telemental health for women veterans [21] and the potential for it to reach this growing segment of VHA users [22] who have unique mental health needs. In a national CVT program for veterans with bipolar disorder, 19% of the participants were women [23]. Thus, our finding suggests there is an opportunity to increase the CVT

engagement of women veterans through programs that address their specific needs.

Our findings suggest the presence of economic and race or ethnicity disparities in the use of eHealth technologies during the study period (2007-2012). Although the odds of CVT engagement were not associated with patient economic need, high economic need was inversely associated with MHV and dual use. African American and Latino race or ethnicity and high economic need were consistently associated with lower odds of use of MHV adoption or CVT engagement, even after adjusting for rural or urban residence and other demographic characteristics and diagnoses. The VHA experience is similar to that outside the VHA, as Roblin et al reported a gap in adoption of personal health records in the Kaiser Permanente Georgia patient population [24].

There were differences in MHV use and CVT engagement across mental health diagnoses. This study expands beyond prior work [13] that found veterans who had a diagnosis of major depression, PTSD, or bipolar disorder more often used any MHV feature and used MHV to refill prescriptions on the Web than veterans with other mental health diagnoses. Veterans with major depression or PTSD were high users of secure messaging as compared with other veterans with other mental health diagnoses. Patients who had diagnoses of schizophrenia or schizoaffective disorder were consistently less likely than patients with other mental health diagnoses to use any MHV feature, prescription refill, or secure messaging. They were also least likely to be dual users.

There is some evidence that CVT benefits veterans with serious mental illness. An analysis of a national CVT program for veterans with bipolar disorder showed positive effects across several domains, including patient engagement, clinical impact, and quality of care [23]. There is limited data available on the adoption and use of CVT in veterans with schizophrenia or schizoaffective disorder and the benefits of CVT on the outcomes of such patients. Kasckow et al conducted a review of telepsychiatry assessment and treatment in patients with schizophrenia (not just veterans) [25]. They reviewed internet, telephone, and video-based approaches. There were a limited number of studies in each modality, including a handful that explored video-based interventions. The CVT studies reviewed had limitations, and per the authors, the video-based modality showed initial promise with individuals diagnosed with schizophrenia.

Future Research

It is important to uncover and address barriers patients face using eHealth technologies before new technologies are designed and implemented and before clinical interventions are delivered using those technologies. Future research could explore and then address differences in the barriers to use of patient-facing eHealth technologies, especially among veterans with mental health conditions. Such research should include exploring patterns of use of specific features or services across patients with different mental health diagnoses.

Understanding barriers and disparities could inform outreach efforts designed to increase adoption and use. For example, promotion and communication materials could be tailored to address specific barriers and disparities, and outreach efforts could include personalized and novel methods of training and education, including (1) members of a veteran's clinical team teaching, reinforcing, and encouraging use of specific MHV features relevant to the veteran's care such as accessing important illness, self-management, and provider visit content (eg, a medication list, a symptom checklist, or laboratory results from a recent visit); (2) creation of brief, step-by-step instructional videos of key MHV features and dissemination of these videos through the MHV portal and other means, including the VHA's social media and public video-sharing websites; and (3) incorporating videoconferencing and/or screen sharing into customer service to help veterans learn and use MHV.

Future research should also explore more recent patterns of MHV use and CVT engagement, including dual use and complementary use among patients with mental health conditions. Since this study period (2007-2012), there have been major changes to MHV, including the full implementation of secure messaging and increases in patient adoption and use. In addition, the implementation of and engagement in CVT programs has grown. It would also be important to explore whether disparities in dual use and complementary use have changed. Future research should examine whether there are differences in mental health outcomes (eg, reduction in mental health symptoms, improvements in quality of life, and improvements in medication adherence) for different eHealth technology user groups.

Complementary use of eHealth technologies might support patient care, for example, prescribing providers (or their team) who use CVT might use secure messaging to remind patients about follow-up for important laboratory tests related to their treatment. As MHV allows users to track and self-enter a variety of information (eg, activity and food journals, vitals and readings, and allergies), providers and veterans who use manualized evidence-based treatments delivered using CVT might use MHV to support clinical treatment. MHV could host clinical program materials, and providers and veterans could share program materials using MHV secure messaging and could incorporate seamlessly MHV self-management tools (or program specific tools hosted on MHV) into clinical encounters.

Limitations

There are several limitations to this study. Inherent differences in access to the health information technologies examined influenced our findings. MHV is a Web-based patient portal available to all VHA users, although actual access depends on factors such as computer or cell phone use and internet access. Unlike MHV, CVT is not available to all VHA users. It is behind the VHA firewall and not a publicly available Web-based tool. CVT programs are specific and vary by site; participation in them often requires travel (a known barrier to care) to a nearby VHA medical center or outpatient facility. Though there are CVT services conducted into the home, they comprise a smaller percentage of CVT telemental health encounters. Providers refer veterans to CVT programs, so veterans are not self-referred. As we did not have information about these veterans' access to or familiarity with computers and cell phones, we could not assess

their effects, if any, on MHV only or dual use. We were also unable to capture access to internet, which may be less in rural areas where the availability of broadband is much more variable. At the time of this study, information about MHV feature use was limited to the secure messaging and prescription refill features. Other features were not available or there was no tracking or capturing of information about their use. We also did not have information about the content or purpose of secure messaging such as the purpose of its use (ie, administrative or for mental health or medical condition management). Finally, the MHV secure messaging rollout to VHA mental health providers occurred in earnest after the study period; thus, the lack of mental health provider's access to and use of MHV secure messaging before the rollout may have impacted adoption and use of MHV.

Conclusion

Dual use has the potential to be highly beneficial for promoting access to care for patients with mental health diagnoses. Although CVT makes it possible to receive clinical care from remote clinicians, patient portal functionality such as secure messaging assists patients in communicating with their clinical providers between in-person or telehealth visits. Prescription refill functionality can support the clinical management of prescription medications and has the potential to improve medication adherence. Appointment viewing and scheduling functionalities can help patients schedule and obtain the in-person or CVT care they need. During this study, dual use was still exceeding rare. However, since 2012, both MHV use and the number of CVT encounters have increased. In fiscal year 2012, there were over 218,000 unique CVT encounters, and in fiscal year 2017, there were over 470,000 unique CVT encounters [14]. In addition, the VHA has introduced new technologies to support patient care, including VA Video Connect, a free new desktop computer, smartphone or tablet app. There is a nationwide implementation of this technology, which allows for telemental health encounters to take place from anywhere. VA Video Connect will exponentially increase the number of home telehealth visits in the coming years. The VHA is also implementing a secure SMS text messaging app for patient use in the self-management of numerous conditions. The goal is to enhance self-care using reminders, motivational texts, and educational protocols for numerous conditions (eg, weight loss and smoking cessation). Other mental health *apps* can be used independently or as part of therapy. Given the number of eHealth technologies, additional research is needed to understand how they can best be used in conjunction with one another to facilitate patient treatment. It will also be important to continue to monitor eHealth technology adoption and use by vulnerable patients to identify barriers to access and devise solutions to those barriers in the years to come.

Acknowledgments

A US Department of Veterans Affairs Health Services Research and Development (HSR&D) Quality Enhancement Research Initiative (QUERI) Rapid Response Proposal (RRP 11-404) supported this study. EAA was supported by a VA Office of Academic Affiliations Advanced Postdoctoral Fellowship in Medical Informatics. SLS was supported by the VA HSR&D QUERI Career Development Award (CDA 10-210). The views expressed in this paper are those of the authors and do not necessarily represent the views of the US Department of Veterans Affairs. The authors would like to thank Hua Feng, PhD, for her contribution to dataset preparation.

Conflicts of Interest

None declared.

References

1. Shore P, Goranson A, Ward MF, Lu MW. Meeting veterans where they're @: a VA Home-Based Telemental Health (HBTMH) pilot program. Int J Psychiatry Med 2014 Oct 29;48(1):5-17. [doi: 10.2190/PM.48.1.b] [Medline: 25354923]
2. United States Department of Veterans Affairs: Office of Rural Health. 2017. America's Veterans Thrive in Rural Communities URL: https://www.ruralhealth.va.gov/docs/ORH_2015-2019Strategic_Plan_508compliant_FINAL.pdf [WebCite Cache ID 73VuF0tht]
3. Buzza C, Ono SS, Turvey C, Wittrock S, Noble M, Reddy G, et al. Distance is relative: unpacking a principal barrier in rural healthcare. J Gen Intern Med 2011 Nov;26 Suppl 2(10):648-654 [FREE Full text] [doi: 10.1007/s11606-011-1762-1] [Medline: 21989617]
4. Weeks WB, Wallace AE, West AN, Heady HR, Hawthorne K. Research on rural veterans: an analysis of the literature. J Rural Health 2008;24(4):337-344. [doi: 10.1111/j.1748-0361.2008.00179.x] [Medline: 19007387]
5. United States Department of Veterans Affairs. 2016. VA Telehealth Services: Real-Time Clinic Based Video Telehealth URL: https://www.telehealth.va.gov/real-time/ [WebCite Cache ID 73jYXczTt]
6. Chumbler NR, Haggstrom D, Saleem JJ. Implementation of health information technology in Veterans Health Administration to support transformational change: telehealth and personal health records. Med Care 2011 Dec;49 Suppl:S36-S42. [doi: 10.1097/MLR.0b013e3181d558f9] [Medline: 20421829]
7. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academies Press; 2001.

8. Hogan TP, Wakefield B, Nazi KM, Houston TK, Weaver FM. Promoting access through complementary eHealth technologies: recommendations for VA's Home Telehealth and personal health record programs. J Gen Intern Med 2011 Nov;26 Suppl 2:628-635 [FREE Full text] [doi: 10.1007/s11606-011-1765-y] [Medline: 21989614]
9. 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 Sep;17(2):203-211 [FREE Full text] [doi: 10.1136/jamia.2009.000240] [Medline: 20190065]
10. United States Department of Veterans Affairs. 2016. My HealtheVet Website - My HealtheVet Account Types URL: https://www.myhealth.va.gov/mhv-portal-web/anonymous.portal?_nfpb=true&_nfto=false&_pageLabel=spotlightArchive&contentPage=ipa/mhv_account_status-definitions.html#advanced&WT.ac=splash_account_advanced
11. United States Department of Veterans Affairs. MHV Program MyHealtheVet Statistics (through July 2018). 2018. URL: https://www.myhealth.va.gov/mhv-portal-web/webusers [accessed 2018-08-28]
12. Tsai J, Rosenheck RA. Use of the internet and an online personal health record system by US veterans: comparison of Veterans Affairs mental health service users and other veterans nationally. J Am Med Inform Assoc 2012 Jun;19(6):1089-1094 [FREE Full text] [doi: 10.1136/amiajnl-2012-000971] [Medline: 22847305]
13. Shimada SL, Brandt CA, Feng H, McInnes DK, Rao SR, Rothendler JA, et al. Personal health record reach in the Veterans Health Administration: a cross-sectional analysis. J Med Internet Res 2014 Dec 12;16(12):e272 [FREE Full text] [doi: 10.2196/jmir.3751] [Medline: 25498515]
14. United States Department of Veterans Affairs. VHA Support Service Centers (VSSC), Virtual Care and Telehealth (CVT & SFT Telehealth Data). Washington, DC: United States Department of Veterans Affairs; 2018.
15. Whealin JM, Seibert-Hatalsky LA, Howell JW, Tsai J. E-mental health preferences of Veterans with and without probable posttraumatic stress disorder. J Rehabil Res Dev 2015;52(6):725-738 [FREE Full text] [doi: 10.1682/JRRD.2014.04.0113] [Medline: 26562090]
16. Whealin JM, Jenchura EC, Wong AC, Zulman DM. How Veterans With Post-Traumatic Stress Disorder and Comorbid Health Conditions Utilize eHealth to Manage Their Health Care Needs: A Mixed-Methods Analysis. J Med Internet Res 2016 Oct 26;18(10):e280 [FREE Full text] [doi: 10.2196/jmir.5594] [Medline: 27784650]
17. Goulet JL, Erdos J, Kancir S, Levin FL, Wright SM, Daniels SM, et al. Measuring performance directly using the veterans health administration electronic medical record: a comparison with external peer review. Med Care 2007 Jan;45(1):73-79 [FREE Full text] [doi: 10.1097/01.mlr.0000244510.09001.e5] [Medline: 17279023]
18. Lurie N, Popkin M, Dysken M, Moscovice I, Finch M. Accuracy of diagnoses of schizophrenia in Medicaid claims. Hosp Community Psychiatry 1992 Jan;43(1):69-71. [Medline: 1544654]
19. Walkup JT, Wei W, Sambamoorthi U, Crystal S. Sensitivity of an AIDS case-finding algorithm: who are we missing? Med Care 2004 Aug;42(8):756-763. [Medline: 15258477]
20. United States Department of Veterans Affairs: Office of Rural Health. 2018. Rural Veterans: Rural Definition URL: https://www.ruralhealth.va.gov/aboutus/ruralvets.asp [accessed 2018-10-28] [WebCite Cache ID 73jYqeglg]
21. Moreau JL, Cordasco KM, Young AS, Oishi SM, Rose DE, Canelo I, et al. The Use of Telemental Health to Meet the Mental Health Needs of Women Using Department of Veterans Affairs Services. Womens Health Issues 2018 Mar;28(2):181-187. [doi: 10.1016/j.whi.2017.12.005] [Medline: 29339013]
22. United States Department of Veterans Affairs,. National Center for Veterans Analysis and Statistics. 2017. Quick Facts: VA Utilization Profile, FY 2016 URL: https://www.va.gov/vetdata/docs/QuickFacts/VA_Utilization_Profile.PDF [accessed 2018-10-28] [WebCite Cache ID 73Vw5StFo]
23. Bauer MS, Krawczyk L, Miller CJ, Abel E, Osser DN, Franz A, et al. Team-Based Telecare for Bipolar Disorder. Telemed J E Health 2016 Dec;22(10):855-864. [doi: 10.1089/tmj.2015.0255] [Medline: 26906927]
24. Roblin DW, Houston 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 Sep;16(5):683-689 [FREE Full text] [doi: 10.1197/jamia.M3169] [Medline: 19567790]
25. Kasckow J, Felmet K, Appelt C, Thompson R, Rotondi A, Haas G. Telepsychiatry in the assessment and treatment of schizophrenia. Clin Schizophr Relat Psychoses 2014 Apr;8(1):21-27A [FREE Full text] [doi: 10.3371/CSRP.KAFE.021513] [Medline: 23428781]

Abbreviations

CVT: Clinical Video Telehealth
eHealth: electronic health
HSR&D: Health Services Research and Development
MHV: My HealtheVet
OR: odds ratio
PTSD: posttraumatic stress disorder
QUERI: Quality Enhancement Research Initiative

TCT: telehealth clinical technician
VA: Veterans Affairs
VHA: US Department of Veterans Affair Health Administration
VISN: Veterans Integrated Service Networks
SMS: short message service

Edited by G Eysenbach; submitted 19.06.18; peer-reviewed by J Whealin, K Nazi, D Klein; comments to author 12.07.18; revised version received 03.09.18; accepted 12.09.18; published 07.11.18.

Please cite as:
Abel EA, Shimada SL, Wang K, Ramsey C, Skanderson M, Erdos J, Godleski L, Houston TK, Brandt CA
Dual Use of a Patient Portal and Clinical Video Telehealth by Veterans with Mental Health Diagnoses: Retrospective, Cross-Sectional Analysis
J Med Internet Res 2018;20(11):e11350
URL: http://www.jmir.org/2018/11/e11350/
doi:10.2196/11350
PMID:30404771

Original Paper

New Integrated Model Approach to Understand the Factors That Drive Electronic Health Record Portal Adoption: Cross-Sectional National Survey

Jorge Tavares[1], MSc, PhD; Tiago Oliveira[1], PhD

NOVA Information Management School, Universidade NOVA de Lisboa, Lisboa, Portugal

Corresponding Author:
Jorge Tavares, MSc, PhD
NOVA Information Management School
Universidade NOVA de Lisboa
Campus de Campolide
Lisboa, 1070 312
Portugal
Phone: 351 213 828 610
Email: d2012072@novaims.unl.pt

Abstract

Background: The future of health care delivery is becoming more patient-focused, and electronic health record (EHR) portals are gaining more attention from worldwide governments that consider this technology as a valuable asset for the future sustainability of the national health care systems. Overall, this makes the adoption of EHR portals an important field to study.

Objective: The aim of this study is to understand the factors that drive individuals to adopt EHR portals.

Methods: We applied a new adoption model that combines 3 different theories, namely, extended unified theory of acceptance and use of technology, health belief model, and the diffusion of innovation; all the 3 theories provided relevant contributions for the understanding of EHR portals. To test the research model, we used the partial least squares causal modeling approach. We executed a national survey based on randomly generated mobile phone numbers. We collected 139 questionnaires.

Results: Performance expectancy (beta=.203; t=2.699), compatibility (beta=.530; t=6.189), and habit (beta=.251; t=2.660) have a statistically significant impact on behavior intention (R^2=76.0%). Habit (beta=.378; t=3.821), self-perception (beta=.233; t=2.971), and behavior intention (beta=.263; t=2.379) have a statistically significant impact on use behavior (R^2=61.8%). In addition, behavior intention (beta=.747; t=10.737) has a statistically significant impact on intention to recommend (R^2=69.0%), results demonstrability (beta=.403; t=2.888) and compatibility (beta=.337; t=2.243) have a statistically significant impact on effort expectancy (R^2=48.3%), and compatibility (beta=.594; t=6.141) has a statistically significant impact on performance expectancy (R^2=42.7%).

Conclusions: Our research model yields very good results, with relevant R^2 in the most important dependent variables that help explain the adoption of EHR portals, behavior intention, and use behavior.

(J Med Internet Res 2018;20(11):e11032) doi:10.2196/11032

KEYWORDS

electronic health records; adoption; eHealth; patients; patient portals

Introduction

Overview

The electronic health record (EHR) portal or an EHR patient portal is a technology that combines an EHR system and a patient portal where patients can communicate with their health care providers (eg, send messages, schedule medical appointments, and request prescription refills online) and access their EHR and medical exams results [1-3]. EHR is a repository of patient data in a digital form, stored and exchanged securely. EHRs may include a range of data such as medical history, medication and allergies, immunization status, laboratory test results, radiology images, vital signs, personal statistics like age and weight, and billing information [4]. EHR portals have received great attention at the governmental level worldwide

[2,3,5]. In the United States, the support given to EHRs, via a meaningful use program, led the federal government to commit unparalleled resources to support the adoption of EHRs through incentive payments that can reach up to US $27 billion over 10 years [5,6]. EHR portals are a relevant topic not only in the United States but also in Europe through several projects such as the European Patients Smart Open Services (EpSOS) initiative promoted by the European Union Commission [3]. EpSOS focuses on developing a practical information and communication technology infrastructure that will enable secure access to patient information, including EHR among different European countries [3].

Understanding the adoption and use of EHR portals by patients is a very relevant topic with clear benefits for the society and future sustainability of the different health care systems in the world [4,7]. The warning signs are that the number of patients with chronic diseases is projected to grow by 45% between 2007 and 2025, and the health care providers' workforce will be 10% smaller [8]. Combining these 2 trends, there will be less health care professionals available in the future to provide support to the patients. EHR portals may help patients carry out self-management activities, making the use of the health care system more effective and sustainable, not only from the patient care standpoint but also from the financial perspective due to the increasing cost of the health care budget in different countries [3,4,8-10]. Regarding EHR portal users' sociodemographic characteristics, there is a consistent trend to be younger and more educated than the population average [11-13].

Most of the EHR portals' usage in the developing countries ranges between 5% and 10% of the total annual target population that they aim to reach [3,14]. Most of the EHR portals are implemented at an organizational or health care unit level, but there are some examples of EHR portals that have been implemented at the national level [3,14]. Probably the most successful nationwide implementation of an EHR portal is the Sundhed portal in Denmark with 1.1 million unique registered users, approximately 20% coverage of the Danish population [14]. In Portugal, both public and private health care institutions have EHR portals [3]. The most relevant public EHR portal is the National Health Service (NHS) portal that in its first implementation was not very successful, but in its new release, which was launched recently, provides a higher level of security (2-factor authentication) and broader access to the patients to their clinical information across NHS [3,7]. Recently in Portugal, we also have good examples of investment by private health care groups, like the EHR portal MyCuf [7]. What we also perceive in the private health care institutions such as MyCuf is the exclusive delivery of the medical examination results and other documentation online, increasing the use of this online platform and making it compulsory, something that is not yet a policy in the NHS [3,7]. Due to the fact that some of the private health care providers perceived an efficiency advantage, they have invested in developing more sophisticated portals than the public hospitals, but the new version of the NHS portal has now also started aggregating all the patient information from the different public health care institutions, making it available to the patient at one place [7]. There are differences between the different types of EHR portals in the same country and among different countries, but the most common and frequently used features identified in the literature, which generally apply to an EHR portal, are as follows: management of health information and communication with health providers, medical appointments schedule, check their own EHR, and request for medical prescription renewals [2-4,15-17]. Taking into account the relevance of all the current initiatives that are ongoing in Portugal regarding EHR portals, a nationwide survey using a sample of randomly generated mobile numbers was applied in our study.

The goals of this study are to estimate the percentage of EHR portal users among the Portuguese population and understand the factors that drive health care consumers to adopt and use EHR portals. We apply 3 different theories to build our research model: the extended unified theory of acceptance and use of technology (UTAUT2), the health belief model (HBM) theory, and the diffusion of innovation (DOI) theory. In the Research Model section, a more detailed rationale explaining why we combined these 3 theories is provided.

Theoretical Background

The goal of our study is to focus on the adoption of the EHR portals from the standpoint of the health care consumer. According to the literature, assessing the adoption of eHealth tools by health care consumers still demands more effort due to the persisting low number of studies published to date and in view of the importance of the topic [3,4]. The most frequently used adoption models when studying eHealth adoption by health care professionals are the unified theory of acceptance and use of technology [2,18,19] and the technology acceptance model (TAM) [3,20,21]. When evaluating the studies published in the field of consumer health information technology adoption, most of the research studies use TAM or extensions of TAM [22-25]. Although the studies that used extended TAM used other models and theories with TAM to adapt it to the consumer health technology context (see Table 1), TAM was not envisaged with consumer focus in mind. Rather, we need a model developed for the consumer use setting, and UTAUT2 was developed precisely with this purpose, achieving good results [26]. A recent study using a UTAUT2 extension demonstrated its usefulness in assessing the critical determinants for the adoption of EHR portals in which the construct habit, which is a consumer-specific construct, was the one with the greatest impact on the adoption of EHR portals [2]. This fact shows the importance of using research models that are consumer specific.

Table 1. eHealth patient-focused adoption models.

Theory	Dependent variable	Findings	Reference
TAM[a], integrated model, motivational model	eHealth behavioral intention	• Users' perceived technology usefulness, users' perceived ease of use, intrinsic motivation, and extrinsic motivation have significant positive impact on behavioral intention • Integrated model does not have better results than TAM or motivational model when predicting behavioral intention	[25]
UTAUT2[b] plus CFIP[c] (cross-country analysis: United States vs Portugal)	Behavioral intention and use behavior in EHR[d] portals	• Behavioral intention drivers are performance expectancy, effort expectancy, social influence, hedonic motivation, price value, and habit. The predictors of use behavior are habit and behavioral intention • Social influence, hedonic motivation, and price value are only predictors in the US group • Confidentiality issues do not seem to influence acceptance	[27]
TAM, Trust and Privacy	Intention to adopt eHealth	• Perceived ease of use, perceived technology usefulness, and trust are significant predictors	[22]
UTAUT2	Behavioral intention and use behavior in EHR portals	• The behavioral intention drivers are performance expectancy, effort expectancy, social influence, and habit • Habit and behavioral intention are drivers of use behavior	[7]
DOI[e] (mix of qualitative and quantitative study)	Adoption rate of an e-appointment scheduling service	• The influence of the perceived attributes of the e-appointment scheduling service according to the DOI theory helps explaining the low adoption and use • Low socioeconomic status and lower educational level negatively influence the e-appointment scheduling service adoption rate	[13]
Extended TAM in health information technology	Health information technology behavioral intention	• Perceived ease of use, perceived technology usefulness, and perceived threat significantly influenced health consumer behavioral intention	[23]
UTAUT2 extended model	Behavioral intention and use behavior in EHR portals	• Effort expectancy, performance expectancy, habit, and self-perception are predictors of behavioral intention • Habit and behavioral intention are predictors of use behavior	[2]
Institutional theory and UTAUT[f]	Patient portal use behavior	• Coercive and mimetic pressures significantly influence patient portal use behavior • Normative pressure was found to be not relevant	[28]

[a]TAM: technology adoption model.

[b]UTAUT2: extended unified theory of adoption and use of technology.

[c]CFIP: concern for information privacy.

[d]EHR: electronic health record.

[e]DOI: diffusion of innovation.

[f]UTAUT: unified theory of adoption and use of technology.

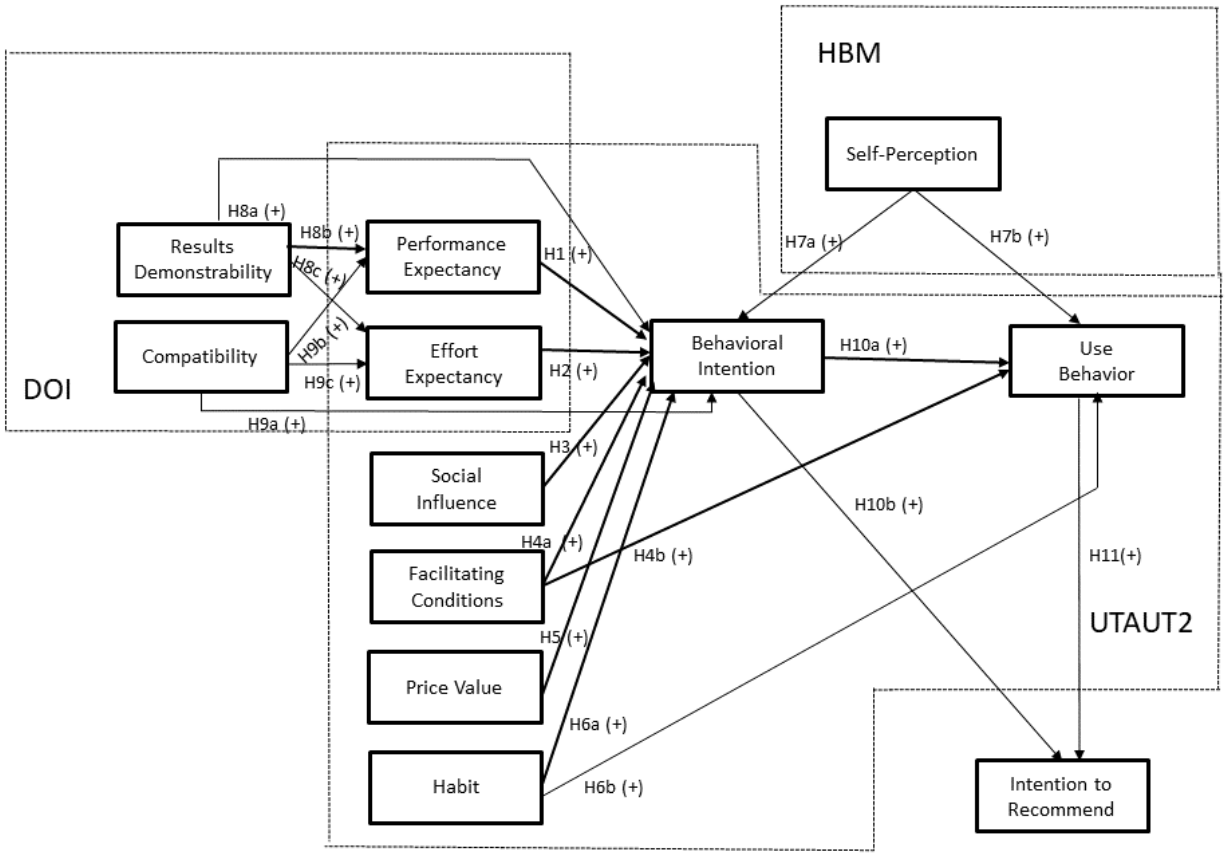
Although EHR portals are consumer-oriented technologies, because a patient can be viewed as a health care consumer, the use of a model like UTAUT2 should not be regarded sufficient to explain the complexity of EHR portal adoption [2,23,26]. Several studies that used constructs or frameworks related to the HBM demonstrated their usefulness and statistical significance in explaining health information consumer adoption [2,23,29]. The HBM advocates that belief in health risk predicts the likelihood of engaging in health behavior, or an alternative way to look into it is to consider that the perceived severity, instead of the real severity, of the health complaint could be the driving force behind the action [23,30]. Evidence in the literature shows that the global usage of EHR portals is still limited [2,5,14,31].

As the rate of adoption is still low in the use of EHR portals, literature that has addressed the eHealth patient technologies under the scope of DOI also mentioned a low level of global use and identified the users as early adopters [13,32]. Earlier studies that focused on understanding eHealth patient-centered technologies and EHR portals identified both performance expectancy and effort expectancy as important predictors of behavioral intention to use [2,7,23,25]. Both performance expectancy and effort expectancy have their equivalents within DOI theory as relative advantage and complexity [32,33], providing another strong argument to use DOI theory when studying EHR portals [7,13]. This study included intention to recommend as a dependent variable. According to our knowledge, this is the first time that intention to recommend is studied in the field of the adoption of EHR portals [2,22,23,27,33]. Understanding whether current users of new technologies that have a low level of adoption can be used to promote them is a valuable asset that should be evaluated [33].

Research Model

As EHR portals are a new technology focused on consumer health [2,3], our research model is a combination of UTAUT2 [26], self-perception, a construct from the HBM [2,23,30,34,35], and a framework based on the DOI model [32,33,36]. On the basis of the extensive literature review and previous studies, the need to have a model with patient-centric focus was identified, something that UTAUT2 provides, with consumer-specific constructs and good results in previous eHealth and EHR portal studies [2,14]. As we are studying the EHR portals from the heath care consumers' perspective and not from the health care providers' standpoint, it is relevant to use UTAUT2, because most of the existing (information technology) IT adoption models are not consumer-specific [2,15]. In addition, previous research identified the relevance of including health care–specific constructs in studying the adoption of EHR portals and eHealth platforms [3,16,17]; therefore, we used a construct related to HBM that already achieved good results within the scope of our study [3]. EHR portals are a new eHealth technology and the DOI model revealed in past studies to be able to explain the adoption of new eHealth tools successfully, therefore making it suitable to be used to study EHR portals adoption [13,18]. We also made some improvements in our research model concerning the theories we used. In the UTAUT2 framework, we did not use the construct hedonic motivation. Hedonic motivation is conceptualized as intrinsic motivation (eg, pleasure or enjoyment) [26]. People use EHR portals frequently when they are ill [1] and that can be viewed by many as not being an enjoyable activity [37]. Recent literature confirms no consistent and relevant results in predicting the adoption of EHR portals with hedonic motivation [2,7,27]. Literature evidence shows that constructs related to the HBM, such as perceived health risk or self-perception, are much better motivation predictors of adoption of EHR portals than hedonic motivation [2,23]. We also used intention to recommend as a dependent variable. This is a variable that has not been used in the literature to explain the adoption of EHR portals [2,27,38]. Instead, it has been used in other technologies to explain adoptions such as mobile payments [33], which were also regarded as relatively new and with a low usage level [33] like EHR portals [2]. In these types of technologies, providers start to rely on current or potential users to recommend them to others [33]. That is why we included intention to recommend in our research model. Figure 1 illustrates the new research model.

Figure 1. The research model. DOI: diffusion of innovation; HBM: health belief model; UTAUT2: extended unified theory of acceptance and use of technology.



Extended Unified Theory of Acceptance and Use of Technology Constructs

Performance expectancy is theorized to be the degree to which using a specific technology provides benefits to consumers in executing particular tasks [26,39]. Overall, patients adopt and use more eHealth tools and EHR portals that provide benefits in executing online health-related activities [2,7,23,25]:

> H1: Performance expectancy will positively influence behavioral intention.

Effort expectancy is the degree of ease connected to consumers' usage of a certain technology [26,39]. The simpler it is for health care consumers to use an EHR portal, the greater is the likelihood that they will use it [2,7,23,25]:

> H2: Effort expectancy will positively influence behavioral intention.

Social influence is the extent to which people acknowledge that others who are significant to them believe they should use a particular technology [2,23,25]. According to the literature, social influence plays a role in the adoption of eHealth and EHR portals, because patients with the same health issues tend to be induced by others sharing the same or similar condition [27,28,40,41]:

> H3: Social influence will positively influence behavioral intention.

Facilitating conditions refers to consumers' awareness of the support and resources available to execute a particular behavior [26,39]. A possible barrier to patients' use of eHealth tools is the nonexistence of resources or support services that enable them to access and use these types of technology, implying that health care consumers with better conditions favor EHR portals usage and adoption [26,27,42]:

> H4(a): Facilitating conditions will positively influence behavioral intention.
>
> H4(b): Facilitating conditions will positively influence use behavior.

If we relate to the consumer environment, price value is a relevant dimension, because consumers usually bear the costs linked with purchasing products and services [26]. If health care consumers can obtain the results of their medical examination online, for example, through an EHR portal, they save time and transportation costs by avoiding an unnecessary trip to a clinic or hospital [27,43]:

> H5: Price value will positively influence behavioral intention.

Habit can be described as the degree to which people tend to perform behaviors automatically due to learning [26]. According to recent literature, habit positively influences the use and adoption of eHealth tools and EHR portals [27,44]:

> H6(a): Habit will positively influence behavioral intention.
>
> H6(b): Habit will positively influence use behavior.

The role of behavioral intention has been recognized in eHealth with the literature affirming that the driver of use and adoption of EHR portals is preceded by the behavioral intention to use them [2,23,25,27]:

> H10(a): Behavioral intention will positively influence use behavior.

Health Behavior Construct

Supporting the concept of self-perception is the HBM. HBM assumes that subjective health concerns determine whether individuals execute a health-related action such as making an appointment with their physician [30]. Self-perception in health [30,34,35] posits that the perceived (rather than the real) severity of the health complaint could be the driving force inducing the action [30,35,45].

There is evidence in the literature that self-perception influences behavior intention to use eHealth tools and EHR portals [2,23]:

> H7(a): Self-perception will positively influence behavioral intention.

There is also evidence in the literature that self-perception can not only drive intentions but also directly influence actions with regard to the use of health-related services [2,23,30]. Often with sensitive topics and particularly with health-related topics, a mismatch between intentions and effective actions may occur [4,27,46]. It is then also relevant to evaluate the potential positive effect of self-perception on use behavior:

> H7(b): Self-perception will positively influence use behavior.

Diffusion of Innovation Constructs

Roger's DOI theory is one of the most acknowledged theories for studying IT adoption [13]. According to DOI, innovation is an idea, technology, or a process that is perceived as unknown or new to a particular group of individuals [13,47]. Diffusion is how the information about the innovation is shared inside the social system [47]. The attributes of an innovation comprise 5 user-perceived qualities: relative advantage, compatibility, complexity, trialability, and observability [47]. Moore and Benbasat [36] expanded the original set of innovation attributes proposed by DOI to be applicable to the IT setting. One example was the construct observability, which was subdivided into results demonstrability and visibility [36]. Subsequent studies have found that results demonstrability is more relevant than visibility in predicting users' intention to use a technology, particularly in IT health care [32]. We did not measure trialability because there was no evidence that our target population has participated in a trial usage of EHR portals [3]. EHR portals should be seen as a new technology that relates to the concept of an innovation in consumer IT within the scope of health care.

Relative advantage is the extent to which the consumer perceives improvements or benefits upon the current technology by adopting an innovation [47]. Relative advantage measures fundamentally the same thing as performance expectancy within the context of DOI [32,33]. Complexity measures the extent to which an innovation is difficult to understand or be used [47]. We also find a commonality between effort expectancy and complexity [32,33]. Both relative advantage and complexity within the context of DOI, according to the literature, may be

regarded as positively influencing the behavioral intention to adopt EHR portals [2,13,27,32].

Results demonstrability is the degree to which the tangible results of adopting and using an innovation can be visible and then communicable [36]. According to the literature, this may have a direct effect on the behavioral intention to use an EHR portal [13,32]. In addition, potential users can better comprehend the benefits of using a new eHealth technology when noticeable results of the tool are directly evident, advocating a positive connection between results demonstrability and performance expectancy [32]. The degree to which a specific individual noticed the results of using an innovation to be demonstrable partially reflects belief in using the tool and more easily achieving the desired outcome [13,32]. Thus, we theorize and ground on the literature that results demonstrability will positively influence effort expectancy:

> H8(a): Results demonstrability will positively influence behavioral intention.
>
> H8(b): Results demonstrability will positively influence performance expectancy.
>
> H8(c): Results demonstrability will positively influence effort expectancy.

Compatibility measures the extent to which an innovation is perceived as being aligned with the existing consumer lifestyle values and current and past experiences [47]. Compatibility has demonstrated to be a predictor of the behavioral intention to adopt a new technology in general, and also in consumer eHealth [13,33]. Compatibility, also like results demonstrability, is an antecedent of performance expectancy and effort expectancy [13,33]. Users may perceive EHR portals to be more compatible if they see advantages in using them to manage specific health care activities without additional complexity [2,13,33]. Compatibility consequently strengthens performance expectancy, effort expectancy, and behavioral intention to use EHR portals [2,13,33]:

> H9(a): Compatibility will positively influence behavioral intention.
>
> H9(b): Compatibility will positively influence performance expectancy.
>
> H9(c): Compatibility will positively influence effort expectancy.

Users' Intention to Recommend Electronic Health Record Portals

IT consumers with a greater intention to adopt a new technology are more likely to become users and to recommend that specific technology to others [33,48]. Often with sensitive topics and particularly with health-related topics, a mismatch between intentions and effective actions may occur [4,27,46], so it is especially relevant to independently measure how the behavioral intention and use behavior may influence the intention to recommend the use of EHR portals:

> H10(b): Behavioral intention will positively influence intention to recommend EHR portals to others.
>
> H11: Use behavior will positively influence intention to recommend EHR portals to others.

Methods

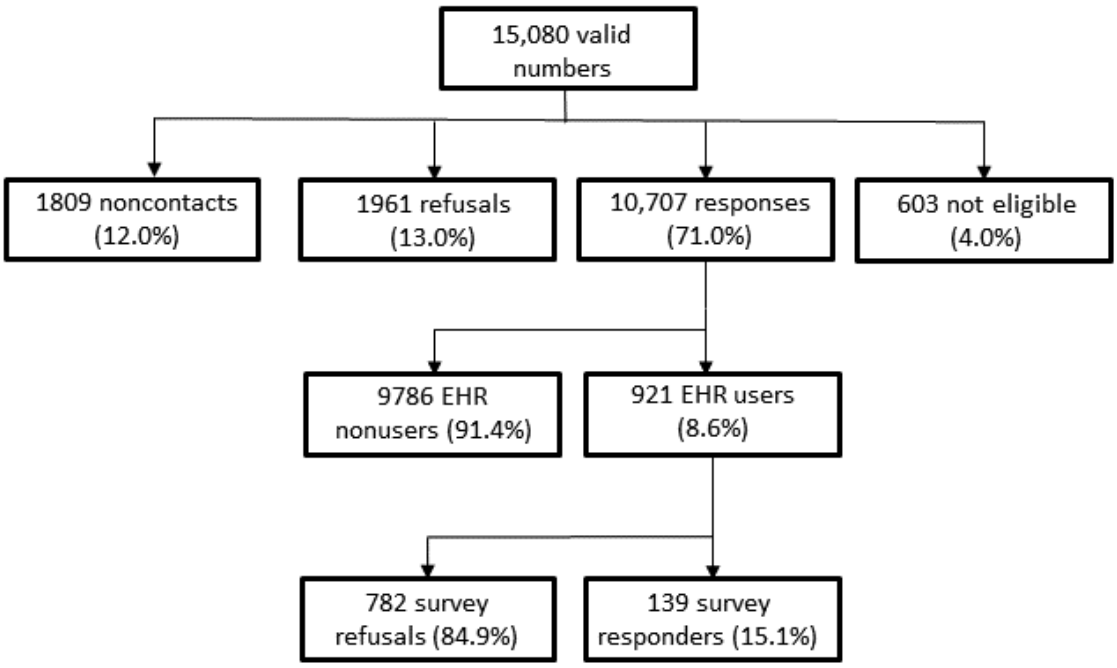
Measurements

All the items were adopted from the studies by Venkatesh et al [26], Wilson and Lankton [25], van de Kar et al [30], Moore and Benbasat [36], and Oliveira et al [33], with minor changes to adapt to EHR portal technology. The items are exhibited in Multimedia Appendix 1. The questionnaire was delivered in Portuguese after being translated by a certified translator. To guarantee that the content did not lose its original meaning, a back-translation was made from Portuguese to English by a different certified translator and compared with the original [49]. The scales' items were measured on a 7-point range scale, ranging from "strongly disagree" (1) to "strongly agree" (7). Use behavior was measured on a different scale. The scale from UTAUT2—from "never" to "many times per day"—was adapted to "never" to "every time I need," because EHR portal usage is not expected to be as regular as mobile internet usage. Sociodemographic questions were also included. Age was measured in years, and gender was coded as a dummy variable (0 or 1), with women represented by 0. Having a private health insurance was also coded as a dummy variable (0 or 1), with its absence represented by 0. Information about the level of education of the respondents was also assessed with 3 different layers (university degree, high school education complete, and high school education incomplete).

Data Collection

A pilot survey was performed to validate the questions and the scale of the survey. From the pilot survey, we had 20 responses. No issues were reported that could question the fact that the questionnaire items were not reliable. However, from the outcome of the pilot survey, there was strong evidence that our nonresponse rate in the main survey could be high (>50%). The data from the pilot survey were not included in the main survey. As one of the goals of our study is to determine the usage prevalence rate of this type of technology, we subdivided our survey into 2 phases. Two-phase sampling designs are frequently used in epidemiological studies, in health care, when a disease is rare, and when the diagnosis of the disease is difficult or expensive [50]. In the first phase, a bigger random sample from the targeted population is screened with less intensive and expensive screening. In the second stage, a random subsample of the individuals is studied more intensively [50]. We used a similar approach; our target population is also infrequent, but in our case, the aim is to handle a potential high nonresponse rate. Specifically, our population of interest is the Portuguese adult population (age≥18 years) who are users of EHR portals. In the first section, we asked the potential respondent if she or he was a Portuguese adult. If the response was positive, we asked if she or he was a user of EHR portals, and only after identifying that she or he was a user, we asked about her or his interest in replying to our main survey. The EHR portal user is a current user of any of the 4 main functionalities that EHR portals can provide in general (management of health information and communication with health providers, medical appointments schedule, check their own EHR, and request for medical prescription renewals) [1,3-7].

Figure 2. Sampling procedure and results. EHR: electronic health record.



To interview our target population, we used a nationwide mobile phone survey. According to the latest research, 94.5% of the Portuguese adult population had a mobile phone by December 2016 [51], making it a valuable approach to conduct this survey due to its high coverage of the target population. The survey was computer-assisted, and all answers were recorded immediately. The mobile phone sample consisted of randomly generated numbers. Portuguese mobile phone numbers are of 9 digits, and the first 2 digits identify the operator [51,52].

The Portuguese Telecommunications Regulation Authority (ANACOM) delivers information concerning the market share of the 3 operators offering mobile services in Portugal [51]. This was used to split the sample into 3 mobile subsamples proportional to the market share [52,53].

Within each 2-digit prefix of the 3 operators, numbers were created by a generator of 7-digit random numbers [52]. Up to 4 additional call attempts were made to each number to establish contact, with the exceptions when the number was identified as nonworking or not attributed (a message from the operator provides this information) [52,53]. The survey took place between July 25, 2017, and October 15, 2017. All study participants were informed about the research purpose, confidentiality protection, and the anonymity of the information collected and that by answering all the questions, they were giving their consent to participate in the survey. In total, we obtained 15,080 valid numbers. From this sample, we obtained a 71.0% response rate regarding the question to identify the users of EHR portals. From the ones that were eligible to answer the survey, we obtained 139 completed questionnaires and a response rate of 15.1% (see Figure 2).

Data Analysis

To test our research model, we applied partial least squares structural equation modeling (PLS-SEM). The motivations for choosing this approach were the model complexity (many constructs and many indicators), formatively measured constructs are part of the structural model, and the fact that the PLS-SEM method is oriented to explain variance of the research model and to detect statistically significant constructs [54-56]. SmartPLS 3 [57] was used to estimate the model. Before evaluating the structural model, we assessed the measurement model to evaluate construct reliability, indicator reliability, convergent validity, and discriminant validity.

Results

Sample Characteristics

The sample characteristics results versus the target population profile are displayed in Table 2.

Age groups are from 2011census data [58], the level of education uses the latest inquiry from the National Institute of Statistics in 2016 [59] as a source, and for the number of people with private health insurance in Portugal, the information is from the Portuguese Association of Insurance Companies from 2016 [60]. Except for the case of gender, all other sample characteristics differ from the target population. We should not generalize these results as representative of the target population due to the high nonresponse rate in the second phase (Figure 2). Early adopters in eHealth are usually younger and more educated than the general population, in line with the findings of our study [13,38,61]. Higher income is also related to eHealth early adopters, which may justify the higher percentage of people in our sample compared with the target population with

private insurance [13,38]. In Portugal, there is a NHS that provides coverage to all citizens, but in the last decade, there was a substantial increase in the number of people obtaining complementary private health insurance [60,62,63]. In Portugal, the main private health care institutions have also implemented measures to encourage the use of eHealth tools, including EHR portals [7].

We also assessed the common method variance initially using Harman one-factor test. If the total variance for a single factor is less than 50%, it suggests that common method variance is not an issue [64]. The greatest variance (47.16%) explained by 1 factor was, in our case by the first one, still lower than 50%. Subsequently, the marker-variable technique was applied, in which we used a theoretical unrelated construct, the marker variable [65]. We found no significant correlation between the research model constructs and the marker variable. Therefore, we can conclude that common method variance was not a serious problem, verified by 2 different and established criteria [64-66].

Usage Results

According to the results in the first stage of our inquiry, 8.6% of the Portuguese adult population uses EHR portals. This value is within the range of 5%-10%, most commonly reported in the literature [2,14]. We obtained a response rate of 71.0% in the first stage. In the case of our survey, we cannot assume that the nonresponses are *missing at random*, and hence, their lack may lead to a bias [67,68]. According to the literature, the ideal value for responses in a survey should be greater than or equal to 80% to make assumptions about the results and be representative of the population [67,69]. The types of nonresponses in our survey are included in Figure 2. They include 4.0% of individuals who were ineligible, mostly because their age was less than 18 years. Overall, according to other surveys in general and surveys for populations of low prevalence, our response rate may be regarded as reasonable [52,53,70,71].

The usage patterns reported in Table 3 show a good adoption and usage by the users. The feature with the least usage is the request for medical prescription renewals; our sample is relatively young (mean age: 36.0 years). The request for prescription renewals is usually related with chronic conditions that are more prevalent among older people [3,72]. The descriptive statistics of the other questionnaire items are provided in Multimedia Appendix 1.

Table 2. Sample characteristics versus target population.

Characteristics	Sample (n=139), n (%)	Population, (n=8,657,240)[a], n (%)	*P* value[b]
Age (in years)			**<.001**
18-34	67 (48.2)	2,243,957 (25.92)	
35-49	58 (41.7)	2,367,755 (27.35)	
50-64	8 (5.8)	2,035,317 (23.51)	
≥65	6 (4.3)	2,010,211 (23.22)	
Gender			**.81**
Male	64 (46.0)	4,072,366 (47.04)	
Female	75 (54.0)	4,584,874 (52.96)	
Private health insurance			**<.001**
Yes	78 (56.1)	2,172,967 (25.10)	
No	61 (43.9)	6,484,273 (74.90)	
Education			**<.001**
University degree	88 (63.3)	1,576,483 (18.21)	
Nonuniversity degree	51 (36.7)	7,080,757 (81.79)	

[a]Portuguese census 2011 adult population.

[b]χ^2 test.

Table 3. Electronic health record portals' usage patterns.

Use indicators	Average	Median	Minimum	Maximum
UB1: Management of personal information and communication with health providers	4.37	5.00	1.00	7.00
UB2: Medical appointments schedule	4.75	5.00	1.00	7.00
UB3: Check their own electronic health record	4.56	5.00	1.00	7.00
UB4: Request for medical prescription renewals	3.34	3.00	1.00	7.00

Measurement Model

Typically, the first criterion to be assessed is construct reliability or internal consistency reliability. It is traditionally evaluated by Cronbach alpha, which delivers an estimation of the reliability grounded on the intercorrelations of the observed indicator variables [54]. Cronbach alpha assumes that all indicators are equally reliable. However, PLS-SEM prioritizes the indicators according to their individual reliability [54]. Due to Cronbach alpha's stated limitations, it is technically more suitable to apply an alternative measure for the same purpose, which is mentioned to as composite reliability [54]. The composite reliability measure takes into account the different indicator variables' outer loadings [54]. Table 4 shows that all constructs have composite reliability higher than .70, showing evidence of internal consistency [26].

The most commonly used PLS-SEM measure to access convergent validity on the construct level is the average variance extracted (AVE) [54,55]. According to the literature, we should aim to an AVE value of .50 or greater, meaning that on average, the construct explains more than 50% of the variance of its indicators [54,55]. The results in Table 4 demonstrate that this criterion is fully achieved. In addition, to evaluate indicator reliability, a well-known rule of thumb is that a latent variable should explain a significant part of each indicator's variance, ideally at least half [54,73]. This means that an indicator's outer loading should be greater than or equal to .70 [54,73]. Nevertheless, indicators with outer loadings between .40 and .70 should be removed only when deleting the indicators leads to an increase in the AVE or the composite reliability above the suggested threshold value [54,55].

Only 1 indicator was removed SP4, with an outer loading below .40. All other indicators have an outer loading higher than .70, and they are shown in Multimedia Appendix 2.

Discriminant validity is the degree to which a construct is truly dissimilar from the other constructs in the model [54]. Traditionally, researchers have relied on 2 measures of discriminant validity [54,55]. One is the Fornell-Larcker criterion that compares the square root of the AVE values with the latent variables' correlations. Particularly, the square root of each construct's AVE should be greater than its highest correlation with any other construct [54,55], and as seen in Table 5, this criterion is met. The other traditional measure of discriminant validity is the cross-loadings. Particularly, an indicator's outer loading on the associated construct should be higher than any of its cross-loadings on other constructs [54,55]. This criterion is also met, as seen in Multimedia Appendix 2. Recent research suggests the use of an alternative criterion, the heterotrait-monotrait ratio (HTMT) of the correlations. HTMT is the ratio of the between-trait correlations to the within-trait correlation [54]. Ideally, the HTMT value should be different from 1; prior research suggests a threshold value of .90 [54]. Ideally, to avoid any ambiguity, the most recent research applied a procedure called bootstrapping to derive a distribution of the HTMT statistic and to determine if it is significantly different from 1 [54]. With this procedure, it is feasible to derive a bootstrap CI (eg, 95%). A CI including the value 1 indicates a lack of discriminant validity. On the contrary, if the value 1 falls outside the interval's range, this advocates that the 2 constructs are empirically different [54]. This criterion is also met for our model, as seen in Multimedia Appendix 2.

Use behavior, which was modeled using 4 formative indicators, is evaluated by specific quality criteria linked to formative indicators [54]. A recently proposed way to evaluate the formative construct's validity is to examine its correlation with an alternative measure of the construct, using a global single item or reflective measures (redundancy analysis). The strength of the path coefficients linking the 2 constructs should be at least .70 [54]. In our study, we used a global single item for use behavior, obtaining a path coefficient of .851, thus confirming the convergent validity for the use behavior formatively measured construct. In addition, we need to assess the formative indicators for potential collinearity issues. As seen in Table 6, all variance inflation factors are below 5, meaning that collinearity is not an issue [54]. An additional relevant criterion for evaluating the contribution of a formative indicator is its weight to be statistically significant, or in case it is not significant, its outer loading must be greater than .50 [54]. All formative indicators comply with these assumptions, as shown in Table 6.

Table 4. Cronbach alpha, composite reliability, and average variance extracted.

Constructs	Cronbach alpha	Composite reliability	Average variance extracted
Behavior intention	.929	.955	.876
Compatibility	.936	.955	.841
Effort expectancy	.897	.929	.767
Facilitating condition	.822	.883	.655
Habit	.876	.924	.803
Intention to recommend	.879	.942	.891
Performance expectancy	.863	.917	.786
Price value	.953	.970	.915
Results demonstrability	.880	.926	.806
Social influence	.958	.973	.923
Self-perception	.817	.893	.739

Table 5. Correlations and square roots of all average variance extracted in the model. Diagonal elements are square roots of all average variance extracted, and off-diagonal elements are correlations.

Constructs	BI	CO	EE	FC	HT	IR	PE	PV	RD	SI	SP	UB
Behavioral intention (BI)	*.936*											
Compatibility (CO)	.809	*.917*										
Effort expectancy (EE)	.561	.645	*.876*									
Facilitating conditions (FC)	.605	.644	.674	*.809*								
Habit (HT)	.703	.616	.541	.534	*.896*							
Intention to recommend (IR)	.826	.779	.610	.593	.585	*.944*						
Performance expectancy (PE)	.695	.651	.481	.468	.537	.648	*.887*					
Price value (PV)	.554	.581	.510	.408	.683	.537	.462	*.956*				
Results demonstrability (RD)	.615	.763	.660	.581	.556	.635	.528	.521	*.898*			
Social influence (SI)	.487	.415	.415	.321	.574	.490	.494	.409	.374	*.961*		
Self-perception (SP)	.514	.432	.224	.333	.552	.401	.494	.243	.449	.380	*.860*	
Use behavior (UB)	.682	.565	.491	.494	.721	.625	.554	.516	.508	.534	.596	formative

Table 6. Formative indicators' quality criteria.

Indicators	Variance inflation factor	*t* value (weights)[a,b]	Outer loadings
UB1: Management of personal information and communication with health providers	1.976	4.923[a]	.892
UB2: Medical appointment schedule	2.432	4.475[a]	.860
UB3: Check their own electronic health record	3.401	0.753	.800
UB4: Request for medical prescription renewals	1.566	1.791	.660

[a]*P*<.01.

[b]*P*<.05.

Considering all the results and findings, all reflective and formative constructs exhibit satisfactory levels of quality. Thus, we can proceed with the evaluation of the structural model.

Structural Model

Structural model path significance levels were estimated using a bootstrap with 5000 iterations of resampling to obtain the maximum possible consistency in the results [54]. We checked the structural model for collinearity issues by examining the variance inflation factor values of all sets of predictor constructs, and all variance inflation factor values are below the threshold of 5. Therefore, collinearity is not a critical issue in the structural model [54]. To assess the structural model we used the R^2, path coefficients significance, and the f^2 effect size [54,55]. The results are shown in Table 7. Overall, the model explains 76.0% of the variance in behavioral intention and 61.8% in use behavior, with these 2 being the most relevant dependent variables in our model. In addition to assessing the R^2 values of all endogenous constructs, the change in the R^2 value when a specific construct is removed from our model can be used to assess whether the construct has a substantial impact on the endogenous constructs [54]. Guidelines for measuring f^2 are that values of .02, .15, and .35, respectively, represent small, medium, and large effects of the exogenous latent variable; values of less than .02 denote that there is a null effect [54]. Taking a particularly important role in our model, compatibility has a medium effect on both behavior intention and performance expectancy and a small effect on effort expectancy, showing the relevance of this construct in our research model. Another construct with a relevant role in our model is behavior intention, with a large effect on intention to recommend and a small effect on use behavior. Finally, habit is a construct that has a medium effect size on use behavior and a small effect size on behavior intention. With only small effect sizes, we have the effect of performance expectancy on behavior intention, self-perception on use behavior, results demonstrability on effort expectancy, and use behavior on intention to recommend; however, the last one is without a statistically significant path coefficient.

Table 7. Structural model results and findings regarding hypotheses.

Dependent and independent variables	f^2	beta	t_{beta}	Hypothesis	Results	R^2	R^2_{adj}
Behavioral intention						**.760**	**.743**
Performance expectancy	.081	.203	2.699[a]	H1	Supported		
Effort expectancy	.001	−.022	.311	H2	Not supported		
Social influence	.002	.025	.450	H3	Not supported		
Facilitating conditions	.014	.086	1.547	H4(a)	Not supported		
Price value	.000	−.015	.277	H5	Not supported		
Habit	.079	.251	2.660[a]	H6(a)	Supported		
Self-perception	.008	.062	.916	H7(a)	Not supported		
Results demonstrability	.015	−.102	1.357	H8(a)	Not supported		
Compatibility	.328	.530	6.189[a]	H9(a)	Supported		
Use behavior						**.618**	**.607**
Facilitating conditions	.005	.056	.727	H4(b)	Not supported		
Habit	.165	.378	3.821[a]	H6(b)	Supported		
Self-perception	.095	.233	2.971[a]	H7(b)	Supported		
Behavioral intention	.075	.263	2.379[b]	H10(a)	Supported		
Intention to recommend						**.690**	**.685**
Behavioral intention	.962	.747	10.737[a]	H10(b)	Supported		
Use behavior	.023	.116	1.565	H11	Not supported		
Effort expectancy						**.483**	**.476**
Compatibility	.092	.337	2.243[b]	H9(c)	Supported		
Results demonstrability	.131	.403	2.888[a]	H8(c)	Supported		
Performance expectancy						**.427**	**.418**
Compatibility	.257	.594	6.141[a]	H9(b)	Supported		
Results demonstrability	.004	.075	.561	H8(b)	Not supported		

[a]$P<.01$.

[b]$P<.05$.

Discussion

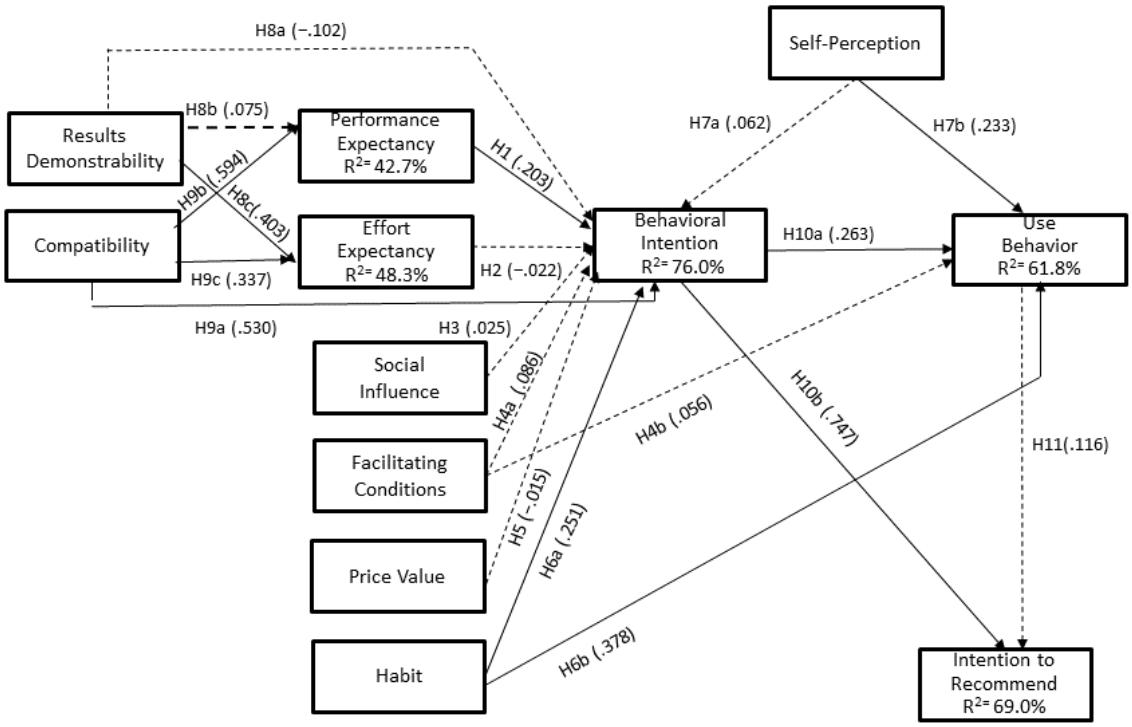
Principal Findings

The results advocate that using our new research model in an eHealth-related area—EHR portal acceptance by patients—yields very good results, explaining 76.0% of the variance on behavioral intention and 61.8% of the variance in use behavior, the most relevant dependent variables in our model [26]. We also obtained an R^2 of 69.0% in intention to recommend, also a very good result [26,33]. Overall, the use of the 3 theories, UTAUT2, HBM, and DOI, was a successful strategy because in all of them we had constructs with statistically significant impact on explaining the adoption of EHR portals (see Figure 3). The constructs with the highest effect size in the model were compatibility, habit, and behavioral intention.

Theoretical Implications

In our model, performance expectancy has a statistically significant effect on behavior intention, suggesting that individuals care about the results and advantages that EHR portals can bring for them to manage their own health more effectively, supporting *H1*. This finding is supported by previous studies [25,27]. In regard to effort expectancy, there is no statistically significant impact, not supporting *H2*. This finding contradicts results from earlier studies that used effort expectancy as part of UTAUT2 [2,7], but in other studies also with new technologies and within health care, when effort expectancy is evaluated as part of DOI, it also obtained nonsignificant results [32,33]. A possible explanation, also supported by the literature, is that early adopters of new technologies have a higher cognitive ability and are more used to manage complexity and that they do not perceive it as an obstacle to use EHR portals [32,47].

Figure 3. Structural model results. Note: path coefficients that are not statistically significant are in dashed arrows.



In our research model, social influence did not show a statistically significant effect on behavioral intention, thus not supporting *H3*. Previous studies have shown potential differences, with results differing among countries, with its positive significance being more consistent in the United States [2,7,27,28]. Potential cultural differences may explain the different behaviors. In our study, our early adopters of EHR portals seem to be more driven by their own individual willingness to try a new technology than to be influenced by what the society generally does. This is also an assumption supported by DOI theory [47]. The nonconfirmation of the facilitating conditions hypothesis, *H4(a)* and *H4(b)*, advocates that the individuals in our study believe that the resources or know-how to use EHR portals are not an issue. This can be justified by the ability of having access to a computer and the internet and is aligned with recent literature findings [2,44]. In Europe, access to most of the eHealth services is free of charge; so, the value that is provided to the patients is to permit them to execute specific activities more efficiently online. Unfortunately, that fact seems to be not acknowledged by the patients and *H5* was rejected. Habit has a statistically significant impact on both behavior intention and use behavior, supporting both *H6(a)* and *H6(b)*. Habit is a consumer-specific construct with a very significant role in our model, showing how important it is to have models tailored with consumer-specific constructs and not just general IT adoption constructs [26], and it is also supported by recent literature findings [26,27,44].

Self-perception has a statistically significant impact on use behavior, supporting *H7(b)*, and a nonsignificant impact on behavior intention, not supporting *H7(a)*. Often with sensitive topics and particularly with health-related topics, mismatch between intentions and effective actions occur [4,27,46]. In fact, this is the case with self-perception. Although it does not drive the intentions, self-perception directly influences actions in the usage of EHR portals. Results demonstrability has a statistically significant impact on effort expectancy, supporting *H8(c)*, and a nonsignificant impact on both performance expectancy *H8(b)* and behavior intention *H8(a)*, not supporting these 2 last hypotheses. Our results point out that when an innovation produces results that are readily discernible, perceptions of how easy it is to use a technology are considerably affected (this finding is in accordance with the literature [32]), but not the perceptions related with performance expectancy or a direct influence on behavior intention. Compared with results demonstrability and also from DOI, compatibility has a much greater effect in our research model demonstrated not only by the f^2 but also by having all its paths in the model statistically significant. Compatibility has a statistically significant impact on behavior intention *H9(a)*, performance expectancy *H9(b)*, and effort expectancy *H9(c)*, supporting these 3 hypotheses. The results indicate that behavior intention *H9(a)*, performance expectancy *H9(b)*, and effort expectancy *H9(c)* are greater when the heath care consumer perceives the technology to be compatible. Our study's results are in line with other studies in

this regard [13,32]. Behavioral intention positively influences use behavior, supporting *H10(a)*. This finding is in accordance with the literature suggesting that using EHR portals and eHealth tools is preceded by the intention to use them [23,26,27,44]. Behavioral intention also positively influences intention to recommend, supporting *H10(b)*. Our model explains 69.0% of the variance in recommendation, and the findings validate the significant influence of behavioral intention over it. Nevertheless, use behavior does not have a significant impact on intention to recommend, not supporting *H11*. A probable explanation might be that being a high user does not necessarily link to higher recommendation, but that a strong intention to use, independent of the usage level, is a stronger predictor of intention to recommend.

Managerial Implications

The study identifies areas that may influence EHR portal adoption, regarding its conceptualization, implementation, and redesign. Performance expectancy is a significant adoption driver of EHR portals. Therefore, while conceiving and promoting EHR portals, it is relevant to emphasize the advantages that they provide to the users in managing their health-related activities more efficiently. It is also important when conceiving an EHR portal that results are easily demonstrable because perceptions of how easy a technology is to use are affected by them. Compatibility is a very important construct in our model, and it is important to develop EHR portals that fit the health care customers' lifestyle. A good example is the providers that are already developing mobile versions of their EHR portals, allowing people to access their data everywhere [7]. In addition to the automatic and direct effect of habit on usage, habit also operates as a stored intention path to influence behavior [26]. This requires more communication effort to reinforce both the stored intention and its link to behavior [26]. As habit has been defined as the degree to which individuals tend to execute behaviors automatically due to learning [26], it is advisable that EHR portals have customer support services to help and provide support to the users with the platform.

Another relevant outcome is that the construct that is specific to health care—self-perception—also has a statistically significant role on the EHR portals usage. Self-perception is linked to the fact that the perceived, rather than the real, severity of the health problem is the driving force behind the action [30]. Health care interventions that enable the patient to be more conscious of her or his health condition may also endorse the usage of the EHR portal. In addition, the inclusion of educational health materials in the EHR portals may encourage patients to use the platform. Another important contribution of our study is to be able to demonstrate the influence of the intention to recommend in the adoption of EHR portals. Social network marketing and the opinions shared by friends and relatives are influential ways to help in the promotion and successful adoption of EHR portals. The managerial implications stated here are relevant not only for enhancing the adoption of EHR portals but also for growing the usage frequency of current users. These can be done by developing new EHR portals or by making improvements to existing ones.

Limitations and Future Research

Unfortunately, our study had a very high nonresponse rate concerning people that refused to answer the main questionnaire. With this high nonresponse rate, it is difficult to make direct assumptions related with the users in the Portuguese population. Nevertheless, earlier literature indicates that users and early users of eHealth tools and EHR portals are younger and more educated than the population average [2,7,13,27,38,61], in line with our study findings. The use of SEM is usually linked with the need of having questionnaires that are not short, making it more difficult for people to answer this questionnaire, especially by phone [52,54,73]. The use of gifts and other incentives may be a useful strategy to overcome the issue of the high nonresponse rate [26]. Testing the research model with samples of EHR users from other countries may also be an interesting path to follow, as the literature has shown that multicountry assessment provides interesting and diverse insights [15,22,27]. We used PLS-SEM instead of covariance based-SEM for the following reasons [54,73]: we have a complex model (many constructs and many indicators), we had the goal of identifying key *driver* constructs, and we also verified that our data were non-normally distributed. We acknowledge that future research may go in the direction of using covariance based-SEM, which allows using global goodness-of-fit criteria, but due to the circumstances and the study goals, we adopted PLS-SEM in our research [54,73].

Conclusions

Although we acknowledge that we had a very high nonresponse rate in the second stage of our sampling procedure, the much lower nonresponse rate in the first stage provides an estimate of 8.6% usage of these types of platforms in Portugal, a valuable contribution from our study. Our respondents' demographics follow the same trend as reported in other similar studies in the literature [13,38,61], providing additional support to our findings. Overall, the use of the 3 theories, UTAUT2, HBM, and DOI, to support our research was a successful strategy because in all of them, we had constructs with statistically significant impact on explaining the adoption of EHR portals. We were also able to demonstrate that consumers with a greater intention to adopt a new technology are more likely to become users and to recommend that specific technology to others. The new research model obtained very good results, with relevant R^2 in the most important dependent variables that help to explain the adoption of EHR portals, behavior intention (76.0%), and use behavior (61.8%).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questionnaire items.

[PDF File (Adobe PDF File), 103KB - jmir_v20i11e11032_app1.pdf]

Multimedia Appendix 2

Partial least squares loadings and cross-loadings plus CI for heterotrait-monotrait ratio.

[PDF File (Adobe PDF File), 139KB - jmir_v20i11e11032_app2.pdf]

References

1. 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]
2. Tavares J, Oliveira T. Electronic health record patient portal adoption by health care consumers: an acceptance model and survey. J Med Internet Res 2016 Mar;18(3):e49 [FREE Full text] [doi: 10.2196/jmir.5069] [Medline: 26935646]
3. Tavares J, Oliveira T. Electronic health record portals definition and usage. In: Cruz Cunha MM, Miranda IM, Martinho R, Rijo R, editors. Encyclopedia of E-Health and Telemedicine. Hershey, PA, USA: IGI Global; 2016:555-562.
4. Angst CM, Agarwal R. Adoption of electronic health records in the presence of privacy concerns: the elaboration likelihood model and individual persuasion. MIS Q 2009;33(2):339-370 [FREE Full text]
5. Kern LM, Edwards A, Kaushal R, HITEC Investigators. The meaningful use of electronic health records and health care quality. Am J Med Qual 2015 Aug;30(6):512-519. [doi: 10.1177/1062860614546547] [Medline: 25122006]
6. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med 2010 Aug 5;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]
7. Tavares J, Goulão A, Oliveira T. Electronic health record portals adoption: empirical model based on UTAUT2. Inform Health Soc Care 2018 Mar;43(2):109-125. [doi: 10.1080/17538157.2017.1363759] [Medline: 29035646]
8. Alpay LL, Henkemans OB, Otten W, Rövekamp TA, Dumay AC. E-health applications and services for patient empowerment: directions for best practices in The Netherlands. Telemed J E Health 2010 Sep;16(7):787-791. [doi: 10.1089/tmj.2009.0156] [Medline: 20815745]
9. McKee M, Karanikolos M, Belcher P, Stuckler D. Austerity: a failed experiment on the people of Europe. Clin Med (Lond) 2012 Aug;12(4):346-350. [Medline: 22930881]
10. Ancker JS, Brenner S, Richardson JE, Silver M, Kaushal R. Trends in public perceptions of electronic health records during early years of meaningful use. Am J Manag Care 2015 Aug;21(8):e487-e493 [FREE Full text] [Medline: 26625503]
11. Ancker JS, Osorio SN, Cheriff A, Cole CL, Silver M, Kaushal R. Patient activation and use of an electronic patient portal. Inform Health Soc Care 2015 Apr;40(3):254-266. [doi: 10.3109/17538157.2014.908200] [Medline: 24786648]
12. Smith SG, O'Conor R, Aitken W, Curtis LM, Wolf MS, Goel MS. Disparities in registration and use of an online patient portal among older adults: findings from the LitCog cohort. J Am Med Inform Assoc 2015 Jul;22(4):888-895 [FREE Full text] [doi: 10.1093/jamia/ocv025] [Medline: 25914099]
13. Zhang X, Yu P, Yan J, Ton AM Spil I. 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]
14. Gheorghiu B, Hagens S. Use and maturity of electronic patient portals. Stud Health Technol Inform 2017;234:136-141. [Medline: 28186030]
15. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. BMC Public Health 2007;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]
16. Weingart SN, Rind D, Tofias Z, Sands DZ. Who uses the patient internet portal? The PatientSite experience. J Am Med Inform Assoc 2006 Jan;13(1):91-95 [FREE Full text] [doi: 10.1197/jamia.M1833] [Medline: 16221943]
17. Daim TU, Behkami NA, Basoglu N, Kök OM, Hogaboam L. Healthcare Technology Innovation Adoption: Electronic Health Records and Other Emerging Health Information Technology Innovations. New York: Springer; 2016.
18. Chang I, Hwang H, Hung W, Li Y. Physicians' acceptance of pharmacokinetics-based clinical decision support systems. Expert Syst Appl 2007 Aug;33(2):296-303. [doi: 10.1016/j.eswa.2006.05.001]
19. Vanneste D, Vermeulen B, Declercq A. Healthcare professionals' acceptance of BelRAI, a web-based system enabling person-centred recording and data sharing across care settings with interRAI instruments: a UTAUT analysis. BMC Med Inform Decis Mak 2013 Nov 27;13:129 [FREE Full text] [doi: 10.1186/1472-6947-13-129] [Medline: 24279650]
20. Dünnebeil S, Sunyaev A, Blohm I, Leimeister JM, Krcmar H. Determinants of physicians' technology acceptance for e-health in ambulatory care. Int J Med Inform 2012 Nov;81(11):746-760. [doi: 10.1016/j.ijmedinf.2012.02.002] [Medline: 22397989]

21. Ketikidis P, Dimitrovski T, Lazuras L, Bath PA. Acceptance of health information technology in health professionals: an application of the revised technology acceptance model. Health Informatics J 2012 Jun;18(2):124-134. [doi: 10.1177/1460458211435425] [Medline: 22733680]
22. Hoque MR, Bao YK, Sorwar G. Investigating factors influencing the adoption of e-Health in developing countries: a patient's perspective. Inform Health Soc Care 2017 Jan;42(1):1-17. [doi: 10.3109/17538157.2015.1075541] [Medline: 26865037]
23. Kim J, Park H. Development of a health information technology acceptance model using consumers' health behavior intention. J Med Internet Res 2012 Oct;14(5):e133 [FREE Full text] [doi: 10.2196/jmir.2143] [Medline: 23026508]
24. Lemire M, Paré G, Sicotte C, Harvey C. Determinants of internet use as a preferred source of information on personal health. Int J Med Inform 2008 Nov;77(11):723-734. [doi: 10.1016/j.ijmedinf.2008.03.002] [Medline: 18434246]
25. Wilson EV, Lankton NK. Modeling patients' acceptance of provider-delivered e-health. J Am Med Inform Assoc 2004 Apr;11(4):241-248 [FREE Full text] [doi: 10.1197/jamia.M1475] [Medline: 15064290]
26. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the Unified Theory of Acceptance and Use of Technology. MIS Q 2012;36(1):157-178 [FREE Full text]
27. Tavares J, Oliveira T. Electronic health record portal adoption: a cross country analysis. BMC Med Inform Decis Mak 2017 Jul 05;17(1):97 [FREE Full text] [doi: 10.1186/s12911-017-0482-9] [Medline: 28679423]
28. Bozan K, Davey B, Parker K. Social influence on health IT adoption patterns of the elderly: an institutional theory based use behavior approach. Procedia Comput Sci 2015;63:517-523. [doi: 10.1016/j.procs.2015.08.378]
29. Ahadzadeh AS, Pahlevan SS, Ong FS, Khong KW. Integrating health belief model and technology acceptance model: an investigation of health-related internet use. J Med Internet Res 2015 Feb;17(2):e45 [FREE Full text] [doi: 10.2196/jmir.3564] [Medline: 25700481]
30. van de Kar A, Knottnerus A, Meertens R, Dubois V, Kok G. Why do patients consult the general practitioner? Determinants of their decision. Br J Gen Pract 1992 Aug;42(361):313-316 [FREE Full text] [Medline: 1457150]
31. Nøhr C, Parv L, Kink P, Cummings E, Almond H, Nørgaard JR, et al. Nationwide citizen access to their health data: analysing and comparing experiences in Denmark, Estonia and Australia. BMC Health Serv Res 2017 Aug 07;17(1):534 [FREE Full text] [doi: 10.1186/s12913-017-2482-y] [Medline: 28784173]
32. Yi MY, Jackson JD, Park JS, Probst JC. Understanding information technology acceptance by individual professionals: toward an integrative view. Inf Manag 2006 Apr;43(3):350-363. [doi: 10.1016/j.im.2005.08.006]
33. Oliveira T, Thomas M, Baptista G, Campos F. Mobile payment: understanding the determinants of customer adoption and intention to recommend the technology. Comput Hum Behav 2016 Aug;61:404-414. [doi: 10.1016/j.chb.2016.03.030]
34. Chan KM, Pang WS, Ee CH, Ding YY, Choo P. Self-perception of health among elderly community dwellers in Singapore. Ann Acad Med Singapore 1998 Jul;27(4):461-467. [Medline: 9791647]
35. Kaleta D, Polańska K, Dziankowska-Zaborszczyk E, Hanke W, Drygas W. Factors influencing self-perception of health status. Cent Eur J Public Health 2009 Sep;17(3):122-127 [FREE Full text] [Medline: 20020600]
36. Moore GC, Benbasat I. Development of an instrument to measure the perceptions of adopting an information technology innovation. Inf Syst Res 1991 Sep;2(3):192-222. [doi: 10.1287/isre.2.3.192]
37. Lee CJ, Gray SW, Lewis N. Internet use leads cancer patients to be active health care consumers. Patient Educ Couns 2010 Dec;81 Suppl:S63-S69 [FREE Full text] [doi: 10.1016/j.pec.2010.09.004] [Medline: 20889279]
38. Or CK, Karsh BT. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009 Jul;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]
39. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):157-178. [doi: 10.2307/30036540]
40. Fisher J, Clayton M. Who gives a tweet: assessing patients' interest in the use of social media for health care. Worldviews Evid Based Nurs 2012 Apr;9(2):100-108. [doi: 10.1111/j.1741-6787.2012.00243.x] [Medline: 22432730]
41. Thackeray R, Crookston BT, West JH. Correlates of health-related social media use among adults. J Med Internet Res 2013 Jan 30;15(1):e21 [FREE Full text] [doi: 10.2196/jmir.2297] [Medline: 23367505]
42. Keselman A, Logan R, Smith CA, Leroy G, Zeng-Treitler Q. Developing informatics tools and strategies for consumer-centered health communication. J Am Med Inform Assoc 2008 Jul;15(4):473-483 [FREE Full text] [doi: 10.1197/jamia.M2744] [Medline: 18436895]
43. Peek ST, Wouters EJ, van Hoof J, Luijkx KG, Boeije HR, Vrijhoef HJ. Factors influencing acceptance of technology for aging in place: a systematic review. Int J Med Inform 2014 Apr;83(4):235-248. [doi: 10.1016/j.ijmedinf.2014.01.004]
44. Yuan S, Ma W, Kanthawala S, Peng W. Keep using my health apps: discover users' perception of health and fitness apps with the UTAUT2 model. Telemed J E Health 2015 Sep;21(9):735-741. [doi: 10.1089/tmj.2014.0148] [Medline: 25919238]
45. Menec VH, Chipperfield JG, Perry RP. Self-perceptions of health: a prospective analysis of mortality, control, and health. J Gerontol B Psychol Sci Soc Sci 1999 Mar;54(2):85-93. [Medline: 10097770]
46. Baumgartner H. The Handbook of Marketing Research: Uses, Misuses, and Future Advances. Thousand Oaks, California: SAGE Publications, Inc; 2006.
47. Rogers EM. Diffusion of Innovations, 5th ed. New York: The Free Press/ Simon & Schuster, Inc; 2003.

48. Miltgen CL, Popovič A, Oliveira T. Determinants of end-user acceptance of biometrics: integrating the "Big 3" of technology acceptance with privacy context. Decision Support Systems 2013 Dec;56:103-114. [doi: 10.1016/j.dss.2013.05.010]
49. Brislin RW. Back-translation for cross-cultural research. J Cross Cult Psychol 1970 Sep 01;1(3):185-216. [doi: 10.1177/135910457000100301]
50. Gao SJ, Hui SL, Hall KS, Hendrie HC. Estimating disease prevalence from two-phase surveys with non-response at the second phase. Stat Med 2000 Aug 30;19(16):2101-2114 [FREE Full text] [Medline: 10931514]
51. Anacom.: O Sector das Comunicações '16. Edited by Comunicações ANd; 2016. The Communications Sector 2016 URL: https://www.anacom.pt/render.jsp?contentId=1411094&languageId=1 [accessed 2017-05-05] [WebCite Cache ID 6yvdbi6Wj]
52. Vicente P, Reis E. Telephone surveys using mobile phones: an analysis of response rates, survey procedures and respondents' characteristics. Aust J Mark Soc Res 2009;17(2):49-56.
53. Aanerud M, Braut H, Wentzel-Larsen T, Eagan TM, Bakke PS. Non-response in telephone surveys of COPD patients does not introduce bias. J Telemed Telecare 2013 Jan;19(1):40-44. [doi: 10.1177/1357633X12474960] [Medline: 23449960]
54. Hair J, Hult GT, Ringle C, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM), Second edition. Thousand Oaks: SAGE Publications, Inc; 2017.
55. Hair JF, Sarstedt M, Ringle CM, Mena JA. An assessment of the use of partial least squares structural equation modeling in marketing research. J Acad Mark Sci 2012;40(3):414-433. [doi: 10.1007/s11747-011-0261-6]
56. Ringle CM, Sarstedt M, Straub DW. A critical look at the use of PLS-SEM in MIS Quarterly. MIS Q 2012;36(1):3-14. [doi: 10.1108/IMR-04-2014-0148]
57. Ringle CM, Wende S, Becker JM. Smartpls. SmartPLS 3 URL: https://www.smartpls.com/downloads [accessed 2018-04-24] [WebCite Cache ID 6yvcS6KNb]
58. Censo.ine. Censos 2011 URL: http://censos.ine.pt/xportal/xmain?xpid=CENSOS&xpgid=censos_quadros_populacao [WebCite Cache ID 6yvch9RVO]
59. Pordata. Thousands of Statistics about Municipalities, Portugal and Europe URL: https://www.pordata.pt/en/Home [accessed 2018-04-24] [WebCite Cache ID 6yvcsiqUn]
60. Mateus A, Ramalho E, Oliveira H, Rodrigues H, Ferreira R. Aphp-pt. Sector Privado da Saúde em Portugal 2017 URL: http://www.aphp-pt.org/pdf/Estudo-Sector_Privado_da_Sa%C3%BAde_em_Portugal.pdf [accessed 2018-04-24] [WebCite Cache ID 6yvc3pmLm]
61. Ronda MC, Dijkhorst-Oei LT, Gorter KJ, Beulens JW, Rutten G. Differences between diabetes patients who are interested or not in the use of a patient Web portal. Diabetes Technol Ther 2013 Jul;15(7):556-563 [FREE Full text] [doi: 10.1089/dia.2013.0023] [Medline: 23777369]
62. Böhm K, Schmid A, Götze R, Landwehr C, Rothgang H. Five types of OECD healthcare systems: empirical results of a deductive classification. Health Policy 2013 Dec;113(3):258-269. [doi: 10.1016/j.healthpol.2013.09.003] [Medline: 24095274]
63. 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]
64. Podsakoff PM, MacKenzie SB, Lee J, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 2003 Oct;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]
65. Lindell MK, Whitney DJ. Accounting for common method variance in cross-sectional research designs. J Appl Psychol 2001 Feb;86(1):114-121. [Medline: 11302223]
66. Malhotra NK, Kim SS, Patil A. Common method variance in IS Research: a comparison of alternative approaches and a reanalysis of past research. Manage Sci 2006 Dec;52(12):1865-1883. [doi: 10.1287/mnsc.1060.0597]
67. Altman DG, Bland JM. Missing data. Br Med J 2007 Feb 24;334(7590):424 [FREE Full text] [doi: 10.1136/bmj.38977.682025.2C] [Medline: 17322261]
68. Powney M, Williamson P, Kirkham J, Kolamunnage-Dona R. A review of the handling of missing longitudinal outcome data in clinical trials. Trials 2014 Jun 19;15:237 [FREE Full text] [doi: 10.1186/1745-6215-15-237] [Medline: 24947664]
69. Evans SJ. Good surveys guide. Br Med J 1991 Feb 09;302(6772):302-303 [FREE Full text] [Medline: 2001503]
70. Kalton G, Anderson DW. Sampling rare populations. J R Stat Soc Ser A 1986;149(1):65-82. [doi: 10.2307/2981886]
71. Picot SJ, Samonte J, Tierney JA, Connor J, Powel LL. Effective sampling of rare population elements: black female caregivers and noncaregivers. Res Aging 2001 Nov 01;23(6):694-712. [doi: 10.1177/0164027501236004]
72. Osborn CY, Mayberry LS, Wallston KA, Johnson KB, Elasy TA. Understanding patient portal use: implications for medication management. J Med Internet Res 2013 Jul 03;15(7):e133 [FREE Full text] [doi: 10.2196/jmir.2589] [Medline: 23823974]
73. Hair JF, Ringle CM, Sarstedt M. PLS-SEM: indeed a silver bullet. J Mark Theory Pract 2011;19(2):139-151. [doi: 10.2753/MTP1069-6679190202]

Abbreviations

AVE: average variance extracted
CFIP: concern for information privacy
DOI: diffusion of innovation
EHR: electronic health record
epSOS: European patients smart open services
HBM: health belief model
HTMT: heterotrait-monotrait ratio
IT: information technology
NHS: National Health Service
PLS-SEM: partial least squares structural equation modeling
TAM: technology acceptance model
UTAUT: unified theory of acceptance and use of technology
UTAUT2: extended unified theory of acceptance and use of technology

Edited by G Eysenbach; submitted 13.05.18; peer-reviewed by S Dias, H Dehghan; comments to author 04.07.18; revised version received 17.07.18; accepted 17.07.18; published 19.11.18.

Please cite as:
Tavares J, Oliveira T
New Integrated Model Approach to Understand the Factors That Drive Electronic Health Record Portal Adoption: Cross-Sectional National Survey
J Med Internet Res 2018;20(11):e11032
URL: https://www.jmir.org/2018/11/e11032/
doi:10.2196/11032
PMID:30455169

Original Paper

Automated Extraction of Diagnostic Criteria From Electronic Health Records for Autism Spectrum Disorders: Development, Evaluation, and Application

Gondy Leroy[1], PhD; Yang Gu[1], BComm; Sydney Pettygrove[1], PhD; Maureen K Galindo[1], RN, MS; Ananyaa Arora[1], MS; Margaret Kurzius-Spencer[1], PhD

University of Arizona, Tucson, AZ, United States

Corresponding Author:
Gondy Leroy, PhD
University of Arizona
1130 E Helen Street
McClelland Hall
Tucson, AZ, 85721
United States
Phone: 1 5206214106
Email: gondyleroy@email.arizona.edu

Abstract

Background: Electronic health records (EHRs) bring many opportunities for information utilization. One such use is the surveillance conducted by the Centers for Disease Control and Prevention to track cases of autism spectrum disorder (ASD). This process currently comprises manual collection and review of EHRs of 4- and 8-year old children in 11 US states for the presence of ASD criteria. The work is time-consuming and expensive.

Objective: Our objective was to automatically extract from EHRs the description of behaviors noted by the clinicians in evidence of the diagnostic criteria in the Diagnostic and Statistical Manual of Mental Disorders (DSM). Previously, we reported on the classification of entire EHRs as ASD or not. In this work, we focus on the extraction of individual expressions of the different ASD criteria in the text. We intend to facilitate large-scale surveillance efforts for ASD and support analysis of changes over time as well as enable integration with other relevant data.

Methods: We developed a natural language processing (NLP) parser to extract expressions of 12 DSM criteria using 104 patterns and 92 lexicons (1787 terms). The parser is rule-based to enable precise extraction of the entities from the text. The entities themselves are encompassed in the EHRs as very diverse expressions of the diagnostic criteria written by different people at different times (clinicians, speech pathologists, among others). Due to the sparsity of the data, a rule-based approach is best suited until larger datasets can be generated for machine learning algorithms.

Results: We evaluated our rule-based parser and compared it with a machine learning baseline (decision tree). Using a test set of 6636 sentences (50 EHRs), we found that our parser achieved 76% precision, 43% recall (ie, sensitivity), and >99% specificity for criterion extraction. The performance was better for the rule-based approach than for the machine learning baseline (60% precision and 30% recall). For some individual criteria, precision was as high as 97% and recall 57%. Since precision was very high, we were assured that criteria were rarely assigned incorrectly, and our numbers presented a lower bound of their presence in EHRs. We then conducted a case study and parsed 4480 new EHRs covering 10 years of surveillance records from the Arizona Developmental Disabilities Surveillance Program. The social criteria (A1 criteria) showed the biggest change over the years. The communication criteria (A2 criteria) did not distinguish the ASD from the non-ASD records. Among behaviors and interests criteria (A3 criteria), 1 (A3b) was present with much greater frequency in the ASD than in the non-ASD EHRs.

Conclusions: Our results demonstrate that NLP can support large-scale analysis useful for ASD surveillance and research. In the future, we intend to facilitate detailed analysis and integration of national datasets.

(J Med Internet Res 2018;20(11):e10497) doi:10.2196/10497

KEYWORDS

parser; natural language processing; complex entity extraction; Autism Spectrum Disorder; DSM; electronic health records; decision tree; machine learning

Introduction

Based on data from autism spectrum disorder (ASD) surveillance, it is estimated that the prevalence of ASD is approximately 1.5% [1]. In the second half of the 20th century, it was estimated at slightly more than 5 cases per 10,000 people. Since the 1990s, however, measured prevalence has increased [2]. In 2000, prevalence estimates ranged from 4.5 to 9.9 cases per 1000 children and increased to 1 in 110 children in 2006 [3] and 1 in 59 in 2014 [4]. The reasons for this trend are uncertain, but the following factors have been proposed: increased public awareness, changing diagnostic criteria, and substitution of ASD eligibility for other special education eligibilities as well as the possibility that the true prevalence of ASD is increasing [3,4].

Data on long-term trends, symptoms, diagnoses, evaluations, and treatments are critical for planning interventions and educational and health services. To understand and act upon such trends, large-scale studies are needed that can evaluate trends over time, integrate different types of data, and review large datasets. In recent years, data have been increasingly electronically encoded in electronic health records (EHRs) in structured fields and free text. Collection of such EHRs enables analyses that compare and contrast ASD prevalence in relation to other variables and over time.

Much of the published work on ASD leverages information in the structured fields of the EHRs such as gender, medication taken by the mother, birth complications, scores on a variety of tests, and others. The structured data portions are relatively easy to extract and are useful for large-scale studies. However, the results of the analysis are commonly limited to reviews and counts of the presence of conditions in certain populations [5]. For example, Clements et al [6] evaluated the relationship between autism and maternal use of prenatal antidepressants using EHRs.

EHR of people with ASD contain extensive free text fields with important information that is often complementary to and more detailed and explanatory than the structured data. This is because in the absence of any biological laboratory test, diagnosis is generally made in person using specific test instruments, history, and differential diagnosis, and much of this information is recorded as narrative. Automatically extracting this information from the EHRs requires natural language processing (NLP). So far, a few NLP approaches have been used to analyze language generated by people on the spectrum [7,8], but there has been little focus on the text in EHRs.

The existing projects that focus on the text in EHRs fall into two groups. The first group focuses on using all the free text combined with structured fields to automatically assign case status (classification of patients as cases of autism or not) to an entire record. A variety of machine learning algorithms are useful for this task. Using a subset of the EHRs for training, these algorithms create a model that can be applied to future EHRs to assign case status. These models can be human-interpretable, such as decision trees, or can be black box approaches, such as neural networks. In our own work [9], we compared decision trees (C5.0) and a feedforward backpropagation artificial neural network using only the information contained in the free text. Our best approach used the decision tree and was 87% accurate in case assignment. Similarly, promising results (86% accuracy) were achieved by Maenner et al [10] using a random forest algorithm. Lingren et al [11] used International Classification of Diseases (ICD)-9 codes combined with concepts extracted from the free text and compared rule-based and machine learning algorithms with similarly good results.

In addition to case status assignment, more detailed use of the information contained in the free text would be helpful for large-scale analysis, for example, cultural comparisons as suggested by Mandy et al [12], and for combination with other data. To extract this information automatically, a comprehensive set of tools is needed [13,14] for standard NLP tasks, such as part-of-speech (POS) tagging and grammatical parsing. For more specialized tasks, such as concept detection, entity and relation extraction, and coreference resolution, we work on entity extraction algorithms. In medicine, existing entity extraction algorithms focus on different types of text (eg, published research abstracts or clinical narrative) and can be rule-based or use machine learning techniques. The entities themselves have been predominantly single terms or relationships composed of single terms. For example, several projects focus on annotating diseases or genes and proteins [15-18] and the biomedical relationships between them [19,20]. When working with free text from EHRs, a variety of entities have been the focus. For example, NLP for safety surveillance by extracting information on postoperative complications [21] and adverse drug effects from psychiatric records [22], clinical event detection (eg, fever, change in output) for transcriptions of the handoff communication between nurses during shift changes [23,24], and even the creation of new data such as veterans' employment information [25].

In this project, we aimed to extract the expressions of the behaviors indicative of individual diagnostic criteria as described in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR) [26]. Such expressions are more complex than single terms, and each diagnostic criterion can be expressed by a variety of different behaviors described in a diverse manner in the text. We envision that our parser will be useful for two types of work. First, it will enable autism surveillance to speed up and increase the scope when processing school and health records. Currently, the Centers for Disease Control and Prevention (CDC) surveillance effort is limited to 11 states and a subset of schools, the catchment area, in these states. By automating the review of records, an efficient approach will allow all records to be reviewed, many case decisions will be automated, and only a subset of records will require review by experienced clinicians. A second important use that we foresee is in the creation of large sets of structured data from the information available in the EHRs. This will facilitate secondary analysis of data on its own and in combination with other data sources at a previously unseen scale.

We first describe the development and evaluation of the parser, including important decisions on using off-the-shelf tools and machine learning algorithms. Then, we present a case study

where we applied our parser to almost 5000 EHRs to demonstrate usefulness for detailed analysis over time.

Methods

Overview

Our parser uses human-interpretable rules to match complex patterns that represent the DSM diagnostic criteria. These rule-based algorithms rely on the creation of patterns of terms, grammatical relationships, and the surrounding text to recognize the entities of interest in text.

Records and the Diagnostic and Statistical Manual of Mental Disorders Criteria

We work with EHRs created by the Arizona Developmental Disabilities Surveillance Program (ADDSP) as part of the CDC multicenter Autism and Developmental Disability Monitoring Network surveillance. Our ADDSP records are collected from educational and clinical data sources in 11-15 school districts for 8-year olds. From 2000 to 2010, a total of 27,515 records were reviewed and 6176 records were abstracted that included any of the 32 social behavioral triggers consistent with ASD as listed in the Abstraction Manual developed by the CDC. These records referred to 4491 children. The identified records for each child were further evaluated by trained clinical reviewers who applied standardized criteria to highlight criteria and determine ASD case status. This yielded 2312 confirmed cases.

We have access to the records and the case status of each child as determined through expert review of the information. For this study, we leveraged a subset of these records (n=93) that have been printed and have the diagnostics criteria annotated on this paper version. The electronic version does not include markings indicating the criteria. Therefore, we first created an electronic gold standard with all information combined. Records were loaded using WebAnno [27], and the annotations made by clinicians on the paper versions were added to the electronic versions. In 1 hour, 1-3 records could be annotated depending on the length of the record.

We intend to automate the extraction of the DSM-IV-TR [26] criteria for ASD. Textbox 1 shows example criteria rules. The DSM specifies the combination of criteria needed to assign ASD case status. While other instruments exist for diagnosing autism, as well as different versions of the DSM, we focus on matching to the DSM-IV-TR because this is an approach that is used worldwide and that is sometimes used for matching to billing codes (ICD-9 and ICD-10). It is also available with a large set of records for training and testing. Later, we will work with the DSM-V, the newest version.

Textbox 1. Example rules in Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision, to diagnose autistic disorder.

Rules

A: A total of 6 or more items from (1), (2), and (3), with at least 2 from (1) and 1 each from (2) and (3):

- 1: Qualitative impairment in social interaction, as manifested by at least 2 of the following:
  - A1a: Marked impairment in the use of multiple nonverbal behaviors such as eye-to-eye gaze, facial expression, body postures, and gestures to regulate social interaction
  - A1b: Failure to develop peer relationships appropriate to developmental level
  - A1c: A lack of spontaneous seeking to share enjoyment, interests, or achievements with other people
  - A1d: Lack of social or emotional reciprocity
- 2: Qualitative impairments in communication as manifested by at least 1 of the following:
  - A2a: Delay in, or total lack of, the development of spoken language (not accompanied by an attempt to compensate through alternative modes of communication such as gesture or mime)
  - A2b: In individuals with adequate speech, marked impairment in the ability to initiate or sustain a conversation with others
  - A2c: Stereotyped and repetitive use of language or idiosyncratic language
  - A2d: Lack of varied, spontaneous make-believe play or social imitative play appropriate to developmental level
- 3: Restrictive, repetitive, and stereotyped patterns of behaviors, interests, and activities, as manifested by at least 1 of the following:
  - A3a: Encompassing preoccupation with 1 or more stereotyped and restricted patterns of interest that is abnormal either in intensity or focus
  - A3b: Apparently inflexible adherence to specific, nonfunctional routines or rituals
  - A3c: Stereotyped and repetitive motor mannerisms
  - A3d: Persistent preoccupation with parts of objects

Design Choices

Component Selection Rationale

To our knowledge, no parsers exist that identify DSM criteria in EHRs. As part of our development, we evaluated MetaMap [28,29] for use as an off-the-shelf building block. MetaMap maps terms to the Unified Medical Language System (UMLS) Metathesaurus concepts and semantic types.

Using 2 EHRs from our development set, we analyzed MetaMap's outcome in the context of ASD. From the 2 EHRs, a total of 259 phrases were extracted and mapped to 632 UMLS concepts. Overall, 46.5% (294/632) of all candidate mappings for those phrases were correct and useful for our domain; 55.2% (143/259) of phrases were given a single candidate mapping to UMLS concepts, and for those single matches, the accuracy was high, with 81.1% (116/143) correct and useful matches for our domain. However, when the number of matched semantic types increases, it becomes increasingly complicated to identify the correct concept and associated semantic type. Furthermore, the majority of semantic types do not apply to our domain. Using a very lenient approach, we consider approximately 31 semantic types useful to match to DSM criteria (eg, Activity, Anatomical Structure, Behavior, Body Part, Organ or Organ Component, Body Substance, Clinical Attribute, Conceptual Entity, and Daily or Recreational Activity, among others). Although the 259 phrases we analyzed are restricted to 31 relevant semantic types, this is not enough to distinguish ASD diagnostic criteria from rest of the text: only 27.0% (70/259) phrases intersect with ASD diagnostic criteria. Because the number of types that are immediately useful is small and this MetaMap outcome would require significant development to adjust for our purpose, building an extraction system on top of this is impractical. Therefore, we decided to build all the components in-house.

Rule-based Versus Machine Learning Rationale

When developing a new entity extraction artifact, a rule-based or machine learning approach is chosen as the starting point. Both can be combined in ensemble methods later. We performed a baseline test using a decision tree, which was chosen because it is a human-interpretable machine learning algorithm.

We formulated the problem as a multiclass sentence classification problem (12 diagnostic labels or null label). We used Stanford CoreNLP (version 3.7.0) for NLP processing. We used a standard bag-of-words approach with and trained the algorithm on 120 records containing 19,428 sentences. Because our records contain approximately 0.5%-5% sentences describing a DSM criterion, we undersampled negative examples during training to improve recall: for each positive example, we sampled 30 negative samples (except criteria A2a and A2b, which occurred frequently enough to use on the entire training data). Our features were lemmas, as determined by CoreNLP, which appeared more than 5 times in the training data (2913 terms). We used a pruned decision tree (Weka version 3.8.0) with a pruning confidence threshold of 0.25. Size of the vocabulary, undersampling ratio, and pruning threshold were determined based on the best values we found during exploration.

Table 1 shows the results for classification at the sentence level. For comparison, we applied the model to the same EHR test set (not used during decision tree training) used in our parser evaluation below. Overall, *F*-score was <0.5. Neither precision nor recall was high using this approach.

This machine learning approach will require significant work to improve performance. We believe this cannot be attained with simple changes in the input, such as word embedding, or by changing algorithms. It will require more sophisticated features and a much larger dataset. We, therefore, first created a rule-based parser, which may provide better results overall as well as insights related to lexicons and features useful for future combinations with machine learning in a classifier ensemble.

Table 1. Decision tree evaluation for sentence classification.

Rule	Count of positive cases	% positive cases (of all sentences)	Precision	Recall	F-score	Specificity
A1a	120	0.021	0.70	0.52	0.59	0.99
A1b	91	0.016	0.50	0.42	0.45	0.99
A1c	35	0.006	0.16	0.17	0.17	0.99
A1d	160	0.029	0.54	0.14	0.22	1.00
A2a	388	0.069	0.71	0.39	0.50	1.00
A2b	321	0.057	0.69	0.37	0.48	0.99
A2c	120	0.021	0.54	0.47	0.51	0.99
A2d	62	0.011	0.34	0.19	0.25	1.00
A3a	64	0.011	0.20	0.09	0.13	1.00
A3b	123	0.022	0.81	0.47	0.59	1.00
A3c	66	0.012	0.70	0.32	0.44	1.00
A3d	27	0.005	0.27	0.30	0.28	1.00
Microaverage	1577	0.024	0.60	0.35	0.45	0.99

Table 2. Lexicon overview.

Pattern use of lexicons	Lexicons	Number of terms	Example lexicon	Example terms
All rules	11	345	Body_parts	arm, eye, hair, teeth, toe, tongue, finger, fingers, nose
Group A1	7	105	A1_interact	interact, interactions, communicate, relationship
Group A2	3	72	A2_positive	severe, significant, pervasive, marked
Group A3	2	72	A3_object	door, toys, vacuum, blocks, book, television, lights
A1a	4	42	A1a_nonVerbalBehavior	eye contact, eye-to-eye gaze, gestures, nonverbal cues
A1b	2	11	A1b_consistent	good, consistent, appropriately, satisfactory
A1c	5	61	A1c_affect	excitement, feelings, satisfaction, concerns
A1d	12	159	A1d_engage	recognize, recognizes, reacts, respond, regard, attend
A2a	4	117	A2a_gained	gained, used, had, obtained, said, spoke
A2b	8	240	A2b_recepLang	direction, instructions, questions, conversations
A2c	7	145	A2c_idiosyncratic	breathy, echolalia, jargon, neologism, reduced
A2d	7	83	A2d_actions	actions, routines, play, signs, gestures, movements
A3a	7	106	A3a_obsess	obsessed, obsessive, perseverates, preoccupation
A3b	7	119	A3b_nonFunctionalPlay	stack, stacks, lines, lined, nonfunctional, arrange
A3c	3	67	A3c_abnormal	grind, grinds, rocks, twirls, spin, tap, clap, flap
A3d	3	43	A3d_sensitive	defensiveness, sensitivity, hypersensitivities
Total	92	1787	N/A[a]	N/A

[a]N/A: not applicable.

Parser Development

We developed a rule-based parser to extract all A1, A2, and A3 rules as listed in the DSM. Each DSM group contains 4 specific rules that are representative of the criterion (A1a-d, A2a-d, and A3a-d). Our tool comprises 2 components: (1) annotation of relevant ASD trigger words in free text and (2) recognition of diagnostic criteria based on a pattern of trigger words.

The parser was developed through collaboration between NLP experts and clinicians. Annotations from EHRs were translated into patterns by NLP experts. Then, extensions, abstractions, and generalizations were discussed and the patterns augmented and expanded. This iterative process was continued until changes in patterns provided little or no improvement but increased error rates. Several development rounds were completed, and the EHRs were taken from the 2002 to 2010 surveillance years, with 53% of records having an ASD case status. The ASD label itself is of little consequence because both development and testing are done at the sentence level (not the record level). For testing, new EHRs were used that were not seen in previous development rounds. EHRs were selected randomly from those available to us.

Lexicons

Identifying ASD diagnostic criteria in text requires recognizing important trigger words (ie, words describing typical behaviors of ASD). We capture these words, as well as synonyms and singular or plural variants, in lexicons. Approximately 90 lexicons with about 20 terms each were manually created. Table 2 provides an overview with the examples of lexicons and the terms they contain. We used a lexical lookup for each term found in the text and annotated it with the lexicon's label. These labels form part of the patterns used to describe DSM criteria (see the following text). Multiple patterns are needed to capture the different free text expressions for each DSM criterion.

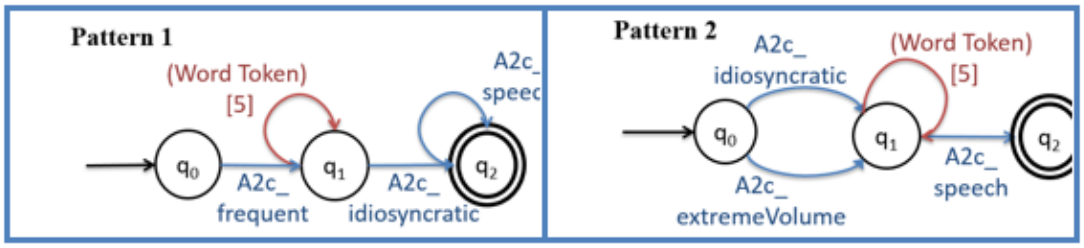
The lexicons are optimized for patterns for each DSM criterion, so the same terms may appear in multiple lexicons. However, a few lexicons are shared by all patterns and used for different DSM criteria. Currently, there are 11 lexicons shared by all patterns (eg, the lexicons containing body parts). In addition, the patterns for the A1, A2, and A3 criteria share, respectively, 7, 3, and 2 lexicons. For example, DSM rules A1a, A1b, A1c, and A1d all require identification of "impairment in social interaction," and the relevant terms for this trigger are combined in the lexicon "A1_interact." In addition to these shared patterns, each DSM pattern requires additional individual lexicons optimized for that pattern.

Pattern Extraction

We used the General Architecture for Text Engineering (GATE) [30,31], a Java-based developer environment, to process the free text from EHRs. We chose GATE because it includes several standard NLP tools as well as the availability of its Java Annotation Pattern Engine (JAPE) to efficiently annotate patterns over text. First, standard processing is applied to all text:

- Tokenizer: recognize individual tokens in the text.
- Sentence splitter: set boundaries on sentences so that parts of speech can be deduced for each word in a sentence.
- POS tagger [32]: assign POS to each word, eg, noun, verb, and adjective. We used the Stanford Tagger.

Figure 1. Visualization of 2 (of 7 existing) patterns for Diagnostic Manual of Mental Disorders criteria A2c.



After processing all free text, terms are annotated using gazetteer lookup. Using the term's POS tags and lexical labels from the 92 lexicons, the annotated text is processed to identify matching patterns. POS tags help narrow down candidate terms, for example, "object" fits in our lexicons when it is a noun but not when it is a verb. Using 43 annotated records from the ADDSP containing 4732 sentences, we developed 12 sets of patterns (total 104 patterns) for the 12 DSM criteria (see Textbox 1).

Figure 1 shows two example finite state automata visualizing patterns for criterion A2c. Because we have many patterns of varying complexity, these examples are included to convey the general principle. A pattern starts in the q_0 state, and when an appropriate input is presented, it proceeds to the next state. It is completed when a final state is reached. If no progress is possible, a sink state is reached and the process discarded. Each label on an arc (eg, A2c_speech) represents the lexicon of terms (terms indicating "speak" as relevant to rule A2c). For example, Pattern 1 would match the text "$[often]_{A2c\_frequent}$ $[speaks]_{word\ token}$ $[with]_{word\ token}$ $[reduced]_{A2c\_idiosyncratic}$ $[volume]_{A2c\_speech.}$" The word "often" matches the lexicon "A2c_frequent," along with words such as "frequently," to indicate that this behavior is a regular occurrence. This match enables the transition from q_0 to q_1. The phrase "speaks with" is accepted in state q_1, which accepts up to 5 word tokens that do not match other arcs transitioning out of this state. We empirically decided to allow 5 intervening terms to avoid the patterns becoming too specific while ensuring that elements in a pattern are still related in one underlying sentence. Then, "reduced" is accepted in the transition from q_1 to q_2 because it is included in "A2c_idiosyncratic" as one of the words to indicate abnormally low volume. (Lexicon "A2c_idiosyncratic" includes words that suggest atypical or idiosyncratic patterns of speech.) Finally, the word "volume" is accepted in q_2 because it is one of the words related to speech that is included in lexicon "A2c_speech."

All patterns are specified in a JAPE file. A JAPE file is a file where patterns to be annotated in the text can be described using GATE-specific formatting. GATE "reads" the JAPE files and applies them to text. When a pattern in the JAPE is recognized in the text, the text matching the pattern is annotated with the labels specified in the JAPE file.

Results

Parser Evaluation

Testbed

Our testbed consists of the 50 new EHRs, not used during development, containing 6634 sentences. The EHRs were randomly sampled from the 2000-2008 surveillance years, with 68% of records having positive ASD case status. Because evaluation is done at the sentence level and does not take record-level information into account, the case label itself is of little consequence. These are records that were annotated by the clinical experts and the text and annotation stored by us in electronic format. Of the entire set, 20.45% (1357/6634) sentences contained annotations, with some sentences containing more than 1 annotation.

A human-created gold standard, such as our testbed, is seldom completely perfect and consistent: entities may have been missed by the human annotators. We noticed such inconsistencies in prior work by us [33] and by others [21]. In this testbed, we encountered a few omissions, that is, annotations that were identified by our parser but not by the human annotators. This may be an oversight, or it may reflect the annotator's intention not to annotate all phrases when they are almost identical and represent the same DSM criterion. While this may suffice for manual review, a complete gold standard is needed for automated evaluation of an algorithm. Therefore, we ensured that in our test phase, we could rely on a complete gold standard. In addition to the phrases annotated, we requested additional expert evaluation to verify whether patterns discovered by the parser but not annotated (false positives) should be part of the gold standard. Of the 366 plausible patterns identified by our system, 277 were identified by the experts as part of the gold standard. We added these missed annotations to our gold standard. Table 3 provides an overview of the number of annotations in the gold standard.

Metrics

Similar to evaluation standards by others [34], we accept partial matches, defined as machine annotations that are considered correct if they contain any part of a gold standard annotation. For example, in the sentence "He also exhibited poor eye contact with the examiner," our tool annotated "exhibited poor eye contact," while the human expert annotated "poor eye contact with the examiner." We accepted these annotations because this region of text can identify meaningful information relevant to criterion A1a (nonverbal behaviors). These adjustments in our evaluation criteria reflect the high variety in expert annotations,

which tend to be inconsistent in their inclusion of subjects and verbs within the boundaries of the annotations.

For our evaluation, we calculated 4 metrics. Precision provides an indication of how correct the annotations made by the parser are; in other words, if the parser annotates sentences with a DSM label, this refers to the percentage of the labels that are correct. Recall (also referred to as sensitivity) provides an indication of how many of the annotations the parser is able to capture; in other words, of all the sentences that received a DSM label by the human annotators, what percentage does the parser also label correctly. We also calculate the *F*-measure, which is the harmonic mean of recall and precision. The scores for the *F*-measure indicate how balanced an approach is: when recall and precision are similarly high, the *F*-measure will be high; however, if one of them is low, the *F*-measure will reflect this with a low F-score. Finally, we also calculate specificity, which indicates how well our parser can ignore sentences that are not an expression of DSM criteria.

Precision (or PPV) = (True Positive)/(True Positive + False Positive)

Recall (or Sensitivity) = (True Positive) / (True Positive + False Negatives)

F-measure = 2 × (Precision × Recall)/(Precision + Recall)

Specificity = (True Negatives)/(True Negatives + False Positive)

We calculate these metrics at the annotation and at the sentence level. A true positive at the *annotation level* means that the system identified a criterion-specific annotation within a sentence also present in the gold standard. If a record or sentence contains 2 annotations for the same criterion, both should be identified individually. This is a stringent evaluation. For example, the sentence "He makes minimal eye contact with adults and struggles with turn-taking in conversations" is evaluated separately for criteria A1a (minimal eye contact) and criteria A2b (turn-taking in conversations).

We also apply the *sentence-level* evaluation for information extraction. In this case, a true positive is defined as identifying the sentence that contains gold standard annotations for a criterion, and the system has identified at least 1 annotation for the same rule. This evaluation can be more forgiving when a sentence contains more than 1 annotation.

Results of Parser Evaluation

Table 4 shows the results. At the annotation level, we achieved 74% precision and 42% recall on average. We took the microaverage, which combines the true and false positive counts across all rules. For individual criteria, precision was higher (≥75%) for most except two (criterion A1d and A3d). Recall was also particularly low for these two criteria, along with A1b and A1c. The best precision and recall were achieved for criterion A1a, with more than half of the annotations (57% recall) identified and with very few errors (96% precision).

Table 3. Gold standard overview.

Diagnostic and Statistical Manual of Mental Disorders diagnostic criteria		Gold standard	
Rule	Theme	Total in records	Average per record
A1a	Nonverbal behaviors	126	2.52
A1b	Peer relationships	91	1.82
A1c	Seeking to share	37	0.74
A1d	Emotional reciprocity	165	3.3
A2a	Spoken language	406	8.12
A2b	Initiate or sustain conversation	333	6.66
A2c	Stereotyped or idiosyncratic language	127	2.54
A2d	Social imitative play	66	1.32
A3a	Restricted patterns of interest	62	1.24
A3b	Adherence to routines	135	2.7
A3c	Stereotyped motor mannerisms	68	1.36
A3d	Preoccupation with parts of objects	28	0.56
Total	N/A[a]	1644	32.88

[a]N/A: not applicable.

Table 4. Annotation-level results.

Annotations[a]	Total in gold standard (number of annotations[b])	Evaluation		
		Precision	Recall	F-measure
A1a	126	0.96	0.57	0.72
A1b	91	0.63	0.27	0.38
A1c	37	0.78	0.19	0.30
A1d	165	0.62	0.27	0.37
A2a	406	0.69	0.44	0.53
A2b	333	0.79	0.44	0.57
A2c	127	0.68	0.36	0.47
A2d	66	0.79	0.56	0.65
A3a	62	0.83	0.40	0.54
A3b	135	0.75	0.51	0.61
A3c	68	0.82	0.41	0.55
A3d	28	0.53	0.29	0.37
Microaverage	N/A[c]	0.74	0.42	0.53

[a]Based on 6634 sentences.

[b]Total annotations=1644.

[c]N/A: not applicable.

Table 5. Sentence-level results.

Sentences[a]	Total in gold standard (number of sentences)[b]	Evaluation			
		Precision	Recall	F-measure	Specificity
A1a	120	0.97	0.59	0.74	1.00
A1b	90	0.68	0.30	0.42	1.00
A1c	35	0.78	0.20	0.32	1.00
A1d	158	0.63	0.28	0.39	1.00
A2a	391	0.71	0.45	0.55	0.99
A2b	329	0.83	0.47	0.60	1.00
A2c	121	0.67	0.37	0.48	1.00
A2d	65	0.83	0.58	0.68	1.00
A3a	61	0.73	0.36	0.48	1.00
A3b	123	0.74	0.52	0.61	1.00
A3c	64	0.82	0.42	0.56	1.00
A3d	28	0.53	0.29	0.37	1.00
Microaverage	1585	0.76	0.43	0.55	1.00
Any Rule	1357	0.82	0.46	0.59	0.97

[a]Based on 6634 sentences.

[b]Sentences with annotations =1357.

The results are very similar to those for the sentence-level evaluation (Table 5). Both metrics are slightly higher, with average precision at 76% and average recall at 43%. For the A1a criterion, more than half of the required sentences were identified (recall 59%) with minimal errors (97% precision). Using a sentence as a unit of analysis, it is also possible to compute specificity, or true negative rate, which was not possible with annotation-level evaluation because we would have to predefine in advance how many possible annotations (ie, sentence segments) there are in the EHRs. However, specificity is not a very interesting metric for this task. We achieve nearly perfect specificity because only 0.5%-5% of all

sentences contain true annotations for each rule, and our system reports very few false positives (high precision).

We conducted a final, more lenient approach by evaluating whether the system can identify the relevant sentences for DSM criteria, regardless of which criterion they represent. In this case, we found that our parser achieves 82% precision and 46% recall in identifying the 1357 sentences that were annotated for autism-like behavior (Table 5, last line "Any rule").

Discussion

Principal Findings

Overall, the rule-based approach resulted in a better performance than the machine learning approach. Some criteria, such as A1c and A3d, showed very large differences in precision between the two approaches, while others, like A1d and A3a, showed a large difference in recall. This may be due to the sparsity of the examples available for training. Furthermore, we chose to evaluate decision trees because of their interpretability. More sophisticated algorithms will be tested when larger datasets become available, and these may provide better results.

As is expected with the development of a rule-based extraction system, the results for precision are higher than those for recall. False negatives represent the annotations that were missed by our algorithm and lowered recall. We noticed 3 types of false negatives due to annotations not seen in the training data. First, there are new examples of behaviors; for example, "being a picky eater" is an A1a criterion, but it did not appear in our training data. To solve this, we will write additional JAPE rules. Second, there are sometimes different lexical variants of behaviors (ie, synonyms or related terms) used to describe behaviors. To solve this, we will look into expanding our lexicons, for example, through using word embeddings. Third, sometimes complex language or longer interstitial text is used that is not captured by our patterns. For example, "Eye contact, while it was also present, was limited at times." The solution will require further augmenting the patterns. In addition, some false negatives are the results of localized patterns. The criteria annotated in the EHR are sometimes determined by the clinicians using information in the EHR context or the surrounding text. This is not covered by JAPE patterns because it does not appear in the same sentence or neighboring text.

False positives usually occur for either of two reasons: accidental matches to nonsensical sentence fragments or plausible phrases with insufficient context. For example, Pattern 1 in Figure 1 also incorrectly matches to "$[\text{frequently}]_{\text{A2c\_frequent}}$ $[\text{will}]_{\text{word token}}$ $[\text{exhibit}]_{\text{word token}}$ $[\text{difficulty}]_{\text{word token}}$ $[\text{handling}]_{\text{word token}}$ $[\text{loud}]_{\text{A2c\_idiosyncratic}}$," a fragment in our training dataset (while not part of the DSM criteria, the phrase is commonly found in the records) that obviously cannot be a correct annotation. Meanwhile, the machine annotation "difficulty communicating with the teacher and peers" appears to indicate a failure to develop peer relationships as described in criterion A1b, but it is not accepted by domain experts because the criterion refers to challenges in social interactions, while this text fragment focuses on verbal communication.

We see large differences between the various DSM-IV criteria. For example, criterion A1c, which refers to "a lack of spontaneous seeking to share enjoyment, interests, or achievements with other people," is expressed completely differently in the test set and was not captured by our rules. This is not surprising because A1c is the criterion for which we have the least amount of training data (averaging 0.5 annotations per record). Additionally, the criterion covers a wide range of behaviors that can be expressed in many different ways. The variations and lack of data make describing patterns very difficult. Criterion A1a, which is related to nonverbal communication, obtained relatively high precision and recall. This is because clinicians tend to describe nonverbal communication in unambiguous, self-contained phrases, such as "eye contact" and "nonverbal communication," for which we can create precise patterns. For a similar reason, we also obtained good results for criteria A3a, A3c, and A3d, which are about abnormal interests, stereotypical actions, and tactile sensitivities, respectively. Some of the criteria (eg, A3c) have precision near 90%. Criteria A2a and A2b, which describe expressive and receptive language issues, are most prevalent among the rules. Combined, they account for >40% of the gold standard annotations. Taking advantage of the large sample of gold standard annotations, we were able to develop many and obtain relatively stable performance from development to testing.

In our case, we believe lower recall does not preclude useful applications of the parser. While some particular expression of a DSM criterion may be missed, it will be rare that all expressions of that particular DSM criterion in one record would be missed and, so, the detected DSM criterion would be taken into account for case assignment. Moreover, because of the high precision of the parser, when an expression of a DSM criterion is flagged, it is unlikely to be a false positive. As a result, large-scale analyses that focus on patterns of different criteria can be performed.

Parser Application: Case Study

Testbed

Given the high precision of our parser, we conducted a case study that shows insights into and the potential of the parser for future work. Our goal is to provide a broad overview of DSM criteria patterns found in existing EHRs over a 10-year span.

For our case study, we analyzed 4480 records available electronically from the ADDSP. These records have not been used during the development of the parser and contain a minimum of text (40 characters was empirically determined as the cutoff in this set; this represented about 10 words or a complete sentence, which is required for a complete annotation). We focus only on the free text fields and the results from applying our parser. Figure 2 shows the descriptive statistics. Records were collected every 2 years, starting in 2000 and ending (for our analysis) in 2010. In the first 3 collection periods, fewer records were collected; however, in each of the last 3 collection periods, around 1000 records were collected. The prevalence of autism in the records was lower in the first year (39%). This is associated with the relative inexperience of the data collection team who abstracted more records than

necessary to avoid missing cases. In subsequent years, data collection was more efficiently focused on records that included information consistent with an autism diagnosis, and the proportion of children abstracted who were determined to have met the case definition was commonly between 50% and 60%.

Abstractor training has been consistent over the years with the goal to enter only the necessary information to meet the project deadlines. Even so, the average length of the free text has increased over the years: the average number of words before 2006 was 1427 and has increased to 2450 from 2006 until 2010, nearly double the number.

Results

The records contained on average 5.76 different DSM criteria. We performed our analysis separately for records of children with ASD and of those labeled as non-ASD. All counts are normalized by record length: the number of criteria found is divided by the number of words in the document. This normalization avoids increasing the count of criteria solely due to having longer records, for example, when a child is seen multiple times for evaluation and the resulting EHR is longer, but the diversity of criteria may remain the same. Figure 3 shows the word count for the EHRs used in this evaluation. The word counts for ASD and non-ASD cases follow a similar trend. After 2004, there has been an increase in the length of records, which levels off the next year.

We first focus on the A1 DSM criteria. These criteria describe impairments in social interaction. For children with ASD (Figure 4, left graph), the A1d criterion (social or emotional reciprocity) is the most common criterion found in the records. The least commonly found criterion was the A1c criterion (shared interest). In the last 4 years, the average number of A1a, A1b, and A1d criteria described in the records has increased, but no similar increase in the average number of records containing A1c was observed.

We performed the same analysis for children without ASD (Figure 4, right graph). The results show, as expected, that fewer criteria are recorded in their records; the patterns are also different. The number of criteria recorded shows a decreasing trend over the last 4 years of records.

We repeat the same analysis for A2 DSM criteria (Figure 5), which focus on impairments in communication. The changes for A2 criteria are very small over the years. The most commonly found criterion is A2a (spoken language), and the least commonly found criteria are A2c (stereotyped or repetitive or idiosyncratic language) and A2d (imaginative play). There is a slight increase in 2002 and 2004 for the records of children with ASD, but few changes over the collection years. The total number of these criteria is higher than that of A1 criteria (see y-axis). Interestingly, there is little difference between the number of criteria found in EHRs labeled as ASD versus non-ASD.

Finally, we show the analysis for A3 DSM criteria (Figure 6), which focus on restricted, repetitive, and stereotyped behavior patterns. For the records labeled as ASD, the most commonly found criterion is A3b (adherence to routines), with the other three criteria being less common and comparable to each other. Overall, fewer criteria are found in the non-ASD-labeled records.

Figure 2. Descriptive information on 4480 records available electronically from the Arizona Developmental Disabilities Surveillance Program.

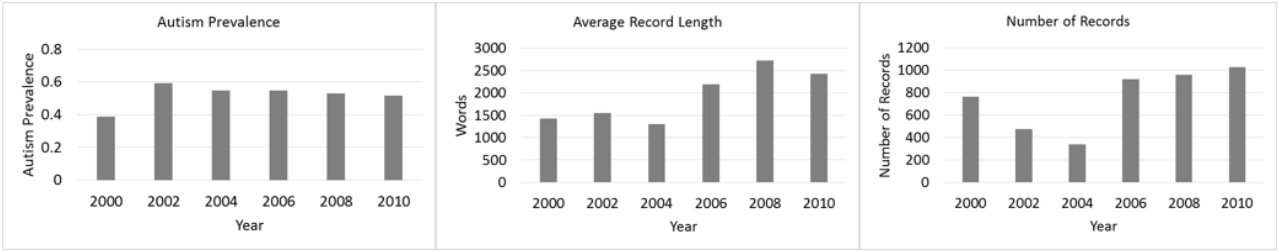


Figure 3. Electronic health record word count for autism spectrum disorder (ASD) and non-ASD cases.

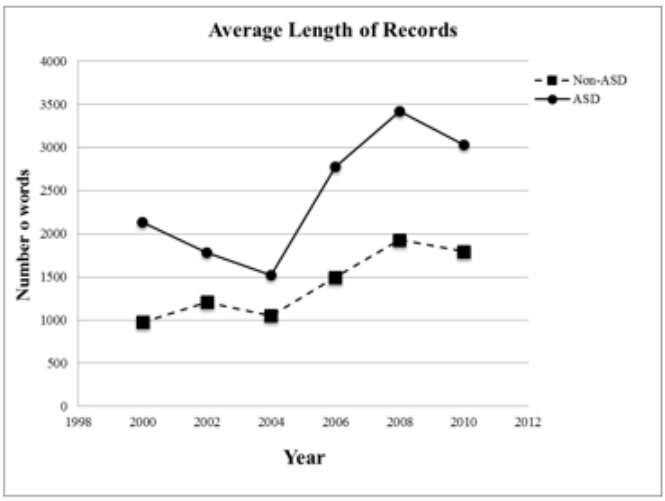


Figure 4. Average A1 criteria per record. ASD: autism spectrum disorder; EHR: electronic health record.

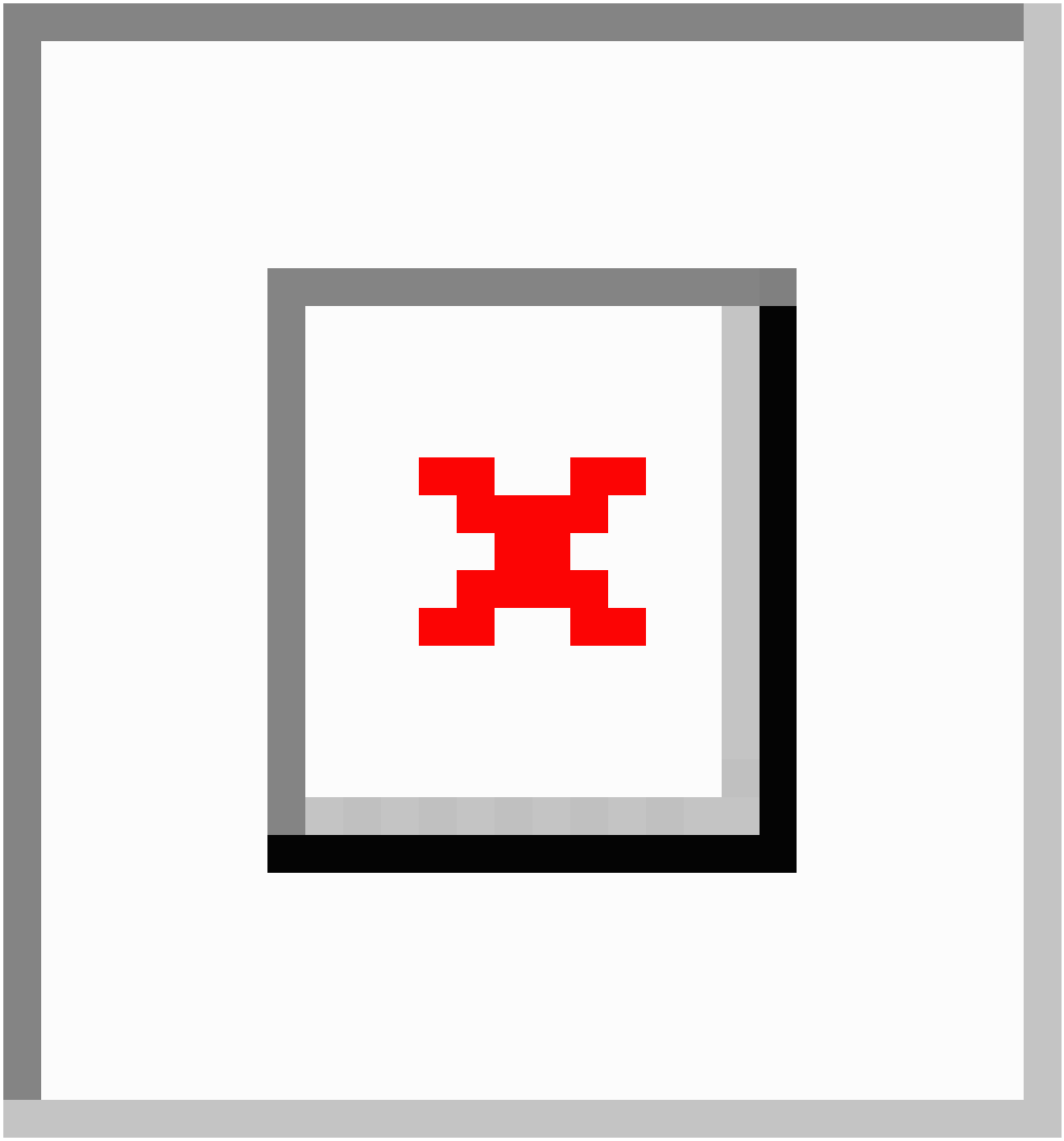


Figure 5. Average A2 criteria per record. ASD: autism spectrum disorder; EHR: electronic health record.

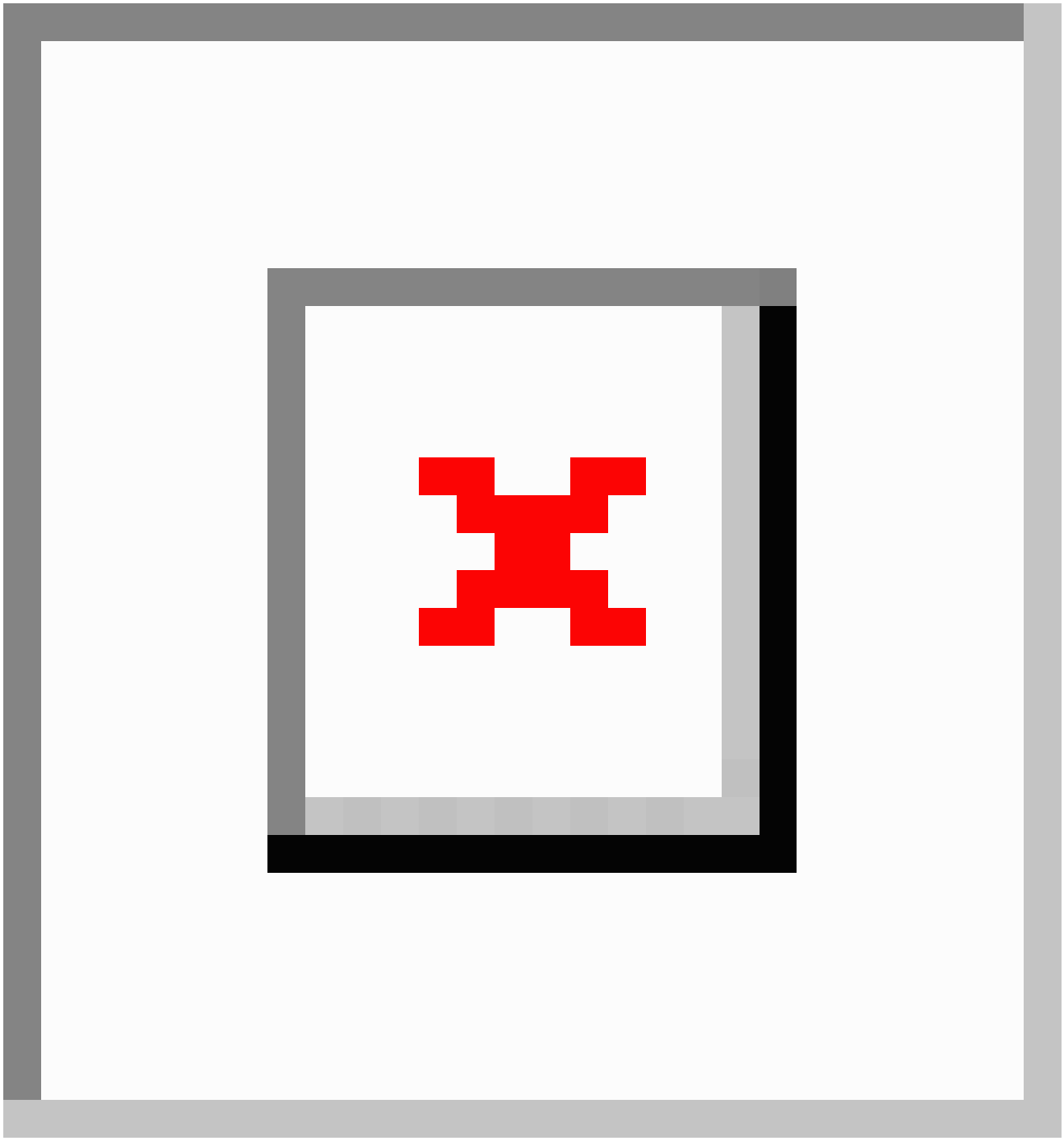
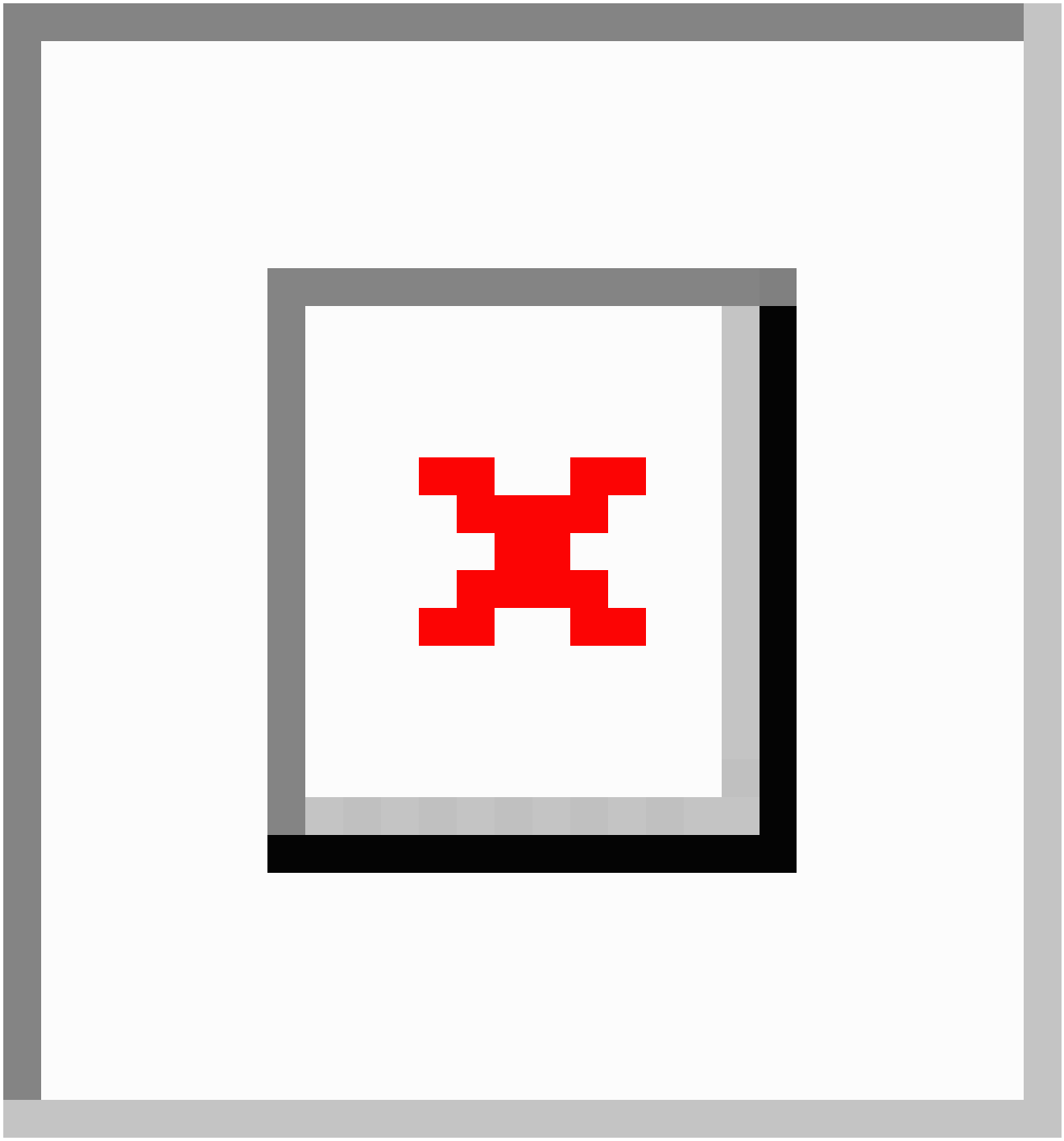


Figure 6. Average A3 criteria per record. ASD: autism spectrum disorder; EHR: electronic health record.



Discussion

The presence of a criterion in a record depends first on its presence in the child, second on whether the evaluator notes that criterion in the child, and third on whether the evaluator notes it in the record. The criteria that we identified with the greatest frequency were A2a (spoken language) and A2b (initiate or sustain conversation). Issues with language acquisition are the most frequently noted first cause of parental concern [35-39]. It is standard to make a note of the reason for the evaluation, and this would be expected to typically capture the first parental concern. Furthermore, these A2 criteria are frequent in all records because most children evaluated for ASD exhibit some type of impairment in communication. Criteria that we found least frequently include A2c (stereotyped or repetitive or idiosyncratic language) and A2d (imaginative play). While these are classic characteristics of ASD, they are less well noted by parents.

The frequencies of criteria in the ASD case records were not as different from the non-ASD records as may be expected. However, all children whose records were included in data collection had some type of diagnosis or special education qualification; no typically developing children are included in these data [40]. Individual criteria for ASD may be seen in children with other developmental disabilities, but it is the constellation of criteria that defines ASD.

Some changes across the years of data collection were observed. The first was an increase in the number of words per record. This increase is likely to reflect a true increase in the words rather than any changes in data collection procedures, as increasing numbers of records to review have motivated efforts to improve efficiency and eliminate the collection of superfluous text. An increase in the number of criteria included will necessarily mean that more words are collected. Next was an increase in the frequency of some specific criteria among the ASD cases. Changes through time in the frequency of a specific criterion may reflect more children who exhibit the criterion or that evaluators may have a heightened awareness of the criterion and are, therefore, more likely to note it. Criteria that were increasing in frequency included A1a (nonverbal behaviors), A1d (social or emotional reciprocity), and A3b (adherence to routine), but the increase in A3b was noted only in the most recent year.

The increases in the frequencies of some criteria in this dataset contrast with results from a study in Sweden, which found fewer autism symptoms among children diagnosed in 2014 than among those diagnosed in 2004 [41]. Arvidsson et al have suggested that clinical diagnoses of autism are being made in the year 2014 for cases that are less severe and would not have been given that diagnosis in 2004. They further suggested that this may explain some of the increase in the estimated prevalence of ASD. Increases in the estimated prevalence in the ADDSP dataset, from 6.5 per 1000 in 2000 to 15.7 per 1000 in 2010, are not susceptible to this decrease in severity as our criteria for determining case status has been consistent over the time period. In fact, we observe an increased proportion of cases with certain criteria and an increase in the average number of criteria over time. The increased prevalence that we have estimated would reflect a decrease in the severity of the condition only if evaluators in the recent years are making a notation of symptoms that are so mild they would not have noted them earlier.

The trend of increasing frequency of criteria A1a (nonverbal behaviors) and A1d (social or emotional reciprocity) in ASD-labeled records and the decreasing trend in those same criteria in non-ASD-labeled records may represent improvements in evaluators' awareness of these as symptoms of ASD and the importance of documenting these criteria for children who have the characteristics of ASD cases.

Conclusion

We described the design and development of a rule-based NLP tool that can identify DSM criteria in text. In comparison to a baseline machine learning approach that used decision trees, the rule-based approach performed better. We evaluated our approach at the annotation level (ie, matching to each rule within a sentence) and at the sentence level (ie, matching to the correct sentence). The system performed reasonably well in identifying individual DSM rule matches, with approximately half of all individual criteria-specific annotations discovered (44% recall) with few errors (79% precision). As expected with manually developed rules, precision was high, while recall was lower. In future work, we intend to increase both lexicons and patterns using machine learning approaches while retaining human-interpretable rules. This will increase the recall of our system. Furthermore, we intend to add negation as an explicit feature, which we believe will be necessary to maintain high precision.

We demonstrated our parser on almost 5000 records and compared the presence of different DSM criteria across several years. Changes in document length as well as in the presence of different DSM criteria are clear. Our analysis also showed that some DSM criteria are almost equally present in both ASD and non-ASD cases. In the future, we intend to increase the size of our records and combine the information extracted (ie, the DSM criteria matches) with other data from the structured fields in those EHRs as well as combine the information with external databases containing environmental and other types of data.

Our future work will be 2-fold. First, we will investigate the integration of our system into the surveillance workflow. For maximum usefulness, we will aim at extreme precision or extreme recall (while both are desirable, there tends to be a trade-off). With extremely high precision, the extracted diagnostic criteria can be used to make case decisions with high precision. Labeling a case as ASD can be automated for a large set of EHRs; only the set where no ASD label is assigned would require human review (due to low recall). In contrast, with extremely high recall, cases where diagnostic criteria are not extracted can be labeled as non-ASD with high confidence and only the cases where a label of ASD is assigned would need review (due to low precision). Second, because the development time of a rule-based system is substantial and application to a new domain would require starting over, we will investigate leveraging lessons learned from the parser to a machine learning approach that can transfer to different domains in mental health.

Acknowledgments

The data presented in this paper were collected by the CDC and Prevention Autism and Developmental Disabilities Monitoring Network supported by CDC Cooperative Agreement Number 5UR3/DD000680. This project was supported by grant number R21HS024988 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality

Conflicts of Interest

None declared.

References

1. Christensen DL, Baio J, Van Naarden Braun K, Bilder D, Charles J, Constantino JN, et al. Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. Surveillance Summaries 2016 Apr;65(3):1-23. [doi: 10.15585/mmwr.ss6503a1] [Medline: 27031587]
2. Abney S, Schapire RE, Singer Y. Boosting Applied to Tagging and PP Attachment. In: Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora. 1999 Presented at: Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora; June 21-22; College Park, MD, USA URL: https://aclanthology.coli.uni-saarland.de/events/emnlp-1999
3. Principal Investigators. Prevalence of Autism Spectrum Disorders --- Autism and Developmental Disabilities Monitoring Network. Surveillance Summaries 2009;58(SS10):1-20 [FREE Full text]
4. Developmental Disabilities Monitoring Network Surveillance Year 2010 Principal Investigators, Centers for Disease Control and Prevention (CDC). Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2010. MMWR Surveill Summ 2014 Mar 28;63(2):1-21 [FREE Full text] [Medline: 24670961]
5. Tyler C, Schramm S, Karafa M, Tang AS, Jain A. Electronic Health Record Analysis of the Primary Care of Adults with Intellectual and Other Developmental Disabilities. J Policy Pract Intellect Disabil 2010 Sep;7(3):204-210 [FREE Full text] [doi: 10.1111/j.1741-1130.2010.00266.x] [Medline: 26113866]
6. Clements CC, Castro VM, Blumenthal SR, Rosenfield HR, Murphy SN, Fava M, et al. Prenatal antidepressant exposure is associated with risk for attention-deficit hyperactivity disorder but not autism spectrum disorder in a large health system. Mol Psychiatry 2015 Jun;20(6):727-734 [FREE Full text] [doi: 10.1038/mp.2014.90] [Medline: 25155880]
7. Luo SX, Shinall JA, Peterson BS, Gerber AJ. Semantic mapping reveals distinct patterns in descriptions of social relations in adults with autism spectrum disorder. Autism Res 2016 Aug;9(8):846-853. [doi: 10.1002/aur.1581] [Medline: 26613541]
8. Aramaki E, Shikata S, Miyabe M, Usuda Y, Asada K, Ayaya S, et al. Understanding the Relationship between Social Cognition and Word Difficulty. A Language Based Analysis of Individuals with Autism Spectrum Disorder. Methods Inf Med 2015;54(6):522-529. [doi: 10.3414/ME15-01-0038] [Medline: 26391807]
9. Gondy Leroy, Sydney Pettygrove, Margaret Kurzius-Spencer. Automated Lexicon Feature Construction Using Word Embedding Clustering for Classification of ASD Diagnoses Using EHR. In: 22nd International Conference on Natural Language & Information Systems (NLDB 2017), Liège, Belgium. 2017 Presented at: International Conference on Applications of Natural Language to Information Systems; 21-23 June 2017; Liège, Belgium. [doi: 10.1007/978-3-319-59569-6_4]
10. Maenner MJ, Yeargin-Allsopp M, Van Naarden BK, Christensen DL, Schieve LA. Development of a Machine Learning Algorithm for the Surveillance of Autism Spectrum Disorder. PLoS One 2016 Dec;11(12):e0168224 [FREE Full text] [doi: 10.1371/journal.pone.0168224] [Medline: 28002438]
11. Lingren T, Chen P, Bochenek J, Doshi-Velez F, Manning-Courtney P, Bickel J, et al. Electronic Health Record Based Algorithm to Identify Patients with Autism Spectrum Disorder. PLoS One 2016 Jul;11(7):e0159621 [FREE Full text] [doi: 10.1371/journal.pone.0159621] [Medline: 27472449]
12. Mandy W, Charman T, Puura K, Skuse D. Investigating the cross-cultural validity of DSM-5 autism spectrum disorder: evidence from Finnish and UK samples. Autism 2014 Jan;18(1):45-54. [doi: 10.1177/1362361313508026] [Medline: 24113342]
13. 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]
14. Kreimeyer K, Foster M, Pandey A, Arya N, Halford G, Jones SF, et al. Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review. J Biomed Inform 2017 Dec;73:14-29 [FREE Full text] [doi: 10.1016/j.jbi.2017.07.012] [Medline: 28729030]
15. Wu C, Schwartz J, Brabant G, Peng S, Nenadic G. Constructing a molecular interaction network for thyroid cancer via large-scale text mining of gene and pathway events. BMC Syst Biol 2015;9 Suppl 6:S5 [FREE Full text] [doi: 10.1186/1752-0509-9-S6-S5] [Medline: 26679379]
16. Raja K, Subramani S, Natarajan J. A hybrid named entity tagger for tagging human proteins/genes. Int J Data Min Bioinform 2014;10(3):315-328. [Medline: 25946866]
17. Bagewadi S, Bobić T, Hofmann-Apitius M, Fluck J, Klinger R. Detecting miRNA Mentions and Relations in Biomedical Literature. F1000Res 2014 Oct;3:205 [FREE Full text] [doi: 10.12688/f1000research.4591.3] [Medline: 26535109]
18. Cejuela JM, McQuilton P, Ponting L, Marygold SJ, Stefancsik R, Millburn GH, FlyBase Consortium. tagtog: interactive and text-mining-assisted annotation of gene mentions in PLOS full-text articles. Database (Oxford) 2014;2014:bau033 [FREE Full text] [doi: 10.1093/database/bau033] [Medline: 24715220]
19. Leroy G, Chen H, Martinez JD. A shallow parser based on closed-class words to capture relations in biomedical text. Journal of Biomedical Informatics 2003 Jun;36(3):145-158. [doi: 10.1016/S1532-0464(03)00039-X]
20. Kwak M, Leroy G, Martinez JD, Harwell J. Development and evaluation of a biomedical search engine using a predicate-based vector space model. J Biomed Inform 2013 Oct;46(5):929-939 [FREE Full text] [doi: 10.1016/j.jbi.2013.07.006] [Medline: 23892296]

21. FitzHenry F, Murff HJ, Matheny ME, Gentry N, Fielstein EM, Brown SH, et al. Exploring the frontier of electronic health record surveillance: the case of postoperative complications. Med Care 2013 Jun;51(6):509-516 [FREE Full text] [doi: 10.1097/MLR.0b013e31828d1210] [Medline: 23673394]
22. Iqbal E, Mallah R, Jackson RG, Ball M, Ibrahim ZM, Broadbent M, et al. Identification of Adverse Drug Events from Free Text Electronic Patient Records and Information in a Large Mental Health Case Register. PLoS One 2015 Aug;10(8):e0134208 [FREE Full text] [doi: 10.1371/journal.pone.0134208] [Medline: 26273830]
23. Jane Carrington, Mihai Surdeanu, Angus Forbes. Augmented Communication Tools (ACTS) to Enhance Nurse Decision-Making. In: College of Nursing Resident Intensive Summer Experience (RISE). 2014 Presented at: College of Nursing Resident Intensive Summer Experience (RISE); 2014; Arizona.
24. Angus Forbes, Mihai Surdeanu, Peter Janson, Jane Carrington. Transmitting Narrative: An Interactive Shift-Summarization Tool for Improving Nurse Communication. In: IEEE Workshop on Interactive Visual Text Analytics. 2013 Oct Presented at: IEEE Workshop on Interactive Visual Text Analytics; October 14; Atlanta, Georgia, USA URL: http://vialab.science.uoit.ca/textvis2013/papers/Forbes-TextVis2013.pdf
25. Dillahunt-Aspillaga C, Finch D, Massengale J, Kretzmer T, Luther SL, McCart JA. Using information from the electronic health record to improve measurement of unemployment in service members and veterans with mTBI and post-deployment stress. PLoS One 2014 Dec;9(12):e115873 [FREE Full text] [doi: 10.1371/journal.pone.0115873] [Medline: 25541956]
26. Association A. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR). Arlington, VA: American Psychiatric Publishing, Incorporated; 2000.
27. Yimam M, Eckart de Castilho R, Gurevych I, Biemann C. Automatic Annotation SuggestionsCustom Annotation Layers in WebAnno. 2014 Jun Presented at: 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations; June 23-24; Baltimore, Maryland USA.
28. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc AMIA Symp 2001:17-21 [FREE Full text] [Medline: 11825149]
29. Aronson AR, Lang F. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc 2010;17(3):229-236 [FREE Full text] [doi: 10.1136/jamia.2009.002733] [Medline: 20442139]
30. Cunningham H, Maynard D, Bontcheva K, Tablan V. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. 2002 Presented at: 40th Anniversary Meeting of the Association for Computational Linguistics; July; Philadelphia, Pennsylvania.
31. Cunningham H, Tablan V, Roberts A, Bontcheva K. Getting more out of biomedical documents with GATE's full lifecycle open source text analytics. PLoS Comput Biol 2013;9(2):e1002854 [FREE Full text] [doi: 10.1371/journal.pcbi.1002854] [Medline: 23408875]
32. Toutanova K, lein D, Manning C, inger Y. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In: HLT-NAACL. 2003 Presented at: 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; May-June; Edmonton.
33. Leroy G, Rindflesch TC. Effects of information and machine learning algorithms on word sense disambiguation with small datasets. Int J Med Inform 2005 Aug;74(7-8):573-585. [doi: 10.1016/j.ijmedinf.2005.03.013] [Medline: 15897005]
34. Tsai RT, Wu S, Chou W, Lin Y, He D, Hsiang J, et al. Various criteria in the evaluation of biomedical named entity recognition. BMC Bioinformatics 2006 Feb 24;7:92 [FREE Full text] [doi: 10.1186/1471-2105-7-92] [Medline: 16504116]
35. De Giacomo A, Fombonne E. Parental recognition of developmental abnormalities in autism. Eur Child Adolesc Psychiatry 1998 Sep;7(3):131-136. [Medline: 9826299]
36. Karp EA, Ibañez LV, Warren Z, Stone WL. Brief Report: What Drives Parental Concerns About Their 18-Month-Olds at Familial Risk for Autism Spectrum Disorder? J Autism Dev Disord 2017 Dec;47(5):1535-1541. [doi: 10.1007/s10803-017-3060-1] [Medline: 28236100]
37. Kozlowski AM, Matson JL, Horovitz M, Worley JA, Neal D. Parents' first concerns of their child's development in toddlers with autism spectrum disorders. Dev Neurorehabil 2011;14(2):72-78. [doi: 10.3109/17518423.2010.539193] [Medline: 21410398]
38. Matheis M, Matson JL, Burns CO, Jiang X, Peters WJ, Moore M, et al. Factors related to parental age of first concern in toddlers with autism spectrum disorder. Dev Neurorehabil 2017 May;20(4):228-235. [doi: 10.1080/17518423.2016.1211186] [Medline: 27715357]
39. Wong V, Yu Y, Keyes ML, McGrew JH. Pre-diagnostic and Diagnostic Stages of Autism Spectrum Disorder: A Parent Perspective. Child Care in Practice 2016 Nov 28;23(2):195-217. [doi: 10.1080/13575279.2016.1199537]
40. Rice CE, Baio J, Van Naarden BK, Doernberg N, Meaney FJ, Kirby RS. A public health collaboration for the surveillance of autism spectrum disorders. Paediatr Perinat Epidemiol 2007 Mar;21(2):179-190. [doi: 10.1111/j.1365-3016.2007.00801.x] [Medline: 17302648]
41. Arvidsson O, Gillberg C, Lichtenstein P, Lundström S. Secular changes in the symptom level of clinically diagnosed autism. J Child Psychol Psychiatry 2018 Jul;59(7):744-751. [doi: 10.1111/jcpp.12864] [Medline: 29377119]

Abbreviations

ADDSP: Arizona Developmental Disabilities Surveillance Program
ASD: autism spectrum disorder
CDC: Centers for Disease Control and Prevention
DSM-IV-TR: Diagnostic Statistical Manual of Mental Disorders, Fourth Edition, Text Revision
EHR: electronic health record
GATE: General Architecture for Text Engineering
ICD: International Classification of Diseases
JAPE: Java Annotation Pattern Engine
NLP: natural language processing
POS: part-of-speech
UMLS: Unified Medical Language System

Edited by G Eysenbach; submitted 25.03.18; peer-reviewed by H Kilicoglu, Y Wang, M Torii; comments to author 03.05.18; revised version received 18.06.18; accepted 10.07.18; published 07.11.18.

Please cite as:
Leroy G, Gu Y, Pettygrove S, Galindo MK, Arora A, Kurzius-Spencer M
Automated Extraction of Diagnostic Criteria From Electronic Health Records for Autism Spectrum Disorders: Development, Evaluation, and Application
J Med Internet Res 2018;20(11):e10497
URL: https://www.jmir.org/2018/11/e10497/
doi:10.2196/10497
PMID:30404767

Original Paper

An Interpretable and Expandable Deep Learning Diagnostic System for Multiple Ocular Diseases: Qualitative Study

Kai Zhang[1,2], MS; Xiyang Liu[1,3,4], PhD; Fan Liu[3], BSc; Lin He[1], BSc; Lei Zhang[1], BSc; Yahan Yang[2], BSc; Wangting Li[2], BSc; Shuai Wang[3], BSc; Lin Liu[1], BSc; Zhenzhen Liu[2], MD, PhD; Xiaohang Wu[2], MD, MPH; Haotian Lin[2], MD, PhD

[1]School of Computer Science and Technology, Xidian University, Xi'an, China

[2]State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China

[3]School of Software, Xidian University, Xi'an, China

[4]Institute of Software Engineering, Xidian University, Xi'an, China

Corresponding Author:
Haotian Lin, MD, PhD
State Key Laboratory of Ophthalmology
Zhongshan Ophthalmic Center
Sun Yat-sen University
54 Xian Lie South Road
Guangzhou,
China
Phone: 86 (020)87330
Email: haot.lin@hotmail.com

Abstract

Background: Although artificial intelligence performs promisingly in medicine, few automatic disease diagnosis platforms can clearly explain why a specific medical decision is made.

Objective: We aimed to devise and develop an interpretable and expandable diagnosis framework for automatically diagnosing multiple ocular diseases and providing treatment recommendations for the particular illness of a specific patient.

Methods: As the diagnosis of ocular diseases highly depends on observing medical images, we chose ophthalmic images as research material. All medical images were labeled to 4 types of diseases or normal (total 5 classes); each image was decomposed into different parts according to the anatomical knowledge and then annotated. This process yields the positions and primary information on different anatomical parts and foci observed in medical images, thereby bridging the gap between medical image and diagnostic process. Next, we applied images and the information produced during the annotation process to implement an interpretable and expandable automatic diagnostic framework with deep learning.

Results: This diagnosis framework comprises 4 stages. The first stage identifies the type of disease (identification accuracy, 93%). The second stage localizes the anatomical parts and foci of the eye (localization accuracy: images under natural light without fluorescein sodium eye drops, 82%; images under cobalt blue light or natural light with fluorescein sodium eye drops, 90%). The third stage carefully classifies the specific condition of each anatomical part or focus with the result from the second stage (average accuracy for multiple classification problems, 79%-98%). The last stage provides treatment advice according to medical experience and artificial intelligence, which is merely involved with pterygium (accuracy, >95%). Based on this, we developed a telemedical system that can show detailed reasons for a particular diagnosis to doctors and patients to help doctors with medical decision making. This system can carefully analyze medical images and provide treatment advices according to the analysis results and consultation between a doctor and a patient.

Conclusions: The interpretable and expandable medical artificial intelligence platform was successfully built; this system can identify the disease, distinguish different anatomical parts and foci, discern the diagnostic information relevant to the diagnosis of diseases, and provide treatment suggestions. During this process, the whole diagnostic flow becomes clear and understandable to both doctors and their patients. Moreover, other diseases can be seamlessly integrated into this system without any influence on existing modules or diseases. Furthermore, this framework can assist in the clinical training of junior doctors. Owing to the rare high-grade medical resource, it is impossible that everyone receives high-quality professional diagnosis and treatment service.

This framework can not only be applied in hospitals with insufficient medical resources to decrease the pressure on experienced doctors but also deployed in remote areas to help doctors diagnose common ocular diseases.

(J Med Internet Res 2018;20(11):e11144) doi:10.2196/11144

KEYWORDS

deep learning; object localization; multiple ocular diseases; interpretable and expandable diagnosis framework; making medical decisions

Introduction

Although there have been many artificial intelligence-based automatic diagnostic platforms, the diagnostic results produced by such computer systems cannot be easily understood. Artificial intelligence that obtains diagnostic results from the computational perspective cannot provide the reason that is depicted as clinical practice for a given diagnosis. Some researchers have attempted to make the conclusion obtained from artificial intelligence methods explainable, such as Raccuglia et al used a decision tree to understand the classification result from the support vector machine [1]. Hazlett et al used a deep belief network, a reverse trackable neural network, to find diagnostic evidence of autism [2]. Zhou et al used the output of the last full-connected layer of the convolution neural network to infer which part of an image causes the final classification result, which also provides the evidence of classification [3]. In addition, Zeiler et al used occlusion test to study which parts of images produce a given classification result [4]. These studies made great achievements in explainable artificial intelligence, but readily explainable automatic diagnostic systems are still rare. The primary cause is that these explainable methods did not explain their result according to human thought patterns. Therefore, this research aims to make additional progress based on previous studies.

There are many existing works about the automatic diagnosis of different types of diseases with medical imaging, but all these works are isolated; those cannot regard all diseases shown in a specific format of medical images with a unified perspective, which is common in natural image processing and practical medical scenes. On the other hand, once all diseases are regarded as unified, the extensibility for integrating other types of medical imaging or disease will be easy. The diagnosis of ophthalmic diseases is highly dependent on observing medical images, so this work selected ophthalmic images that represent multiple ocular diseases as material and treated them with a consistent view. Of note, the unified automatic diagnostic procedure is the simulation of the work flow of doctors. An explainable artificial intelligence-based automatic diagnosis platform offers many advantages. First, it can increase the confidence in the diagnostic results. Second, it assists doctors to perfect the diagnosing thinking. Third, it helps medical students deepen the medical knowledge. Finally, it can clear a path toward diagnosing higher numbers of diseases from a unified perspective.

Besides, doctors can diagnose diseases by observing medical images, but doctors from many specialties and subspecialties cannot tackle all diseases. If a patient suffers from more than one type of disease, the system can tackle these diseases simultaneously. This work plans to integrate the experience of doctors from many subspecialties to construct an omnipotent ophthalmologist.

Thus, to create an explainable automatic diagnostic system with artificial intelligence, we simulated the workflow of doctors to help artificial intelligence follow the patterns of human thought. This research aims to apply artificial intelligence techniques to fully simulate the diagnostic process of doctors so that reasons for a given diagnosis can be illustrated directly to doctors and patients.

In this research, we designed an interpretable and expandable framework for multiple ocular diseases. There are 4 stages in this diagnostic framework: primary classification of disease, detection of each anatomical parts and foci, judging the conditions of anatomical parts, and foci and providing treatment recommendations. The accuracies of all stages surpass 93%, 82%-87%, 79%-98%, and 95%, respectively. Not only is this system an interpretable diagnostic tool for doctors and patients but it also facilitates the accumulation of medical knowledge for medical students. Moreover, this system can be enriched to cover more ophthalmic diseases or more diseases of other specialties to provide more services as the workflow of doctors. Telemedicine [5] can combine medical experts and patients with considerable low cost. This research develops an interpretable and expandable telemedical artificial intelligence diagnostic system, which can also effectively improve the undesirable condition that medical resource with high quality is not adequate and the distribution of it is not even. Finally, the health level of people all over the world and the medical condition of underdeveloped countries can be improved with the help of a computer network.

Methods

Data Preparation

Data are important for data-driven research [6]. The dataset is examined by all members of our team. Besides, we developed some programs to facilitate the examination of data. All images were collected in the Sun Yat-sen University Zhongshan Ophthalmic Center, which is the leading ophthalmic hospital in China [7]. In order to simulate the experience and diagnostic process of doctors, all images were segmented into several parts according to anatomical knowledge or diagnostic experiences and, then, were annotated. Next, multiple attributes of all parts were classified as the actual states of these parts (including foci). All the relevant aspects of the data (images, coordinates of each part, and the attribute information) were used to train an artificial intelligence system. This data preparation process can not only help simulate the diagnostic process of doctors but also facilitate many follow-up studies such as medical image segmentation,

clinical experience mining, and integration of refined diagnosing of multiple diseases.

We collected 1513 images that can be classified into 5 classes (normal, pterygium [9], keratitis [10], subconjunctival hemorrhage [11], and cataract [12]). Figure 1 lists the number of images of each class. Furthermore, the examples of objects to be detected in images are shown in Figure 2; for fundus images (the last row), the localized objects include an artery (blue), vein (green), the macula (black), the optic disc (light purple), hard exudate (yellow), and so on. For other types of images, the objects to be localized include the eyelid (red), eyelash (green), keratitis focus (yellow), cornea and iris zone with keratitis (pink), the pupil zone (blue), conjunctiva and sclera zone with hyperemia (orange), the conjunctiva and sclera zone with edema (light blue), the conjunctiva and sclera zone with hemorrhage (brown), the pupil zone with cataracts (white), the slit arc of the cornea (black), cornea and iris zone (dark green), the conjunctiva and sclera zone (purple), pterygium (gray), the slit arc of keratitis focus (dark red), and the slit arc of the iris (light brown). Table 1 lists the detailed diagnostic attributes to be classified, and each diagnostic information corresponds to a classification problem. The diagnostic information in Table 1 is corresponding to stage 3 (see Methods). This information is essential and fundamental for diagnosing and providing treatment advice and will be determined in stage 3 of the interpretable artificial intelligence system (see Methods). All information (object annotation and diagnostic information) was double-blind marked by the annotation team, which consisted of 5 experienced ophthalmic doctors and 20 medical students. The annotation of fundus images was completed; however, the experiments on fundus images were not finished. Because of the intrinsic characteristics of the fundus image, the output of the annotation method for fundus image is suitable for semantic segmentation.

Methodology

The framework consists of 4 functional stages as follows: (1) judging the class of disease, preliminary diagnosis that is completed with original image without any processing; (2) detecting each part of image, localization of anatomical parts, and foci that are used to discern different parts with different appearance so that more careful checking can be guaranteed; (3) classifying the attributes of each part, severity and illness assessment, which is closely connected to the second stage, is used to determine the condition of the illness; and (4) providing treatment advice according to the results from the first, second, and third stages, except for the treatment advice of a pterygium is from artificial intelligence, whereas treatment advice of other diseases is from experiences of doctors. First, the disease is primarily identified during stage 1. Second, all anatomical parts and foci are localized during stage 2, and important parts (cornea and iris zone with keratitis and pterygium) are segmented for the analysis in stage 3. Then, the attributes of all anatomical parts and foci are determined during stage 3. Then, the treatment advice is provided in stage 4. The whole process imitates the diagnostic procedure of doctors so that the reasons for a given diagnosis can be tracked and used to construct an evidence-based diagnostic report. Finally, treatment advice can be provided according to the full workflow presented above. Figure 3 shows the flowchart of this system. The analysis of fundus images is coming soon and will be easily integrated into this system quickly as the same idea with existing images. The first, second, and third function is fully based on artificial intelligence, which is trained with dataset; the fourth function is dependent on both artificial intelligence and the experience of doctors.

Figure 1. Information of image dataset.

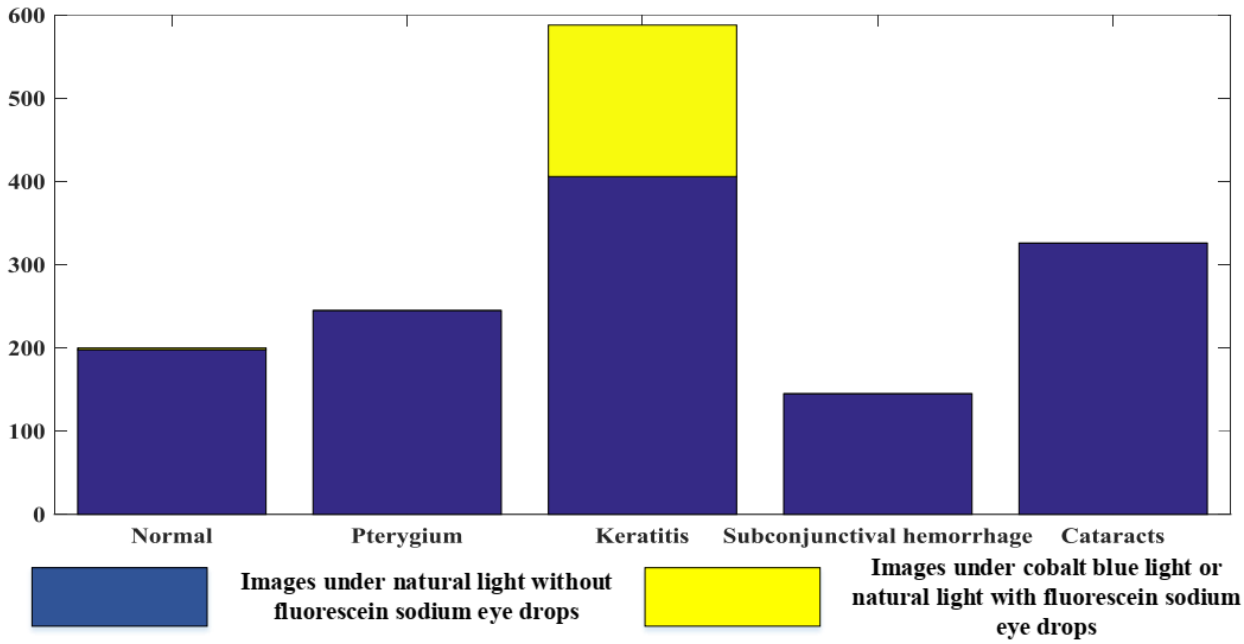


Figure 2. Examples of each object in terms of each type of disease or normal eye.

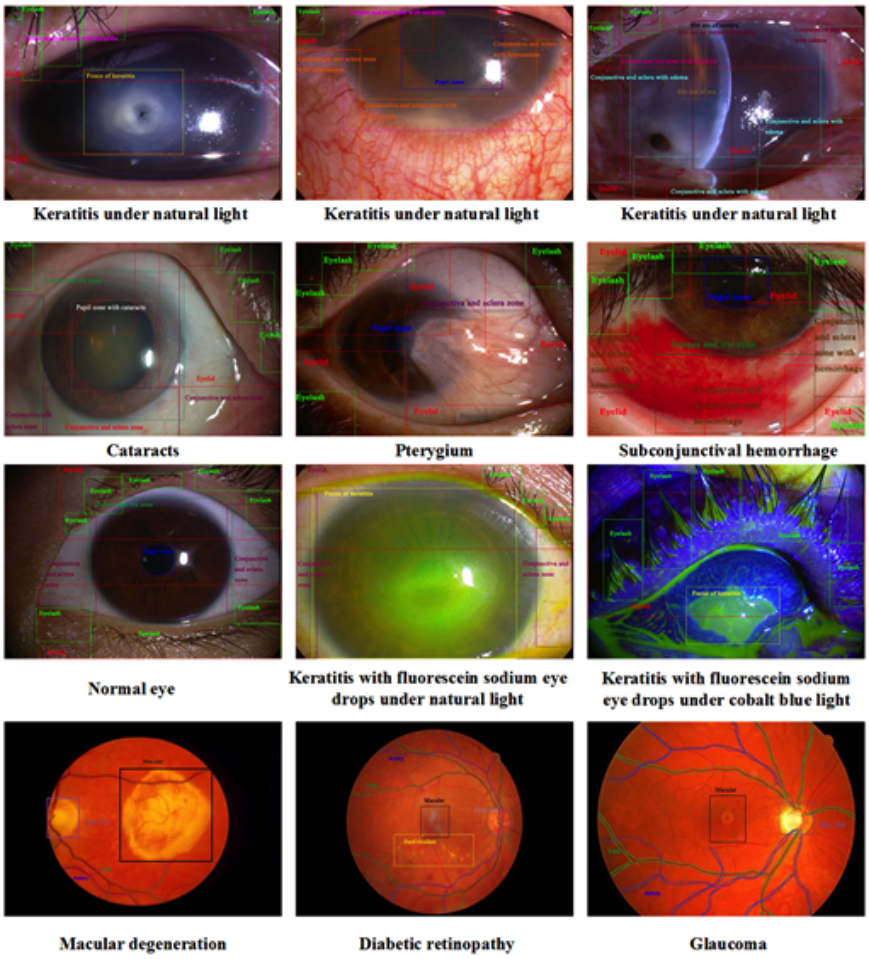


Table 1. Detailed diagnostic information regarding the dataset.

Disease	Diagnostic information (Number of classification problems)	Values of diagnostic information	Type of image
Pterygium	1. Whether the body of the pterygium is hypertrophied 2. Whether pseudo pterygium is present 3. Whether the head of the pterygium is uplifted 4. Whether the head and body of the pterygium is hyperemic 5. Whether the pterygium is in the progressive period	Yes or no	Images under natural light without fluorescein sodium eye drops
Keratitis	1. Turbidity degree of the cornea 2. Stage of keratitis 3. Corneal neovascularization 4. Edge of foci is clear 5. The condition of illness based on dyeing	1. Pupil zone is invaded by turbidity or not 2. Infiltration stage and ulcer stage, perforation stage, or convalescence 3. Yes or no 4. No dyeing and dot staining, sheet dyeing, or dyeing with coloboma	Images under cobalt blue light or natural light with fluorescein sodium eye drops [8]

Machine learning, especially deep learning technique represented by the convolutional neural network (CNN), is becoming the effective computer vision tool for automatically diagnosing diseases using biomedical images. It has been widely applied in the medical image classification and automatic diagnosis of disease, such as the diagnosis of attention deficit hyperactivity disorder with functional magnetic resonance imaging [13]; gradation of brain tumor [14], breast cancer [15],

and lung cancer [16]; and diagnosis of skin disease [17], kidney disease [18], and ophthalmic diseases [19-23]. In this research, inception_v4 [24] and residual network (Resnet) [25] (101 layers) were used to carry out stage 1 and stages 3 and 4, respectively. While stage 1 (inception_v4) can give a general diagnostic conclusion, stages 3 (Resnet) and 4 (Resnet) can provide further information about diseases and treatment recommendations. In this research, cost-sensitive CNN was adopted because the imbalanced classification is common in this research. Inception_v4 is a wider and deeper CNN that is suitable for careful classification (the difference between all classes is easily neglected sometimes). Resnet is a type of thin CNN, the architecture of which is full of cross-layer connections. The objective function is transformed to fit the residual function so that the performance of Resnet is improved considerably. In addition, Resnet is suitable for rough classification (the difference between all classes does not need to be carefully analyzed). In addition, we chose Resnet with 101 layers whose volume is adequate for the classification problems in this research. Stage 1 is a 5-classes classification, with some classes being very similar in color and shape; thus, inception_v4 is chosen in stage 1. As other classification problems are limited in one specific disease, Resnet is selected in stages 3 and 4. Furthermore, the chain rule of derivatives based on the stochastic gradient descent algorithm [26] was used to minimize the loss function.

Figure 3. Architecture of the overall framework for interpretable diagnosis of multiple ocular diseases. AI: artificial intelligence.

Faster-region based convolutional neural network (RCNN), an effective and efficiency approach, was adopted to localize the anatomical parts and foci (Stage 2). Faster-RCNN [27] is developed on the basis of RCNN [28] and Fast-RCNN [29], which originally applied superpixel segmentation algorithm to produce proposal regions, whereas Faster-RCNN uses an anchor mechanism to generate region proposals quickly and then adopts 2-stage training to obtain the transformations of bounding box repressor and classifier. The first stage of Faster-RCNN is region proposal network, which is responsible for generating region proposals. Then, whether the proposals are objects or not are judged, and the coordinates of each object are primary regressed. The second stage is judging the class of each object and eventually regressing the coordinate of each object, which is the same as RCNN and Fast-RCNN. In this research, pretrained ZF (Zeiler and Fergus [4]) network was exploited to save training time.

Experimental Settings

This system was implemented with convolutional architecture for fast feature embedding [30] (Berkeley Vision and Learning Center deep learning framework) and Tensorflow [31]; all models were trained in parallel on four NVIDIA TITAN X GPUs. For the classification problem, indicators applied to evaluate the performance are as follows:

Precisioni= TPi/(TPi+ FPi [TPi+ FPi]

SensitivityiTPR, RecallTPi/(TPi+ FNi [TPi+ FNi]

FNRifalse-negative rateFNiTPFNi [TP + FNi]

Specificityi= TNi/TNi+ FPi [TNi+ FPi]

FPRi(false-positive rate) = FPi/TNiFPi [TNi + FPi]

where N is the total number of samples; P_i indicates the number of correctly classified samples of i th class; k is the number of classes in specific classification problem;TP_i denotes the number of samples that are correctly classified as i th class; FP_i is the number of samples that are wrongly recognized as i th class; FN_i denotes the number of samples that are classified as j th class, j [1,c]/i; TN_i is the number of samples recognized as negative j th class, j [1,c]/i. All the above performance indicators can be computed with a confusion matrix. In addition, the receiver operating characteristics (ROC) curve, which indicates how many samples of i th class are recognized conditioned on a specific number of j th class (j [1,c]/i), are classified as i th class, PR (precision recall) curve, which illustrates how many samples of j th class are recognized as samples of i th class conditioned on a specific number of j th class (j [1,c]/i), are classified as i th class and area under the ROC curve (AUC), which means the area of the zone under the ROC curve was also adopted to assess the performance [32]. The indicators (precision, sensitivity, specificity, ROC curve with AUC, and PR curve) were only used to evaluate the performance of binary classification problems. Furthermore, accuracy and confusion matrix were used to evaluate the performance of multiclass classification problems.

For object localization problem, the interpolated average precision is always used to evaluate the performance [33]. The interpolated average precision is computed with the PR curve using the equation presented below:

In the equation, p(η) is the measured precision at specific recall η. In this research, 4-fold cross-validation was used to evaluate the performance of this system firmly for all classification problems and localization problems. The application of the cost-sensitive CNN is dependent on the distribution of the dataset in specific classification problems. Except for the classification problems 1, 6, and 8, other classification problems in stages 3 and 4 were completed with the cost-sensitive CNN.

Results

Performance of Stages 1 and 2

All stages and the whole work flow of this system were completed with acceptable performance. The 4 stages in the framework were separately trained and validated, and all relevant results in stages 1 and 2 are shown in Figures 4 and 5. The rows and columns of all heat maps stand for ground truth labels and predicted labels, respectively. Figure 4 shows the heat map of stage 1; the accuracy reaches 92%. Figure 5 shows the detection performance of Faster-RCNN in recognizing anatomical parts and foci; the mean value of average precision over all classes surpasses 82% and 90% for images under natural light without fluorescein sodium eye drops, and images under cobalt blue light or natural light with fluorescein sodium eye drops, respectively. The left image in Figure 5 is the performance for localizing objects in images without fluorescein sodium eye drops during stage 2, where I-VX represent the cornea and iris zone with keratitis, the focus of keratitis, the conjunctiva and sclera zone, the slit arc of the cornea, the slit arc of keratitis focus, the eyelid, the slit arc of the iris, the conjunctiva and sclera zone with hyperemia, the conjunctiva and sclera zone with edema, cornea and iris zone, pterygium, eyelash, pupil zone, the conjunctiva and sclera zone with hemorrhage, and the pupil zone with cataracts, respectively. The right image in Figure 5 presents the performance for localizing the objects in images with fluorescein sodium eye drops during stage 2, where I-VII represent the cornea and iris zone with keratitis, the focus of keratitis, the slit arc of the cornea, the slit arc of keratitis focus, the slit arc of the iris, the eyelid, and the eyelash, respectively. The statistical results of stage 2 are shown in Multimedia Appendix 1.

Figure 4. Performance of stage 1.

Figure 5. Performance of stage 2. AP: average precision.

Performance of Stages 3 and 4

Stage 3 was decomposed into 10 classification problems, and the relevant results are shown in Figure 6, including the boxplots for the accuracy, specificity and sensitivity, ROC curve with the AUC, PR curve for all binary classification problems, and the heat maps with accuracy for all multiclass classification problems. Figure 6 also shows the classification performance of stage 4, which includes boxplot for the accuracy, sensitivity and specificity, ROC curve with the AUC value and PR curve. The only one classification problem addressed by stage 4 is whether a patient who suffers from pterygium needs surgery. In stage 2, the detection rate of some objects is low because Faster-RCNN cannot effectively detect some small objects. We will overcome this issue by adjusting the parameters of Faster-RCNN. In spite of this, stage 3 will not be affected by this drawback because the detection rate of the cornea and iris zone with keratitis and pterygium (the relevant anatomical parts and foci), which is involved with stage 3, is considerably high. In addition, the detection performance of the pupil zone, which is related to vision is also satisfactory. In stage 3, the specificity of classification problems 1, 3, 4, and 5 is slightly low; the application scene of this system is hospitals where doctors pay more attention to sensitivity than specificity. The result of all classification problems is satisfactory and acceptable. Furthermore, the performance of classification problems 1, 3, 4, and 5 can be improved with more samples under the circumstance of Web-based learning. The statistical results of stages 3 and 4 are shown in Multimedia Appendix 1.

Performance of Stage 3 and 4 with Original Images

To study which anatomical parts are essential for automatic diagnostic, stages 3 and 4 were repeated with original medical images without processing; all parameters were same as the original parameters used in stages 3 and 4. The relevant results are shown in Figure 7. The classification performance close to that of the classification with anatomical parts and foci. In other words, the important parts, the cornea and iris zone with keratitis and pterygium, are essential for automatic diagnosis. The statistical results of stages 3 and 4 with original images are shown in Multimedia Appendix 1.

Figure 6. Performance of stage 3 and 4. PR: precision recall; ROC: receiver operating characteristics.

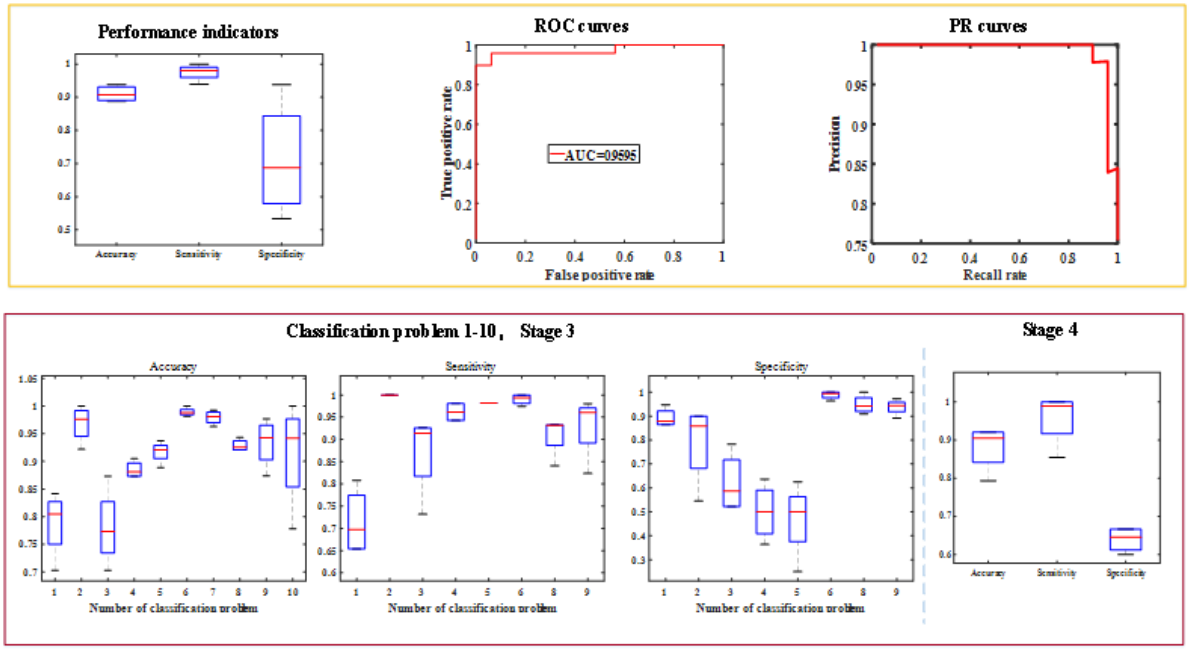
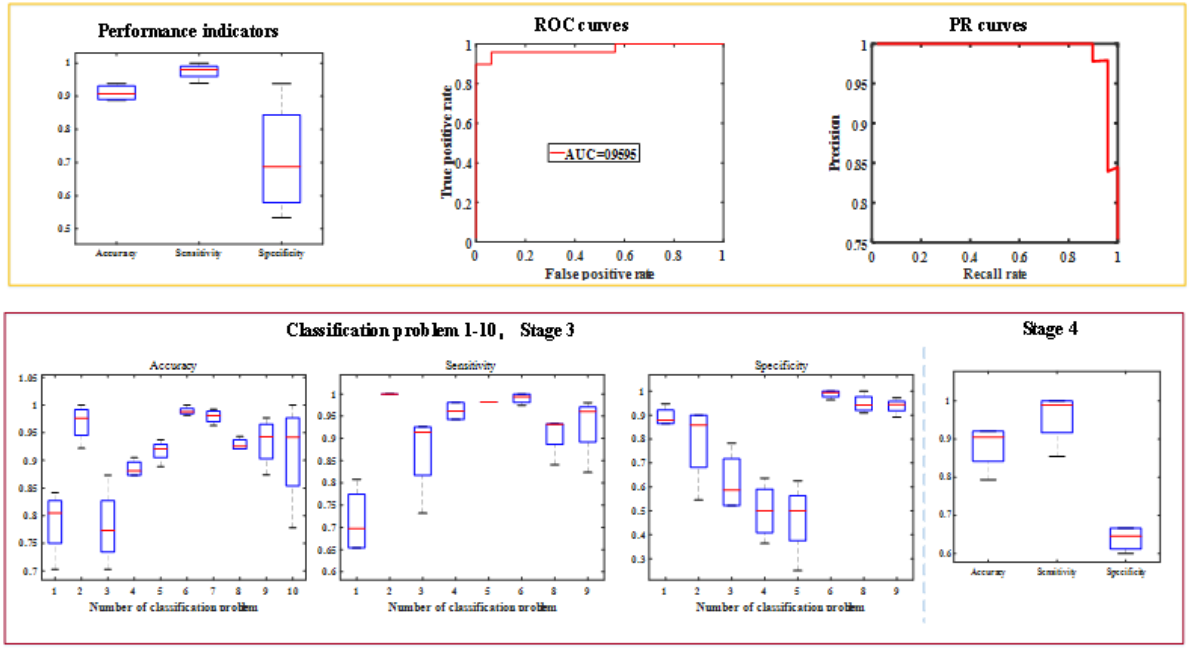


Figure 7. Performance of stage 3 and 4 with original images. PR: precision recall; ROC: receiver operating characteristics; AUC: area under the curve.



Web-Based Automatic Diagnostic System

We applied Django framework [34] to develop a telemedical decision-making and automatic diagnosing system to facilitate doctors and patients; this system can analyze inputted medical images, show the diagnostic result as the working process of doctors, and provide treatment advice by producing an examination report. In addition, this telemedical system can finely analyze medical images and provide treatment advice with a diagnostic report (a PDF file) that includes treatment suggestion according to the analysis result and the consultation between a doctor and a patient. The format of the diagnostic report is shown in Multimedia Appendix 1. All diagnostic information can be shown to a doctor and a patient by storing into a database. Administrators and doctors can handle all information and contact patients conveniently. Furthermore, this system can be deployed in multiple hospitals and medical centers to screen common diseases and collect more medical data, which can be used to improve the diagnosis performance. The website is available in Multimedia Appendix 1.

Discussion

In this study, we constructed an explainable artificial intelligence system for the automatic diagnosis of multiple ophthalmic diseases. This system carefully mimics the work flow of doctors so that reasons for specific diagnosis can be explained to doctors and patients with high performance. Besides, this system accelerates the application of telemedicine with the assistance of computer network and helps develop the health level and medical condition. Moreover, this system can be easily expanded to cover more diseases as long as the diagnostic processes of other diseases are simulated seamlessly. In addition, this system can help medical students to understand diagnosis and diseases. In the future, considerable progress can be made in this field. In this research, we did not consider a multilabel classification for those patients with multiple diseases. In the future, multiple-label classification can be adopted to make this system closer to real clinical circumstances. Moreover, because the bound box is not suitable for some anatomical parts, semantic segmentation can be applied in this system for segmenting medical images more accurately.

Acknowledgments

This study was funded by the National Key Research and Development Program (2018YFC0116500); the National Science Foundation of China (#91546101, #61472311, #11401454, #61502371, and #81770967), National Defense Basic Research Project of China (jcky2016110c006), the Guangdong Provincial Natural Science Foundation (#YQ2015006, #2014A030306030, #2014TQ01R573, and #2013B020400003), the Natural Science Foundation of Guangzhou City (#2014J2200060), The Guangdong Provincial Natural Science Foundation for Distinguished Young Scholars of China (2014A030306030), the Science and Technology Planning Projects of Guangdong Province (2017B030314025), the Key Research Plan for the National Natural Science Foundation of China in Cultivation Project (#91546101), the Ministry of Science and Technology of China Grants (2015CB964600), and

the Fundamental Research Funds for the Central Universities (#16ykjc28). We gratefully thank the volunteers of AINIST (medical artificial intelligence alliance of Zhongshan School of Medicine, Sun Yat-sen University).

Authors' Contributions

XL and HL designed the research; KZ conducted the study; WL, ZL, and XW collected the data and prepared the relevant information; KZ, FL, LH, LZ, LL, and SW were responsible for coding; LH and LZ developed the Web-based system; KZ analyzed and completed the experimental results; and KZ, WL, HL, and XL cowrote the manuscript. HL critically revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Relevant material.

[DOCX File, 42KB - jmir_v20i11e11144_app1.pdf]

References

1. Raccuglia P, Elbert KC, Adler PDF, Falk C, Wenny MB, Mollo A, et al. Machine-learning-assisted materials discovery using failed experiments. Nature 2016 May 05;533(7601):73-76. [doi: 10.1038/nature17439] [Medline: 27147027]
2. Hazlett HC, Gu H, Munsell BC, Kim SH, Styner M, Wolff JJ, IBIS Network, Clinical Sites, Data Coordinating Center, Image Processing Core, Statistical Analysis. Early brain development in infants at high risk for autism spectrum disorder. Nature 2017 Dec 15;542(7641):348-351 [FREE Full text] [doi: 10.1038/nature21369] [Medline: 28202961]
3. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning Deep Features for Discriminative Localization. 2016 Jun 26 Presented at: Internaltional Conference on Computer Vision and Pattern Recogintion; June 26th - July 1st, 2016; Las Vegas p. 2015-2019.
4. Zeiler M, Fergus R. Visualizing and Understanding Convolutional Networks. 2014 Sep 6 Presented at: European Conference on Computer Vision; September 6th-12th, 2014; Zurich.
5. Perednia DA, Allen A. Telemedicine technology and clinical applications. JAMA 1995 Feb 08;273(6):483-488. [Medline: 7837367]
6. Vibhu A, Zhang L, Zhu J, Fang S, Cheng T, Hong C, et al. Impact of Predicting Health Care Utilization Via Web Search Behavior: A Data-Driven Analysis. Journal of Medical Internet Research 2016 Sep 21;18(9):A.
7. Lin H, Long E, Chen W, Liu Y. Documenting rare disease data in China. Science 2015 Sep 04;349(6252):1064. [doi: 10.1126/science.349.6252.1064-b] [Medline: 26339020]
8. Amparo F, Wang H, Yin J, Marmalidou A, Dana R. Evaluating Corneal Fluorescein Staining Using a Novel Automated Method. Investigative Ophthalmology & Visual Science 2017 Jul 15;58(6):BIO168-BIO173.
9. Hirst LW. Treatment of pterygium. Aust N Z J Ophthalmol 1998 Nov;26(4):269-270. [Medline: 9843252]
10. Prajna N, Krishnan T, Rajaraman R, Patel S, Shah R, Srinivasan M, Mycotic Ulcer Treatment Trial Group. Predictors of Corneal Perforation or Need for Therapeutic Keratoplasty in Severe Fungal Keratitis: A Secondary Analysis of the Mycotic Ulcer Treatment Trial II. JAMA Ophthalmol 2017 Sep 01;135(9):987-991 [FREE Full text] [doi: 10.1001/jamaophthalmol.2017.2914] [Medline: 28817744]
11. Welch JF, Dickie AK. Red Alert: diagnosis and management of the acute red eye. J R Nav Med Serv 2014;100(1):42-46. [Medline: 24881426]
12. Lin H, Ouyang H, Zhu J, Huang S, Liu Z, Chen S, et al. Lens regeneration using endogenous stem cells with gain of visual function. Nature 2016 Mar 17;531(7594):323-328 [FREE Full text] [doi: 10.1038/nature17181] [Medline: 26958831]
13. Riaz A, Asad M, Alonso E, Slabaugh G. Fusion of fMRI and non-imaging data for ADHD classification. Comput Med Imaging Graph 2018 Apr;65:115-128. [doi: 10.1016/j.compmedimag.2017.10.002] [Medline: 29137838]
14. Mohan G, Subashini MM. MRI based medical image analysisurvey on brain tumor grade classification. Biomedical Signal Processing & Control 2018;39(1):139-161.
15. Golden JA. Deep Learning Algorithms for Detection of Lymph Node Metastases From Breast Cancer: Helping Artificial Intelligence Be Seen. JAMA 2017 Dec 12;318(22):2184-2186. [doi: 10.1001/jama.2017.14580] [Medline: 29234791]
16. Yu K, Zhang C, Berry GJ, Altman RB, Ré C, Rubin DL, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. Nat Commun 2016 Dec 16;7:12474 [FREE Full text] [doi: 10.1038/ncomms12474] [Medline: 27527408]
17. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Dec 02;542(7639):115-118. [doi: 10.1038/nature21056] [Medline: 28117445]

18. Kolachalama VB, Singh P, Lin CQ, Mun D, Belghasem ME, Henderson JM, et al. Association of Pathological Fibrosis With Renal Survival Using Deep Neural Networks. Kidney Int Rep 2018 Mar;3(2):464-475 [FREE Full text] [doi: 10.1016/j.ekir.2017.11.002] [Medline: 29725651]
19. Long E, Lin H, Liu Z, Wu X, Wang L, Jiang J, et al. An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. Nature Biomedical Engineering 2017 Jan 30;1(2):0024.
20. Gargeya R, Leng T. Automated Identification of Diabetic Retinopathy Using Deep Learning. Ophthalmology 2017 Dec;124(7):962-969. [doi: 10.1016/j.ophtha.2017.02.008] [Medline: 28359545]
21. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA 2016 Dec 13;316(22):2402-2410. [doi: 10.1001/jama.2016.17216] [Medline: 27898976]
22. Giancardo L, Meriaudeau F, Karnowski TP, Li Y, Garg S, Tobin KW, et al. Exudate-based diabetic macular edema detection in fundus images using publicly available datasets. Med Image Anal 2012 Jan;16(1):216-226. [doi: 10.1016/j.media.2011.07.004] [Medline: 21865074]
23. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. JAMA 2017 Dec 12;318(22):2211-2223 [FREE Full text] [doi: 10.1001/jama.2017.18152] [Medline: 29234807]
24. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. 2017 Feb 4 Presented at: AAAI Conference on Artificial Intelligence; Febuary 4th- 9th, 2017; San Fransisco, USA.
25. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 Jun 26 Presented at: International Conference on Computer Vision and Pattern Recognition; June 26th - July 1st, 2016; Las Vegas, USA.
26. Ketkar N. Parallelized stochastic gradient descent. 2010 Presented at: Advances in neural information processing systems; December 6TH -11th, 2010; Vancouver， Canada.
27. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. : International Conference on Neural Information Processing Systems; 2015 Presented at: Advances in Neural Information Processing Systems; December 7th -12th, 2015; Montreal， Canada.
28. Girshick R, Donahue J, Darrell T, Malik J. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. IEEE Trans Pattern Anal Mach Intell 2016 Jan;38(1):142-158. [doi: 10.1109/TPAMI.2015.2437384] [Medline: 26656583]
29. Girshick R. Fast R-CNN. 2015 Dec 13 Presented at: IEEE International Conference on Computer Vision; December 13th - 16th, 2015; Santiago, Chile.
30. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, et al. Caffe: Convolutional Architecture for Fast Feature Embedding. In: Caffe: Convolutional Architecture for Fast Feature Embedding. Caffe: Acm International Conference on Multimedia; 2014 Nov 3 Presented at: Acm International Conference on Multimedia; November 3rd - 7th, 2014; Orlando, Florida, USA.
31. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. eprint arXiv.0. 2016. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems URL: https://arxiv.org/abs/1603.04467 [accessed 2018-09-24] [WebCite Cache ID 72fPxz3sU]
32. Shi C, Pun CM. Superpixel-based 3D deep neural networks for hyperspectral image classification. Pattern Recognition 2018 Feb;74:600-616. [doi: 10.1016/j.patcog.2017.09.007]
33. Everingham M, Zisserman A, Williams C, Gool L, Allan M, Bishop C, et al. The 2005 PASCAL Visual Object Classes Challenge. In: The 2005 PASCAL Visual Object Classes Challenge. Berlin, Heidelberg: Springer; 2006:117-176.
34. Holovaty A, Kaplan-Moss J. The Definitive Guide to Django: Web development done right. In: The Definitive Guide to Django: Web development done right. California, American: Apress; 2009.

Abbreviations

AUC: area under receiver operating characteristics curve
CNN: Convolutional Neural Network
PR: precision recall
RCNN: region based convolutional neural network
Resnet: residual network
ROC: receiver operating characteristics

Edited by G Eysenbach; submitted 26.05.18; peer-reviewed by M Banf, L Zhang, Q Pan; comments to author 12.07.18; revised version received 31.07.18; accepted 02.08.18; published 14.11.18.

Please cite as:
Zhang K, Liu X, Liu F, He L, Zhang L, Yang Y, Li W, Wang S, Liu L, Liu Z, Wu X, Lin H
An Interpretable and Expandable Deep Learning Diagnostic System for Multiple Ocular Diseases: Qualitative Study
J Med Internet Res 2018;20(11):e11144
URL: http://www.jmir.org/2018/11/e11144/
doi:10.2196/11144
PMID:30429111

Viewpoint

Measuring Engagement in eHealth and mHealth Behavior Change Interventions: Viewpoint of Methodologies

Camille E Short[1], PhD; Ann DeSmet[2], PhD; Catherine Woods[3], PhD; Susan L Williams[4], PhD; Carol Maher[5], PhD; Anouk Middelweerd[6], PhD; Andre Matthias Müller[7,8], PhD; Petra A Wark[9], PhD; Corneel Vandelanotte[4], PhD; Louise Poppe[2], MSc; Melanie D Hingle[10], PhD; Rik Crutzen[11], PhD

[1]Freemasons Foundation Centre for Men's Health, School of Medicine, University of Adelaide, Adelaide, Australia

[2]Department of Movement and Sports Sciences, Ghent University, Brussels, Belgium

[3]Health Research Institute, Centre for Physical Activity and Health, Department of Physical Education and Sport Sciences, University of Limerick, Limerick, Ireland

[4]Physical Activity Research Group, Appleton Institute, School of Health, Medical and Applied Sciences, Central Queensland University, Rockhampton, Australia

[5]Alliance for Research in Exercise, Nutrition and Activity, Sansom Institute, School of Health Sciences, University of South Australia, Adelaide, Australia

[6]Department of Rheumatology, Erasmus Medical Center, Rotterdam, Netherlands

[7]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore

[8]Centre for Sport and Exercise Sciences, University of Malaya, Kuala Lumpur, Malaysia

[9]Centre for Innovative Research Across the Life Course, Faculty of Health and Life Sciences, Coventry University, Coventry, United Kingdom

[10]Department of Nutritional Sciences, College of Agriculture & Life Sciences, University of Arizona, Tucson, AZ, United States

[11]Department of Health Promotion, Care and Public Health Research Institute, Maastricht University, Maastricht, Netherlands

Corresponding Author:
Camille E Short, PhD
Freemasons Foundation Centre for Men's Health
School of Medicine
University of Adelaide
Level 7, South Australian Health and Medical Research Institute
North Terrace
Adelaide, 5000
Australia
Phone: 61 61883130532
Email: camille.short@adelaide.edu.au

Abstract

Engagement in electronic health (eHealth) and mobile health (mHealth) behavior change interventions is thought to be important for intervention effectiveness, though what constitutes engagement and how it enhances efficacy has been somewhat unclear in the literature. Recently published detailed definitions and conceptual models of engagement have helped to build consensus around a definition of engagement and improve our understanding of how engagement may influence effectiveness. This work has helped to establish a clearer research agenda. However, to test the hypotheses generated by the conceptual modules, we need to know how to measure engagement in a valid and reliable way. The aim of this viewpoint is to provide an overview of engagement measurement options that can be employed in eHealth and mHealth behavior change intervention evaluations, discuss methodological considerations, and provide direction for future research. To identify measures, we used snowball sampling, starting from systematic reviews of engagement research as well as those utilized in studies known to the authors. A wide range of methods to measure engagement were identified, including qualitative measures, self-report questionnaires, ecological momentary assessments, system usage data, sensor data, social media data, and psychophysiological measures. Each measurement method is appraised and examples are provided to illustrate possible use in eHealth and mHealth behavior change research. Recommendations for future research are provided, based on the limitations of current methods and the heavy reliance on system usage data as the sole assessment of engagement. The validation and adoption of a wider range of engagement measurements and their thoughtful application to the study of engagement are encouraged.

(J Med Internet Res 2018;20(11):e292) doi:10.2196/jmir.9397

KEYWORDS

telemedicine; internet; health promotion; evaluation studies; treatment adherence and compliance; outcome and process assessment (health care)

Introduction

Electronic health (eHealth) and mobile health (mHealth) behavioral interventions offer wide-reaching support at a low cost, while retaining the capacity to provide comprehensive, ongoing, tailored, and interactive support necessary for improving public health [1,2]. Although there is evidence that eHealth and mHealth behavior change interventions can be effective, low levels of adherence and high levels of attrition have been commonly reported [1-3]. In response, there have been calls to design and implement more engaging interventions to address these concerns [4-6].

It is generally agreed that a certain level of engagement is necessary for intervention effectiveness. However, there is a lack of clarity on how to conceptualize engagement. Some researchers have defined engagement solely as a psychological process relating to user perceptions and experience, whereas others consider engagement a purely behavioral construct, synonymous with intervention usage [4,7]. Consequently, it is often confused with adherence, which refers to whether the intervention is used as intended by the developers [3,8,9]. There have also been interdisciplinary differences. Behavioral scientists tend to characterize *good* engagement as high acceptability, satisfaction, or intervention adherence, whereas computer scientists tend to consider high engagement as a mental state associated with increased attention and enjoyment [4]. To consolidate these viewpoints and provide a less fragmented foundation for future research, 2 new conceptual models of engagement have been proposed [4,5].

Using a process of expert consensus, Yardley et al [5] proposed distinguishing between micro- and macrolevel engagement when examining the relationships between the user experience, usage, and behavior change. Microlevel engagement refers to the moment-to-moment engagement with the intervention, including the extent of use of the intervention (eg, number of activities completed) and the user experience (eg, level of user interest and attention when completing activities). Macrolevel engagement is defined as the depth of involvement with the behavior change process (eg, extent of motivation for changing behavior) and is linked to the behavioral goals of the intervention. The timing and relationship between micro and macro forms of engagement depend on the intervention, the user, and the broader context. Yardley's model suggests that after a period of *effective* engagement at the microlevel, the user may disengage from the platform but still be immersed in the behavior change process. Perski et al [4] offer a similar but more extensive framework based on a systematic review. Similar to Yardley et al, they define engagement as both the extent of usage and a subjective experience but refine this further by characterizing the subjective experience as being related specifically to attention, interest, and affect. These constructs are said to capture the cognitive and emotional aspects of engagement as they are described in computer science disciplines (eg, flow, immersion, and presence), all of which relate to a level of absorption and preoccupation (see Table 1 for definitions of these constructs). According to Perksi et al [4], high engagement influences behavior change through its influence on the determinants of behavior (similar to macroengagement, as described by Yardley et al). Engagement itself is hypothesized to be influenced by intervention features such as content, mode of delivery, and contextual features such as the physical environment (eg, internet access) and individual characteristics (eg, internet self-efficacy).

Both Perski et al and Yardley et al extend previous models [6,9-11] by considering the interaction between usage and psychological processes. By doing so, both models suggest that intervention usage may be a useful indicator of overall engagement with the intervention but is not a valid indicator of engagement in the behavior change process per se. Perski et al also highlight potential moderators and mediators of the engagement process and outline possible pathways in which engagement can influence overall intervention efficacy. These models serve as useful tools to refine and test hypotheses about how to influence engagement and how engagement impacts efficacy, which is necessary if we are to advance eHealth and mHealth behavioral science. However, an understanding of how to measure engagement is needed to test these models.

Basic overviews of the types of measures to assess engagement in eHealth and mHealth interventions have been provided by Yardley et al [5] as well as Perski et al [4]. Yardley et al briefly described the potential usefulness of different measurement types, including qualitative measures, self-report questionnaires, ecological momentary assessment, system usage data, sensor data, and psychophysiological measures. Perski et al identified over 100 studies related to engagement and noted the data collection methods used (eg, survey, website logs, and face-to-face interviews) in each study. Our aim is to extend their work by providing a comprehensive overview of the measurement options currently available. Our overall goal is to summarize and appraise measures of engagement used in eHealth and mHealth research and to highlight future areas of research when evaluating engagement in eHealth and mHealth behavior change interventions. We anticipate this will serve as a useful primer for those interested in the study of engagement and help to advance the field of eHealth and mHealth and behavior change by facilitating the use and validation of a wider range of engagement measurements and their thoughtful application to the study of engagement.

Overview of Methods Used to Identify and Assess Engagement Measures

We used a snowballing approach to identify relevant engagement measures. To begin, we extracted measures

identified by Perski et al [4] as well as other systematic reviews and published articles known to us through our former work in the field [12-14]. A data extraction table (see Multimedia Appendix 1) focusing on measurement type, engagement domain, and validity information was used to extract, sort, and explore measurement information to aid synthesis. During the writing and revision process, we searched for additional articles using Google Scholar and reran Perski's [4] original search strategy on MEDLINE and PsycINFO to identify more recent relevant literature. Readers should, therefore, consider this as a comprehensive, but not exhaustive, overview of the literature.

In line with Yardley et al's suggestions [5], our overview focuses on a wide range of methods to measure engagement. These include qualitative measures, self-report questionnaires, ecological momentary assessment, psychophysiological measures, as well as the analysis of system usage data, sensor data, and social media data. Methods that capture microlevel constructs were included in our synthesis if they were related to emotional, cognitive, or behavioral aspects of the user experience that could be characterized as interest, attention, affect, or intervention usage. This includes the constructs of flow, cognitive absorption, presence, and immersion, which have been commonly used in other disciplines. An overview of definitions for each of these constructs is provided in Table 1. Macrolevel measures were included if they related specifically to engagement in the behavior change process because of the digital intervention or its features. A single author initially drafted each section below, with all other authors providing a critical review.

Table 1. Definitions for constructs used to describe the emotional, cognitive, or behavioral aspects of engagement in previous literature.

Construct	Description
Interest	Individual interest is an enduring preference for certain topics and activities. It is impacted by pre-existing knowledge, personal experiences, and emotions. Situational interest is an emotional state brought about by situational stimuli (eg, the unexpectedness of information). It is evoked spontaneously and is presumed to be transitory. Both types of interest are related to liking and willful engagement in a cognitive activity that affects the use of specific learning strategies and how we allocate attention [15,16].
Attention	A state of focused awareness of specific perceptual information [17]. Focalization and concentration of consciousness are the essence of attention. Paying attention implies withdrawal from some perceptual information to deal effectively with others [18].
Affect	Affect is an intrinsic part of the sensory experience. It represents how an object or situation impacts how a person feels. It can be described by 2 psychological properties: hedonic valence (pleasure/displeasure) and arousal (activation/sleepy). It can be a central or background feature of consciousness, depending on where and how attention is applied [19,20].
Flow	Flow refers to an optimal state that arises when an individual is deeply absorbed in a task. It is characterized by enjoyment, focused attention, absorption, and distorted time perception and is considered intrinsically rewarding. It assumes the complete absence of negative affect [21].
Cognitive absorption	Cognitive absorption is a state of deep involvement, similar to flow, though it does not assume intrinsic motivation or the complete absence of negative affect. Cognitive absorption may still occur when a user is frustrated (and, therefore, the experience is not optimal) or extrinsically motivated (eg, by winning a competition with friends; [22]).
Immersion	Immersion is also similar to cognitive absorption and flow, though it is often used to describe a less extreme experience of engagement, one where one may still have some awareness of one's surroundings [22,23].
Presence	The term presence has been popular since the development of virtual reality technologies. Definitional consensus for presence is still emerging, though it is often described as the psychological sense of being *there* [24].
Intervention usage	The extent to which the intervention has been observed or interacted with by the user. It is made up of several components, including frequency of use, time spent on the intervention, and the type of interaction participated in. This is distinct from intended usage, which is the way in which users should utilize the intervention to derive the minimum benefit, as defined by the intervention developers [3].

Overview of Engagement Measures

Qualitative Methods

Focus Areas

Qualitative measures enable evaluation of micro- and macrolevel engagement and include methods such as focus groups, observations, interviews, and think-aloud activities (Table 2). At the microlevel, they allow for an in-depth account of the users' experience of the intervention. At the macrolevel, they can be used to explore the users' perceptions of how the intervention has helped them to engage in the behavior change process.

Current Use and Future Directions

Qualitative methodologies are commonly employed in the digital health setting to inform the development of interventions (ie, usability testing) and as an evaluation measure (eg, [25-29]). In most cases, the focus of the evaluation has been on perceptions of usability and acceptability, rather than engagement. However, there are some notable exceptions. For example, some studies have used think-aloud measures to understand cognitive processes and emotional reactions when navigating the intervention and viewing intervention content in real time [30-33]. Others have explored users' flow experiences, adherence and lived experience of technology using qualitative interviews [34-36], focus groups [37], or a combination of think-aloud and interview methods [32].

Along with exploring the direct user experience, qualitative measures are also often used to probe the perceived usefulness

of the intervention experience. Although this can relate to macroengagement (eg, by providing insights into how the intervention may have helped the user to achieve behavioral goals), efforts to explore the users' experience of the behavior change process in more depth are recommended. For example, researchers could explore how certain intervention features impact intentions and self-efficacy and how the relationship between intervention features and changes in psychosocial factors relate to use or disuse. This could be achieved using simple methods such as open-ended items in a questionnaire or more elaborate methods such as postintervention focus groups, which may help users to reflect on how the intervention has or has not engaged them in the behavior change process in more detail. Assessing these constructs at different time points may be particularly fruitful, especially given the cyclical nature of behavior change [38]. Exploring users' real-time engagement in the behavior change process was achieved in 1 recent study by thematically analyzing participant responses to intervention text messages [39]. By doing so, the authors were able to demonstrate that the study participants frequently gained positive cognitive and behavioral benefits from the text messages.

Considerations

A limitation of qualitative measures is that the results can be difficult to compare between studies. Results are also often not generalizable, mostly due to sampling bias. Qualitative measures are often used to collect rich data rather than representative data. For this reason, qualitative methods may be particularly suited to help generate hypotheses about engagement including how engagement relates to efficacy and effectiveness. They may also be useful for exploring hypotheses, especially when the focus is on understanding engagement on an individual level such as in n-of-1 studies [40]. In instances where representative data can be collected, such as in the text messaging study described above [39], hypothesis testing at the group level may be possible. However, the time and expertise needed to analyze data, which would ideally involve more than 1 person, is a barrier. This may be overcome in the future using machine learning tools to automate the coding of qualitative data [41].

Table 2. Overview of qualitative approaches to assessing engagement with considerations and example questions.

Qualitative approach	Description	Example items	Considerations (pros/cons)
Semistructured interviews	Provide an opportunity for sharing of lived experiences and feelings to uncover concealed perceptions related to digital health intervention or the technology; includes informal conversational interviews (spontaneous-suited to ethnographic research), semistructured interviews (interview guide used to steer otherwise spontaneous conversation), or standardized open-ended interviews (worded questions used for all participants).	Microlevel: Tell us what you think about the content; How did completing that module make you feel?; Please explain your pattern of use?; Why did you log on when you did?; Macrolevel: Did you notice any change to your thinking as a result of using the …("app")?; What impact did using the ... ("website") have on how you are going about changing your behavior?	Pros: inform modifications to increase acceptability, interactivity and tailor to end-user needs; identify a range of issues associated with use (both short and long term); augment interpretation of quantitative evaluation; generally small sample sizes. Cons: subject to bias (eg, recall and social desirability), especially if leading questions are asked; time consuming to collect and transcribe; time consuming to analyze and often requires more than 1 person to decide on and confirm themes.
Think aloud	Aim to capture the experience of using the technology in real time. The user is provided with a specific task to complete and is observed while they perform the task. The user is prompted to think aloud throughout the process.	Microlevel: Tell me what you are thinking; What are you looking at?; What's on your mind?; How are you feeling?; Why did you click on this?; Why did you frown/smile/sigh?; Macrolevel: Are you learning anything new?	Pros: can be used at various stages of development and implementation to understand how intervention features impact on engagement; occurs in real time, so less subject to recall bias. Cons: subject to observer bias; can be cognitively difficult for participants and requires practice; may require additional resources such as video or sound recording equipment to obtain a comprehensive picture. Acquired data can be time consuming and complex to analyze; may be most useful for exploring microlevel engagement.
Focus groups	Used to identify the social and contextual factors in specific population subgroups that influence engagement with digital health intervention and needs for technological characteristics and operations that promote user alignment and functional utility.	Microlevel: What did you think of the intervention?; Which components caught your attention the most?; What about them caught your attention?; Were there any components that caused frustration?; Did any aspects make you feel guilty? Macrolevel: How often did you think of the intervention during the week?; Was the intervention in the back of your mind?; How did the intervention help or hinder you reach your goals?	Pros: allow for spontaneous discussion of topics and subsequent voicing of ideas and perceptions that may go unnoticed in semistructured or structured interviews; Can obtain rich data from multiple people at the same time. Cons: subject to group or social desirability bias; some participants may not express themselves as fully in a group situation; requires practice to manage group discussion; can take a long time to transcribe due to interruptions/butting in; time consuming to analyze and often requires more than 1 person to decide on and confirm themes.

To facilitate the use of qualitative measures in the future, a brief overview of example questions by qualitative method type, as well as key considerations are provided in Table 2.

Self-Report Questionnaires

Focus Areas

Questionnaires can be used to assess both experiential and behavioral aspects of microlevel engagement as well as aspects of macrolevel engagement.

Current Use and Future Directions

Self-report questionnaires have most often been used to gain insight into users' subjective experience of digital platforms. Although questionnaire items have often been purpose-built and not subjected to psychometric testing (see Multimedia Appendix 1), there are a number of more rigorously developed scales. An overview of scales identified by our search [4,12-14] is presented in Multimedia Appendix 2. In brief, most scales have been developed to assess subjective experiential engagement with e-commerce websites or video games. Only 2 scales developed specifically for the eHealth and mHealth setting were identified (ie, the eHealth Engagement Scale [42] and the Digital Behavior Change Intervention Engagement Scale [43]), and only 1 of these has been validated [42], whereas validation of the other is currently underway [43]. Of note, some of the available scales assess attributes posited to predict engagement (eg, aesthetic appeal and usability experience [44-46]) as well as attributes considered to be a part of engagement (interest, attention, and affect). This is particularly the case for scales developed in the e-commerce setting and raises some validity concerns. Several of the scales are also quite long, which may place an undue burden on participants. The development and evaluation of high-quality short questionnaires relevant to eHealth and mHealth are therefore encouraged.

Questionnaires have also been used to assess behavioral aspects of engagement (ie, intervention usage). Although objective behavioral data are often available (see usage data below), questionnaires have been used when this is not the case. For example, a study comparing the relative efficacy of 2 *off-the-shelf* apps used questionnaires to assess the frequency and time of app use [47]. Although there are several scales with reasonable psychometric properties available for assessing the users' subjective experience (Multimedia Appendix 2), scales for assessing behavioral aspects of engagement in eHealth and mHealth interventions are lacking. Perski et al's self-report measure [43], which includes 2 items on behavioral engagement, is an exception. However, the validity of the measure is still being investigated. Perski's items and the purpose-built item used by other researchers usually have reasonable face-validity (eg, "how many times per week did you use the app?") but might lead to over- or underreporting depending on how items are phrased [48,49]. The validity of the chosen scale should be considered when interpreting the findings of self-reported behavioral data, and we recommend efforts to test the psychometric properties of developed items before use, if not yet available. This could be achieved by comparing the self-reported data with objectively collected data in a controlled setting (eg, [43]). The development of self-reported usage questionnaires that complement and provide useful context for objective usage measures should be considered. For example, if time on site or using an app is of interest, questionnaire data may identify cases where the user has left the program running in the background but has not been actively engaged. Likewise, information on behavioral cues at the point of engagement (eg, "what were you doing before you logged your steps using the app?") may complement usage data and provide a more comprehensive measure of usage patterns. Lessons may be gleaned from the scales developed to assess social networking intensity [50].

The third use of questionnaires relevant to the study of engagement at the macrolevel is the repeated assessment of psychological mechanisms hypothesized to account for behavioral changes (eg, self-efficacy). The assessment of change in these mechanisms and the conduction of a formal mediation analysis have been increasingly encouraged in the behavioral sciences [51,52] to investigate whether interventions are working as intended (ie, that the selected eHealth and mHealth strategies are indeed influencing determinants and changes in determinants are influencing behavior, eg, [53]). This methodology can be adopted to study engagement. Arguably, a user who demonstrates favorable changes in 1 or more of these determinants can be considered engaged in the behavior change process (eg, self-efficacy significantly increases over time). Furthermore, someone demonstrating changes at a prespecified cut point or where changes are associated with behavioral outcomes could be said to be engaged effectively. There are a number of pre-existing scales that can be used to assess changes in psychological determinants of behavior (eg, [54-56]) as well as guides for constructing purpose-built questions if existing scales are not suitable (eg, [57,58]). Decisions regarding what psychological constructs to assess changes in should be based on the theoretical underpinning of the intervention and the key intervention objectives and strategies used to achieve them.

Considerations

Overall, questionnaires can be a useful tool for measuring various aspects of engagement in a systematic, standardized, and convenient way. This can allow for easy comparison across studies and between experimental arms [5]. Limitations include questionnaire length (and, therefore, duration of completion); a lack of experiential measures designed and tested within a health context; a lack of focus on the behavioral aspects of engagement; and in some cases, the inclusion of items that measure predictors of engagement within engagement scales.

To select an appropriate scale, an understanding of the different constructs used to describe engagement across disciplines will be necessary (see Table 1). Reviewing the wording of the items and assessing how they will fit within the context of one's project may further help with scale selection. To this end, example items for each scale summarized above are provided in Multimedia Appendix 3. Most items will need to be adapted for a health setting, and not all scales will be applicable across study types or useful for assessing all aspects of engagement (ie, interest, attention, affect, intervention usage, and involvement in behavior change process). In some cases, it may

be necessary to generate completely new items or a completely new scale. In such cases, researchers are encouraged to report a measure of internal consistency (preferably McDonald omega) and present factor-analytic evidence confirming the dimensionality of the scale [59]. Attention to the length of the scale should also be given. This will likely be necessary to minimize missing data. The perceptions of those who drop out of the study are currently often not captured in evaluations of eHealth and mHealth interventions, which is problematic as those who drop out are usually those who have used the intervention the least. Ecological momentary assessments (EMAs; described in more detail below) may be useful to assess relevant engagement parameters regularly during the intervention and give a better impression of engagement throughout use [60]. Alternatively, selecting a representative subsample to administer surveys to and reimbursing them for their time might be a viable solution.

Ecological Momentary Assessments

Focus Area

EMAs can be used to assess both experiential and behavioral aspects of microlevel engagement as well as aspects of macrolevel engagement. The main objective of EMAs is to assess behaviors, perceptions, or experiences in real time and as they occur in their natural setting [61]. By prompting users to self-report data at varying times per day, EMAs allow these phenomena to be studied in different contexts and times.

Current Use and Future Directions

In EMAs, short surveys can either be accessed by the user on demand (eg, when logging a recent behavior), sent at specific or random intervals (eg, every 2 hours per day: *time-based sampling*), or they can be triggered by a certain event (eg, only when an activity tracker indicates the user is performing moderate to vigorous physical activity: *event-based sampling*). The latter is especially useful to capture rare behaviors, perceptions, or experiences. EMAs are often conducted on smartphone screens, but wearable devices can also be used (eg, CamNtech ProDiary, Philips Actiwatch Spectrum Plus, or Samsung Gear Life) [62].

EMAs have mostly been applied in eHealth and mHealth studies to measure health behavior and determinants (eg, [63,64]). We identified 1 study from previous reviews that used EMA to measure user engagement. This study [65] used event-based sampling to assess the breaks in levels of presence with a shooter game (not intended to improve health). The events that were sampled consisted of several parts of game play. No validity or reliability information for the slider was explicitly provided.

Despite the limited application of EMAs to measure engagement so far, EMAs may be well suited to study moment-to-moment or microlevel engagement with an intervention [5]. EMAs could provide data-driven insights into reasons for low adherence or dropout. EMAs are usually conducted over a short period with regular measurements over the day or week. However, it is also possible to adjust the timing and measurement intervals to collect longer-term insights into engagement. Contextual data and determinant data provided in EMA may enrich intervention usage data obtained from other sources to provide further insights into reasons for dropout.

Considerations

EMA surveys are intended to be very brief, because the purpose is to capture experiences in the moment and often to collect many data points over time, which can pose a burden to users [61]. Ensuring measures are brief is, therefore, important for both validity and for promoting adherence to the EMA protocol. Recent reviews of adherence to EMA protocols in health settings [66,67] suggest that compliance rates (proportion of EMAs completed) are reasonable (>70%), especially when sampling protocols are easy to follow. This speaks to the feasibility of utilizing this measurement approach; however, data analysis can be challenging for those unfamiliar with intensive longitudinal datasets (for a discussion regarding the challenges of EMA and example analysis approaches, see [68-72]). Advantages of EMAs include less recall bias than retrospective self-reports and potential for high ecological validity, as it studies behavior or effects in real-world contexts [60,61].

System Usage Data

Focus Area

System usage data quantitatively capture how the intervention is physically used by each participant. This relates to the behavioral component of microlevel engagement. When paired with other data sources, system usage data can provide insights into how usage patterns, intervention dose, and different adherence rates relate to other aspects of engagement (eg, interest, attention, affect, and changes in determinants) and efficacy and effectiveness outcomes (eg, [73-76]).

Current Use and Future Directions

System usage data are the most commonly collected and reported measures of engagement in eHealth and mHealth interventions [4]. Although the focus has predominantly been on nonusage attrition and overall adherence to the intervention [3,8], more recent studies have begun to explore the multidimensional nature of usage data [77-79], focusing on the depth and type of engagement as well as frequency measures. As the field progresses, it would be helpful to have shared ways of conceptualizing these data, as recent reviews have tended to categorize types of usage data differently using an inductive approach [4,78]. The FITT acronym [80], which stands for frequency, intensity, time, and type, and is commonly used in physical activity research, might be a useful tool in this sense, especially for considering usage data as an engagement measure *a priori*. Specific examples of how usage data could be categorized using this principle are given in Table 3. Frequency provides information on how often a participant visits the intervention site or uses the app. Intensity measures the strength or depth of engagement with the intervention, for example, the proportion of the intervention site or app features used out of the total available features [4]. Type refers to the type of engagement, for example, this could be categorized as reflective (eg, self-reporting behavior change), altruistic (eg, helping others), or gamified (eg, participating in a challenge) in nature. Type can also be divided into “active” (eg, active input such as when responding to a quiz, self-monitoring, or writing an action

plan) or "passive" (eg, an individual can view the intervention without having to interact with it) categories. Time is a measure of the duration of engagement during any single visit or a measure to assess level of exposure as an aggregate over the intervention period.

Examining usage data by aggregating data across the FITT categories can provide greater insights into engagement than focusing on any one domain [77,79,88]. For example, although the total time on site for users may appear similar (time data), their intensity data could be meaningfully different, which could lead to differences in engagement profiles (eg, attention, elaboration, and experience [79,88]). Separating users with similar data for time on site but markedly different patterns of use in terms of the type of activities may be helpful for identifying what aspects of the intervention are more engaging than others [92]; what aspects may be more influential for achieving behavior change, and in addition, whether this is moderated by user profiles (eg, [88]). The insight obtained from careful examination of system usage data in this way can assist intervention developers with data-driven solutions to encourage engagement [93].

Table 3. Examples of system usage data and type of information recorded.

Frequency, intensity, time, and type (FITT) principle	Example application
Frequency of engagement with the intervention	[81-83]
Log-in (number of log-ins recorded per participant, average log-ins per unit of time or total for intervention duration)	
Visits to the site (number of visits/hits per participant, average per unit of time or total)	
Intensity of engagement	[84-87]
Pages viewed (number)	
Lessons or modules viewed (total number, % of prescribed)	
Posts viewed (eg, lurking)	
Number of emails sent	
Number of posts written	
Accessed "Expert forum" (Ask the Expert) to pose a question/seek advice (number)	
Action plan created	
Number of quizzes attempted	
Time or duration of engagement with the program	[88,89]
Amount of time spent at each visit per participant (average and total minutes)	
Number of days between first and last log-in (duration or intervention *stickiness*)	
Type of engagement	[90,91]
Reflective (eg, participant recording of behavior or health status)	
Gamified (eg, accepting challenges and sending gifts)	
Altruistic (eg, helping others) or malevolent (eg, trolling others)	
Didactic (eg, reading posts and taking quizzes)	
Active (eg, recording behavior) versus Passive (eg, reading posts).	

Considerations

User behavior in digital health interventions can be tracked by embedding programming code as part of the development process or by using third-party services. For both methods, it is important during software design (or selection) to consider the type of data desired or needed to track behavioral engagement and ensure the data are adequately captured and can be extracted easily. The most commonly used third-party service is Google Analytics, a service that can be implemented by connecting to the Google Analytics application programming interface. Google Analytics can be used to collect information on the users' environment (location, browser, and connection speed), and the users' behavior (eg, number of page visits, time on site, where users came from, and which page they visited last before exiting [94]). Capturing usage data more specific to the intervention platform, such as participation in a quiz or percentages of answers correct, require, as in Google Analytics, intentional programming and capture at the level of the software. Before programming, considerable thought should be given to how the usage data will be analyzed, as good tracking generates a large amount of data (ie, every navigational move that every participant has ever made and even the moves they did not make) that can be hard to make sense of; therefore, an *a priori* analysis plan is recommended. Visualization tools [82] and engagement indices such as those discussed by Baltierra et al [79] and Couper et al [88], or consideration of new data analyses techniques may be useful to get insights into data [95,96]. Although system usage data are often considered objective and reliable, some caution interpreting data is recommended. The increasing use of dynamic internet protocol (IP) addresses and virtual private networks (which change or hide your IP address),

the use of IP addresses shared by multiple users (eg, via the family computer and internet cafes), and typical browsing behavior (eg, leaving multiple tabs open) may obscure usage data, especially for applications that do not require a unique log-in. This may be less of an issue for mobile apps compared with websites.

Intervention developers should, wherever possible, collect and analyze system usage data. Compared with the usage of other behavioral interventions (eg, a printed booklet), these data can be easily collected with early planning and good data capture techniques. Although usage data does not provide direct information on the psychological form of user engagement [4,5], it can provide some information to help us to understand what is engaging about an intervention, and what is not, in an unobtrusive way. There is also some evidence of predictive validity, with technology usage generally correlating with positive behavior change or health outcomes [81,91,97,98]. However, more research to establish the predictive validity of system usage data is needed, especially given that most analyses to date have lacked a suitable control group.

As with analyzing intensive longitudinal EMA data, the analysis of system usage data can be challenging. This is due to the intensive longitudinal and multidimensional nature of the data as well as the pattern of missingness (which tends to be nonrandom and nonignorable). Recognizing this, a comprehensive analysis plan should be developed before the commencement of the study. Exploration of the data visualization tools, composite engagement metrics, and analysis approaches referenced above might assist with the development of this plan.

It is also recommended that developers consider and outline the intended usage of the intervention. Intended usage is the way in which individuals should experience the intervention to derive maximum benefit, based on the conceptual framework informing intervention design (ie, developers' views on how the intervention should work best for who). Notably, intended usage may not be the same for all individuals (eg, in adaptive interventions [99,100]). By specifying intended usage *a priori* and comparing this with observed usage, we can establish whether individuals have adhered to the intervention and, in turn, the impact of adherence on efficacy [3].

Sensor Data

Focus Area

Sensors such as global positioning systems (GPS), cameras (eg, facilitating eye tracking analyses), microphones, and accelerometers can unobtrusively monitor users' behavior and the physical context in which this behavior takes place. They can be provided by the investigator, but many of them are embedded in smartphones or trackers. This relates to the behavioral component of microlevel (eg, information on intervention fidelity) and macrolevel (eg, tracking behavior in real-life settings) engagement.

Current Use and Future Directions

Analyzing sensor data presents an unobtrusive way of measuring engagement that requires no additional time effort from users other than the time spent engaging with the program. Their value lies in being able to track behavior of many users [101] and to enrich usage information in real-life situations or combining them with other user engagement measures such as EMA. There are calls for a different evaluation of eHealth and mHealth behavior change interventions than traditional interventions, to more nimbly respond to rapidly changing technologies and user preferences for functionalities [102-105]. Adaptations to eHealth and mHealth interventions are likely to be needed soon after first design and again after first implementation. Information from sensors that automatically track usage in real-life situations can help in measuring engagement with these interventions and distinguishing between successful mastery of intervention goals or need for continued engagement [5]. For example, in physical activity interventions, accelerometer information could continuously monitor the current activity level and indicate whether lower adherence to the intervention should be considered as a successful completion or disengagement. Sensor data paired with usage information may thus provide insights in macrolevel engagement as a mediator of positive intervention outcomes. In a similar vein, GPS information can enrich macrolevel engagement measures. GPS gives information on where people use the intervention and where it is less often used. For example, an app designed to facilitate healthy food choices may be used at home or at grocery stores but shows lower usage in restaurants. The GPS data give further insight into offline engagement with the intervention goals.

Sensors can also provide an indication of intervention fidelity. For example, distance traveled as measured by GPS and phone cameras taking pictures of meals can indicate whether the intervention is used in the appropriate manner and context [106]. The combination of usage and commonly included sensors can provide more detailed measures of real-life user engagement than usage information by itself. Sensor data can, moreover, trigger the event-based form of EMA. For example, users may be prompted to indicate their engagement with the intervention when the accelerometer shows the person is physically inactive or assess user engagement when GPS data show the person is in a certain physical context (eg, at a bar where there is a personal risk of smoking or alcohol consumption).

Considerations

A challenge of using GPS data for this purpose is the time-intensive nature of GPS data preparation and analysis. This will likely get easier in the future as new analysis packages become available to facilitate automation. Sensors, moreover, have the advantage of presenting a low level of respondent burden. However, especially with context-aware sensing using GPS, users are concerned about privacy issues [107,108]. In addition, sensors integrated in smartphones tend to negatively impact the battery life of the mobile device, and users may, therefore, be less compliant with running these sensors on their phones. This may especially be the case when users are skeptical toward the accuracy and relevance of context-aware smartphone sensing [25]. Therefore, communicating research findings about the validity of such measures [109,110]) and conducting pilot tests and validity studies of new measures may be necessary to

increase their use in future interventions and optimize uptake among participants.

Social Media

Focus Area

Another unobtrusive, low-burden approach to capturing engagement with the intervention is to analyze users' social media patterns. In social media, users create online communities (eg, social networking sites) via which they share information, opinions, personal messages, or visual material. Despite the interest of behavior change professionals in using social media to increase intervention effectiveness (see eg, [111]), to our knowledge, little research is available on the use of social media to measure engagement with eHealth and mHealth. The available resources mostly come from marketing and media audience research [112,113]. Social media message threads may provide useful information on user experience (microlevel engagement with the intervention) but might also provide insights in macrolevel engagement (eg, wall posts on behavioral achievements).

Current Use and Future Directions

One study examined the number of wall posts made over time as an indication of engagement with a social networking physical activity intervention [114]. An approach to reduce the burden in analysis is to use *markers* that are previously nonexisting words launched exclusively within the intervention [115]. These markers are used to trace any conversation that takes place on social media in relation to the intervention and are a way to measure social proliferation associated with the intervention content. An example comes from a video intervention on cognitive problems that may result from being a victim of violence [115]. To clearly identify all conversations and mentions on social media that would result from this topic, they launched the word *falterhead* to describe how the main character experienced the negative effects on his brain functioning after being violently attacked. This marker allowed a quick identification of all social media content related to the program, as this nonexisting word is unlikely to occur for content unrelated to the intervention. Several social media sources are then searched with text- and data-mining tools (eg, HowardsHome Finchline) for the occurrence and content of messages that contain these markers. The messages are next analyzed in terms of quantity (eg, Is the intervention being talked about?; What are patterns of social proliferation over time?) and quality (eg, How is the topic mentioned or discussed?; Is this how we wished viewers would think and talk about the intervention?). Social media messages relating to the eHealth and mHealth intervention might also be analyzed for their occurrence of certain profiles in social media engagement. On a continuum from passive and uninterested to more active and engaged, profiles of lurkers, casuals, actives, committed, and loyalists can be distinguished. Although to our knowledge, this has not yet been applied to analyze engagement with eHealth and mHealth behavior change interventions, interventions showing more actives, committed, and loyalists on social media might indicate higher user engagement than those receiving more lurkers and casuals [116]. This might especially be useful to assess comments on engagement in behavior change programs in real-life settings.

Considerations

The vast amount of social media content may make it difficult to extract what is relevant to the intervention. Markers mentioned earlier and audit tools are useful to facilitate such social media analyses. Examples of free audit tools to analyze social media are Sprout Social Simply Measured, Instagram Insights, and Union Metrics. The free statistical software program R also has many packages to analyze social media data. The analysis of these social media patterns requires a combination of qualitative techniques to assess discussion or post sentiment and topic, and quantitative methods, for example, to assess reach by combining number of followers for each mention on social media [117]. Text analytic tools available in many statistical packages such as R and SAS may also be useful here.

Psychophysiological Measures

Focus Area

Psychophysiological methods of measurement are used to examine the relationship between physiology and overt behavior or cognitive processes and variables. Psychophysiological measures are operationalization of cognitive processes or variables, just as self-reported questionnaires are used to measure processes or variables derived from theory [118]. They have been shown to be valuable approaches for measuring the experiential aspects of microengagement [119].

Current Use and Future Directions

There are several types of psychophysiological measures used to study cognitive and affective processes (for a comprehensive overview of measures used in human-computer interaction and user experience research, see [119-121]). We describe the 2 most common methods with a strong temporal resolution (ie, electroencephalography [EEG] and eye-tracking). A strong temporal solution (ie, precision of measurement with respect to time) is warranted to investigate engagement over time. It needs to be stressed, however, that other methods show promising results as well [122-127]. For example, predicting engagement using a novel visual analysis approach to recognize affect performed significantly better or on par with using self-reports [125]. The methods presented here are noninvasive but obtrusive in comparison with, for example, most measurements of system usage data. These methods are mostly used in laboratory settings and during intervention development (eg, pretesting of a website), but the opportunities to use them in field settings are increasing (eg, [128]). Moreover, it is also possible to use these methods in parallel with a trial or afterward to gain more insight into user engagement and, thereby, shed more light on trial findings.

EEG records electrical activity in the brain using small, flat metal discs (electrodes) attached to a person's scalp. Using this method requires adequate expertise, both in terms of measurement [129] and analysis [130] of data. Event-related potentials (ERPs) are the *average* changes in the EEG signal in response to a stimulus, and characteristic ERP responses are

referred to as components [131]. For example, Leiker et al [132], in a study on motion-controlled video games, focused on the amplitude of a specific component (labeled eP3a), which is a reliable index of attentional reserve [133,134]. This study revealed that participants who reported higher levels of engagement (as measured by the Intrinsic Motivation Inventory) showed a smaller eP3a, which is indicative of paying more attention to the primary task (eg, playing the game). Another study revealed that late negative slow wave components of the ERP were indicative of attention, which was partly confirmed by findings from self-reports (ie, the Immersive Experience Questionnaire) [123].

Eye-tracking is based on the strong association between eye movements and attention [135]. It is a suitable method to assess the course of attention over time [136]. For example, fixation data of an experimental study revealed that participants' eye movements in the immersive condition decreased over time, which is indicative of increased attention [137]. Another example is a study comparing a video with a text condition of a physical activity intervention. This study revealed that participants in the video condition displayed greater attention to the physical activity feedback in terms of gaze duration, total fixation duration, and focusing on feedback [138]. Another study using eye-tracking found that participants focused more on certain experimentally manipulated aspects of a health-related website (ie, in terms of frequency and duration), but this did not affect usage data (ie, the number of pages visited or the time on the website) [139]. It might be that these aspects attract attention, but there is a trade-off in the sense that participants then focus less on other aspects of the website. However, it could also be that attention only partly predicts engagement.

Considerations

With regard to both EEG and eye-tracking, it is important to note that attention is only the first appraisal in the process of engagement [139]. There are other psychophysiological methods besides EEG and eye-tracking that are mostly focused on measuring arousal. A previous study, for example, recorded electrodermal activity (EDA) and facial muscle activity (electromyography [EMG]) in addition to a Game Experience Questionnaire [140]. The association between these measures, however, was not straightforward. For example, EMG orbicularis oculi (periocular) is usually used to indicate positive emotions and high arousal but was negatively correlated to competence (which is a positive dimension of the Game Experience Questionnaire). Another study measured engagement in 5 different ways: self-reports using 4 dimensions of the Temple Presence Inventory, content analyses of user videos, EDA, mouse movements, and click logs (the latter 2 are measurements of usage data) [124]. These 5 measures correlated in limited ways. The authors concluded that "engagements as a construct is more complex than is captured in any of these measures individually and that using multiple methods to assess engagement can illuminate aspects of engagement not detectable by a single method of measurement" [124].

This is indicative of the complexity of engagement as a construct and reflects recent calls from the human-computer interaction field for future studies to identify valid combinations of psychophysiological measures that more fully capture the multidimensional nature of engagement [119].

Discussion

It is generally agreed that some form of engagement is necessary for eHealth and mHealth behavior change interventions to be effective. However, cohesive and in-depth knowledge about how to develop engaging interventions and the pathways between engagement and efficacy are lacking. Several models of engagement have been proposed in the literature to address this deficit, but little testing of the models has been conducted. To support research in this area and progress the science of user engagement, we aimed to provide a comprehensive overview of the measurement options available to assess engagement in an eHealth and mHealth behavioral intervention setting. The overview should not be treated as exhaustive; however, it should serve as a useful point of reference when considering engagement measures for behavioral eHealth and mHealth research.

The best measurement approach will likely depend on the stage of research and the specific research context, although there are benefits from using multiple methods and pairing the data (eg, self-report data relating to interest, attention and affect combined with system usage data). It is also important to make an inventory—before data collection—to check whether the available expertise for using different methods (eg, EEG) is available. Given the complexity of engagement as a construct, using multiple methods may be necessary to illuminate it fully [119,124]. At present, most studies in the eHealth and mHealth behavioral intervention space rely on system usage data only. Although system usage data is undoubtedly a valuable engagement marker, it is not considered a valid measure of micro- or macroengagement on its own [4,5]. Greater efforts are needed to also assess the psychological aspects of engagement to better understand the interplay between perceptions, usage, and efficacy.

Questionnaires are perhaps the most accessible way to assess microlevel engagement in terms of cost. However, there is currently a lack of validated self-report questionnaires specific to the eHealth and mHealth behavior change intervention context. This is reflected in the large number of purpose-built questionnaires (ie, questionnaires designed for a specific study) that have been used to date [4]. As the main benefit of questionnaires is that they allow for the collection of subjective data in a standardized way, greater efforts are needed to develop and implement standard items. Although not yet validated, the questionnaire developed by Perski et al [44] is promising in this regard, as it includes constructs related to both psychological and behavioral aspects of engagement and only focuses on engagement constructs. The other questionnaires identified focus only on the psychological aspects of engagement, and some include constructs more aligned with standard acceptability items (eg, perceived credibility), rather than the constructs of interest, attention, and affect. It may be best to avoid these questionnaires when testing models that hypothesize that acceptability markers influence engagement parameters.

There are several other measures of engagement that may also be used to test engagement models (eg, sensors, social media data, EMA, and psychophysiological measures). Despite their potential advantages, little research has been conducted exploring their use (and validity) in the digital behavior change setting. This is likely due to higher cost, time, and data analysis requirements relative to other measures. To mitigate this, behavioral researchers are increasingly drawing on expertise across other relevant disciplines (eg, informatics, human-computer interaction, experimental, and cognitive psychology). It is hoped that this paper will help to facilitate this research, especially research establishing the criterion, as well as divergent and predictive validity of these measures.

Overall, establishing the validity of engagement measures across multiple settings and learning how to triangulate measures in a complementary way are necessary next steps to advance the field. This will allow us to thoroughly test contemporary models of user engagement and hence, deepen our understanding of the interplay between intervention perceptions, usage, and efficacy across different settings.

Acknowledgments

The authors would like to thank Celine Chong for her assistance extracting data presented in the literature and reviewing engagement measures. Celine was supported by a Freemasons Foundation Centre for Men's Health summer scholarship. CES was supported by a National Health and Medical Research Council (NHMRC) Early Career Research fellowship (ID 1090517). LP is funded by the Research Foundation—Flanders. CV (ID 100427) is funded through a Future Leader Fellowship from the National Heart Foundation of Australia. CM is supported by an NHMRC Career Development fellowship (ID 1125913). AD is supported by a Research Foundation Flanders grant (FWO16/PDO/060, 12H6717N).

Authors' Contributions

CES conceived of the idea for this viewpoint. CES, AD, RC, CW, and SLW defined the scope of the manuscript and drafted the initial sections and revisions. CM, AMM, AM, PAW, CV, LP, and MDH provided critical review, refined the scope, and contributed to redrafting and editing of the manuscript.

Conflicts of Interest

None declared

Multimedia Appendix 1
Initial data extraction table.
[XLSX File (Microsoft Excel File), 35 KB - jmir_v20i11e292_app1.xlsx]

Multimedia Appendix 2
Self-report questionnaires for measuring microlevel engagement.
[PDF File (Adobe PDF File), 78 KB - jmir_v20i11e292_app2.pdf]

Multimedia Appendix 3
Example items in self-report questionnaires for measuring microlevel engagement.
[PDF File (Adobe PDF File), 51 KB - jmir_v20i11e292_app3.pdf]

References

1. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, present, and future of eHealth and mHealth research to improve physical activity and dietary behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-228.e1. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
2. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
3. Kelders M, 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]
4. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Dec;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
5. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Dec;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]

6. Short CE, Rebar A, Plotnikoff RC, Vandelanotte C. Designing engaging online behaviour change interventions: a proposed model of user engagement. Health Psychol Rev 2015;17(1):32-38 [FREE Full text]
7. Walton H, Spector A, Tombor I, Michie S. Measures of fidelity of delivery of, and engagement with, complex, face-to-face health behaviour change interventions: a systematic review of measure quality. Br J Health Psychol 2017 Dec;22(4):872-903 [FREE Full text] [doi: 10.1111/bjhp.12260] [Medline: 28762607]
8. 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]
9. Ryan C, Bergin M, Wells JS. Theoretical perspectives of adherence to web-based interventions: a scoping review. Int J Behav Med 2018 Dec;25(1):17-29. [doi: 10.1007/s12529-017-9678-8] [Medline: 28730402]
10. O'Brien H, Toms EG. What is user engagement? A conceptual framework for defining user engagement with technology. J Am Soc Inf Sci Technol 2008 Apr;59(6):938-955. [doi: 10.1002/asi.20801]
11. Crutzen R, Ruiter R. Interest in behavior change interventions: a conceptual model. The European Health Psychologist 2015;17(1):a-11.
12. Saket B. Stasko, Beyond Usability Performance: A Review of User Experience-focused Evaluations in Visualization, in Proceedings of the Sixth Workshop on Beyond Time Errors on Novel Evaluation Methods for Visualization. 2016 Presented at: BELIV '16 Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization; October 24, 2016; Baltimore, MD, USA p. A-142 URL: http://bahadorsaket.com/publication/BELIV2016.pdf
13. Denisova A, Nordin AI, Cairns P. The Convergence of Player Experience Questionnaires. In: Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play. 2016 Presented at: CHI PLAY '16; October 16 - 19, 2016; Austin, Texas, USA p. 33-37.
14. Boyle EA, Connolly TM, Hainey T, Boyle JM. Engagement in digital entertainment games: a systematic review. Comput Human Behav 2012 May;28(3):771-780. [doi: 10.1016/j.chb.2011.11.020]
15. Schiefele U. Interest, learning, and motivation. Educ Psychol 1991 Jun;26(3-4):299-323. [doi: 10.1080/00461520.1991.9653136]
16. Schraw G, Lehman S. Situational interest:a review of the literature and directions for future research. Educ Psychol Rev 2001;13(1):23-52. [doi: 10.1023/A:1009004801455]
17. Gerrig R, Zimbardo P. Psychology and Life. CA, United States: Pearson; 2012.
18. James W. In: Miller G, editor. The Principles of Psychology. New York: Dover Publications; 1950.
19. Duncan S, Barrett LF. Affect is a form of cognition: a neurobiological analysis. Cogn Emot 2007 Sep;21(6):1184-1211 [FREE Full text] [doi: 10.1080/02699930701437931] [Medline: 18509504]
20. Barrett LF, Russell JA. The structure of current affect: controversies and emerging consensus. Curr Dir Psychol Sci 2016 Jun 22;8(1):10-14. [doi: 10.1111/1467-8721.00003]
21. Csikszentmihalyi M. Flow: The Psychology of Optimal Experience. New York: Harper Perennial; 1990.
22. Brockmyer JH, Fox CM, Curtiss KA, McBroom E, Burkhart KM, Pidruzny JN. The development of the Game Engagement Questionnaire: a measure of engagement in video game-playing. J Exp Soc Psychol 2009 Jul;45(4):624-634. [doi: 10.1016/j.jesp.2009.02.016]
23. Baños RM, Botella C, Alcañiz M, Liaño V, Guerrero B, Rey B. Immersion and emotion: their impact on the sense of presence. Cyberpsychol Behav 2004 Dec;7(6):734-741. [doi: 10.1089/cpb.2004.7.734] [Medline: 15687809]
24. Lombard M, Ditton T. At the heart of it all: the concept of presence. J Comput Mediat Commun 1997 Sep 1;3(2). [doi: 10.1111/j.1083-6101.1997.tb00072.x]
25. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
26. Milward J, Khadjesari Z, Fincham-Campbell S, Deluca P, Watson R, Drummond C. User preferences for content, features, and style for an app to reduce harmful drinking in young adults: analysis of user feedback in app stores and focus group interviews. JMIR Mhealth Uhealth 2016 May 24;4(2):e47 [FREE Full text] [doi: 10.2196/mhealth.5242] [Medline: 27220371]
27. Kernot J, Olds T, Lewis LK, Maher C. Usability testing and piloting of the Mums Step It Up program--a team-based social networking physical activity intervention for women with young children. PLoS One 2014;9(10):e108842 [FREE Full text] [doi: 10.1371/journal.pone.0108842] [Medline: 25272039]
28. Short C, James EL, Rebar AL, Duncan MJ, Courneya KS, Plotnikoff RC, et al. Designing more engaging computer-tailored physical activity behaviour change interventions for breast cancer survivors: lessons from the iMove More for Life study. Support Care Cancer 2017 Dec;25(11):3569-3585. [doi: 10.1007/s00520-017-3786-5] [Medline: 28624949]
29. Kirwan M, Duncan MJ, Vandelanotte C, Mummery WK. Design, development, and formative evaluation of a smartphone application for recording and monitoring physical activity levels: the 10,000 Steps “iStepLog”. Health Educ Behav 2013 Apr;40(2):140-151. [doi: 10.1177/1090198112449460] [Medline: 22984196]

30. Perski O, Blandford A, Ubhi HK, West R, Michie S. Smokers' and drinkers' choice of smartphone applications and expectations of engagement: a think aloud and interview study. BMC Med Inform Decis Mak 2017 Dec 28;17(1):25 [FREE Full text] [doi: 10.1186/s12911-017-0422-8] [Medline: 28241759]
31. Bradbury K, Morton K, Band R, van Woezik A, Grist R, McManus RJ, et al. Using the person-based approach to optimise a digital intervention for the management of hypertension. PLoS One 2018;13(5):e0196868 [FREE Full text] [doi: 10.1371/journal.pone.0196868] [Medline: 29723262]
32. Crane D, Garnett C, Brown J, West R, Michie S. Factors influencing usability of a smartphone app to reduce excessive alcohol consumption: Think Aloud and Interview Studies. Front Public Health 2017;5:39 [FREE Full text] [doi: 10.3389/fpubh.2017.00039] [Medline: 28421175]
33. Alkhaldi G, Modrow K, Hamilton F, Pal K, Ross J, Murray E. Promoting engagement with a dimgital health intervention (HeLP-Diabetes) using email and text message prompts: mixed-methods study. Interact J Med Res 2017 Aug 22;6(2):e14 [FREE Full text] [doi: 10.2196/ijmr.6952] [Medline: 28829328]
34. El-Hilly A, Iqbal SS, Ahmed M, Sherwani Y, Muntasir M, Siddiqui S, et al. Game On? Smoking cessation through the gamification of mHealth: a longitudinal qualitative study. JMIR Serious Games 2016 Oct 24;4(2):e18 [FREE Full text] [doi: 10.2196/games.5678] [Medline: 27777216]
35. Hwang M, Hong J, Hao Y, Jong J. Elders' usability, dependability, and flow experiences on embodied interactive video games. Educ Gerontol 2011 Aug;37(8):715-731. [doi: 10.1080/03601271003723636]
36. Morrison L, Moss-Morris R, Michie S, Yardley L. Optimizing engagement with Internet-based health behaviour change interventions: comparison of self-assessment with and without tailored feedback using a mixed methods approach. Br J Health Psychol 2014 Nov;19(4):839-855 [FREE Full text] [doi: 10.1111/bjhp.12083] [Medline: 24308806]
37. Horsch C, Lancee J, Beun RJ, Neerincx MA, Brinkman WP. Adherence to technology-mediated insomnia treatment: a meta-analysis, interviews, and focus groups. J Med Internet Res 2015 Sep 04;17(9):e214 [FREE Full text] [doi: 10.2196/jmir.4115] [Medline: 26341671]
38. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]
39. Irvine L, Melson AJ, Williams B, Sniehotta FF, McKenzie A, Jones C, et al. Real time monitoring of engagement with a text message intervention to reduce binge drinking among men living in socially disadvantaged areas of Scotland. Int J Behav Med 2017 Dec;24(5):713-721 [FREE Full text] [doi: 10.1007/s12529-017-9666-z] [Medline: 28702758]
40. McDonald S, Quinn F, Vieira R, O'Brien N, White M, Johnston DW, et al. The state of the art and future opportunities for using longitudinal n-of-1 methods in health behaviour research: a systematic literature overview. Health Psychol Rev 2017 Dec;11(4):307-323. [doi: 10.1080/17437199.2017.1316672] [Medline: 28406349]
41. Kevin C, Xiaozhong L, Eileen E. Machine learning and rule-based automated coding of qualitative data. 2010 Presented at: ASIS&t '10 Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem; October 22-27, 2010; Pittsburgh, Pennsylvania p. 1-2 URL: https://pdfs.semanticscholar.org/2f4a/60cb9abf8da062f362c4c47819cc16471bcb.pdf
42. Lefebvre C, Tada Y, Hilfiker SW, Baur C. The assessment of user engagement with eHealth content: the eHealth engagement scale. J Comput Mediat Commun 2010;15(4):666-681. [doi: 10.1111/j.1083-6101.2009.01514.x]
43. Perski O. Osf. 2017. Study protocol: Development and psychometric evaluation of a self-report instrument to measure engagement with digital behaviour change interventions URL: https://osf.io/cj9y7/ [accessed 2018-07-13] [WebCite Cache ID 70sBGZe9w]
44. O'Brien HL, Toms EG. The development and evaluation of a survey to measure user engagement. J Am Soc Inf Sci 2009 Oct 19;61(1):50-69. [doi: 10.1002/asi.21229]
45. Laugwitz B, Held T, Schrepp M. Construction and Evaluation of a User Experience Questionnaire. 2008 Presented at: Symposium of the Austrian HCI and Usability Engineering Group USAB 2008: HCI and Usability for Education and Work; November 20-21, 2008; Graz, Austria p. 63-76. [doi: 10.1007/978-3-540-89350-9_6]
46. Jackson S, Marsh HW. Development and validation of a scale to measure optimal experience: the flow state scale. J Sport Exerc Psychol 1996 Mar;18(1):17-35. [doi: 10.1123/jsep.18.1.17]
47. Direito A, Jiang Y, Whittaker R, Maddison R. Apps for IMproving FITness and increasing physical activity among young people: the AIMFIT pragmatic randomized controlled trial. J Med Internet Res 2015 Aug 27;17(8):e210 [FREE Full text] [doi: 10.2196/jmir.4568] [Medline: 26316499]
48. Boase J, Ling R. Measuring mobile phone use: self-report versus log data. J Comput-Mediat Comm 2013 Jun 10;18(4):508-519. [doi: 10.1111/jcc4.12021]
49. Scharkow M. The accuracy of self-reported internet use—a validation study using client log data. Commun Methods Meas 2016 Mar 24;10(1):13-27. [doi: 10.1080/19312458.2015.1118446]
50. Sigerson L, Cheng C. Scales for measuring user engagement with social network sites: A systematic review of psychometric properties. Comput Human Behav 2018 Jun;83:87-105. [doi: 10.1016/j.chb.2018.01.023]
51. MacKinnon D, Fairchild AJ, Fritz MS. Mediation analysis. Annu Rev Psychol 2007;58:593-614 [FREE Full text] [doi: 10.1146/annurev.psych.58.110405.085542] [Medline: 16968208]

52. Murray J, Brennan SF, French DP, Patterson CC, Kee F, Hunter RF. Mediators of behavior change maintenance in physical activity interventions for young and middle-aged adults: a systematic review. Ann Behav Med 2018 May 18;52(6):513-529. [doi: 10.1093/abm/kay012] [Medline: 29672666]
53. Rhodes R, Pfaeffli LA. Mediators of physical activity behaviour change among adult non-clinical populations: a review update. Int J Behav Nutr Phys Act 2010 May 11;7:37 [FREE Full text] [doi: 10.1186/1479-5868-7-37] [Medline: 20459781]
54. Dewar DL, Lubans DR, Morgan PJ, Plotnikoff RC. Development and evaluation of social cognitive measures related to adolescent physical activity. J Phys Act Health 2013 May;10(4):544-555. [doi: 10.1186/1479-5868-9-36] [Medline: 22975481]
55. Rhodes R, Hunt Matheson D, Mark R. Evaluation of social cognitive scaling response options in the physical activity domain. Meas Phys Educ Exerc Sci 2010 Jul 28;14(3):137-150. [doi: 10.1080/1091367X.2010.495539]
56. Hall E, Chai W, Koszewski W, Albrecht J. Development and validation of a social cognitive theory-based survey for elementary nutrition education program. Int J Behav Nutr Phys Act 2015 Apr 09;12:47 [FREE Full text] [doi: 10.1186/s12966-015-0206-4] [Medline: 25886631]
57. Francis J, Johnston M, Eccles M, Walker A, Grimshaw JM, Foy R, et al. Constructing questionnaires based on the theory of planned behaviour: a manual for Health Services Researchers. Newcastle upon Tyne, UK: Centre for Health Service Research; 2004. URL: http://openaccess.city.ac.uk/1735/1/TPB%20Manual%20FINAL%20May2004.pdf [WebCite Cache ID 73dOb6cML]
58. Bandura A. Guide for constructing self-efficacy scales. J Phys Act Res 2006:307-337. [doi: 10.12691/jpar-2-1-2]
59. Crutzen R, Peters GJ. Scale quality: alpha is an inadequate estimate and factor-analytic evidence is needed first of all. Health Psychol Rev 2017 Dec;11(3):242-247. [doi: 10.1080/17437199.2015.1124240] [Medline: 26602990]
60. Doherty K, Doherty G. The construal of experience in HCI: uUnderstanding self-reports. Int J Hum Comput Stud 2018 Feb;110:63-74. [doi: 10.1016/j.ijhcs.2017.10.006]
61. Reis HT. Why Researchers Should Think “Real World”: A Conceptual Rationale. In: Mehl MR, Conner TS, editors. Handbook of Research Methods for Studying Daily Life. New York: The Guilford Press; 2013:3-22.
62. Hernandez J, McDuff D, Infante C, Maes P, Quigley K, Picard R. Wearable ESM: differences in the experience sampling method across wearable devices. 2016 Presented at: MobileHCI '16 Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services; September 6 to 9, 2016; Florance, Italy p. 195-205.
63. Dunton GF, Liao Y, Intille SS, Spruijt-Metz D, Pentz M. Investigating children's physical activity and sedentary behavior using ecological momentary assessment with mobile phones. Obesity (Silver Spring) 2011 Jun;19(6):1205-1212 [FREE Full text] [doi: 10.1038/oby.2010.302] [Medline: 21164502]
64. Fanning J, Mackenzie M, Roberts S, Crato I, Ehlers D, McAuley E. Physical activity, mind wandering, affect, and sleep: an ecological momentary assessment. JMIR Mhealth Uhealth 2016 Aug 31;4(3):e104 [FREE Full text] [doi: 10.2196/mhealth.5855] [Medline: 27580673]
65. Chung J, Gardner HJ. Temporal presence variation in immersive computer games. Int J Hum Comput Stud 2012 Aug;28(8):511-529. [doi: 10.1080/10447318.2011.627298]
66. Wen C, Schneider S, Stone AA, Spruijt-Metz D. Compliance with mobile ecological momentary assessment protocols in children and adolescents: a systematic review and meta-analysis. J Med Internet Res 2017 Dec 26;19(4):e132 [FREE Full text] [doi: 10.2196/jmir.6641] [Medline: 28446418]
67. Cain AE, Depp CA, Jeste DV. Ecological momentary assessment in aging research: a critical review. J Psychiatr Res 2009 Jul;43(11):987-996 [FREE Full text] [doi: 10.1016/j.jpsychires.2009.01.014] [Medline: 19272611]
68. Modecki KL, Mazza GL. Are we making the most of ecological momentary assessment data? A comment on Richardson, Fuller-Tyszkiewicz, O'Donnell, Ling, & Staiger, 2017. Health Psychol Rev 2017 Dec;11(3):295-297. [doi: 10.1080/17437199.2017.1347513] [Medline: 28651484]
69. Richardson B, Fuller-Tyszkiewicz M, O'Donnell R, Ling M, Staiger PK. Regression tree analysis of ecological momentary assessment data. Health Psychol Rev 2017 Dec;11(3):235-241. [doi: 10.1080/17437199.2017.1343677] [Medline: 28618838]
70. Ginexi EM, Riley W, Atienza AA, Mabry PL. The promise of intensive longitudinal data capture for behavioral health research. Nicotine Tob Res 2014 May;16 Suppl 2:S73-S75 [FREE Full text] [doi: 10.1093/ntr/ntt273] [Medline: 24711629]
71. Hamaker E, Wichers M. No Time Like the Present. Curr Dir Psychol Sci 2017 Feb 08;26(1):10-15. [doi: 10.1177/0963721416666518]
72. Burke L, Shiffman S, Music E, Styn MA, Kriska A, Smailagic A, et al. Ecological momentary assessment in behavioral research: addressing technological and human participant challenges. J Med Internet Res 2017 Dec 15;19(3):e77 [FREE Full text] [doi: 10.2196/jmir.7138] [Medline: 28298264]
73. Kelders S, Van Gemert-Pijnen JE, Werkman A, Nijland N, Seydel ER. Effectiveness of a Web-based intervention aimed at healthy dietary and physical activity behavior: a randomized controlled trial about users and usage. J Med Internet Res 2011 Apr 14;13(2):e32 [FREE Full text] [doi: 10.2196/jmir.1624] [Medline: 21493191]
74. Danaher BG, Boles SM, Akers L, Gordon JS, Severson HH. Defining participant exposure measures in Web-based health behavior change programs. J Med Internet Res 2006 Aug 30;8(3):e15 [FREE Full text] [doi: 10.2196/jmir.8.3.e15] [Medline: 16954125]

75. Graham ML, Strawderman MS, Demment M, Olson CM. Does usage of an eHealth intervention reduce the risk of excessive gestational weight gain? Secondary analysis from a randomized controlled trial. J Med Internet Res 2017 Dec 09;19(1):e6 [FREE Full text] [doi: 10.2196/jmir.6644] [Medline: 28069560]
76. Mattila E, Lappalainen R, Välkkynen P, Sairanen E, Lappalainen P, Karhunen L, et al. Usage and dose response of a mobile acceptance and commitment therapy app: secondary analysis of the intervention arm of a randomized controlled trial. JMIR Mhealth Uhealth 2016 Jul 28;4(3):e90 [FREE Full text] [doi: 10.2196/mhealth.5241] [Medline: 27468653]
77. Taki S, Lymer S, Russell CG, Campbell K, Laws R, Ong K, et al. Assessing user engagement of an mHealth intervention: development and implementation of the growing healthy app engagement index. JMIR Mhealth Uhealth 2017 Jun 29;5(6):e89 [FREE Full text] [doi: 10.2196/mhealth.7236] [Medline: 28663164]
78. McCallum C, Rooksby J, Gray CM. Evaluating the impact of physical activity apps and wearables: interdisciplinary review. JMIR Mhealth Uhealth 2018 Mar 23;6(3):e58 [FREE Full text] [doi: 10.2196/mhealth.9054] [Medline: 29572200]
79. Baltierra NB, Muessig KE, Pike EC, LeGrand S, Bull SS, Hightow-Weidman LB. More than just tracking time: Complex measures of user engagement with an internet-based health promotion intervention. J Biomed Inform 2016 Feb;59:299-307 [FREE Full text] [doi: 10.1016/j.jbi.2015.12.015] [Medline: 26732997]
80. Barisic A, Leatherdale ST, Kreiger N. Importance of frequency, intensity, time and type (FITT) in physical activity assessment for epidemiological research. Can J Public Health 2011;102(3):174-175. [Medline: 21714314]
81. Funk KL, Stevens VJ, Appel LJ, Bauck A, Brantley PJ, Champagne CM, et al. Associations of internet website use with weight change in a long-term weight loss maintenance program. J Med Internet Res 2010 Jul 27;12(3):e29 [FREE Full text] [doi: 10.2196/jmir.1504] [Medline: 20663751]
82. Arden-Close EJ, Smith E, Bradbury K, Morrison L, Dennison L, Michaelides D, et al. A visualization tool to analyse usage of web-based interventions: the example of positive online weight reduction (POWeR). JMIR Hum Factors 2015 May 19;2(1):e8 [FREE Full text] [doi: 10.2196/humanfactors.4310] [Medline: 27026372]
83. Manwaring JL, Bryson SW, Goldschmidt AB, Winzelberg AJ, Luce KH, Cunning D, et al. Do adherence variables predict outcome in an online program for the prevention of eating disorders? J Consult Clin Psychol 2008 Apr;76(2):341-346. [doi: 10.1037/0022-006X.76.2.341] [Medline: 18377129]
84. van Mierlo T. The 1% rule in four digital health social networks: an observational study. J Med Internet Res 2014 Feb 04;16(2):e33 [FREE Full text] [doi: 10.2196/jmir.2966] [Medline: 24496109]
85. Forbes C, Blanchard CM, Mummery WK, Courneya KS. Feasibility and preliminary efficacy of an online intervention to increase physical activity in Nova Scotian cancer survivors: a randomized controlled trial. JMIR Cancer 2015 Nov 23;1(2):e12 [FREE Full text] [doi: 10.2196/cancer.4586] [Medline: 28410166]
86. Short CE, Rebar A, James EL, Duncan MJ, Courneya KS, Plotnikoff RC, et al. How do different delivery schedules of tailored web-based physical activity advice for breast cancer survivors influence intervention use and efficacy? J Cancer Surviv 2017 Feb;11(1):80-91. [doi: 10.1007/s11764-016-0565-0] [Medline: 27498099]
87. Hales SB, Davidson C, Turner-McGrievy GM. Varying social media post types differentially impacts engagement in a behavioral weight loss intervention. Transl Behav Med 2014 Dec;4(4):355-362 [FREE Full text] [doi: 10.1007/s13142-014-0274-z] [Medline: 25584084]
88. Couper MP, Alexander GL, Zhang N, Little RJ, Maddy N, Nowak MA, et al. Engagement and retention: measuring breadth and depth of participant use of an online intervention. J Med Internet Res 2010 Nov 18;12(4):e52 [FREE Full text] [doi: 10.2196/jmir.1430] [Medline: 21087922]
89. Mohr DC, Duffecy J, Ho J, Kwasny M, Cai X, Burns MN, et al. A randomized controlled trial evaluating a manualized TeleCoaching protocol for improving adherence to a web-based intervention for the treatment of depression. PLoS One 2013;8(8):e70086 [FREE Full text] [doi: 10.1371/journal.pone.0070086] [Medline: 23990896]
90. Cussler EC, Teixeira PJ, Going SB, Houtkooper LB, Metcalfe LL, Blew RM, et al. Maintenance of weight loss in overweight middle-aged women through the Internet. Obesity (Silver Spring) 2008 May;16(5):1052-1060 [FREE Full text] [doi: 10.1038/oby.2008.19] [Medline: 18309301]
91. Glasgow RE, Christiansen SM, Kurz D, King DK, Woolley T, Faber AJ, et al. Engagement in a diabetes self-management website: usage patterns and generalizability of program use. J Med Internet Res 2011 Jan 25;13(1):e9 [FREE Full text] [doi: 10.2196/jmir.1391] [Medline: 21371992]
92. Davies C, Corry K, Van Itallie A, Vandelanotte C, Caperchione C, Mummery WK. Prospective associations between intervention components and website engagement in a publicly available physical activity website: the case of 10,000 Steps Australia. J Med Internet Res 2012 Jan 11;14(1):e4 [FREE Full text] [doi: 10.2196/jmir.1792] [Medline: 22260810]
93. Kim JY, Wineinger NE, Taitel M, Radin JM, Akinbosoye O, Jiang J, et al. Self-Monitoring Utilization Patterns Among Individuals in an Incentivized Program for Healthy Behaviors. J Med Internet Res 2016 Dec 17;18(11):e292 [FREE Full text] [doi: 10.2196/jmir.6371] [Medline: 27856407]
94. Crutzen R, Roosjen JL, Poelman J. Using Google Analytics as a process evaluation method for Internet-delivered interventions: an example on sexual health. Health Promot Int 2013 Mar;28(1):36-42. [doi: 10.1093/heapro/das008] [Medline: 22377974]

95. Scherer EA, Ben-Zeev D, Li Z, Kane JM. Analyzing mHealth Engagement: joint models for intensively collected user engagement data. JMIR Mhealth Uhealth 2017 Jan 12;5(1):e1 [FREE Full text] [doi: 10.2196/mhealth.6474] [Medline: 28082257]
96. Fan W, Bifet A. Mining big data. SIGKDD Explor Newsl 2013 Apr 30;14(2):1-5. [doi: 10.1145/2481244.2481246]
97. Cugelman B, Thelwall M, Dawes P. Online interventions for social marketing health behavior change campaigns: a meta-analysis of psychological architectures and adherence factors. J Med Internet Res 2011 Feb 14;13(1):e17 [FREE Full text] [doi: 10.2196/jmir.1367] [Medline: 21320854]
98. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
99. Kidwell K, Hyde LW. Adaptive interventions and SMART designs: application to child behavior research in a community setting. Am J Eval 2016 Sep;37(3):344-363 [FREE Full text] [doi: 10.1177/1098214015617013] [Medline: 28239254]
100. Collins LM, Murphy SA, Bierman KL. A conceptual framework for adaptive preventive interventions. Prev Sci 2004 Sep;5(3):185-196 [FREE Full text] [doi: 10.1023/B:PREV.0000037641.26017.00] [Medline: 15470938]
101. Althoff T, Sosič R, Hicks JL, King AC, Delp SL, Leskovec J. Large-scale physical activity data reveal worldwide activity inequality. Nature 2017 Dec 20;547(7663):336-339 [FREE Full text] [doi: 10.1038/nature23018] [Medline: 28693034]
102. Mohr D, Schueller SM, Riley WT, Brown CH, Cuijpers P, Duan N, et al. Trials of intervention principles: evaluation methods for evolving behavioral intervention technologies. J Med Internet Res 2015 Jul 08;17(7):e166 [FREE Full text] [doi: 10.2196/jmir.4391] [Medline: 26155878]
103. Jacobs M, Graham A. Iterative development and evaluation methods of mHealth behavior change interventions. Curr Opin Psychol 2016 Jun;9:33-37. [doi: 10.1016/j.copsyc.2015.09.001]
104. Vandelanotte C, Duncan MJ, Kolt GS, Caperchione CM, Savage TN, Van Itallie A, et al. More real-world trials are needed to establish if web-based physical activity interventions are effective. Br J Sports Med 2018 Jul 03 Epub ahead of print. [doi: 10.1136/bjsports-2018-099437] [Medline: 29970409]
105. Collins L, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007 May;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
106. Shaw R, Steinberg DM, Zullig LL, Bosworth HB, Johnson CM, Davis LL. mHealth interventions for weight loss: a guide for achieving treatment fidelity. J Am Med Inform Assoc 2014;21(6):959-963 [FREE Full text] [doi: 10.1136/amiajnl-2013-002610] [Medline: 24853065]
107. Cornet VP, Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. J Biomed Inform 2018 Jan;77:120-132. [doi: 10.1016/j.jbi.2017.12.008] [Medline: 29248628]
108. Barkhuus L, Dey A. Location-Based Services for Mobile Telephony: a Study of Users' Privacy Concerns (2003). 2003 Presented at: 9TH IFIP TC13 International Conference On Human-Computer Interaction, Interact 2003; 1st-5th September, 2003; Zürich, Switzerland.
109. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. J Am Med Assoc 2015 Feb 10;313(6):625-626. [doi: 10.1001/jama.2014.17841] [Medline: 25668268]
110. Gordon BA, Bruce L, Benson AC. Physical activity intensity can be accurately monitored by smartphone global positioning system 'app'. Eur J Sport Sci 2016 Aug;16(5):624-631. [doi: 10.1080/17461391.2015.1105299] [Medline: 26505223]
111. Maher C, 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 14;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
112. Hale TM, Pathipati AS, Zan S, Jethwani K. Representation of health conditions on Facebook: content analysis and evaluation of user engagement. J Med Internet Res 2014 Aug 04;16(8):e182 [FREE Full text] [doi: 10.2196/jmir.3275] [Medline: 25092386]
113. D'heer E, Verdegem P. What social media data mean for audience studies: a multidimensional investigation of Twitter use during a current affairs TV programme. Inform Comm Soc 2014 Sep 22;18(2):221-234. [doi: 10.1080/1369118X.2014.952318]
114. Ryan J, Edney S, Maher C. Engagement, compliance and retention with a gamified online social networking physical activity intervention. Transl Behav Med 2017 Dec;7(4):702-708 [FREE Full text] [doi: 10.1007/s13142-017-0499-8] [Medline: 28523603]
115. Bouman M, Drossaert CH, Pieterse ME. Mark My Words: the design of an innovative methodology to detect and analyze interpersonal health conversations in web and social media. J Technol Hum Serv 2012 Jul;30(3-4):312-326. [doi: 10.1080/15228835.2012.743394]
116. Delahaye Paine K. Measure What Matters: Online Tools for Understanding Customers, Social Media, Engagement, and Key Relationships. United Kingdom: John Wiley and Sons Ltd, United Kingdom; 2018.
117. Murdough C. Social Media Measurement. J Interact Advert 2009 Sep;10(1):94-99. [doi: 10.1080/15252019.2009.10722165]
118. Peters GY, Crutzen R. Pragmatic nihilism: how a Theory of Nothing can help health psychology progress. Health Psychol Rev 2017 Dec;11(2):103-121. [doi: 10.1080/17437199.2017.1284015] [Medline: 28110627]

119. Dirican A, Göktürk M. Psychophysiological measures of human cognitive states applied in human computer interaction. Procedia Comp Sci 2011;3:1361-1367. [doi: 10.1016/j.procs.2011.01.016]
120. Cowley B, Filetti M, Lukander K, Torniainen J, Henelius A, Ahonen L, et al. The Psychophysiology Primer: a guide to methods and a broad review with a focus on human-computer interaction. Found Trends Hum Comp Interact 2015;9(3-4):151-308.
121. Ganglbauer E, Schrammel J, Deutsch S, Tscheligi M. Applying psychophysiological methods for measuring user experience: possibilities, challenges and feasibility. 2009 Presented at: User Experience Evaluation Methods in Product Development (UXEM'09) in conjunction with Interact'09; August 24-28, 2009; Sweden.
122. Harmat L, de Manzano Ö, Theorell T, Högman L, Fischer H, Ullén F. Physiological correlates of the flow experience during computer game playing. Int J Psychophysiol 2015 Jul;97(1):1-7. [doi: 10.1016/j.ijpsycho.2015.05.001] [Medline: 25956190]
123. Burns C, Fairclough SH. Use of auditory event-related potentials to measure immersion during a computer game. Int J Hum Comput Stud 2015 Jan;73(Supplement C):107-114. [doi: 10.1016/j.ijhcs.2014.09.002]
124. Martey R, Kenski K, Folkestad J, Feldman L, Gordis E, Shaw A, et al. Measuring Game Engagement. Simul Gaming 2014 Nov 04;45(4-5):528-547. [doi: 10.1177/1046878114553575]
125. Dhamija S, Boult TE. Automated mood-aware engagement prediction. 2017 Presented at: Seventh International Conference on Affective Computing Intelligent Interaction (ACII); 23-26 October, 2017; San Antonio, TX, USA.
126. Huynh S, Kim S, Ko J, Balan RK, Lee Y. EngageMon. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Mar 26;2(1):1-27. [doi: 10.1145/3191745]
127. Bevilacqua F, Engström H, Backlund P. Changes in heart rate and facial actions during a gaming session with provoked boredom and stress. Entertainment Computing 2018 Jan;24:10-20. [doi: 10.1016/j.entcom.2017.10.004]
128. Reinecke K, Cordes M, Lerch C, Koutsandréou F, Schubert M, Weiss M, et al. From lab to field conditions: a pilot study on EEG methodology in applied sports sciences. Appl Psychophysiol Biofeedback 2011 Dec;36(4):265-271. [doi: 10.1007/s10484-011-9166-x] [Medline: 21800184]
129. Kayser J, Tenke CE. Issues and considerations for using the scalp surface Laplacian in EEG/ERP research: A tutorial review. Int J Psychophysiol 2015 Sep;97(3):189-209 [FREE Full text] [doi: 10.1016/j.ijpsycho.2015.04.012] [Medline: 25920962]
130. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 2004 Mar 15;134(1):9-21. [doi: 10.1016/j.jneumeth.2003.10.009] [Medline: 15102499]
131. Luck SJ. An Introduction to the Event-Related Potential Technique. Cambridge, MA: MIT Press; 2005.
132. Leiker A, Miller M, Brewer L, Nelson M, Siow M, Lohse K. The relationship between engagement and neurophysiological measures of attention in motion-controlled video games: a randomized controlled trial. JMIR Serious Games 2016 Apr 21;4(1):e4 [FREE Full text] [doi: 10.2196/games.5460] [Medline: 27103052]
133. Dyke FB, Leiker AM, Grand KF, Godwin MM, Thompson AG, Rietschel JC, et al. The efficacy of auditory probes in indexing cognitive workload is dependent on stimulus complexity. Int J Psychophysiol 2015 Jan;95(1):56-62. [doi: 10.1016/j.ijpsycho.2014.12.008] [Medline: 25528402]
134. Takeda Y, Okuma T, Kimura M, Kurata T, Takenaka T, Iwaki S. Electrophysiological measurement of interest during walking in a simulated environment. Int J Psychophysiol 2014 Sep;93(3):363-370. [doi: 10.1016/j.ijpsycho.2014.05.012] [Medline: 24892726]
135. Rayner K. Eye movements in reading and information processing: 20 years of research. Psychol Bull 1998 Nov;124(3):372-422. [Medline: 9849112]
136. Hermans D, Vansteenwegen D, Eelen P. Eye movement registration as a continuous index of attention deployment: data from a group of spider anxious students. Cognition & Emotion 1999 Jul;13(4):419-434. [doi: 10.1080/026999399379249]
137. Jennett C, Cox A, Cairns P, Dhoparee S, Epps A, Tijs T, et al. Measuring and defining the experience of immersion in games. Int J Hum Comput Stud 2008 Sep;66(9):641-661. [doi: 10.1016/j.ijhcs.2008.04.004]
138. Alley S, Jennings C, Persaud N, Plotnikoff RC, Horsley M, Vandelanotte C. Do personally tailored videos in a web-based physical activity intervention lead to higher attention and recall? - An eye-tracking study. Front Public Health 2014;2:13 [FREE Full text] [doi: 10.3389/fpubh.2014.00013] [Medline: 24575398]
139. Crutzen R, Cyr D, Larios H, Ruiter RA, de Vries NK. Social presence and use of internet-delivered interventions: a multi-method approach. PLoS One 2013;8(2):e57067 [FREE Full text] [doi: 10.1371/journal.pone.0057067] [Medline: 23437310]
140. Nacke L, Grimshaw MN, Lindley CA. More than a feeling: measurement of sonic user experience and psychophysiology in a first-person shooter game. Interact Comput 2010 Sep;22(5):336-343. [doi: 10.1016/j.intcom.2010.04.005]

Abbreviations

EDA: electrodermal activity
EEG: electroencephalography
eHealth: electronic health

EMA: ecological momentary assessment
EMG: electromyography
ERP: event-related potentials
FITT: frequency, intensity, time, and type
IP: internet protocol
mHealth: mobile health
NHMRC: National Health and Medical Research Council

Edited by G Eysenbach; submitted 15.11.17; peer-reviewed by O Perski, V Pekurinen, K Frie, C Yeager; comments to author 15.03.18; revised version received 01.08.18; accepted 10.09.18; published 16.11.18.

Please cite as:
Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, Müller AM, Wark PA, Vandelanotte C, Poppe L, Hingle MD, Crutzen R
Measuring Engagement in eHealth and mHealth Behavior Change Interventions: Viewpoint of Methodologies
J Med Internet Res 2018;20(11):e292
URL: http://www.jmir.org/2018/11/e292/
doi:10.2196/jmir.9397
PMID:30446482

Original Paper

Involuntary Attention Restoration During Exposure to Mobile-Based 360° Virtual Nature in Healthy Adults With Different Levels of Restorative Experience: Event-Related Potential Study

Kyungmi Chung[1,2], PhD; Daeho Lee[3], PhD; Jin Young Park[1,2], MD, PhD

[1]Department of Psychiatry, Yonsei University College of Medicine, Gangnam Severance Hospital, Seoul, Republic of Korea

[2]Institute of Behavioral Science in Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

[3]Department of Interaction Science, Sungkyunkwan University, Seoul, Republic of Korea

Corresponding Author:
Jin Young Park, MD, PhD
Department of Psychiatry
Yonsei University College of Medicine, Gangnam Severance Hospital
611, Eonju-ro
Gangnam-gu
Seoul, 06273
Republic of Korea
Phone: 82 2 2019 3342
Fax: 82 2 3462 4304
Email: empathy@yuhs.ac

Abstract

Background: With the global trend of urbanization, there are increasing reports of a possible association between decreased exposure to nature and increased occurrence of mental disorders. New 360° virtual reality (VR) technology using smartphones and portable VR glasses can overcome spatial and temporal limitations to help people deal with mental fatigue in everyday life.

Objective: On the basis of attention restoration theory (ART), this study aimed to investigate whether the amplitude of the mismatch negativity (MMN)/P3a complex could act as an event-related potential (ERP) biomarker of involuntary attention restoration during exposure to 360° virtual nature in healthy young adults with different levels of restorative VR experience.

Methods: A total of 40 healthy adults completed prequestionnaires on demographics and simulator sickness and postquestionnaires on simulator sickness and perceived restorativeness before and after exposure to virtual nature, respectively. During the VR exposure, brain activity was measured by electroencephalography as participants were asked to conduct a 2-tone passive auditory oddball task.

Results: The amplitude and latency of the MMN/P3a complex were compared between individuals reporting a highly restorative experience and those reporting a less restorative experience. Although viewing a virtual nature environment, the high restorative group (N=19) exhibited significantly reduced P3a amplitudes compared with the low restorative group (N=20; t_{38}=2.57; P=.02; d=0.59). Particularly, a moderate but significant negative correlation was found between the self-reported restorativeness scores and the P3a amplitudes at the fronto-central region (r=−.38; P=.02). However, the latency of the MMN/P3a complex did not significantly differ between the 2 groups (auditory mismatch negativity: t_{38}=−1.47; P=.15 and P3a: t_{38}=−0.31; P=.76).

Conclusions: Considering individuals' restorative experience, the amplitude of the fronto-central MMN/P3a complex can potentially be employed as a distinct ERP component of interest in involuntary attention restoration during virtual nature experience in healthy young adults. The findings for the 360° virtual nature experience seem to be consistent with those of previous ERP studies on the effects of meditation practice. This study extends the findings of previous ART and ERP studies of real-world meditation, restoration, and mental fatigue management into the virtual world created by mobile phone–based VR glasses and 360° video content.

(J Med Internet Res 2018;20(11):e11152) doi:10.2196/11152

KEYWORDS

smartphone; virtual reality; attention; surveys and questionnaires; electroencephalography; evoked potentials

Introduction

Background

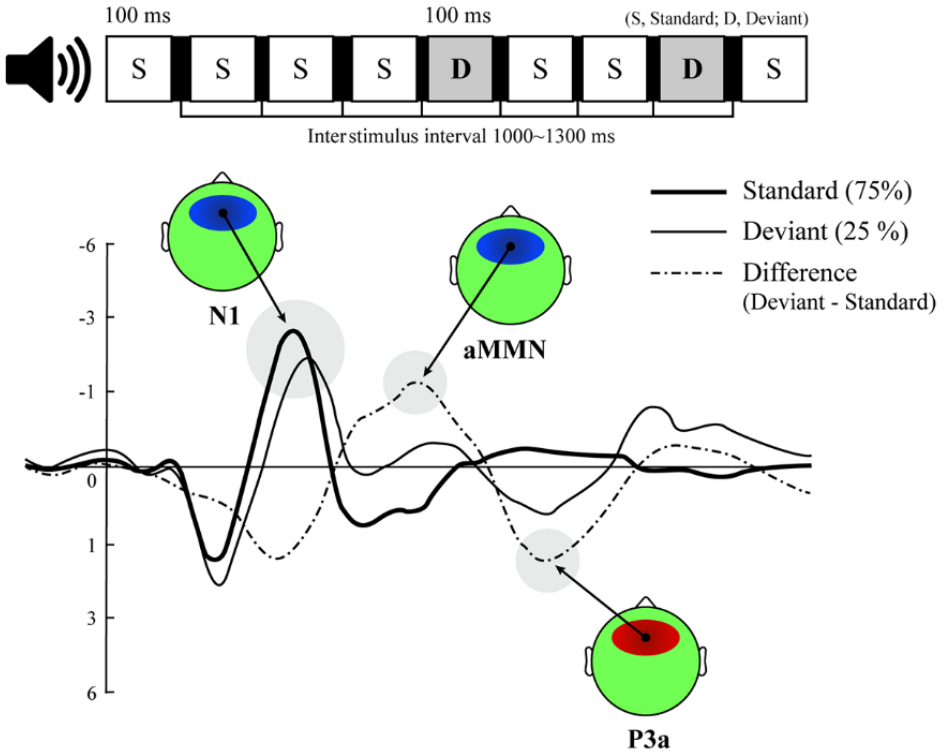
Although some people prefer to enjoy attention-drawing sport or entertainment events to recover from mental fatigue, why do others prefer to walk in or just be exposed to natural environments and why is nature so well known for its restorative effects? From the perspective of the attention restoration theory (ART) [1], exposure to nature serves as a vital ingredient in healthy human functioning, contributing to the replenishment of depleted attentional resources. Involuntary attention, also called *fascination*, is a central component of a restorative experience. Although involuntary attention requires no effort and is resistant to fatigue, directed or focused attention requires a capacity for concentration that can become depleted by taxing task demands, leading to stress responses or impaired performance [1-3]. Individuals can rapidly restore depleted attentional resources by using involuntary attention and allowing directed attention to be in a resting state [4]. Fascination ranges along a continuum from *soft* to *hard* [1]. Soft-fascinating activities (eg, walking in a park or seeing a sunset) are sufficiently distracting to hold one's attention, but they leave room for mental reflection and are aesthetically pleasing, which helps to offset the pain that accompanies reflection on serious matters. In contrast, hard-fascinating activities (eg, sports or entertainment) fill the mind and rivet one's attention in an all-consuming fashion, leaving little or no room for mental reflection [1,5]. A stream of hard-fascinating environmental events is considered as an incoherent collection of impressions and not a restorative experience [1]. Hence, a peaceful and moderate nature experience coupled with aesthetic pleasure fosters a fuller, more deeply restorative experience than a sports or entertainment experience does.

Despite the expected restorative effect of interaction with nature, it is difficult for modern people to deal with mental fatigue at the right time because of increasing urbanization as well as spatial, temporal, and social constraints. Prolonged mental effort results in directed attention fatigue [1]. Coincident with the global trend of urbanization, there is increasing evidence of a possible link between decreased exposure to nature and increased occurrence of mental disorders [6-8]. Direct contact with nature, such as walking in natural environments [9-11] and bringing plants and flowers into residential and office environments [12], leads to recovery from attentional fatigue and improvement in cognitive performance. In addition, people who suffer from directed attention fatigue can benefit from indirect contact with nature such as viewing nature through a window [13] or viewing videos or pictures of natural environments [14,15]. Regardless of whether the contact with nature is direct or indirect, it is important for people to have easy access to it so that they can reflect on their everyday life and restore their depleted attention in a timely manner. In that context, virtual reality (VR) technology can help to overcome possible constraints on access to nature; thus, offering a promising therapeutic alternative for individuals who are vulnerable to cognitive overload and are at a higher risk of mental illness.

In recent ART studies, the event-related potential (ERP) method, which uses ERP components as scalp-recorded voltage changes that appear as a series of positive and negative peaks varying in polarity, amplitude, and duration and reflecting a specific neural or psychological process, has yet to be fully implemented to examine the effect of contact with real nature or even virtual nature. Traditional ART studies measure the restorative effects of natural environments on voluntary attention not only on the basis of subjective, self-reported responses such as the Perceived Restorativeness Scale (PRS) [16] but also on the basis of objective, behavioral responses to cognitive tests, particularly immediately after the nature exposure. Compared with those widely used methods, the ERP method bears the advantages of (1) measuring subjective and objective neuroelectrophysiological responses simultaneously during or after the performance of given tasks and (2) exploring the time courses of both early preattentive and late-attentive sensory processing using topographic and principal component analyses. This study focuses on a specific ERP component called mismatch negativity (MMN)/P3a complex, which is observed at fronto-central sites as a combination of the 2 different preattentive ERP components: MMN and P3a (see Figure 1). The negative-going MMN component (usually peaking at approximately 150-250 ms from the onset of the deviant stimuli in a passive auditory oddball paradigm) is followed by the positive-going P3a component (peaking at approximately 220-280 ms), reflecting a subsequent attention-orienting (or shifting) process as the neurophysiological *handover* from the MMN to the P3a [17-19]. Particularly, utilizing the MMN/P3a complex elicited in the absence of attention is appropriate to investigate the restorative effect of exposure to 360° VR nature on depleted involuntary attention, in that people can simultaneously perform a VR-based visual task in a passive auditory oddball paradigm, without intentionally regulating attention. On the basis of its capability for the early detection of sensory stimuli, the ERP method may provide people vulnerable to repetitive stressful events and cognitive overload with the possibility of early intervention with 360° VR exposure therapy and also a way to systematically evaluate the restorative effect of the therapy.

According to ERP findings during meditation, the mean auditory MMN (aMMN) amplitude of long-term mediators was larger than that of nonmediators, indicating that auditory fatigue or cognitive overload tended to reduce the aMMN amplitude [20]. Cahn and Polich [21] revealed that participants who reported more hours of daily meditation practice produced the strongest meditation-induced decrease in P3a amplitude. Given that the P3a reflects frontal cortical activity elicited by engagement of the focal attentional system [22], the reduced P3a amplitude may reflect the disengagement of attentional networks from stimulus-driven activation during meditation, which meets the goal of meditation to decrease brain reactivity to attention-demanding stimuli and evaluative cognitive processing.

Figure 1. Illustration of expected event-related potential components (N1 and mismatch negativity/P3a complex) and their grand-averaged waveforms and topographic maps at the same region of interest (FCz electrode) evoked by a two-tone passive auditory oddball paradigm. aMMN: auditory mismatch negativity.



Unlike the P3a, the aMMN is less likely to decrease significantly in amplitude with restorative activities because the aMMN, which is independent of attention [23], is not affected by the nature of the visual content and the immersiveness of visual tasks (from traditional reading task to VR tasks) [24]. In fact, the modulation of ERP amplitudes during meditative or restorative activities remains unclear because of a lack of investigations in applied ERP research settings. Furthermore, the aMMN and P3a have been investigated only separately, and previous ERP studies on meditation employed inconsistent experimental protocols.

Objectives

The purpose of this study was to demonstrate the potential role of the amplitude of the MMN/P3a complex [17-19] in reflecting involuntary attention restoration during exposure to restorative nature virtual environments (VEs) in healthy young adults with different levels of restorative experience. ERP amplitudes during exposure to 360° virtual nature might be expected to exhibit patterns similar to those observed during meditative activities [20,21]. We hypothesize that compared with people who have a low level of restorative experience, those who have a high level of restorative experience will show a significant reduction in P3a amplitude but not in aMMN amplitude because of the preattentive nature of the aMMN in both the real world [23] and the virtual world [24]. We further hypothesize that there will be a negative correlation between self-reported PRS scores and P3a amplitudes.

Methods

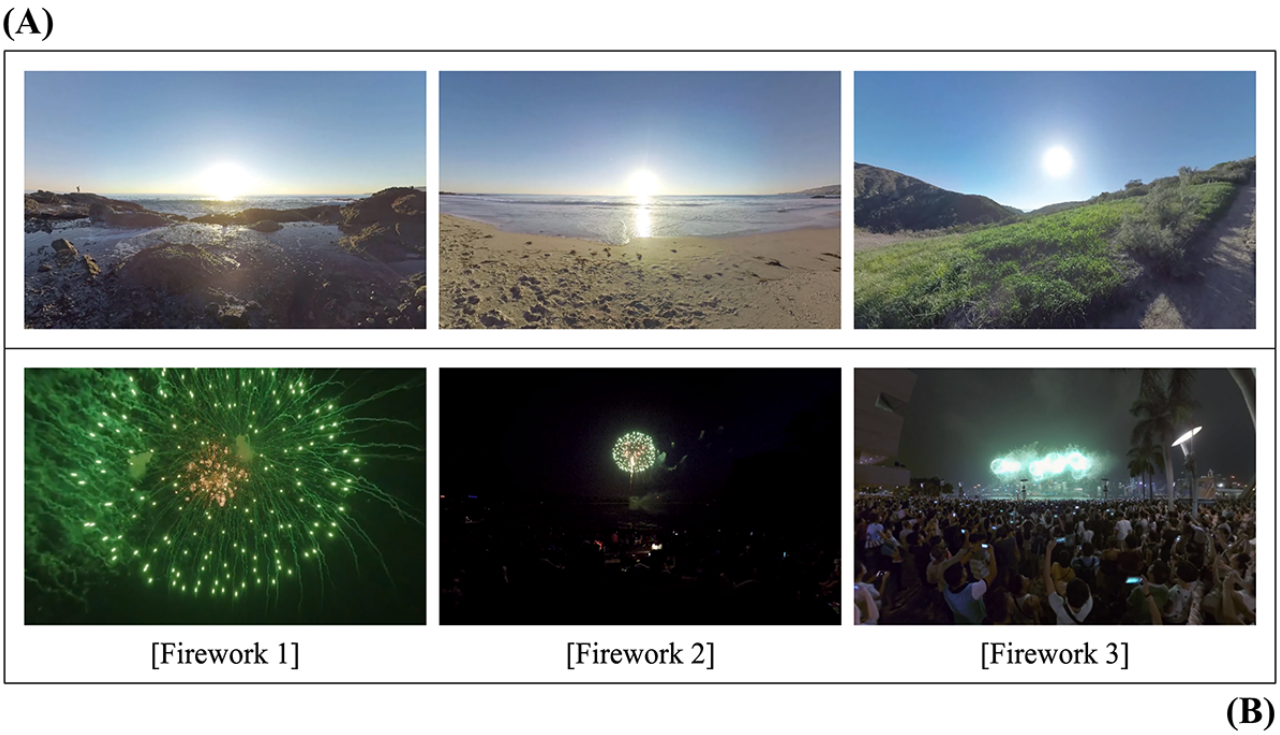
Participants

A total of 40 healthy volunteers (22 males and 18 females) aged from 19 to 36 years (mean 23.78 [SE 0.56]) were recruited in the following 2 ways: (1) an offline advertisement posted on bulletin boards around the Department of Psychiatry at Gangnam Severance Hospital and (2) a Web-based advertisement posted on a single website at Sungkyunkwan University, which has 2 campuses in South Korea. Volunteers who signed informed consent forms approved by the institutional review board (IRB) of Gangnam Severance Hospital, Yonsei University College of Medicine, were enrolled. None of the enrollees subsequently withdrew their participation. All of the enrolled participants had a normal or corrected-to-normal vision and no hearing or color-vision impairments, and none reported brain lesions or previous history of neurological or psychiatric disorders, including any current use of psychotropic medications. Out of 40 participants, 38 were right-handed (38/40, 95%), 1 was ambidextrous (1/40, 3%), and 1 was left-handed (1/40, 3%), as identified by the Edinburgh Handedness Inventory [25]. A total of 14 participants wore glasses, but those participants were required to adjust the VR glasses to their eye condition rather than wear their corrective glasses with the VR glasses. None of the participants reported a prior experience of enjoying 360° VR videos with the LG G5 smartphone-compatible VR glasses used in this study. The detailed demographic information on the enrolled participants is shown in Table 1.

Table 1. Demographic characteristics of the study sample (N=40).

Characteristics	Participants, n (%)
Age (years)	
19-29	38 (95)
30-39	2 (5)
Gender	
Male	22 (55)
Female	18 (45)
Marital status	
Single	39 (98)
Married	1 (3)
Educational level	
Undergraduate	37 (93)
Postgraduate	3 (8)
Employment	
Unemployed	7 (18)
Student	30 (75)
Employed (full-time)	3 (8)
Income	
Low	6 (15)
Lower middle	15 (38)
Upper middle	16 (40)
High	3 (8)
Smoking	
Nonsmoker	34 (85)
Ex-smoker	3 (8)
Smoker	3 (8)
Alcohol drinking	
Never	9 (23)
Once per week	18 (45)
Twice per week	6 (15)
3 times per week	5 (13)
4 times per week	2 (5)
Body mass index (kg/m^2)	
Underweight (<18.50)	1 (3)
Normal (18.50-24.99)	21 (53)
Overweight (23.00-24.99)	11 (28)
Obesity (25.00-29.99)	6 (15)
Extreme obesity (≥30)	1 (3)

Figure 2. Screenshots of (A) three different types of hard-fascinating fireworks virtual environments (VEs) and (B) one mixed soft-fascinating nature virtual environment, including seaside, grassland, and hill locations.



Visual Stimuli

As presented in Figure 2 the restorative nature 360° VE consisted of a variety of seaside, grassland, and hilly scenes. To confirm that the nature VE (with soft fascination) is more restorative than entertainment VEs (with hard fascination), 3 different fireworks VEs were collected (see Figure 2). Fireworks events, such as the annual Seoul International Fireworks festival held in October and the celebration on New Year's Eve, are picturesque, attention-drawing entertainment events in city centers that Koreans and people around the world attend to drive away bad luck and evil spirits and to celebrate the New Year. The fireworks scenes were collected from Google's YouTube. Because only the initial 5 min of a 10-min exposure to nature images was previously found to yield significant physiological responses [15], all of the 4K 360° VR videos were edited to the same running time of 5 min 53 s with the same resolution of 3840×1920 pixels.

Apparatus

Auditory stimuli were delivered via MDR-1A headphones (Sony, Tokyo, Japan) on an OptiPlex 7040 Mini Tower PC (Dell, Round Rock, TX, USA) while the participants focused their attention on the given VR tasks. The experimental paradigm was programmed and presented using E-Prime v.2.0 software (Psychology Software Tools Inc; PST, Pittsburgh, PA, USA). Electroencephalographic (EEG) data were recorded with PST's E-Prime Extensions for Net Station (EENS) v.2.0 and analyzed with the GES 400 system (Electrical Geodesics Inc; EGI, Eugene, OR, USA) using a Net Amps 400 amplifier, a 64-channel HydroCel Geodesic Sensor Net (HCGSN), and Net Station v.5.4 software (ie, Net Station Acquisition / Review / Tools) run by an Apple's MacBook Pro (Apple Inc, Cupertino, CA, USA).

The mobile VR system (LG Electronics, Seoul, South Korea) consisted of an LG G5 smartphone and LG 360° VR glasses (960×720 pixels at 639 ppi per eye), which are only compatible with each other. In terms of comfort and wearability, the LG 360 VR glasses (164.1×185.6×45.9 mm; 134.3 g) were more appropriate in the experimental setting than the Samsung Gear VR released in 2015 (201.9×116.4×92.6 mm; 318 g, headset only), whose total weight was increased to approximately 480 g by the weight of the required smartphone. The LG VR glasses did not press down heavily on the electrodes placed around the forehead, eyes, and ears during pilot tests. For 360° VR videos to be optimally displayed on VR glasses across each participant's vision, all participants were guided to manually adjust the focal length and inter-pupillary distance (the distance between the centers of the pupils of the 2 eyes). After calibrating the VR glasses, the participants were taught how to control the VR display and how to find, play, and view the 360° VR videos in the user interface.

Experimental Paradigm and Visual Task

This ERP study employed a 2-tone passive auditory oddball paradigm in which frequent standard and infrequent deviant tones were presented with probabilities of 75% (180 trials) and 25% (60 trials), respectively. The 750 Hz standard tone and the 1000 Hz deviant tone were randomly presented at 75 dB sound pressure level for 100 ms, and the interstimulus interval varied randomly between 1000 ms and 1300 ms. Accordingly, the experimental paradigm was designed not to exceed a maximum of 5 min 30 s to prevent the VR stimuli from ending before the auditory oddball paradigm completely ended.

While viewing silent 360° VR videos on a revolving chair, the participants were instructed not to intentionally detect sound changes or discriminate between the rare deviant tones and the

frequent standard tones, which allowed them to focus only on the given visual task and ignore all the visual-task-irrelevant sounds and even any sudden noises in the EEG recording room. The 3 fireworks VEs were presented in a random order, followed by the single-nature VE. In accordance with the IRB-approved protocol presented in Figure 3, an experimenter constantly made sure that each time a VR video stimulus ended, the stimulus caused no inconvenience. Participants who showed any adverse reactions to the given stimuli would be withdrawn from the study as well as from the analyses. None of the enrolled participants reported problems with the EEG and mobile VR system or with the VR stimuli.

Self-Report Measures

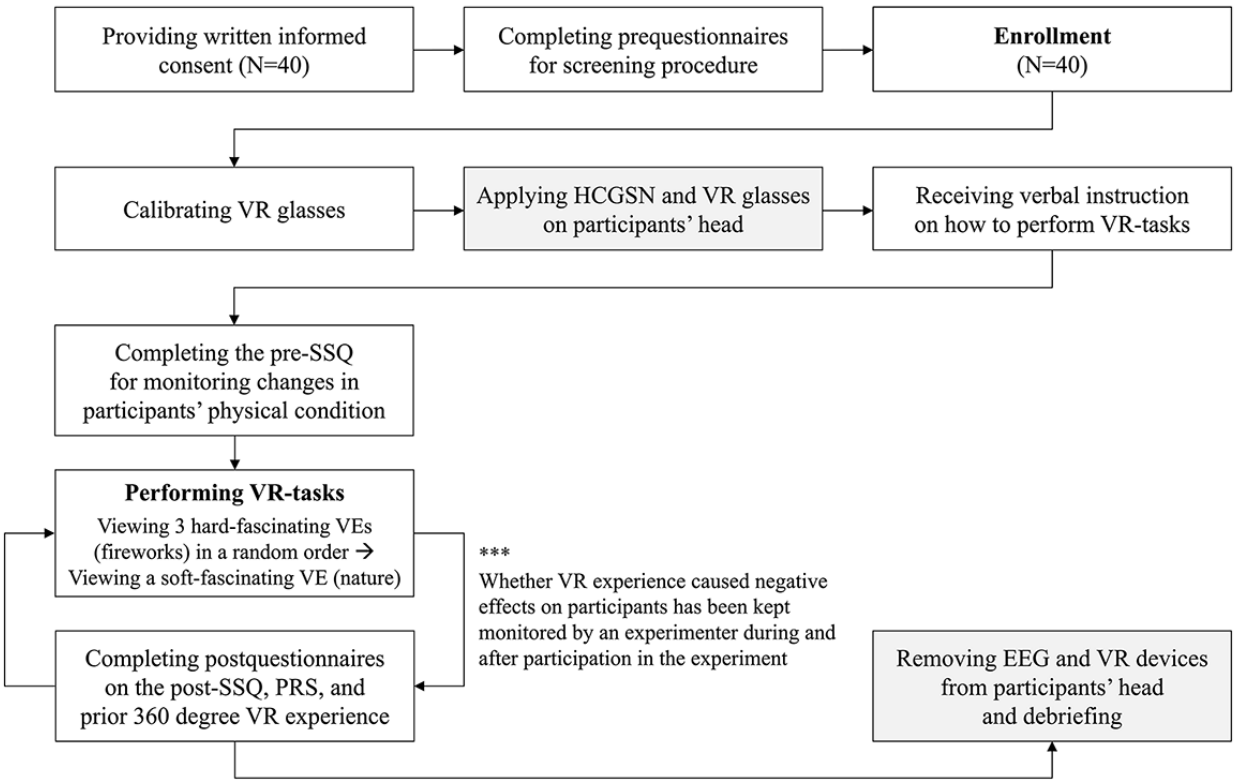
In the context of the 360° VR environment, it is important to determine whether the restorative experience in a nature VE could be threatened by simulator sickness. The Simulator Sickness Questionnaire (SSQ) [26] consists of a total of 16 items that measure the severity of 3 different groups of negative physical symptoms related to the experience of mechanical simulators: nausea (7 items: general discomfort, increased salivation, sweating, nausea, difficulty concentrating, stomach awareness, and burping), oculomotor (7 items: general discomfort, fatigue, headache, eye strain, difficulty focusing, difficulty concentrating, and blurred vision), and disorientation (7 items: difficulty focusing, nausea, fullness of head, blurred vision, dizzy [eye open and eye closed, respectively], and vertigo). Moreover, 3 subfactors of the SSQ contained overlapping symptom items. All items are rated from 0 (none) to 3 (severe). Following the recommendation of Kennedy et al [26], the SSQ was first administered to participants before they performed VR tasks to rule out the possibility of pre-existing symptoms and to measure the baseline physical condition (pre-exposure SSQ). The SSQ was subsequently administered immediately after viewing each 360° VE (postexposure SSQ). In this study, the Korean version of the SSQ was adopted from Min et al [27].

The current version of the PRS is composed of 26 items that measure the following 5 restorative factors: being away (5 items), fascination (8 items, including 2 reversed items), coherence (4 reversed items), compatibility (5 items), and legibility (4 items) [16]. The PRS was administered immediately after each VE to indicate how well certain statements described the participants' restorative experience on a 7-point scale (0=not at all and 6=completely). To compute the total PRS score, the 6 reversed items were inversely coded, and then all the scores for the 5 restorative factors were summed. This study adopted the Korean version of the PRS scale from Yoo et al [28].

Electrophysiological Data Recording and Analysis

During EEG recording, the analog signals were referenced to a single vertex electrode (Cz), filtered with a 0.01 to 400 Hz bandpass filter, and digitized at a sampling rate of 1000 samples per second online. The impedance for all wet electrodes was kept below 30 kΩ. The recommended threshold limit is 50 kΩ to ensure an optimal signal-to-noise ratio for the high-input impedance amplifier [29].

Figure 3. Flowchart of the experimental procedure. VR: virtual reality; HCGSN: HydroCel Geodesic Sensor Net; SSQ: Simulator Sickness Questionnaire; VE: virtual environment; PRS: Perceived Restorativeness Scale; EEG: electroencephalography.



For preprocessing of the EEG signals, all datasets were refiltered with a 0.3 to 30 Hz bandpass filter offline and segmented into epochs ranging from 100 ms before to 500 ms after the onset of each of the 2 stimulus conditions: standard and deviant. On the basis of a timing test to ensure the accuracy of auditory stimulus presentation in E-Prime and EENS, the data were offset by the average offset value of 16 ms. Following the automated algorithm of EGI's Net Station Tools, artifacts such as eye-blinks, eye-movements, and bad channels were detected. If a channel was bad for more than 20% of the segments, the channel was marked bad for all segments. Segments were marked bad if they contained (1) more than 10 bad channels (maximum-minimum>150 μ V for the entire segment, with a moving average of 80 ms), (2) eye-blinks (maximum-minimum>100 μ V, with a moving average of 80 ms), or (3) eye-movements (maximum-minimum>55 μ V, with a moving average of 80 ms). Bad channel replacement was then performed. The data were averaged for individual participants, and baseline corrected from −100 ms to 100 ms. Thereafter, a grand average was calculated using the data from all participants. After that, all of the channels were rereferenced to the average reference offline.

Finally, the deviant-standard difference waves were computed to identify the presence of the MMN/P3a complex component. For a further statistical analysis, the values of the adaptive mean of the aMMN and P3a amplitudes at the FCz electrode site were extracted from the different waveforms within the following time windows: aMMN: 150 to 250 ms; P3a: 220 to 280 ms.

Results

Manipulation Check

For a manipulation check, a one-way repeated measures analysis of variance (RM-ANOVA) was conducted to explore the effect of the type of VE (ie, soft or hard fascination) on the perception of environmental restorativeness. Before the RM-ANOVA, Mauchly test revealed that the assumption of sphericity had been met (Mauchly W=.79; P=.11; ε=.88). There was a significant effect of VE type ($F_{3,117}$=27.18; P<.001; η_p^2=0.41). All 3 Bonferroni-adjusted post hoc comparisons of the PRS scores between the single nature VE and each of the 3 fireworks VEs indicated significant differences, which are nature (mean 101.25 [SE 3.91]) versus fireworks 1 (mean 69.97 [SE 3.57]), P<.001; nature versus fireworks 2 (mean 77.15 [SE 3.93]), P<.001; nature versus fireworks 3 (mean 81.08 [SE 3.85]), P<.001 (see Figure 2). The soft-fascinating nature VE was perceived as more restorative than the averaged hard-fascinating fireworks VE (mean 76.03 [SE 3.28]; t_{39}=7.26; P<.001; Cohen d=0.78).

An ERP analysis of amplitude modulation in the MMN/P3a complex revealed that exposure to the 2 different VEs resulted in significantly different P3a amplitudes (t_{39}=−2.05; P=.048; d=0.35) but not significantly different aMMN amplitudes (t_{39}=0.59; P=.56). The increase in P3a amplitude elicited by unexpected and distracting stimuli was significantly greater during exposure to the soft-fascinating VE (mean 1.67 [SE 0.30]) than during exposure to the hard-fascinating VE (mean 0.91 [SE 0.17]; see Figure 4). However, there was no significant difference in the latency of the MMN/P3a complex between the 2 VEs (aMMN: t_{39}=−1.23; P=.23 and P3a: t_{39}=−0.61; P=.54).

In that respect, the nature VE stimuli were successful in eliciting participants' restorative responses based on both subjective and neurophysiological measures. Furthermore, it was also evident that the aMMN might not depend on attention, and the amplitude of the MMN/P3a complex is more likely than the latency to reflect involuntary attention restoration.

Simulator Sickness as a Control Variable

To clarify the genuine restorative effect of exposure to virtual nature, a one-way RM-ANOVA for the self-rating of simulator sickness before and after VR tasks was performed. As Mauchly test indicated that the assumption of sphericity had been violated (Mauchly W=.54; P=.006), the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε=.80). As no significant difference in all 5 SSQ scores was found ($F_{3.194,124.547}$=2.42; P=.07), it was clear that the participants' physical conditions did not vary with the VR exposure time and stimuli.

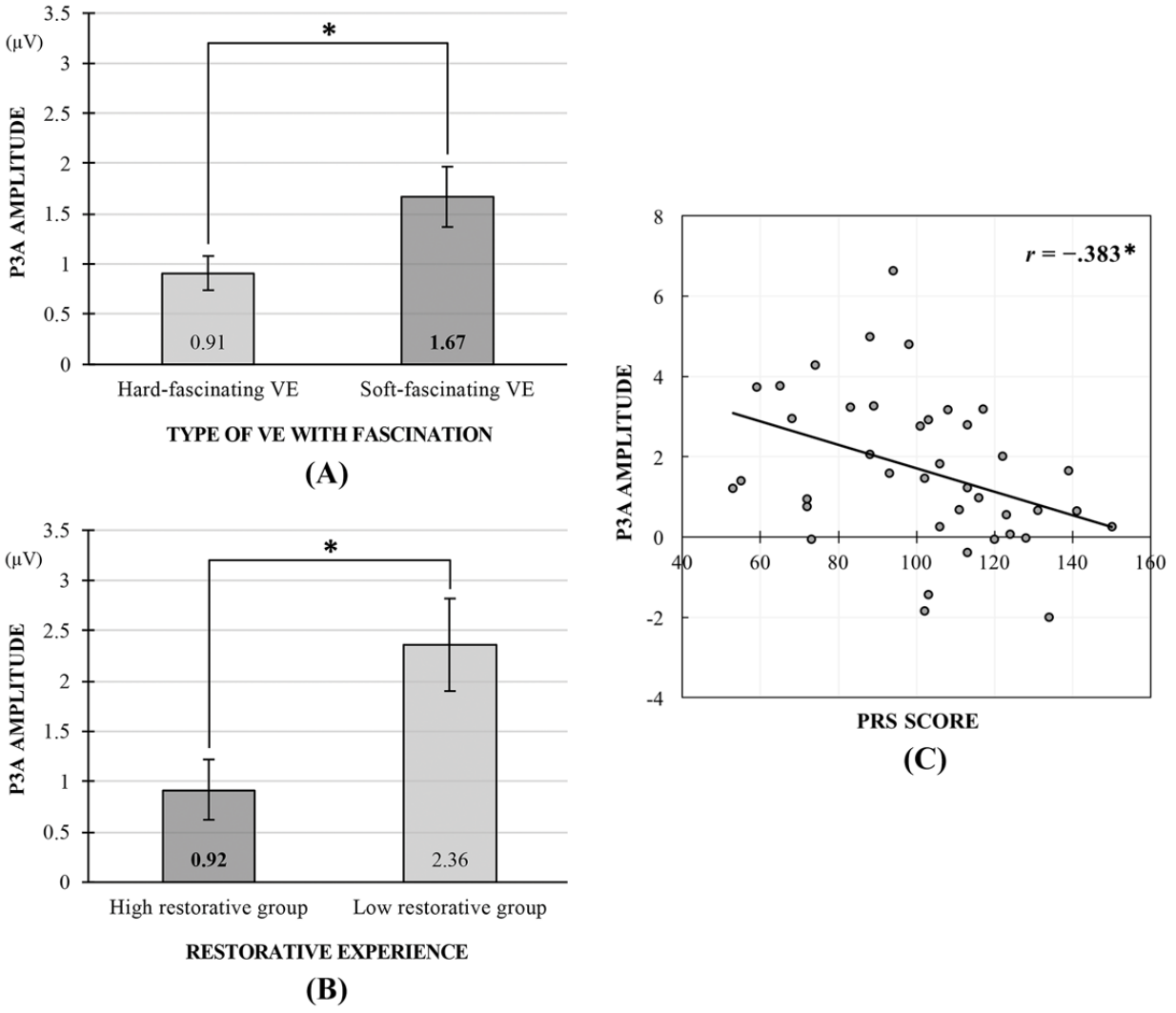
Effect of Group Difference in Perceived Restorativeness on MMN/P3a Complex Modulation

To investigate how individuals' restorative experience during exposure to virtual nature affected the MMN/P3a complex amplitudes, the 40 participants were divided on the basis of the median PRS score (103) into a low restorative group (PRS≤103, N=21; mean 82.62 [SE 3.65]) and a high restorative group (PRS>103, N=19; mean 121.84 [SE 2.88]).

The differences in the MMN/P3a complex amplitudes between the 2 restorative groups were evaluated using an independent samples t test. Whereas the difference in restorative experience failed to induce significantly different aMMN amplitudes (t_{38}=0.91; P=.37), the high restorative group (mean 0.92 [SE 0.30]) showed significantly lower P3a amplitudes than the low restorative group (mean 2.36 [SE 0.46]) when viewing the soft-fascinating nature VE (t_{38}=2.57; P=.02; d=0.59; see Figure 4). Consistent with the result of the manipulation check, the latency of the MMN/P3a complex did not significantly differ between the 2 groups (aMMN: t_{38}=−1.4;, P=.15 and P3a: t_{38}=−0.31; P=.76).

As the ERP reflection of normal cognitive functioning and involuntary attention restoration, the fronto-central N1/aMMN and P3a components were prominently produced by deviant tone detection while viewing 360° virtual nature. In particular, the high restorative group exhibited more attenuated fronto-central P3a activity in the topographical distributions than the low restorative group, which is consistent with decreased involvement of the frontal cortex in response to auditory distractors during meditation (see Figure 5). Figure 5 shows the ERP time courses of the 2 groups with different levels of restorative experience at the FCz electrode site.

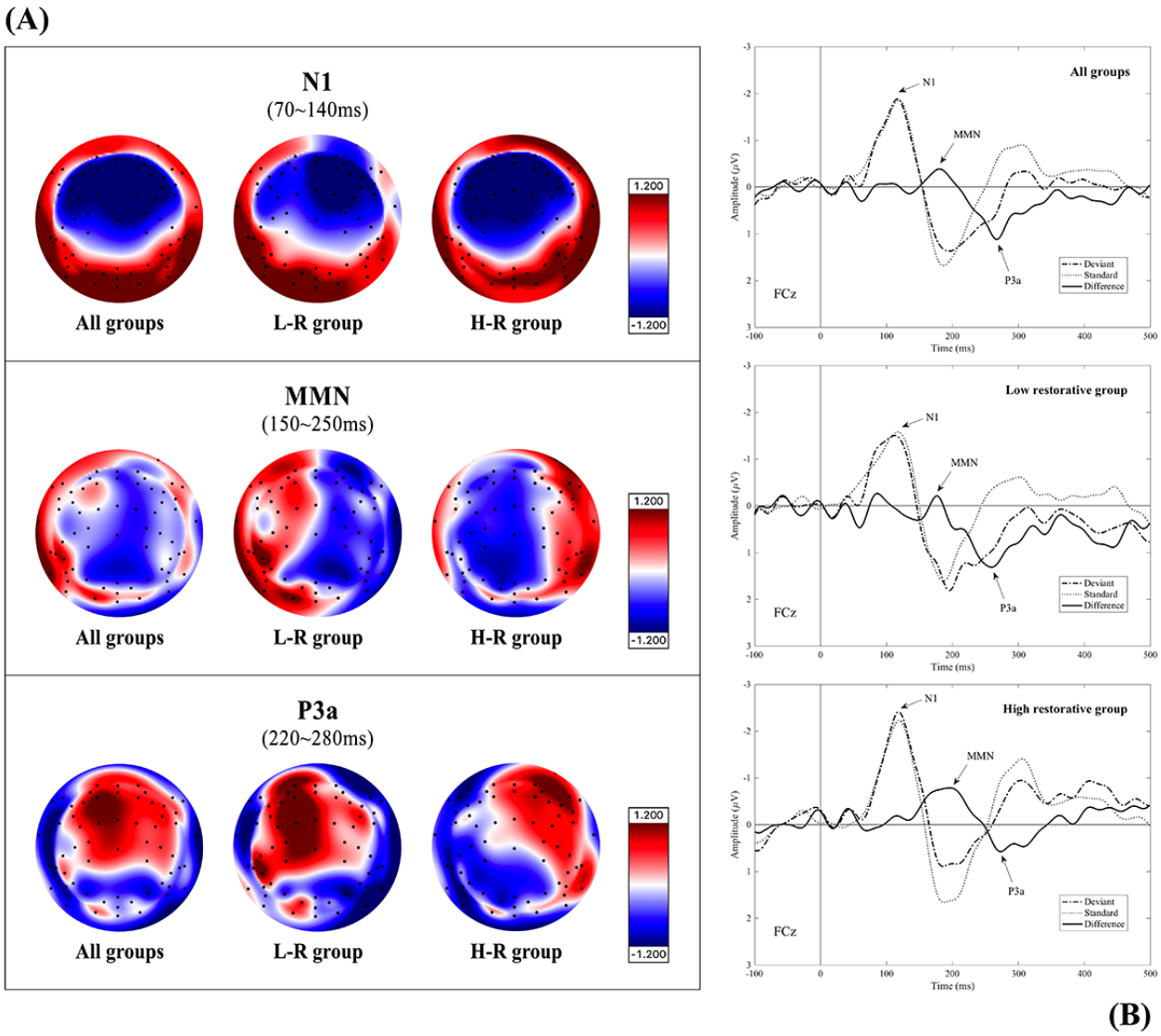
Figure 4. (A) Comparison of the P3a amplitudes while viewing hard-fascinating and soft-fascinating virtual environments. (B) Comparison of the P3a amplitudes for 2 groups with different levels of restorative experience from the soft-fascinating virtual environment. (C) Correlation between self-reported perceived restorativeness scores (PRS) and P3a amplitudes. VE: virtual environment. Asterisk indicates $P<.05$.



Correlations Between Individual's Perceived Restorativeness Scale Scores and P3a Amplitudes

A correlation analysis was conducted to examine the relationship between self-reported PRS scores and P3a amplitudes over the 3 fronto-central midline electrode sites: Fz, FCz, and Cz. Following the interpretation of Cohen guideline [30], there was a moderate but significant negative correlation between the explicit and implicit measures at the FCz site (r=−.38; P=.02; see Figure 4) but not at the Fz (r=−.11; P=.50) and Cz (r=−.28; P=.08) sites. A significantly negative correlation between the PRS score and the P3a amplitude at the FCz electrode is in line with the decreased amplitude of the P3a after meditation practice, which may indicate decreased attentional engagement by the frontal cortex [21,22]. However, the aforementioned studies collected limited EEG data at only the Fz and Cz sites by using a low-density 19-channel electrode cap compared with the 3 electrode sites incorporated in the high-density 64 channel sensor net used in this study. In that context, the result reflects the P3a characteristic as an auditory ERP component measured mainly at the fronto-central region (ie, FCz or Cz).

Figure 5. (A) Event-related potential topographic maps of the N1 for deviant tones and the MMN/P3a complex for the deviant-standard difference wave. (B) Time courses of the N1 and mismatch negativity/P3a complex components for FCz, which were shown by each of the 2 restorative groups during exposure to virtual nature. L-R group: low-restorative group; H-R group: high restorative group.



Discussion

Principal Findings

The aim of this ART-based study is to determine whether the amplitude of the MMN/P3a complex can be a distinct ERP biomarker of involuntary attention restoration during exposure to 360° virtual nature in healthy young adults with different levels of restorative experience. Unlike previous studies that look at aMMN or P3a separately, this study focused on the neurophysiological handover from preattentive information processing (ie, aMMN) to subsequent attention orienting processing (ie, P3a) as indexed by the MMN/P3a complex. The findings of previous ART and ERP studies in the real world extend into those of this study in the virtual world. The theoretical and empirical implications of this study will support the development and assessment of restorative VR content by allowing various stakeholders such as designers, researchers, and clinicians to investigate the restorative responses that potential users have to active relaxation in restorative VEs, particularly at the level of involuntary attention (previous studies were limited to voluntary attention).

As shown by the manipulation check and the hypothesis testing, there were no significant differences in the aMMN amplitudes between the 2 restorative groups or between the 2 fascinating VE conditions, which corresponds to the preattentive nature of the aMMN, beyond the attention debate in the real world [24], and the condition that participants had to focus only on conducting visual tasks and could ignore any environmental sounds. In other words, not only is automatic change detection one of the most important cognitive functions for human survival from an evolutionary perspective but also the normal functioning of the cognitive system is characterized by maintaining a sound balance between goal-directed behavior and involuntary orientation [31,32]. Despite no significant changes in the aMMN amplitudes during exposure to virtual nature, the robustness of the aMMN amplitudes indicates that the study participants qualified as healthy subjects. Mental fatigue during or after prolonged periods of cognitive tasks impairs preattentive auditory processing, as revealed by the fact that aMMN

amplitudes at fronto-central electrode sites can be significantly decreased by mental fatigue [33]. Such impairments in the preattentive processing of auditory change detection also appear in clinical populations, such as patients with schizophrenia [34,35], attention deficit disorder or attention-deficit hyperactivity disorder [36], or major depressive disorder [37]. Given the possible link between decreased exposure to nature and increased occurrence of mental disorders [6-8], it is important to clarify the role of the aMMN amplitude in different groups of people, including those who are vulnerable to mental fatigue, at high risk for mental disorders, or currently suffering from acute or chronic mental disorders.

On the basis of the few ART studies that focused on the concept of fascination [1,5,38], this study compared the restorative effect of entertaining events with that of natural events for the manipulation check. Consistent with the ART [1], the less restorative urban fireworks VEs with hard fascination consumed relatively more involuntary attentional resources and left little space for mental reflection compared with the more restorative nature VEs with soft fascination, which elicited significantly increased P3a amplitudes and PRS scores. As hypothesized, the P3a amplitude was differently affected by individuals' restorative experience during exposure to the nature VE, supporting the restorative effect of virtual nature as indicated by the negative correlation between self-reported PRS scores and P3a amplitudes as well as by the attenuated P3a amplitudes in the high restorative group. As the P3a is an index for frontal neural activity generated by stimulus-driven attention mechanisms [22], a trend for clear P3a amplitude reduction is consistent with decreased attentional engagement by the frontal cortex in response to unexpected and distracting stimuli during meditation [21]. In line with the finding of Pierson et al [39], it is more likely that participants who report greater restorative depth find it easier to enter a state of deep quiescence because of their low sensation-seeking trait, resulting in a positive correlation between restorative depth and frontal P3a amplitude. More importantly, meditation not only reflects first-person experiences that cannot be easily shared [40] but also facilitates attention control mechanisms [41,42]. If the restorative effect of virtual nature experience shows a trend similar to that of meditation practice with the aim of decreasing emotional and cognitive reactivity, individuals' restorative experience in a nature VE could play a key role in renewing the capacity for attention control without an intentional effort to regulate attention [1,4]. To shed light on the similarities between the 2 relaxation techniques, future studies need to determine whether the decreased automated reactivity and inhibited evaluative processing related to the task-irrelevant, attention-demanding auditory stimuli are consistent phenomena in other groups of individuals.

This study has some limitations. Although meditation is well known as an effective relaxation technique with both short-term and long-term effects on attentional function [21,43-45], only the short-term effect of exposure to 360° VR nature were examined by collecting normative data from a relatively small sample with a narrow age range and strict exclusion criteria. To generalize the findings, the restorative effect of short-term and long-term exposure to nature VR in people with neuro or psychiatric disorders as well as in younger or elderly adults need to be examined. Mental and attentional fatigue might make it difficult to meditate without guidance or help from professionals. However, some people might want to meditate for long periods of time or in a flexible manner, so they prefer nonguided meditation or free-style relaxation to guided meditation. For those reasons, a self-chosen 360° VR nature therapy within a short time might be more appealing than a virtually guided scenario-based VR meditation. The mobile VR system might allow mentally fatigued workers and students to feel some relief from real-world stressful factors as if they had actually experienced a period of vacation. In addition, people with disabilities, elderly people, and hospital patients would be able to easily use VR nature therapy in the home and ward environments. Individuals who are vulnerable to simulator sickness could control the 360° VEs with subtle head movements or even without wearing VR glasses and, thus, could act independently in the virtual world. Taken together, the results of this study suggest that VR nature therapy may one day provide a self-help, self-administered treatment for mental and attentional fatigue in everyday life.

Conclusions

The 360° VR nature exposure therapy can help individuals suffering from mental fatigue to manage their attentional resources with minimum effort. About 5 min of exposure to virtual nature restored involuntary attention without causing simulator sickness. The amplitude of the fronto-central MMN/P3a complex can potentially be employed as a distinct ERP biomarker of involuntary attention restoration during virtual nature experience in healthy young adults.

Acknowledgments

This study is part of KC's doctoral dissertation submitted to Sungkyunkwan University. The authors would like to thank KC's committee members, Professor Jang Hyun Kim, Professor Sangwon Lee, and Professor Jungyeon Sung for their valuable comments and suggestions to improve the quality of this paper. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government Ministry of Science, ICT and Future Planning (MSIP) [grant number 2016R1C1B2010739].

Conflicts of Interest

None declared.

References

1. Kaplan S. The restorative benefits of nature: toward an integrative framework. J Environ Psychol 1995 Sep;15(3):169-182. [doi: 10.1016/0272-4944(95)90001-2]
2. Hancock PA, Warm JS. A dynamic model of stress and sustained attention. Hum Factors 1989 Oct;31(5):519-537. [doi: 10.1177/001872088903100503] [Medline: 2625347]
3. Bohnen N, Houx P, Nicolson N, Jolles J. Cortisol reactivity and cognitive performance in a continuous mental task paradigm. Biol Psychol 1990 Oct;31(2):107-116. [Medline: 2103746]
4. Kaplan R, Kaplan S. The experience of nature: A psychological perspective. New York: Cambridge University Press; 1989.
5. Herzog TR, Black AM, Fountaine KA, Knotts DJ. Reflection and attentional recovery as distinctive benefits of restorative environments. J Environ Psychol 1997 Jun;17(2):165-170. [doi: 10.1006/jevp.1997.0051]
6. Lederbogen F, Kirsch P, Haddad L, Streit F, Tost H, Schuch P, et al. City living and urban upbringing affect neural social stress processing in humans. Nature 2011 Jun 22;474(7352):498-501. [doi: 10.1038/nature10190] [Medline: 21697947]
7. Lorenc T, Clayton S, Neary D, Whitehead M, Petticrew M, Thomson H, et al. Crime, fear of crime, environment, and mental health and wellbeing: mapping review of theories and causal pathways. Health Place 2012 Jul;18(4):757-765. [doi: 10.1016/j.healthplace.2012.04.001] [Medline: 22542441]
8. Bratman GN, Hamilton JP, Daily GC. The impacts of nature experience on human cognitive function and mental health. Ann N Y Acad Sci 2012 Feb;1249:118-136. [doi: 10.1111/j.1749-6632.2011.06400.x] [Medline: 22320203]
9. Taylor AF, Kuo FE. Children with attention deficits concentrate better after walk in the park. J Atten Disord 2009 Mar;12(5):402-409. [doi: 10.1177/1087054708323000] [Medline: 18725656]
10. Tyrväinen L, Ojala A, Korpela K, Lanki T, Tsunetsugu Y, Kagawa T. The influence of urban green environments on stress relief measures: a field experiment. J Environ Psychol 2014 Jun;38:1-9. [doi: 10.1016/j.jenvp.2013.12.005]
11. Bratman GN, Daily GC, Levy BJ, Gross JJ. The benefits of nature experience: improved affect and cognition. Landscape Urban Plann 2015 Jun;138:41-50. [doi: 10.1016/j.landurbplan.2015.02.005]
12. Bringslimark T, Hartig T, Patil GG. The psychological benefits of indoor plants: a critical review of the experimental literature. J Environ Psychol 2009 Dec;29(4):422-433. [doi: 10.1016/j.jenvp.2009.05.001]
13. Lottrup L, Stigsdotter UK, Meilby H, Claudi AG. The workplace window view: a determinant of office workers' work ability and job satisfaction. Landscape Research 2013 Oct 21;40(1):57-75. [doi: 10.1080/01426397.2013.829806]
14. Berto R. Exposure to restorative environments helps restore attentional capacity. J Environ Psychol 2005 Sep;25(3):249-259. [doi: 10.1016/j.jenvp.2005.07.001]
15. Brown DK, Barton JL, Gladwell VF. Viewing nature scenes positively affects recovery of autonomic function following acute-mental stress. Environ Sci Technol 2013 Jun 04;47(11):5562-5569 [FREE Full text] [doi: 10.1021/es305019p] [Medline: 23590163]
16. Hartig T, Kaiser F, Bowler P. Institute for Housing Research. Further development of a measure of perceived environmental restorativeness URL: http://www.diva-portal.org/smash/get/diva2:130237/FULLTEXT01.pdf [accessed 2018-05-25] [WebCite Cache ID 6zfcfYYQK]
17. Friedman D, Cycowicz YM, Gaeta H. The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. Neurosci Biobehav Rev 2001 Jun;25(4):355-373. [Medline: 11445140]
18. Light GA, Swerdlow NR, Braff DL. Preattentive sensory processing as indexed by the MMN and P3a brain responses is associated with cognitive and psychosocial functioning in healthy adults. J Cogn Neurosci 2007 Oct;19(10):1624-1632 [FREE Full text] [doi: 10.1162/jocn.2007.19.10.1624] [Medline: 18271737]
19. Chen J, Zhang Y, Wei D, Wu X, Fu Q, Xu F, et al. Neurophysiological handover from MMN to P3a in first-episode and recurrent major depression. J Affect Disord 2015 Mar 15;174:173-179. [doi: 10.1016/j.jad.2014.11.049] [Medline: 25499685]
20. Biedermann B, De Lissa P, Mahajan Y, Polito V, Badcock N, Connors MH, et al. Meditation and auditory attention: an ERP study of meditators and non-meditators. Int J Psychophysiol 2016 Dec;109:63-70. [doi: 10.1016/j.ijpsycho.2016.09.016] [Medline: 27693547]
21. Cahn BR, Polich J. Meditation (Vipassana) and the P3a event-related brain potential. Int J Psychophysiol 2009 Apr;72(1):51-60 [FREE Full text] [doi: 10.1016/j.ijpsycho.2008.03.013] [Medline: 18845193]
22. Polich J. Updating P300: an integrative theory of P3a and P3b. Clin Neurophysiol 2007 Oct;118(10):2128-2148 [FREE Full text] [doi: 10.1016/j.clinph.2007.04.019] [Medline: 17573239]
23. Näätänen R, Gaillard AW, Mäntysalo S. Early selective-attention effect on evoked potential reinterpreted. Acta Psychol (Amst) 1978 Jul;42(4):313-329. [Medline: 685709]
24. Chung K, Park JY. Beyond the real world: attention debates in auditory mismatch negativity. Neuroreport 2018 Apr 11;29(6):472-477. [doi: 10.1097/WNR.0000000000000999] [Medline: 29578987]
25. Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 1971 Mar;9(1):97-113. [Medline: 5146491]
26. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol 1993 Jul;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]
27. Min B, Jeon H, Sung E, Chung S, Kim C. A study on simulator sickness and physiological responses in dynamic driving simulator. J Soc Korea Ind Syst 2003;26(1):22-29 [FREE Full text]

28. Yoo Y, Yeoun P, Shin W. The comparison of PRS (Perceived Restorativeness Scale) on types of urban forest. KFRM 2013;17(1):33-45 [FREE Full text]
29. Ferree TC, Luu P, Russell GS, Tucker DM. Scalp electrode impedance, infection risk, and EEG data quality. Clin Neurophysiol 2001 Mar;112(3):536-544. [Medline: 11222977]
30. Cohen J. Statistical power analysis for the behavioral sciences, 2nd edn. Mahwah: Lawrence Erlbaum Associates; 1988.
31. James W. The Principles of Psychology. New York: Pantianos Classics; 2013.
32. Escera C, Alho K, Schröger E, Winkler I. Involuntary attention and distractibility as evaluated with event-related brain potentials. Audiol Neurootol 2000;5(3-4):151-166. [doi: 10.1159/000013877] [Medline: 10859410]
33. Yang B, Xiao W, Liu X, Wu S, Miao D. Mental fatigue impairs pre-attentive processing: a MMN study. Neurosci Lett 2013 Jan 04;532:12-16. [doi: 10.1016/j.neulet.2012.08.080] [Medline: 22985513]
34. Salisbury DF, Kuroki N, Kasai K, Shenton ME, McCarley RW. Progressive and interrelated functional and structural evidence of post-onset brain reduction in schizophrenia. Arch Gen Psychiatry 2007 May;64(5):521-529 [FREE Full text] [doi: 10.1001/archpsyc.64.5.521] [Medline: 17485604]
35. Turetsky BI, Bilker WB, Siegel SJ, Kohler CG, Gur RE. Profile of auditory information-processing deficits in schizophrenia. Psychiatry Res 2009 Jan 30;165(1-2):27-37 [FREE Full text] [doi: 10.1016/j.psychres.2008.04.013] [Medline: 18990453]
36. Yamamuro K, Ota T, Iida J, Nakanishi Y, Kishimoto N, Kishimoto T. Associations between the mismatch-negativity component and symptom severity in children and adolescents with attention deficit/hyperactivity disorder. Neuropsychiatr Dis Treat 2016;12:3183-3190 [FREE Full text] [doi: 10.2147/NDT.S120540] [Medline: 28003754]
37. Naismith SL, Mowszowski L, Ward PB, Diamond K, Paradise M, Kaur M, et al. Reduced temporal mismatch negativity in late-life depression: an event-related potential index of cognitive deficit and functional disability? J Affect Disord 2012 Apr;138(1-2):71-78. [doi: 10.1016/j.jad.2011.12.028] [Medline: 22301116]
38. Aspinall P, Mavros P, Coyne R, Roe J. The urban brain: analysing outdoor physical activity with mobile EEG. Br J Sports Med 2015 Feb;49(4):272-276. [doi: 10.1136/bjsports-2012-091877] [Medline: 23467965]
39. Pierson A, Le HJ, Fossaert A, Dubal S, Jouvent R. Frontal reactivity and sensation seeking an ERP study in skydivers. Prog Neuropsychopharmacol Biol Psychiatry 1999 Apr;23(3):447-463. [Medline: 10378229]
40. Schmidt S. Opening up meditation for science: the development of a meditation classification system. In: Meditation-Neuroscientific Approaches and Philosophical Implications. Cham: Springer; 2014:137-152.
41. Davidson RJ, Goleman DJ. The role of attention in meditation and hypnosis: a psychobiological perspective on transformations of consciousness. Int J Clin Exp Hypn 1977 Oct;25(4):291-308. [doi: 10.1080/00207147708415986] [Medline: 330418]
42. Shapiro DH. Overview: clinical and physiological comparison of meditation with other self-control strategies. Am J Psychiatry 1982 Mar;139(3):267-274. [doi: 10.1176/ajp.139.3.267] [Medline: 7036760]
43. Cahn BR, Polich J. Meditation states and traits: EEG, ERP, and neuroimaging studies. Psychol Bull 2006 Mar;132(2):180-211. [doi: 10.1037/0033-2909.132.2.180] [Medline: 16536641]
44. Lutz A, Slagter HA, Dunne JD, Davidson RJ. Attention regulation and monitoring in meditation. Trends Cogn Sci 2008 Apr;12(4):163-169 [FREE Full text] [doi: 10.1016/j.tics.2008.01.005] [Medline: 18329323]
45. Raffone A, Srinivasan N. The exploration of meditation in the neuroscience of attention and consciousness. Cogn Process 2010 Feb;11(1):1-7. [doi: 10.1007/s10339-009-0354-z] [Medline: 20041276]

Abbreviations

aMMN: auditory mismatch negativity
ART: attention restoration theory
EEG: electroencephalography
EENS: Extensions for Net Station
EGI: Electrical Geodesics Inc
ERP: event-related potential
HCGSN: HydroCel Geodesic Sensor Net
IRB: institutional review board
MMN: mismatch negativity
PRS: Perceived Restorativeness Scale
PST: Psychology Software Tools Inc
RM-ANOVA: repeated measures analysis of variance
SSQ: Simulator Sickness Questionnaire
VE: virtual environment
VR: virtual reality

Edited by G Eysenbach; submitted 28.05.18; peer-reviewed by SJ Son, T Guetterman; comments to author 19.06.18; revised version received 14.09.18; accepted 25.10.18; published 30.11.18.

Please cite as:
Chung K, Lee D, Park JY
Involuntary Attention Restoration During Exposure to Mobile-Based 360° Virtual Nature in Healthy Adults With Different Levels of Restorative Experience: Event-Related Potential Study
J Med Internet Res 2018;20(11):e11152
URL: http://www.jmir.org/2018/11/e11152/
doi:10.2196/11152
PMID:30504121

Original Paper

Automatic Classification of Online Doctor Reviews: Evaluation of Text Classifier Algorithms

Ryan Rivas[1], BS; Niloofar Montazeri[1], PhD; Nhat XT Le[1], BS; Vagelis Hristidis[1], PhD

Department of Computer Science and Engineering, University of California, Riverside, Riverside, CA, United States

Corresponding Author:
Ryan Rivas, BS
Department of Computer Science and Engineering
University of California, Riverside
363 Winston Chung Hall
900 University Avenue
Riverside, CA, 92521
United States
Phone: 1 951 827 2838
Email: rriva002@ucr.edu

Abstract

Background: An increasing number of doctor reviews are being generated by patients on the internet. These reviews address a diverse set of topics (features), including wait time, office staff, doctor's skills, and bedside manners. Most previous work on automatic analysis of Web-based customer reviews assumes that (1) product features are described unambiguously by a small number of keywords, for example, *battery* for phones and (2) the opinion for each feature has a positive or negative sentiment. However, in the domain of doctor reviews, this setting is too restrictive: a feature such as *visit duration* for doctor reviews may be expressed in many ways and does not necessarily have a positive or negative sentiment.

Objective: This study aimed to adapt existing and propose novel text classification methods on the domain of doctor reviews. These methods are evaluated on their accuracy to classify a diverse set of doctor review features.

Methods: We first manually examined a large number of reviews to extract a set of features that are frequently mentioned in the reviews. Then we proposed a new algorithm that goes beyond bag-of-words or deep learning classification techniques by leveraging natural language processing (NLP) tools. Specifically, our algorithm automatically extracts dependency tree patterns and uses them to classify review sentences.

Results: We evaluated several state-of-the-art text classification algorithms as well as our dependency tree–based classifier algorithm on a real-world doctor review dataset. We showed that methods using deep learning or NLP techniques tend to outperform traditional bag-of-words methods. In our experiments, the 2 best methods used NLP techniques; on average, our proposed classifier performed 2.19% better than an existing NLP-based method, but many of its predictions of specific opinions were incorrect.

Conclusions: We conclude that it is feasible to classify doctor reviews. Automatically classifying these reviews would allow patients to easily search for doctors based on their personal preference criteria.

(J Med Internet Res 2018;20(11):e11141) doi:10.2196/11141

KEYWORDS

patient satisfaction; patient reported outcome measures; quality indicators, health care; supervised machine learning

Introduction

Background

The problem of automatic reviews analysis and classification has attracted much attention because of its importance in ecommerce applications [1-3]. Recently, there has been an increase in the number of sites where users rate doctors. Several works have analyzed the content and scores of such reviews, mostly by examining a subset of them through qualitative and quantitative analysis [4-9] or by applying text-mining techniques to characterize trends [10-12]. However, not much work has studied how to automatically classify doctor reviews.

In this study, our objective was to automatically summarize the content of a textual doctor review by extracting the features it mentions and the opinion of the reviewer for each of these features; for example, to estimate if the reviewer believes that

the wait time or the visit time is long or if the doctor is in favor of complementary medicine methods. We explore the feasibility of reaching this objective by defining a broader definition of the review classification problem that addresses challenges in the domain of doctor reviews and examining the performance of several machine learning algorithms in classifying doctor review sentences.

Previous work on customer review analysis focused on automated extraction of features and the polarity (also referred as opinion or sentiment) of statements about those features [2,13,14]. Specifically, these works tackle the problem in 2 steps: first they extract the features using rules, and then, for each feature, they estimate the polarity using hand-crafted rules or supervised machine learning methods. This works well if (1) the features are *basic*, such as the battery of a phone, which are generally described by a single keyword, for example, *the battery of the camera is poor*, and (2) the opinion is objectively positive or negative but does not support more subjective features like visit time, where for some patients it is positive to be longer, and for some, it is negative. In other words, statements about features in product reviews tend to be more straightforward and unambiguously positive or negative, whereas reviews on service, such as doctor reviews, are often less so, as there may be many ways to express an opinion on some aspect of the service.

In our study, the features may be more complex, for example, the *visit time* feature can be expressed by different phrases such as "spends time with me," "takes his time," "not rushed," and so on. As another example, "appointment scheduling" can be expressed in many different ways, for example, "I was able to schedule a visit within days" or "The earliest appointment I could make is in a month." Other complex classes include *staff* or *medical skills*.

Furthermore, in our study, what is positive for one user may be negative for another. For example, consider the sentence "Dr. Chan is very fast so there is practically no wait time and you are in and out within 20 minutes." The sentiment in this sentence is positive, but a short visit implied by *in and out within 20 minutes* may be negative for some patients. Instead, what we want to measure is long visit time versus short visit time. This is different from work on detecting transition of sentiment [15] because it is not enough to detect the *true* sentiment, but we must also associate it with a class (long visit time vs short visit time).

To address this variation of the review classification problem, we created a labeled dataset consisting of 5885 sentences from 1017 Web-based doctor reviews. We identified several classes of doctor review opinions and labeled each sentence according to the presence and polarity of these opinion classes. Note that our definition of polarity is broader than in previous work as it is not strictly positive and negative but rather takes the subjectivity of patient opinions into account (eg, complementary medicine is considered good by some and bad by others).

We adapt existing and propose new classifiers to classify doctor reviews. In particular, we consider 3 diverse types of classifiers:

1. Bag-of-words classifiers such as Support Vector Machine (SVM) [16,17] and Random Forests [18] that leverage the statistical properties of the review text, such as the frequency of each word.
2. Deep learning methods such as Convolutional Neural Network (CNN) [19], which also consider the proximity of the words.
3. Natural Language Processing (NLP)–based classifiers, which leverage the dependency tree of a review sentence [20]. Specifically, we consider an existing NLP-based classifier [21] and propose a new one, the Dependency Tree-Based Classifier (DTC).

DTC generates the dependency tree for each sentence in a review and applies a set of rules to extract dependency tree–matching patterns. These patterns are then ranked by their accuracy on the training set. Finally, the sentences of a new review are classified based on the highest-ranking matching pattern. This is in contrast to the work by Matsumoto et al [21], which treats dependency tree patterns as features in an SVM classifier.

The results of our study show that classifying doctor reviews to identify patient opinions is feasible. The results also show that DTC generally outperforms all other implemented text classification techniques.

Here is a summary of our contributions:

1. We propose a broader definition for the review classification problem in the domain of doctor reviews, where the features can be complex entities and the polarity is not strictly positive or negative.
2. We evaluated a diverse set of 5 state-of-the-art classification techniques on a labeled dataset of doctor reviews containing a set of commonly used and useful features.
3. We propose a novel decision tree–based classifier and show that it outperforms the other methods; we have published the code on the Web [22].

Literature Review

In this section, we review research in fields related to this study, which we organize into 5 categories:

- Quantitative and qualitative analysis of doctor review ratings and content
- The application of text mining techniques to describe trends in doctor reviews
- Feature and polarity extraction in customer reviews
- Application of dependency tree patterns to sentiment analysis
- Recent work in text classification

Doctor Review Analysis

Several previous works have analyzed Web-based doctor reviews. Gao et al described trends in doctor reviews over time to identify which characteristics influence Web-based ratings [4]. They found that obstetricians or gynecologists and long-time graduates were more likely to be reviewed than other physicians, recent graduates, board-certified physicians, highly rated medical school graduates, and doctors without malpractice claims received higher ratings, and reviews were generally positive. Segal et al compared doctor review statistics with

surgeon volume [5]. They found that high-volume surgeons could be differentiated from low-volume surgeons by analyzing the number of numerical ratings, the number of text reviews, the proportion of positive reviews, and the proportion of critical reviews. López et al performed a qualitative content analysis of doctor reviews [6]. They found that most reviews were positive and identified 3 overarching domains in the reviews they analyzed: interpersonal manner, technical competence, and system issues. Hao analyzed Good Doctor Online, an online health community in China, and found that gynecology-obstetrics-pediatrics doctors were the most likely to be reviewed, internal medicine doctors were less likely to be reviewed, and most reviews were positive [7]. Smith and Lipoff conducted a qualitative analysis of dermatology practice reviews from Yelp and ZocDoc [8]. They found that both the average review scores and the proportion of reviews with 5 out of 5 stars from ZocDoc were higher than those from Yelp. They also found that high-scoring reviews and low-scoring reviews had similar content (eg, physician competency, staff temperament, and scheduling) but opposite valence. Daskivich et al analyzed health care provider ratings across several specialties and found that allied health providers (eg, providers who are neither doctors nor nurses) had higher patient satisfaction scores than physicians, but these scores were also the most skewed [9]. They also concluded that specialty-specific percentile ranks might be necessary for meaningful interpretation of provider ratings by consumers.

Text Mining of Doctor Reviews

Other previous papers have employed text-mining techniques to characterize trends in doctor reviews. Wallace et al designed a probabilistic generative model to capture latent sentiment across aspects of care [10]. They showed that including their model's output in regression models improves correlations with state-level quality measures. Hao and Zhang used topic modeling to extract common topics among 4 specialties in doctor reviews collected from Good Doctor Online [11]. They identified 4 popular topics across the 4 specialties: the experience of finding doctors, technical skills or bedside manner, patient appreciation, and description of symptoms. Similarly, Hao et al used topic modeling to compare reviews between Good Doctor Online and the US doctor review website RateMDs [12]. Although they found similar topics between the 2 sites, they also found differences that reflect differences between the 2 countries' health care systems. These works differ from ours in that they use text-mining techniques to analyze doctor reviews in aggregate, while our goal is to identify specific topics in individual reviews.

Customer Review Feature and Polarity Extraction

As discussed earlier in the Introduction, these works operate on a more limited problem setting where the features are usually expressed by a single keyword, and the sentiment is strictly positive or negative. Hu and Liu extracted opinions of features in customer reviews with a 4-step algorithm [2]. This algorithm consists of applying association rule mining to identify features, pruning uninteresting and redundant features, identifying infrequent features, and finally determining semantic orientation of each opinion sentence. Popescu and Etzioni created an unsupervised system for feature and opinion extraction from product reviews [3]. After finding an explicit feature in a sentence, they applied manually crafted extraction rules to the sentence and extracted the heads of potential opinion phrases. This method only works when features are explicit.

Sentiment Analysis With Dependency Trees

There are number of existing works that use dependency trees or patterns for sentiment analysis. A key difference is that our method does not always capture sentiment but the various class labels (eg, short or long) for each class (eg, visit time). Hence, we cannot rely on external sentiment training data or on hard-coded sentiment rules, but we must use our own training data.

Agarwal et al used several hand-crafted rules to extract dependency tree patterns from sentences [23]. They combined this information with the semantic information present in the Massachusetts Institute of Technology Media Lab ConceptNet ontology and employed the extracted concepts to train a machine learning model to learn concept patterns in the text, which were then used to classify documents into positive and negative categories. An important difference from our method is that their dependency patterns generally consist of only 2 words in certain direct relations, while our patterns can contain several more in both direct and indirect relations.

Wawer induced dependency patterns by using target-sentiment (T-S) pairs and recording the dependency paths between T and S words in the dependency tree of sentences in their corpus [24]. These patterns were supplemented with conditional random fields to identify targets of opinion words. In contrast to our patterns, which can represent a subtree of 2 or more words, the patterns in this work are generated from the shortest path between the T and S words.

Matsumoto et al's work [21] is the closest work to our proposed method, which we experimentally compare in the Results section. They extract frequent word subsequences and dependency subtrees from the training data and use them as features in an SVM sentiment classifier. Their patterns involve frequent words and only include direct relations, whereas our patterns involve high-information gain words and consider indirect relations. Pak and Paroubek follow a similar strategy of extracting dependency tree patterns based on predefined rules and using them as features for an SVM classifier [25]. Matsumoto et al perform better on the common datasets they considered.

Text Classification

Machine learning algorithms are commonly used for text classification. Kennedy et al used a random forest classifier to identify harassment in posts from Twitter, Reddit, and The Guardian [26]. Posts were represented through several features such as term frequency-inverse document frequency (TF-IDF) of unigrams, bigrams, and short character sequences; URL and hashtag token counts; source (whether the post was from Twitter); and sentiment polarity. Gambäck and Sikdar used a CNN to classify hate speech in Twitter posts [27]. The CNN model was tested with multiple feature embeddings, including random values and word vectors generated with Word2Vec

[28]. Lix et al used an SVM classifier to determine patient's alcohol use using text in electronic medical records [29]. Unigrams and bigrams in these records were represented using a bag-of-words model.

Problem Definition

Given a text dataset with a set of classes c_1, c_2, …, c_m that represent features previously identified by a domain expert, each class c_i can take 3 values (polarity):

- c_i^0:Neutral. The sentence is not relevant to the class.
- c_i^x, c_i^y:Yes or no. Note that to avoid confusion, we do not say positive or negative, as for some classes such as *visit time* in doctor reviews, some patients prefer when their visit time is long and some prefer short. In this example, "Yes" could arbitrarily be mapped to *long* and "No" to *short*.

As another example, class c_8 from the doctor review dataset is *wait time* or the time spent waiting to see a doctor. It has 3 possible values: c_8^x, c_8^y, or c_8^0. A sentence with class label c_8^x expresses the opinion that the time spent waiting to see the doctor is short. Examples of c_8^x include "I got right in to see Dr. Watkins," "I've never waited more than five minutes to see him," and "Wait times are very short once you arrive for an appointment." A sentence with class label c_8^y expresses the opinion that the time spent waiting to see the doctor is long. Examples of c_8^y include "There is always over an hour wait even with an appointment," "My biggest beef is with the wait time," and "The wait time was terrible." A sentence with class label c_8^0 makes no mention of wait time. Such sentences may have c_i^x or c_i^y labels from other classes, for example, "This doctor lacks affect and a caring bedside manner" and "His staff, especially his nurse Lucy, go far above what their job requires," or they may instead not be relevant to any class, such as "Dr. Kochar had been my primary care physician for seven years" and "I'll call to reschedule everything." A sentence may take labels from more than one class.

In this study, given a training set T of review sentences with class labels from classes c_1, c_2, …, c_m, we build a classifier for each class c_i to classify new sentences to one of the possible values of c_i. Specifically, we build m training sets T_i corresponding to each class. Each sentence in T_i is assigned a class label c_i^x, c_i^y, or c_i^0.

Methods

Doctor Reviews Dataset

We crawled Vitals [30], a popular doctor review website, to collect 1,749,870 reviews. Each author read approximately 200 reviews and constructed a list of features. Afterward, through discussions, we merged these lists into a single list of 13 features, which we represent by classes as described in the problem definition (Table 1).

To further filter these classes, we selected 600 random reviews to label. We labeled these reviews using WebAnno, a Web-based annotation tool [31] (Figure 1). Specifically, each sentence was tagged (labeled) with 0 or more classes from Table 1 by 2 of the authors. The union of these labels was used as the set of ground-truth class labels of each sentence; that is, if at least one of the labelers labeled a sentence as c_i^x, that sentence is labeled c_i^x in our dataset.

We found that some of these classes were underrepresented. For each underrepresented class, we used relevant keywords to find and label more reviews from the collected set of reviews, for example, *wait* for wait time and *listen* for information sharing, which resulted in a total of *1017 reviews* (417 in addition to the original 600). These 1017 reviews are our labeled dataset used in our experiments.

Table 1. Description of initial opinion classes. For each class, a sentence that does not mention the class is labeled c_i.

Class	c_i^x	c_i^y
Appointment scheduling	Easy to schedule an appointment	Hard to schedule an appointment
Bedside manner	Friendly and caring	Rude and uncaring
Complementary medicine	Promotes complementary medicine	No promotion of complementary medicine
Cost	Inexpensive and billing is simple	Expensive and billing problems
Information sharing	Answers questions and good explanations	Does not answer questions and poor explanations
Joint decision making	Treatment plan accounts for patient opinions	Treatment plan made without patient input
Medical skills	Effective treatment and correct diagnoses	Ineffective treatment and misdiagnoses conditions
Psychological support	Addresses stress and anxiety	Does not address stress and anxiety
Self-management	Encourages active management of care	Does not encourage self-management of care
Staff	Staff is friendly and helpful	Staff is rude and unhelpful
Technology	Uses email, Web-based appointments, and electronic health records	Does not use email and Web-based appointments
Visit time	Spends substantial time with patients	Spends very little time with patients
Wait time	Short time spent waiting to see the doctor	Long time spent waiting to see the doctor

Figure 1. Screenshot of WebAnno's annotation interface with an annotated review.

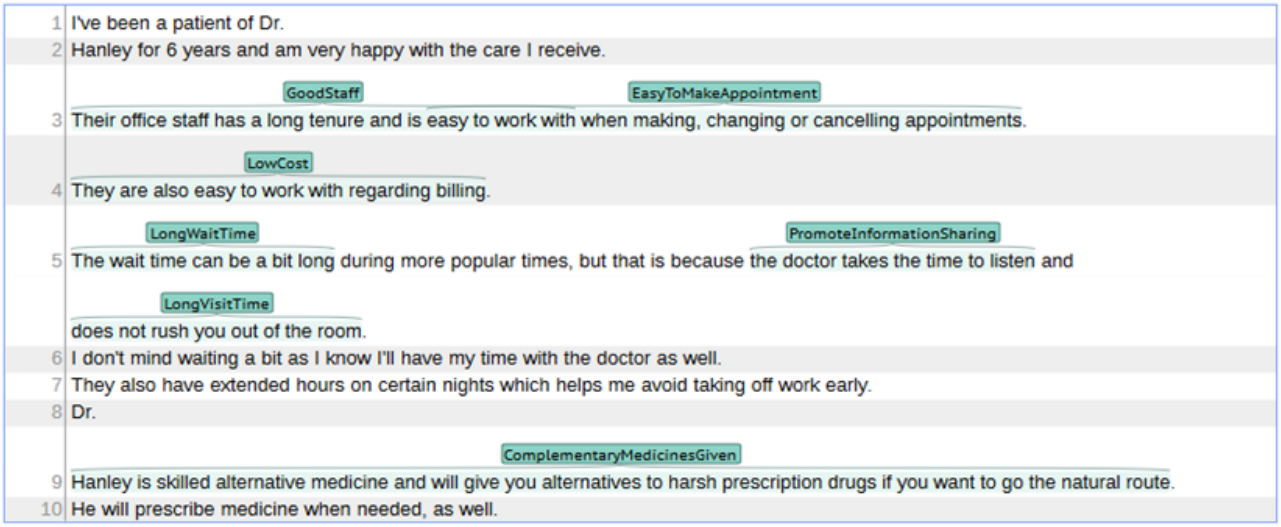


Table 2. Frequency of each class label in the doctor review dataset.

Class	Frequency of c_i^x	Frequency of c_i^y	Frequency of c_i^0
c_1: appointment scheduling	51	84	5750
c_2: bedside manner	569	341	4975
c_3: cost	25	261	5599
c_4: information sharing	316	136	5433
c_5: medical skills	481	232	5172
c_6: staff	262	368	5255
c_7: visit time	143	79	5663
c_8: wait time	48	199	5638

Following this, we found that some classes such as complementary medicine and joint decision making were still underrepresented, which we define as having less than 2% of the dataset's sentences labeled c_i^x or c_i^y, so we omitted them from the dataset. The final dataset consists of 5885 sentences and 8 opinion classes. These classes and the frequency of each of their labels are shown in Table 2.

Background on Dependency Trees

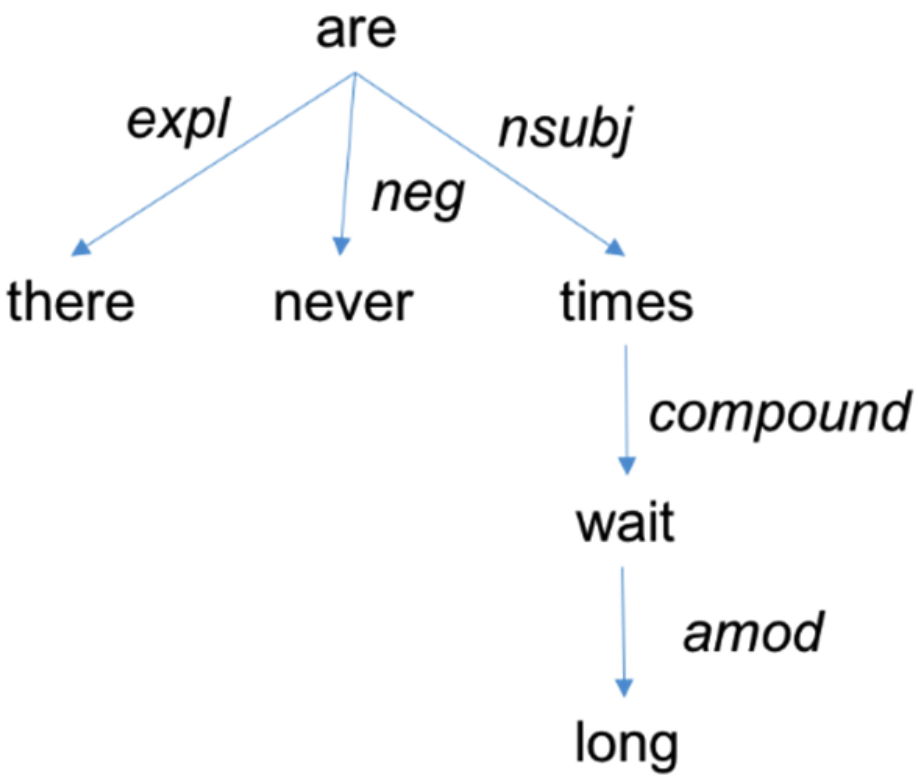
In this section, we describe dependency trees and the semgrex tool that we used for defining matching patterns. Dependency trees capture the grammatical relations between words in a sentence and are produced using a dependency parser and a dependency language. In a dependency tree, each word in a sentence corresponds to a node in the tree and is in one or more syntactic relations between the word or node exactly one other word or node. A dependency tree is a triple $T = \langle N, E, R \rangle$, where

- N is the set of nodes in T where each node n N is a tuple containing one or more string attributes describing a word in the sentence T was built from, such as *word*, *lemma*, or *POS* (part of speech)
- E is the set of edges in T where each edge e E is a triple $e = \langle n_g, r, n_d \rangle$, where
  - n_g N is the governor or parent in relation r
  - r is a syntactic relation between the words represented by n_g and n_d
  - n_d N is the dependent or child in relation r
- R N is the root node of T

Figure 2 shows a sample dependency tree for the sentence "there are never long wait times." The string representation of this tree, including the parts of speech for its words, is as follows:

[are/VBP expl>there/EX neg>never/RB nsubj>[times/NNS compound>[wait/NN amod>long/JJ]]]

Figure 2. A dependency tree for the sentence "There are never long wait times".



To match patterns against dependency trees, we used Stanford semgrex utility [32]. In the following, we explain some of the basics of semgrex patterns that help the reader understand patterns presented in this study using descriptions and examples from the Chambers et al study [32]. Semgrex patterns are composed of nodes and relations between them. Nodes are represented as *{attr1:value1;attr2:value2;…}* where attributes (attr) are regular strings such as *word*, *lemma*, and *pos*, and values can be strings or regular expressions marked by "/"s. For example, *{lemma:run;pos:/VB.*/}* means any verb form of the word *run*. Similar to "." in regular expressions, *{}* means any node in the graph. Relations in a semgrex have 2 parts: the relation symbol, which can be either < or > and optionally the relation type (ie, *nsubj* and *dobj*). In general, *A<reln B* means *A* is the dependent of a relation (*reln)* with *B*, whereas *A>reln B* means *A* is the governor of a *reln* with *B*. Indirect relations can be specified by the symbols >> and <<. For example, *A<<reln B* means there is some node in a dep->gov chain from *A* that is the dependent of a *reln* with *B*. Relations can be strung together with or without using the symbol *&*. All relations are relative to first node in string. For example, *A>nsubj B>dobj D* means *A* is a node that is the governor of both an *nsubj* relation with *B* and a *dobj* relation with *D*. Nodes can be grouped with parentheses. For example, *A>nsubj (B>dobj D)* means *A* is the governor of an *nsubj* relation with *B*, whereas *B* is the governor of a *dobj* relation with *D*. A sample pattern that matches the tree in Figure 2 can be:

{} >neg {} >> ({word:wait} > {word:long})

Using the Stanford CoreNLP Java library [33], our proposed classifier builds a dependency tree from a given sentence and determines whether any of a list of semgrex patterns matches any part of the tree.

Proposed Dependency Tree–Based Classifier

Our DTC algorithm is trained on a labeled dataset of sentences as described in the Problem Definition section. On a high level, given a sentence in training dataset *T*, the classifier generates a dependency tree using the Stanford Neural Network Dependency Parser [34] and extracts semgrex patterns from the dependency tree. These patterns are assigned the same class as the training sentence. When classifying a new sentence, the classifier generates the sentence's dependency tree and assigns a class label to the sentence based on which patterns from the training set match the dependency tree.

In more detail, the classifier's training algorithm generates a sorted list of semgrex patterns, each with an associated class label, from a training dataset *T* and integer parameters n_i^x, n_i^y, and *m*. Parameters n_i^x and n_i^y are the maximum number of terms (words or phrases) that will be used to generate patterns of classes c_i^x and c_i^y, respectively. In this study, we only use words, as dependency trees capture relations between words rather than phrases.

The pattern extraction algorithm described in the Pattern Extraction section below receives as input 2 sets W^x and W^y of high-information gain words, for the "Yes" (c_i^x) and "No" (c_i^y) class labels, respectively, from where we pick nodes for the generated patterns. The intuition is that high-information gain words are more likely to allow a pattern to differentiate between the class labels. Considering all words would be computationally

too expensive, and it does not offer any significant advantage as we have seen in our experiments. The information gain for W^x is determined by a logical copy of training dataset T in which class labels other than c_i^x are given a new class label $c_i^{x\prime}$, as the words in W^x will be used to identify sentences of class c_i^x. This process is repeated for W^y. Parameter m is the maximum number of these selected words that can be in a single pattern.

The final list of (semgrex pattern p and class label c') pairs is sorted by the weighted accuracy of the pair on the training data, which we define below.

We define $Accuracy_c(p, T)$ as the ratio of training instances in T with class label c that were correctly handled by pattern p. Pattern p, which was paired with class label $'$, is correct if it matches an instance with class label c' or it does not match an instance without class label c', but it is incorrect if it matches an instance without class label c' or it does not match an instance with class label c'. $|c_i|$ is the number of class labels in class c_i, which is 3 for all of the classes in this study. Intuitively, weighted accuracy treats all class labels with equal importance regardless of their frequency, so patterns that perform well on sentences of often low-frequency class labels c_i^x and c_i^y are assigned higher rank than they would otherwise. The training algorithm is shown in Textbox 1.

Given a to-be-classified sentence, we compute its dependency tree t and find the highest ranked (pattern p and class label c) pair where p matches t. Then the sentence is classified as c. If no pattern matches the sentence, we provide 2 possibilities: the sentence can be classified as the most common class label in T or it can be classified by a *backup* classifier trained on T.

Parameters Setting

In all experiments, we use $n_i^x=n_i^y=30$, as intuitively it is unlikely that there are more than 30 words for a class that can participate in a discriminative semgrex pattern. We set m to 4 for all experiments, because for $m>4$, it becomes too computationally expensive to compute all patterns.

Pattern Extraction in the Dependency Tree Classifier Algorithm

Overview

Given a dependency tree, we now describe how to extract patterns. Note that we repeat the pattern extraction for the "Yes" and "No" class labels, using W^x and W^y, respectively (W in this section refers to W^x or W^y). We extract semgrex patterns from a dependency tree t with class label c using a set of high-information gain words W and a maximum number of words m. The algorithm returns a set of patterns extracted from t made from up to m words in W.

The rationale for only working with high-information gain words is that we want to generate high-information gain patterns. We also want to preserve negations as they have a great impact to the accuracy of the patterns. If a low information gain word is negated, we replace it by a wildcard (*), which we found to be a good balance for these 2 goals. Each pattern p is associated with c such that a new sentence that matches p is classified as c. Textbox 2 describes the pattern extraction algorithm.

Textbox 1. The dependency tree classifier's training algorithm.

```
train(T, n_i^x, n_i^y, m):
  P=list of semgrex patterns used for classification, initially empty
  for each class label c in {c_i^x, c_i^y}:
    D=set of dependency trees for sentences in T with class label c
    T_c=copy of T with all non-c class labels given a new class label c'
    W=set of top n_c words w in T_c by information gain
    for each tree t in D:
      add all semgrex patterns from extract(W, t, c, m) to P
  test each pattern in P on T
  sort P by the weighted accuracy of each semgrex pattern tested on T in descending order
  return P
```

Textbox 2. Pattern extraction algorithm.

```
extract(W, t, c, m):
  P=set of patterns, initially empty
  S=stack of (tree, word set) pairs, initially empty
  for each combination C of words in W with |C |==min(|W|, m)
    S.push((t, C))
  while S is not empty:
    (t', C)=S.pop()
    t'=prune(t', C)
    n=root of t'
    while n==* and n has exactly 1 child:
      n=child of n
    t'=subtree of t' with root n
    remove each "*" node n' in t' with exactly 1 child c', and make the parent of n' the parent of c' with an indirect relation
    add (pattern(t'), c) to P
    for each combination C' of non-* words in t' with |C'| > 1:
      S.push((t', C'))
  return P

prune(t, W):
  t'=copy of t
  recursively prune from t' leaves that do not start with any word in W and are not in a negation relation
  for each node n in t':
    if n does not start with any word in W:
      n=*
  return t'
```

Details

The algorithm first creates a copy t' of t for each combination C of m words in W and pushes each (t', C) pair onto a stack. For each (t', C) popped from the stack, we execute the following steps:

1. Create initial subtree: Prune t' to keep only words in C, negations, and intermediate "*" nodes connecting them.
2. Remove unimportant nodes: Eliminate "*" nodes from t' starting with the root if it is a "*" node and has exactly 1 child (the child becomes the new root of t' and this repeats until the root no longer meets these criteria). Subsequently, remove each "*" node n' in t' with exactly 1 child and add an indirect relation edge from the parent of n' to the child of n'.
3. Add subpatterns: If (*pattern*(t'), c) is not already in P, add (*pattern*(t'), c) to the set of patterns P, and then push(t', c') onto the stack for each combination C' of 2 or more non-* words in t'.

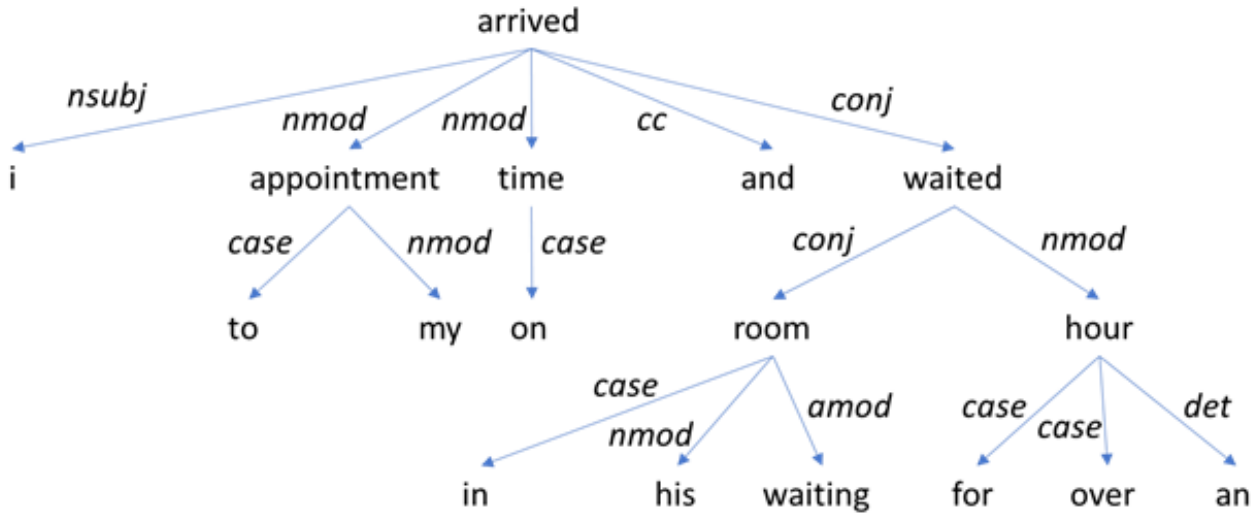
The algorithm then moves on to the next item on the stack. Once the stack is empty, we return the resulting set of patterns and their associated class labels.

The *prune*(t, w) procedure recursively removes leaf nodes that do not start with any word in W and are not in a negation relation with their parents. Intermediate nodes that connect the remaining nodes and do not start with any word in W are replaced by *. The pattern(t) procedure converts a dependency tree t to its semgrex format representation. Each "*" node is represented by an empty node *{}*, and most relations are represented by the generic > or >> relations (for direct and indirect relations, respectively), which match any type of relation. An exception to this is the negation relation, which is preserved in the semgrex pattern as the *>neg* token.

Example

Consider a sentence from the doctor review dataset class c_8 (wait time), "I arrived to my appointment on time and waited in his waiting room for over an hour," which has class label c_8^y (long wait). The dependency tree generated from this sentence is shown in Figure 3.

Figure 3. Dependency tree for the sentence "I arrived to my appointment on time and waited in his waiting room for over an hour".



Among the patterns extracted from this tree are:

1. {} > {word:/time.*/} >> {word:/hour.*/}
2. {word:/arrived.*/} > {word:/time.*/}
3. {} > {word:/time.*/} > ({} > {word:/room.*/} > {word:/hour.*/})
4. {word:/arrived.*/} >> {word:/hour.*/}

Pattern 1 means that some node has a direct descendant *time* and an indirect descendant *hour*. Pattern 2 means that *time* is a direct descendant of *arrived*. Pattern 3 means that some node has 2 direct descendants; 1 is *time* and the other is some other node that has direct descendants *room* and *hour*. Finally, pattern 4 means that *hour* is an indirect descendant of *arrived*.

Results

Classifiers Employed

We consider 3 types of classifiers:

1. *Statistical bag-of-words classifiers*, which view the documents as bags of keywords:
   - Random Forests (*RF*): RF, as implemented in Scikit-learn by Pedregosa et al [35]. Documents are represented with TF-IDF using n grams of 1 to 3 words, a minimum document frequency of 3%, up to 1000 features, stemming, and omission of stop words. The classifier uses 2000 trees. All other parameters are given their default values from [35].
   - *SVM*: C-support vector classifier as implemented in Scikit-learn by Pedregosa et al [35], which is based on the implementation from the study by Chang and Lin [36]. Documents are represented with TF-IDF using the same parameters as with random forest. The parameters for the classifier are given their default values from Scikit-learn by Pedregosa et al [35].
2. *Deep learning classifiers:*
   - *CNN* or *CNN-W* (CNN with Word2Vec): We use 2 variants of the CNN implementation by Britz [37]. Both use the default parameters. The first variant is initialized with a random uniform distribution, as in the CNN implementation by Britz [37]. The second is initialized with values from the Word2Vec model implementation from Gensim by Rehurek and Sojka [38].
   - *D2V-NN* (Doc2Vec Nearest Neighbor): A nearest neighbor classifier that uses the Doc2Vec model [39] implementation from Gensim by Rehurek and Sojka [38]. Documents are converted to paragraph vectors and classified according to the nearest neighbor using cosine similarity as the distance function.

   For CNN-W and D2V-NN, the Word2Vec and Doc2Vec models, respectively, are trained on an unlabeled set of 8,977,322 sentences from the collected doctor reviews that were not used to create the labeled dataset.
3. *NLP classifiers*, which exploit the dependency trees of a review's sentences:
   - *Matsumoto*: We implemented the method described in the study by Matsumoto et al [21] using the best-performing combination of features from their experiment using the Internet Movie Database dataset from the study of Pang and Lee [40], that is, unigrams, bigrams, frequent subsequences, and lemmatized frequent subtrees. For POS tagging before the step in frequent subsequence generation that splits sentences into clauses, our implementation uses the Stanford parser [41]. We use the dependency parser by Chen and Manning [34] to generate dependency trees for frequent subtree generation. For the SVM, we use the implementation from Pedregosa et al's Scikit-learn with a linear kernel and all other parameters given their default values from [35]. All parameters related to frequent subsequence and subtree generation are the same as described in the study by Matsumoto et al [21].
   - *DTC*: As described in the Methods section.

Variants of Dependency Tree Classifier

We consider the following variants of our DTC text classifier:

DTC: as described above, with sentences not matching any pattern classified as the most common class label in the training data.

DTC_{RF}: Sentences not matching any pattern are classified by a random forests classifier trained on the training data for each class.

$DTC_{CNN\text{-}W}$: Sentences not matching any pattern are classified by a CNN-W text classifier (as defined above) trained on the training data for each class.

Experiments

We performed experiments with the classifiers on each class of the doctor review dataset using 10-fold cross validation. To evaluate their performance, we use weighted accuracy. For a trained classifier C and dataset D of class c_i, we define this as shown below.

$Accuracy_c(C, D)$ is the ratio of sentences in D with class label c that were classified correctly by C. As before, $|c_i|$ is 3, the number of class labels in class c_i. We use weighted accuracy in our experiments as it places more importance on less frequent class labels, whereas regular accuracy is often above 90% because of the high number of instances labeled c_i^0 for each c_i.

The results of our experiments are shown below. In Table 3, we see that $DTC_{CNN\text{-}W}$ has better weighted accuracy than at least 4 baselines in each class. On average, it performs 2.19% better than the second-best method, the Matsumoto classifier ([57.05%-55.83%]/55.83%=2.19%). We also observe that both the deep learning classifiers (CNN, CNN-W, and D2V-NN) and NLP classifiers (Matsumoto and DTC variants) tend to perform better than the bag-of-words classifiers (RF and SVM). This is expected as the deep learning and NLP classifiers take advantage of information in sentences such as word order and syntactic structure that cannot be expressed by a bag-of-words vector.

Next, we further examine the performance of the top 3 classifiers, CNN-W, Matsumoto, and $DTC_{CNN\text{-}W}$. Table 4 shows the ratio of review sentences with class label c_i^x or c_i^y that were classified correctly in our experiments. Note that this is the $Accuracy_c(C, D)$ measure described above. $DTC_{CNN\text{-}W}$ generally outperforms the other classifiers with this measure; notable exceptions are c_6^y (bad staff), c_7^x (long visit time), and c_8^y (long wait time), where substantial numbers of sentences with these class labels were misclassified with the opposite label: 26.98% of c_6^y sentences were misclassified as c_6^x (good staff), 38.03% of c_7^x sentences were misclassified as c_7^y (short visit time), and 43.22% of c_8^y sentences were misclassified as c_8^x (short wait time). Finally, Table 5 shows the ratio of review sentences classified as c_i^x or c_i^y (ie, a classifier predicted their class labels as c_i^x or c_i^y) that were classified correctly. By this measure, $DTC_{CNN\text{-}W}$ performs poorly compared with CNN-W and Matsumoto. Although the DTC algorithm's semgrex patterns classify more sentences as c_i^x or c_i^y, many of these classifications are incorrect. In the next section, we discuss reasons for some of these misclassifications.

Table 3. Weighted accuracy of classifiers on doctor review dataset.

Classifier	c_1 (%)	c_2 (%)	c_3 (%)	c_4 (%)	c_5 (%)	c_6 (%)	c_7 (%)	c_8 (%)	Average (%)
CNN[a]	42.06	56.69	42.75	51.45	47.81	61.42	55.38	60.93	52.31
CNN-W[b]	49.89	*59.68*[c]	44.30	53.53	49.71	64.04	54.29	63.51	54.87
D2V-NN[d]	38.83	45.16	38.00	42.25	41.44	42.19	41.04	43.64	41.57
Matsumoto	45.76	59.63	45.89	53.40	49.89	*66.45*	57.24	*68.36*	55.83
RF[e]	40.78	42.00	34.76	37.29	41.62	52.88	45.65	51.66	43.33
SVM[f]	33.33	35.77	33.33	33.33	33.33	48.94	33.33	48.07	37.43
DTC[g]	51.72	50.48	41.27	47.23	38.49	54.31	*60.90*	65.91	51.29
DTC_{RF}	*54.00*	46.64	39.19	47.29	40.20	56.15	60.57	58.05	50.26
$DTC_{CNN\text{-}W}$	53.89	59.37	*48.66*	*57.98*	*50.77*	61.43	56.63	67.67	*57.05*

[a]CNN: Convolutional Neural Network.

[b]CNN-W: Convolutional Neural Network with Word2Vec.

[c]The highest value for each c_i is italicized for emphasis.

[d]D2V-NN: Doc2Vec Nearest Neighbor.

[e]RF: Random Forests.

[f]SVM: Support Vector Machine.

[g]DTC: dependency tree classifier.

Table 4. Per-label accuracy of top 3 classifiers on doctor review dataset for each c_i^x and c_i^y.

Label and classifier	c_1 (%)	c_2 (%)	c_3 (%)	c_4 (%)	c_5 (%)	c_6 (%)	c_7 (%)	c_8 (%)
c_1^x								
CNN-W[a]	31.37%	57.22%	0.00%	47.62%	40.54%	60.69%	45.07%	40.85%
Matsumoto	13.73%	57.04%	*4.00%* [b]	48.57%	41.16%	59.16%	*52.11%*	47.89%
DTC[c]$_{CNN-W}$	*33.33%*	*59.69%*	*4.00%*	*51.11%*	*48.02%*	*64.89%*	39.44%	*71.83%*
c_1^y								
CNN-W	19.05%	27.35%	34.48%	15.44%	13.36%	35.42%	18.99%	50.75%
Matsumoto	23.81%	27.65%	35.00%	13.24%	12.93%	*43.32%*	20.25%	*57.79%*
DTC$_{CNN-W}$	*33.33%*	*48.24%*	*47.51%*	*38.97%*	*25.00%*	27.52%	*35.44%*	35.68%

[a]CNN-W: Convolutional Neural Network with Word2Vec.

[b]For each c_i, the highest value for both c_i^x and c_i^y are italicized for emphasis.

[c]DTC: dependency tree classifier.

Table 5. Ratio of sentences classified by the top 3 classifiers as c_i^x or c_i^y that were classified correctly.

Label and classifier	c_1 (%)	c_2 (%)	c_3 (%)	c_4 (%)	c_5 (%)	c_6 (%)	c_7 (%)	c_8 (%)
c_1^x								
CNN-W[a]	34.78%	*60.19%* [b]	0.00%	62.50%	50.26%	66.81%	57.14%	65.91%
Matsumoto	*46.67%*	43.40%	*50.00%*	*66.23%*	*55.31%*	*71.10%*	*67.27%*	*77.27%*
DTC[c]$_{CNN-W}$	16.04%	41.66%	10.00%	20.69%	22.58%	43.59%	23.73%	21.52%
c_1^y								
CNN-W	40.00%	*41.52%*	50.56%	22.83%	*28.70%*	41.27%	29.41%	59.06%
Matsumoto	*58.82%*	34.18%	*56.52%*	*34.62%*	25.64%	*49.53%*	*53.33%*	*70.99%*
DTC$_{CNN-W}$	10.98%	13.50%	28.57%	13.38%	14.25%	22.90%	14.29%	29.96%

[a]CNN-W: Convolutional Neural Network with Word2Vec.

[b]For each c_i, the highest value for both c_i^x and c_i^y are italicized for emphasis.

[c]DTC: dependency tree classifier.

Discussion

Anecdotal Examples

In this section, we show some specific patterns generated by our algorithm along with some actual review sentences that match these patterns. The semgrex pattern *{} >neg {} >> ({word:/wait.*/} > {word:/long.*/})* was generated from a sentence with class label c_8^x (short wait) in class c_8 (wait time) in the doctor review dataset. It consists of a node that has 2 descendants: another generic node in a direct negation relation and *wait* in an indirect relation. The word *wait* has 1 direct descendant, the word *long*. The following is an example of a correctly matched sentence: "You are known by name and never have to wait long." This is an incorrectly matched one: "As a patient, I was not permitted to complain to the doctor about the long wait, placed on hold and never coming back to answer call." We see that it contains the words *long* and *wait*, as well as a negation (the word *never*); however, the negation is not semantically related to the *long wait* the author mentioned. Providing additional training data to the classifier may prevent such misclassifications by finding a pattern (or improving the rank of an existing pattern) that more appropriately makes such distinctions.

Limitations

In addition to the incorrect handling of negation described above, another limitation of our algorithm is that some sentences of a particular class can be sufficiently similar to sentences from another class, which may lead to misclassifications. Some examples of this can be seen in class c_6 (staff). Specifically, some sentences referring to a doctor (rather than staff members) were incorrectly classified as c_6^x (good staff) or c_6^y (bad staff). For example, "Dr. Fang provides the very best medical care available anywhere in the profession" and "Dr. Overlock treated me with the utmost respect," which clearly refer to doctors rather than staff and should have been classified as c_6^0 (no mention

of staff). The DTC algorithm generated some patterns for c_6^x that focus on positive statements for a person but miss the requirement that this person is staff. In the case of the above sentences, they were matched by *{} >> {word:/dr.*/} >> {word:/best.*/}* and *{} >> {word:/with.*/} >> {word:/dr.*/}*, respectively, which both erroneously include the word *dr*. More work is needed to address such tricky issues.

Conclusions

In this paper, we study the doctor reviews classification problem. We evaluate several existing classifiers and 1 new classifier. A key challenge of the problem is that features may be complex entities, for which polarity is not necessarily compatible with traditional positive or negative sentiment. Our proposed classifier, DTC, uses dependency trees generated from review sentences and automatically generates patterns that are then used to classify new reviews. In our experiments on a real-world doctor review dataset, we found that DTC outperforms other text classification methods. Future work may build upon the DTC classifier by also incorporating other NLP structures, such as discourse trees [42], to better capture the semantics of the reviews.

Acknowledgments

This project was partially supported by the National Science Foundation grants IIS-1447826 and IIS-1619463.

Authors' Contributions

RR built crawlers for collecting doctor reviews, labeled the doctor review dataset, researched related work, built the dependency tree classifier (DTC) algorithm, conducted the experiments, and wrote the manuscript. NM researched related work and wrote the pattern extraction algorithm. NXTL researched related work and provided guidance in building the DTC algorithm and conducting experiments. VH conceived the study, labeled the doctor review dataset, and provided coordination and guidance in the experiments and writing of the manuscript.

Conflicts of Interest

None declared.

References

1. Ding X, Liu B, Yu PS. A holistic lexicon-based approach to opinion mining. New York, NY, USA: Association for Computing Machinery; 2008 Feb 11 Presented at: 2008 International Conference on Web Search and Data Mining; February 11-12, 2008; Palo Alto, CA, USA p. 231-240. [doi: 10.1145/1341531.1341561]
2. Hu M, Liu B. Mining and summarizing customer reviews. New York, NY, USA: Association for Computing Machinery; 2004 Aug 22 Presented at: Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 22-25, 2004; Seattle, WA, USA p. 168-177. [doi: 10.1145/1014052.1014073]
3. Popescu AM, Etzioni O. Extracting product features and opinions from reviews. In: Natural Language Processing and Text Mining. London: Springer; 2007:9-28.
4. Gao GG, McCullough JS, Agarwal R, Jha AK. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. J Med Internet Res 2012 Feb 24;14(1):e38 [FREE Full text] [doi: 10.2196/jmir.2003] [Medline: 22366336]
5. Segal J, Sacopulos M, Sheets V, Thurston I, Brooks K, Puccia R. Online doctor reviews: do they track surgeon volume, a proxy for quality of care? J Med Internet Res 2012 Apr 10;14(2):e50 [FREE Full text] [doi: 10.2196/jmir.2005] [Medline: 22491423]
6. López A, Detz A, Ratanawongsa N, Sarkar U. What patients say about their doctors online: a qualitative content analysis. J Gen Intern Med 2012 Jan 04;27(6):685-692 [FREE Full text] [doi: 10.1007/s11606-011-1958-4] [Medline: 22215270]
7. Hao H. The development of online doctor reviews in China: an analysis of the largest online doctor review website in China. J Med Internet Res 2015 Jun 01;17(6):e134 [FREE Full text] [doi: 10.2196/jmir.4365] [Medline: 26032933]
8. Smith R, Lipoff J. Evaluation of dermatology practice online reviews: lessons from qualitative analysis. JAMA Dermatol 2016 Feb;152(2):153-157. [doi: 10.1001/jamadermatol.2015.3950] [Medline: 26606326]
9. Daskivich T, Luu M, Noah B, Fuller G, Anger J, Spiegel B. Differences in online consumer ratings of health care providers across medical, surgical, and allied health specialties: observational study of 212,933 providers. J Med Internet Res 2018 May 09;20(5):e176 [FREE Full text] [doi: 10.2196/jmir.9160] [Medline: 29743150]
10. Wallace BC, Paul MJ, Sarkar U, Trikalinos TA, Dredze M. A large-scale quantitative analysis of latent factors and sentiment in online doctor reviews. J Am Med Inform Assoc 2014 Jun 10;21(6):1098-1103 [FREE Full text] [doi: 10.1136/amiajnl-2014-002711] [Medline: 24918109]
11. Hao H, Zhang K. The voice of Chinese health consumers: a text mining approach to web-based physician reviews. J Med Internet Res 2016 May 10;18(5):e108 [FREE Full text] [doi: 10.2196/jmir.4430] [Medline: 27165558]

12. Hao H, Zhang K, Wang W, Gao G. A tale of two countries: international comparison of online doctor reviews between China and the United States. Int J Med Inform 2017 Mar;99:37-44. [doi: 10.1016/j.ijmedinf.2016.12.007] [Medline: 28118920]
13. Zhai Z, Liu B, Xu H, Jia P. Clustering product features for opinion mining. New York, NY, USA: Association for Computing Machinery; 2011 Feb 09 Presented at: Fourth ACM International Conference on Web Search and Data Mining; February 9-12, 2011; Hong Kong, China p. 347-354. [doi: 10.1145/1935826.1935884]
14. Liu Q, Gao Z, Liu B, Zhang Y. Automated rule selection for aspect extraction in opinion mining. : AAAI Press; 2015 Jul 25 Presented at: The 24th International Conference on Artificial Intelligence; July 25-31, 2015; Buenos Aires, Argentina p. 1291-1297.
15. Polanyi L, Zaenen A. Contextual valence shifters. In: Computing Attitude and Affect in Text: Theory and Applications. The Information Retrieval Series, vol 20. Dordrecht: Springer; 2006:1-10.
16. Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. New York, NY, USA: Association for Computing Machinery; 1992 Jul 01 Presented at: Fifth Annual Workshop on Computational Learning Theory; July 27-29, 1992; Pittsburgh, PA, USA p. 144-152. [doi: 10.1145/130385.130401]
17. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995 Sep;20(3):273-297. [doi: 10.1007/BF00994018]
18. Breiman L. Random forests. Mach Learn 2001 Oct;45(1):5-32. [doi: 10.1023/A:1010933404324]
19. Kim Y. Convolutional neural networks for sentence classification. Stroudsburg, PA, USA: Association for Computational Linguistics; 2014 Presented at: The 2014 Conference on Empirical Methods in Natural Language Processing; October 25-29, 2014; Doha, Qatar p. 1746-1751 URL: http://www.aclweb.org/anthology/D14-1181
20. Tesnière L. Éléments de Syntaxe Structurale. Paris: Klincksieck; 1959.
21. Matsumoto S, Takamura H, Okumura M. Sentiment classification using word sub-sequences and dependency sub-trees. Berlin: Springer; 2005 Presented at: 9th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD'05); May 18-20, 2005; Hanoi, Vietnam p. 301-311. [doi: 10.1007/11430919_37]
22. Rivas R, Montazeri N, Le NXT, Hristidis V. Github. DTC classifier URL: https://github.com/rriva002/DTC-Classifier [accessed 2018-05-06] [WebCite Cache ID 6zDfHhqEd]
23. Agarwal B, Poria S, Mittal N, Gelbukh A, Hussain A. Concept-level sentiment analysis with dependency-based semantic parsing: a novel approach. Cogn Comput 2015 Jan 20;7(4):487-499. [doi: 10.1007/s12559-014-9316-6]
24. Wawer A. Towards domain-independent opinion target extraction. : Institute of Electrical and Electronics Engineers; 2015 Presented at: 2015 IEEE International Conference on Data Mining; November 14-17, 2015; Atlantic City, NJ, USA p. 1326-1331. [doi: 10.1109/ICDMW.2015.255]
25. Pak A, Paroubek P. Text representation using dependency tree subgraphs for sentiment analysis. Berlin: Springer; 2011 Presented at: 16th International Conference on Database Systems for Advanced Applications (DASFAA 2011); April 22-25, 2011; Hong Kong, China p. 323-332. [doi: 10.1007/978-3-642-20244-5_31]
26. Kennedy G, McCollough A, Dixon E, Bastidas A, Ryan J, Loo C, et al. Hack Harassment: Technology Solutions to Combat Online Harassment. In: Proceedings of the first workshop on abusive language online. 2017 Presented at: Annual Meeting of the Association for Computational Linguistics; July 30-August 4, 2017; Vancouver, Canada p. 73-77 URL: http://www.aclweb.org/anthology/W17-3011
27. Gambäck B, Sikdar UK. Using convolutional neural networks to classify hate-speech. In: Proceedings of the first workshop on abusive language online. 2017 Presented at: Annual Meeting of the Association for Computational Linguistics; July 30-August 4, 2017; Vancouver, Canada p. 85-90 URL: https://pdfs.semanticscholar.org/0dca/29b6a5ea2fe2b6373aba9fe0ab829c06fd78.pdf
28. Mikolov T, Chen K, Corrado G, Dean J. Arxiv. 2013 Sep 07. Efficient estimation of word representations in vector space URL: https://arxiv.org/abs/1301.3781 [accessed 2018-10-19] [WebCite Cache ID 73IPH7ARs]
29. Lix L, Munakala SN, Singer A. Automated classification of alcohol use by text mining of electronic medical records. Online J Public Health Inform 2017 May 02;9(1):e069. [doi: 10.5210/ojphi.v9i1.7648]
30. Vitals. URL: http://www.vitals.com/ [accessed 2016-08-08] [WebCite Cache ID 6jcVjHu5p]
31. Yimam SM, Gurevych I, de Castilho RE, Biemann C. WebAnno: a flexible, Web-based and visually supported system for distributed annotations. 2013 Presented at: The 51st Annual Meeting of the Association for Computational Linguistics; August 4-9, 2013; Sofia, Bulgaria p. 1-6 URL: http://www.aclweb.org/anthology/P13-4001.pdf
32. Chambers N, Cer D, Grenager T, Hall D, Kiddon C, MacCartney B, et al. Learning alignments and leveraging natural logic. Stroudsburg, PA, USA: Association for Computational Linguistics; 2007 Presented at: The ACL-PASCAL Workshop on Textual Entailment and Paraphrasing; June 28-29, 2007; Prague, Czech Republic p. 165-170.
33. Manning CD, Surdeanu M, Bauer J, Finkel J, Bethard SJ, McClosky D. The Stanford CoreNLP natural language processing toolkit. In: Proceedings of the 52nd annual meeting of the association for computational linguistics: system demonstrations. 2014 Presented at: The 52nd Annual Meeting of the Association for Computational Linguistics; June 22-27, 2014; Baltimore, MD, USA p. 55-60 URL: https://nlp.stanford.edu/pubs/StanfordCoreNlp2014.pdf
34. Chen D, Manning CD. A fast and accurate dependency parser using neural networks. 2014 Presented at: The 2014 Conference on Empirical Methods in Natural Language Processing; October 25-29, 2014; Doha, Qatar p. 740-750 URL: https://cs.stanford.edu/~danqi/papers/emnlp2014.pdf

35. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825-2830.
36. Chang C, Lin C. LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2011 Apr 01;2(3):1-27. [doi: 10.1145/1961189.1961199]
37. Britz D. WildML. 2015 Dec 11. Implementing a CNN for text classification in TensorFlow URL: http://www.wildml.com/2015/12/implementing-a-cnn-for-text-classification-in-tensorflow/ [accessed 2018-05-06] [WebCite Cache ID 6zDgJUNLF]
38. Rehurek R, Sojka P. Software framework for topic modelling with large corpora. In: Proceedings of the LREC 2010 workshop on new challenges for NLP frameworks. 2010 Presented at: The Seventh International Conference on Language Resources and Evaluation; May 19-21, 2010; Valleta, Malta p. 45-50 URL: https://radimrehurek.com/gensim/lrec2010_final.pdf
39. Le Q, Mikolov T. Distributed representations of sentences and documents. 2014 Presented at: The 31st International Conference on Machine Learning; June 14-17, 2017; Lugano, Switzerland p. 1188-1196 URL: https://cs.stanford.edu/~quocle/paragraph_vector.pdf
40. Pang B, Lee L. A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. Stroudsburg, PA, USA: Association for Computational Linguistics; 2004 Presented at: The 42nd Annual Meeting of the Association for Computational Linguistics; July 21-26, 2004; Barcelona, Spain p. 271 URL: http://www.aclweb.org/anthology/P04-1035 [doi: 10.3115/1218955.1218990]
41. Klein D, Manning CD. Fast exact inference with a factored model for natural language parsing. Cambridge, MA, USA: MIT Press; 2003 Presented at: The 15th International Conference on Neural Information Processing Systems; December 9-14, 2002; Vancouver, British Columbia, Canada p. 3-10.
42. Joty S, Carenini G, Ng R, Mehdad Y. Combining intra-and multi-sentential rhetorical parsing for document-level discourse analysis. 2013 Presented at: The 51st Annual Meeting of the Association for Computational Linguistics; August 4-9, 2013; Sofia, Bulgaria p. 486-496 URL: http://www.aclweb.org/anthology/P13-1048

Abbreviations

Attr: attribute
CNN: Convolutional Neural Network
CNN-W: Convolutional Neural Network initialized with values from a Word2Vec model
D2V-NN: nearest neighbor classifier that uses the Doc2Vec model
DTC: dependency tree classifier
NLP: natural language processing
POS: part of speech
RF: random forests
S: sentiment
SVM: Support Vector Machine
TF-IDF: term frequency-inverse document frequency
T: target

Edited by G Eysenbach; submitted 30.05.18; peer-reviewed by M Tkáč, H Hao, L Shen; comments to author 26.07.18; revised version received 21.08.18; accepted 04.09.18; published 12.11.18.

Please cite as:
Rivas R, Montazeri N, Le NXT, Hristidis V
Automatic Classification of Online Doctor Reviews: Evaluation of Text Classifier Algorithms
J Med Internet Res 2018;20(11):e11141
URL: http://www.jmir.org/2018/11/e11141/
doi:10.2196/11141
PMID:30425030

Original Paper

Economic Evaluation of an Internet-Based Preventive Cognitive Therapy With Minimal Therapist Support for Recurrent Depression: Randomized Controlled Trial

Nicola S Klein[1,2*], MSc; Claudi LH Bockting[1,3*], PhD; Ben Wijnen[4,5*], PhD; Gemma D Kok[6*], PhD; Evelien van Valen[7*], PhD; Heleen Riper[8,9,10*], PhD; Pim Cuijpers[8,9*], PhD; Jack Dekker[8,11*], PhD; Colin van der Heiden[12,13], PhD; Huibert Burger[3,14*], PhD; Filip Smit[4,8,9*], PhD

[1]Department of Psychology, University of Groningen, Groningen, Netherlands

[2]Top Referent Traumacentrum, GGZ Drenthe, Beilen, Netherlands

[3]Department of Psychiatry, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands

[4]Department of Public Mental Health, Trimbos Institute, Utrecht, Netherlands

[5]Care and Public Health Research Institute School for Public Health and Primary Care, Department of Health Services Research, Maastricht University, Maastricht, Netherlands

[6]Ambulante Forensische Psychiatrie Noord, GGZ Drenthe, Emmen, Netherlands

[7]Department of Geriatrics, University Medical Center Utrecht, Utrecht, Netherlands

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

[9]Amsterdam Public Health, VU University Medical Center, Amsterdam, Netherlands

[10]GGZ inGeest, Amsterdam, Netherlands

[11]Research Department, Arkin Mental Health Institute, Amsterdam, Netherlands

[12]Department of Psychology, Erasmus University Rotterdam, Rotterdam, Netherlands

[13]PsyQ Mental Health Care Center, Rotterdam, Netherlands

[14]Department of General Practice, University Medical Center Groningen, Groningen, Netherlands

[*]these authors contributed equally

Corresponding Author:
Claudi LH Bockting, PhD
Department of Psychiatry
Amsterdam University Medical Centers
University of Amsterdam
Meibergdreef 9
Amsterdam,
Netherlands
Phone: 31 20 891 3610
Email: c.l.bockting@amc.uva.nl

Abstract

Background: Major depressive disorder (MDD) is highly recurrent and has a significant disease burden. Although the effectiveness of internet-based interventions has been established for the treatment of acute MDD, little is known about their cost effectiveness, especially in recurrent MDD.

Objectives: Our aim was to evaluate the cost effectiveness and cost utility of an internet-based relapse prevention program (mobile cognitive therapy, M-CT).

Methods: The economic evaluation was performed alongside a single-blind parallel group randomized controlled trial. Participants were recruited via media, general practitioners, and mental health care institutions. In total, 288 remitted individuals with a history of recurrent depression were eligible, of whom 264 were randomly allocated to M-CT with minimal therapist support added to treatment as usual (TAU) or TAU alone. M-CT comprised 8 online lessons, and participants were advised to complete 1 lesson per week. The economic evaluation was performed from a societal perspective with a 24-month time horizon. The health outcomes were number of depression-free days according to *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*, (DSM-IV) criteria assessed with the Structured Clinical Interview for DSM-IV axis I disorders by blinded interviewers after 3,

12, and 24 months. Quality-adjusted life years (QALYs) were self-assessed with the three level version of the EuroQol Five Dimensional Questionnaire (EQ-5D-3L). Costs were assessed with the Trimbos and Institute for Medical Technology Assessment Questionnaire on Costs Associated with Psychiatric Illness (TiC-P). Incremental cost-effectiveness ratios were calculated and cost-effectiveness planes and cost-effectiveness acceptability curves were displayed to assess the probability that M-CT is cost effective compared to TAU.

Results: Mean total costs over 24 months were €8298 (US $9415) for M-CT and €7296 (US $8278) for TAU. No statistically significant differences were found between M-CT and TAU regarding depression-free days and QALYs (*P*=.37 and *P*=.92, respectively). The incremental costs were €179 (US $203) per depression-free day and €230,816 (US $261,875) per QALY. The cost-effectiveness acceptability curves suggested that for depression-free days, high investments have to be made to reach an acceptable probability that M-CT is cost effective compared to TAU. Regarding QALYs, considerable investments have to be made but the probability that M-CT is cost effective compared to TAU does not rise above 40%.

Conclusions: The results suggest that adding M-CT to TAU is not effective and cost effective compared to TAU alone. Adherence rates were similar to other studies and therefore do not explain this finding. The participants scarcely booked additional therapist support, resulting in 17.3 minutes of mean total therapist support. More studies are needed to examine the cost effectiveness of internet-based interventions with respect to long-term outcomes and the role and optimal dosage of therapist support. Overall, more research is needed on scalable and cost-effective interventions that can reduce the burden of recurrent MDD.

Trial registration: Netherlands Trial Register NTR2503; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=2503 (Archived by WebCite at http://www.webcitation.org/73aBn41r3)

(J Med Internet Res 2018;20(11):e10437) doi:10.2196/10437

KEYWORDS

major depressive disorders; recurrence; cognitive therapy; internet; prevention; cost effectiveness

Introduction

In 2016, an estimated 268 million individuals worldwide were affected with a depressive disorder [1]. Major depressive disorder (MDD) is a highly recurrent disorder [2] with a substantial disease burden [1,3] and formidable economic costs due to health care use and productivity losses [4-7].

To alleviate the burden of MDD, psychological interventions and/or antidepressants are recommended for the acute phase of MDD and as continuation/maintenance therapy to prevent relapse and recurrence [8]. However, health care resources are limited and many individuals fail to seek treatment [9-12]. Because of their flexible and accessible nature, internet-based interventions could be a viable cost-effective solution that reaches a large number of at-risk individuals. The effectiveness of internet-based interventions in acute and residual MDD has been established [13-17], with small to moderate effect sizes for interventions without therapist support and higher effect sizes with therapist support [13,15]. To date, only one study examined the long-term effects of an internet-based relapse prevention program for depression [18] and no study has examined its cost effectiveness. Only a single study aimed at the prevention of MDD examined the cost effectiveness of an internet-based intervention but this study aimed to prevent the first onset of MDD [19]. Thus, little is known about the cost effectiveness of internet-based relapse prevention for recurrent MDD. More information is needed on the health impact and economic costs to inform policy makers and health care providers.

In our randomized controlled trial (RCT), we examined the clinical effectiveness of an internet-based relapse prevention program (mobile cognitive therapy, M-CT) added to treatment as usual (TAU) compared to TAU alone in remitted individuals with recurrent MDD. Results showed that M-CT added to TAU was slightly but not significantly superior to TAU alone after 24 months in terms of cumulative relapse/recurrence rate, number of depressive relapses, and depressive symptoms [20]. In this study, we evaluated the cost effectiveness and cost utility of M-CT to see if the economic case could be made for this low-cost and highly scalable intervention that was added to TAU. We hypothesize that M-CT added to TAU is cost effective compared with TAU alone as it might generate slightly better health outcomes and thereby lower costs for other mental health services.

Methods

Study Design

The economic evaluation was performed alongside a single-blind parallel group 2-arm RCT in which 288 participants aged between 18 and 65 years were eligible, of whom 264 were randomized to either M-CT added to TAU or TAU alone. This trial was registered at Nederlands Trial Register [NTR2503] and approved by Stichting Medisch-Ethische Toetsingscommissie Instellingen Geestelijke Gezondheidszorg, an independent medical ethics committee. The results were reported according to the Consolidated Standards of Reporting Trials–EHEALTH checklist [21] (Multimedia Appendix 1). The economic evaluation was conducted and reported according to the Consolidated Health Economic Evaluation Reporting Standards statement (Multimedia Appendix 2). The study design and results are described in detail elsewhere [20,22] but are summarized briefly here.

Participants and Procedure

The participants were recruited via media, general practitioners, and mental health care institutions and were included between

mid-September 2010 and August 2013 after providing a written informed consent. To be included, the following criteria had to be met: (1) a history of at least 2 major depressive episodes (MDEs) according to *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition,* (DSM-IV) criteria assessed with the Structured Clinical Interview for DSM-IV Disorders (SCID-I) [23] of which the latest MDE occurred within the last 2 years and (2) currently remitted for at least 2 months according to the SCID-I and a score of ≤10 on the Hamilton Rating Scale for Depression (HRSD) [24]. Exclusion criteria were current or past (hypo)mania, bipolar or psychotic disorder, alcohol or drug abuse, or a predominant anxiety disorder. Independent psychologists or research assistants interviewed the potential participants for inclusion and exclusion criteria. Our initial criterion of having experienced at least 2 depressive episodes within 5 years was discarded, as individuals with multiple episodes over a longer period of time are also at risk [25]. We examined whether this affected our primary outcomes, but this was not the case.

Randomization and Masking

Randomization was planned to be stratified by type of aftercare and number of MDEs, but eventually simple randomization was carried out due to a programming error. The participants were randomized (allocation ratio 1:1) by an independent researcher not otherwise involved in the study who was masked for clinical characteristics and who used computer-generated numbers generated in Stata software (StataCorp LLC). An independent researcher not involved in the follow-up interviews assigned the participants to the treatment conditions. The participants were not blinded to treatment allocation due to the nature of the intervention. The interviewers were unaware of the participants' treatment allocation, and the participants were instructed not to inform the interviewer of their treatment allocation. The assessor was replaced by another independent assessor in case the randomization was broken.

Interventions

M-CT is based on preventive cognitive therapy (PCT) [26], a face-to-face therapy that protects against relapse/recurrence in remitted individuals [27-29]. Bockting and Van Valen developed the content of M-CT [30], and it was built into the ePlatform of the Trimbos Institute, a nonprofit organization in the Netherlands with a focus on issues related to mental health and addiction. Participants from previous relapse prevention studies and a patients' association for depression (Depressievereniging) were involved in the development of the research question, outcome measures, design, development, and implementation of M-CT. The content of the program remained unaltered during the evaluation period and logfile analysis was used to monitor the use of the intervention. The program was free of charge for the participants, and an independent researcher provided participants with usernames and passwords to log in. M-CT comprised 8 online modules with minimal therapist support and continued mobile mood monitoring using text messages. The participants were advised to work on one lesson each week and were offered a minimum of 2 and a maximum of 4 telephone contacts with a therapist (maximum duration: 30 minutes per contact). Two therapist contacts were prebooked and 2 optional contacts could be booked by the participants additionally. Participating therapists were supervised by an experienced clinical psychologist. The primary task of the therapists was to work through the M-CT program. The participants received a reminder by email or text message if they did not log in to the website for 6 weeks. Feedback from the participants on the intervention was obtained by giving them the opportunity to evaluate each specific lesson. Participants randomized to M-CT and TAU continued to receive usual care, which could include, for example, antidepressants, counseling, or no care.

Costs

The economic evaluation was performed according to the Dutch guidelines [31] in which a societal perspective is recommended, implying that costs both inside and outside the health care sector are assessed. Health care costs included medication use and inpatient, outpatient, and primary care. Costs directly related to the M-CT intervention included costs of training and supervision of therapists during the study, the duration of contacts between participants and therapists (telephone and email contact), and costs related to information and communication technologies. The last category mainly consisted of costs related to developing, upgrading, and maintaining the M-CT website. In addition, various types of costs outside the health care sector were examined. Patient and family costs included informal care (ie, the monetary valuation of time invested by relatives or acquaintances in helping or assisting the participant), travel expenses, and psychiatric home care. Costs of productivity losses due to illness-related absence from work were estimated as were costs related to changes in efficiency while at work and changes in the amount of voluntary (unpaid) work conducted by the participants. Costs of productivity losses were estimated with the friction cost method [32]. This method takes the employer's perspective and calculates the time (the friction period) until another employee has replaced the worker that is absent.

Cost data were collected with the Trimbos and Institute for Medical Technology Assessment Questionnaire on Costs Associated with Psychiatric Illness (TiC-P) [33]. This questionnaire was administered online to all participants every 3 months, starting at baseline. Since there was variation in the maximum number of days medication use could be filled out by the participants, we extrapolated all medication use to 3 months. In order to facilitate comparisons with other economic evaluations, unit prices (ie, the price of one unit of each included cost type) were based on Dutch standard prices for the year 2014 [34]. Full economic cost prices of used resources were computed when standard prices were not available.

Outcomes

The economic evaluation comprised cost-effectiveness and cost-utility analyses. The health outcome measure of the cost-effectiveness analysis was the number of depression-free days based on DSM-IV criteria assessed with a telephone version of the SCID-I after 3, 12, and 24 months.

Quality-adjusted life years (QALYs) were used as a health outcome measure of the cost-utility analysis using the area under the curve method. The QALY is a health measure that combines

quality of life and the amount of time spent in a health condition, where one QALY is equal to 1 year lived in perfect health. The quality component of the QALY was derived from the three level version of the EuroQol Five Dimensional Questionnaire (EQ-5D-3L) administered online at 3-month intervals starting from baseline [35] by using Dutch tariffs to obtain utilities for specific health states [36]. The EQ-5D-3L is a commonly applied self-administered instrument that measures the generic health status and consists of 5 questions covering the following 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Both costs and health outcomes were discounted at 1.5% for health outcomes and 4% for costs according to the Dutch guidelines [31].

Economic Evaluation

The power calculation of the primary outcome is described elsewhere [20,22]. Since the study was only powered to detect differences in health outcomes and not in costs, as in most economic evaluations, we used probabilistic and medical decision-making techniques to draw inferences about the cost effectiveness. The intention-to-treat principle was used, in which all participants were analyzed according to their randomized condition, irrespective of their actual treatment. In our main analysis, we used multiple imputations by chained equations with predictive mean matching to account for missing data. The use of this technique may avoid bias associated with complete case analyses and makes optimal use of available data. Baseline variables predictive of clinical and cost outcomes and of a variable being missing were incorporated in the imputation model as recommended by White et al [37] to enhance the precision of the model and correct for possible bias. To account for participants with extremely high costs resulting in unstable imputation estimates, winsorizing was used for the main analyses. Using winsorizing, extreme values are instead replaced by certain percentiles, in this case the 97.5th percentile as opposed to trimming in which the extreme values are merely deleted [38,39].

Costs and outcomes were used to calculate the incremental cost-effectiveness ratio (ICER) of M-CT relative to TAU alone [40]. The following formula was used for calculating the ICER:

...where $C_{M\text{-}CT}$ and C_{TAU} are the mean costs, and $QALY_{M\text{-}CT}$ and $QALY_{TAU}$ are the mean QALYs in M-CT and TAU, respectively. The ICER is interpreted as the additional costs per QALY gained when M-CT is offered rather than TAU. The bootstrap method [41] was applied to account for the uncertainty in the economic evaluation by repeated random sampling with replacement from the original dataset. Seemingly unrelated regression equations were bootstrapped (5000 times) to allow for correlated residuals of the cost and utility equations and adjust for baseline differences in one of the sensitivity analyses. In each bootstrap step, the mean cost differences and the mean outcome differences were computed and these were plotted in the cost-effectiveness plane [42]. Finally, cost-effectiveness acceptability curves (CEACs) [43] were graphed, taking into account the relative placement of the bootstrap replications. CEACs inform decision makers on the likelihood that an intervention is deemed to be cost effective given a range of willingness-to-pay ceilings for gaining an additional unit of health (ie, gaining one QALY and gaining one depression-free day). The analyses were conducted using Stata (StataCorp LLC).

Sensitivity Analyses

Due to the amount of missing data, we used multiple imputations in the main analysis to handle missing data. To ascertain the robustness of our findings, we performed several sensitivity analyses, each handling missing data in a different way. It should be noted that in the main analysis, 29 participants were not included since they dropped out immediately after randomization and no follow-up information was available. Nevertheless, we performed an additional sensitivity analysis in which all participants were included for a full-fledged intention-to-treat analysis. The main analysis was repeated again but now restricted to individuals for whom at least 50% of the cost data were available. A final analysis to evaluate the impact of drop-out was based on complete cases. At baseline, we observed a slight imbalance between both conditions with respect to gender, severity of the last MDE, and baseline costs. Studies suggest that gender is not associated with relapse or recurrence but that severity of the last MDE might be [25]. Therefore, we repeated the main analysis now adjusting for the small baseline imbalances in severity of the last MDE and baseline costs.

Results

Sample Characteristics

The participant flow during the study is displayed in Multimedia Appendix 3. In total, 288 participants were eligible of whom 264 were randomized to either M-CT added to TAU (n=132) or TAU alone (n=132). In total, 29 participants dropped out immediately after randomization and 24 were lost to follow-up during the study. Overall, the baseline clinical and demographic characteristics of all participants (Table 1) and participants with any follow-up data were similar and equally distributed over the treatment conditions, suggesting no systematic bias due to drop-out of the 29 individuals with no follow-up data. Complete cost and effect data (available for all measurements) were available for 19.1 % (45/235) participants. At least one measurement of cost data after baseline was available for 83.0% (195/235) of the participants. For 54.9% (129/235) of the participants, at least half of the cost measurements were available.

Table 1. Sample characteristics by condition at baseline. Values may not add up to 100% because of missing data.

Characteristics	M-CT[a] (n=132)	TAU[b] (n=132)
Age (years), mean (SD)	45.6 (10.9)	47.1 (10.7)
Female, n (%)	105 (79.5)	92 (69.7)
Country of birth, The Netherlands, n (%)	116 (88.5)	121 (92.4)
Marital status, n (%)		
Single	39 (29.8)	32 (24.2)
Married or cohabiting	82 (62.6)	87 (65.9)
Divorced/widowed	10 (7.7)	13 (9.9)
Education, n (%)		
Primary and/or secondary education	17 (12.9)	22 (16.7)
Vocational education	30 (22.7)	34 (25.8)
Higher education	85 (64.4)	76 (57.6)
Employed, n (%)	87 (66.4)	90 (68.7)
Treatment as usual, n (%)		
No treatment	46 (34.8)	39 (30.0)
General practitioner	34 (25.8)	43 (33.1)
Specialized mental health aftercare	52 (39.4)	48 (36.9)
Treatment with antidepressants, n (%)	72 (55.4)	65 (50.8)
Previous episodes MDD[c], median (IQR[d])	4 (2.8)	4 (2.0)
Total HRSD[e], mean (SD)	3.7 (3.1)	3.4 (2.9)
Depressive symptoms (IDS-SR[f]), mean (SD)	16.5 (10.3)	16.3 (9.7)
Severity past episode, n (%)		
Mild	37 (28.0)	25 (18.9)
Moderate	73 (55.3)	71 (53.8)
Severe	22 (16.7)	36 (27.3)
Baseline utilities (EQ-5D-3L[g]), mean (SD)	0.86 (0.12)	0.84 (0.17)
Baseline costs (€), mean (SD)	1729 (3699)	1552 (3216)

[a]M-CT: mobile cognitive therapy.

[b]TAU: treatment as usual.

[c]MDD: major depressive disorder.

[d]IQR: interquartile range.

[e]HRSD: Hamilton Rating Scale for Depression.

[f]IDS-SR: Inventory of Depressive Symptomatology, Self-Report.

[g]EQ-5D-3L: three level version of the EuroQol Five Dimensional Questionnaire.

Costs

The various types of costs generated by both groups during the 24 months of the study are presented in Multimedia Appendix 4. Presented costs were based on the data of participants for whom at least one cost measurement after the baseline assessment was available during the study. Mean costs of the M-CT intervention were €73 (US $83) per participant. These costs were mainly related to the training and supervision of therapists, contacts between therapists and participants, and information and communication technology costs (development and periodically upgrading the software and server costs). In both conditions, the costs of hospital admissions and outpatient care provided by mental health care services contributed considerably to the overall costs within the health care sector. Costs related to productivity losses had the largest impact on overall societal costs. Mean total costs accrued over 24 months were €8298 (US $9410) for M-CT and €7296 (US $8274) for TAU.

Outcomes

The mean number of depression-free days within the 24 months of the study was 661 in M-CT and 656 in TAU. Mean QALYs were 1.65 for both M-CT and TAU. No statistically significant

differences in depression-free days and QALYs were found (t=0.43, P=.67 and t=0.18, P=.86, respectively).

Economic Evaluation

According to the main analysis, M-CT resulted in an extra 5.6 gain in depression-free days (95% CI 5.3-6.0) and a 0.004 QALY gain (95% CI 0.004-0.005)—but these health gains were achieved at higher costs (€1008 [US $1143], 95% CI €983-€1034). For both health outcomes, most of the bootstrapped ICERs were located in the northeast quadrant (55.5% for depression-free days and 46.5% for QALYs), indicating that the probability that M-CT is deemed cost effective depends on the willingness-to-pay for an additional health gain (see Figures 1 and 2). When the willingness-to-pay per additional depression-free day is zero, M-CT has approximately a 40% probability of being cost effective. When the willingness-to-pay per additional gain in depression-free days increases, the probability that M-CT is cost effective also increases but does not rise above 65%. For QALYs, increased willingness-to-pay only leads to slight increases in the probability that M-CT will be considered cost effective and the probability that M-CT is cost effective does not rise above 40%.

Sensitivity Analyses

Table 2 displays the main analysis and sensitivity analyses. The sensitivity analysis including all randomized participants (n=264) overall yielded similar results. When taking into account participants for whom at least 50% of the data were available, TAU dominated M-CT in terms of depression-free days, and results were roughly similar to the main analysis for QALYs. In the complete case analysis, TAU dominated M-CT. Adjustments for imbalanced baseline variables yielded similar results.

Figure 1. Incremental cost-effectiveness plane (left) and cost-effectiveness acceptability curve (right) for the 5000 bootstrapped incremental costs per depression-free day gained. Reps: bootstrap replication; PE-line: line which represents the point estimate of the ICER (average cost/effect of bootstrap replications).

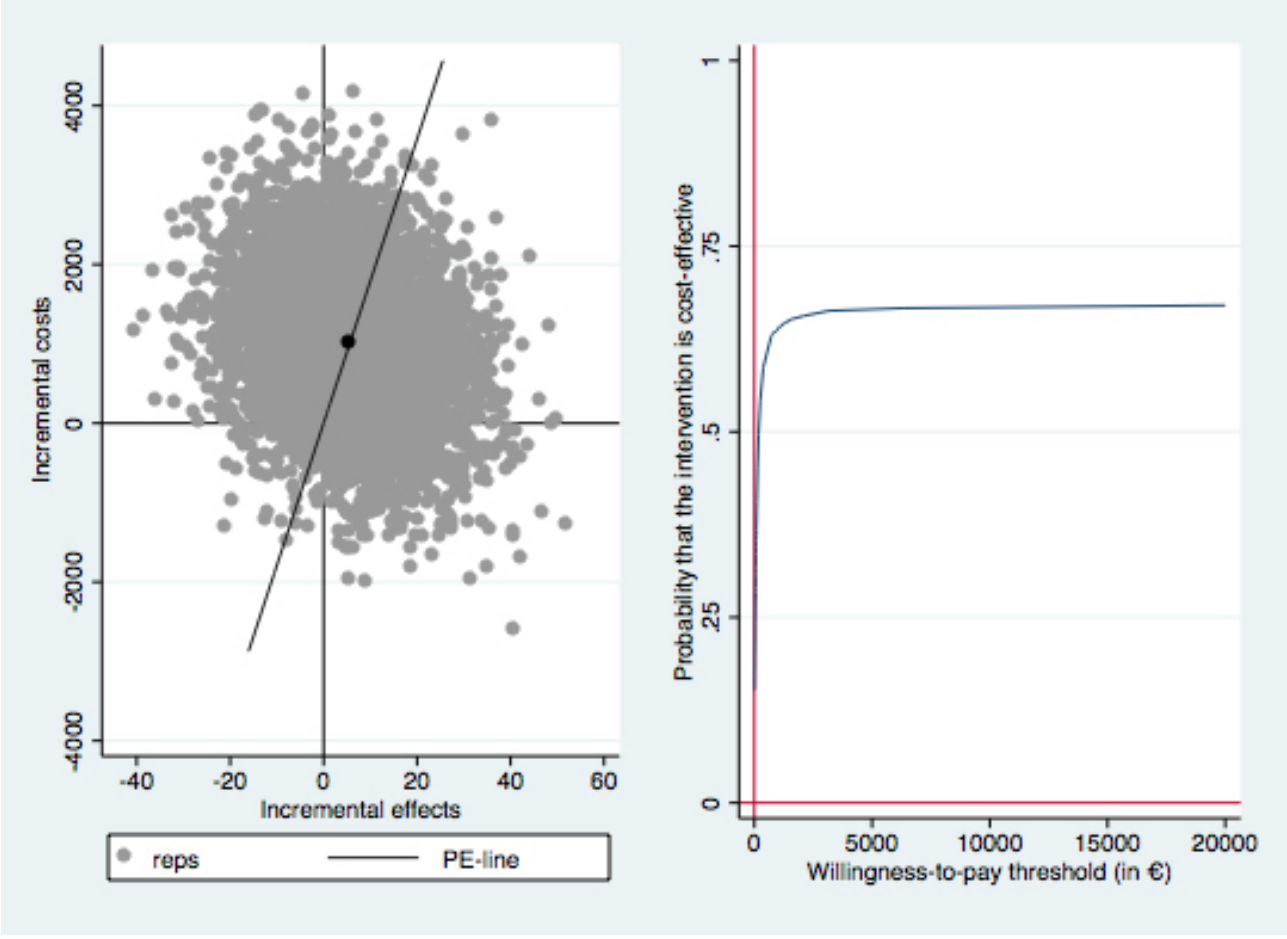


Figure 2. Incremental cost-effectiveness plane (left) and cost-effectiveness acceptability curve (right) for the 5000 bootstrapped incremental costs per quality-adjusted life years gained. Reps: bootstrap replication; PE-line: line which represents the point estimate of the ICER (average cost/effect of bootstrap replications).

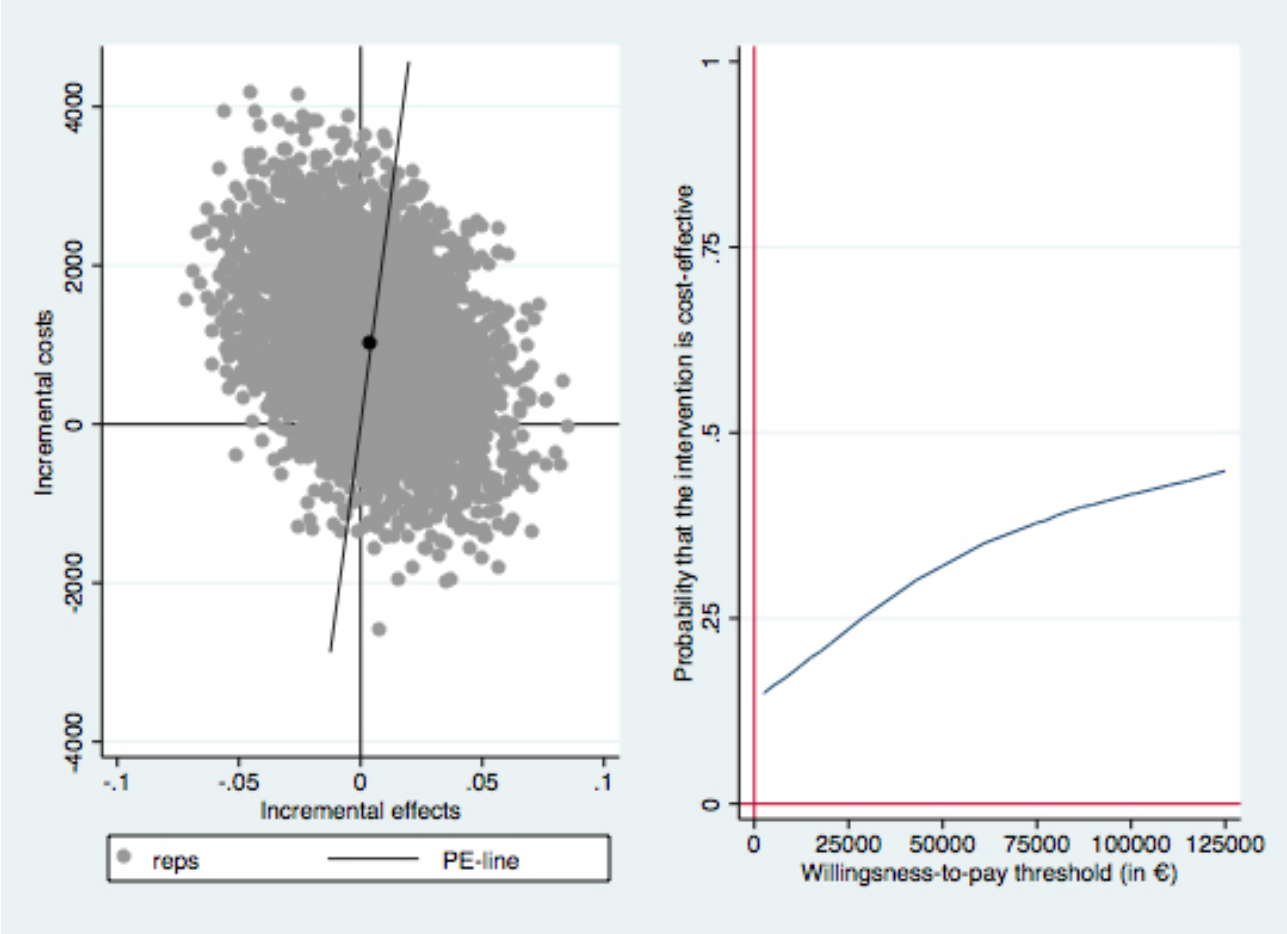


Table 2. Sensitivity analyses with all estimates based on 5000 bootstrap replications.

Characteristics	Costs and effects		Mean ICER[a] (€)	Distribution of cost- effectiveness plane			
	Incremental costs (€)	Incremental effects		NE[b] (%)	SE[c] (%, dominant)	SW[d] (%)	NW[e] (%, inferior)
Cost effectiveness, depression-free days							
Main analysis[f] (n=235)	1008 (983-1034)	5.6 (5.3 to 6.0)	179	55.5	11.6	2.6	30.2
Main analysis all participants[g] (n=264)	1033 (1000-1065)	4.7 (4.4 to 5.0)	140	51.7	15.5	3.8	29.1
≥50% follow-up data (n=129)	1562 (1508-1617)	–34.7 (–35.2 to –34.1)	(dominated)	1.8	1.3	20.3	76.6
Complete case (n=45)	2086 (1990-2181)	–51.7 (–52.5 to –50.9)	(dominated)	1.8	1.3	20.3	76.6
Cost utility, QALYs[h]							
Main analysis (n=235)	1008 (983-1034)	0.004 (0.004 to 0.005)	230,816	46.5	11.2	3.0	39.3
Main analysis all participants (n=264)	1033 1000-1065)	0.02 (0.02 to 0.02)	53,583	63.9	16.5	1.6	18.1
≥50% follow-up data (n=129)	1562 (1508-1617)	0.02 (0.02 to (0.02)	87,676	48.8	20.0	2.6	29.6
Complete case (n=45)	2086 (1990-2181)	–0.04 (–0.05 to –0.04)	(dominated)	15.6	11.9	15.7	56.8

[a]ICER: incremental cost-effectiveness ratio.

[b]NE: northeast.

[c]SE: southeast.

[d]SW: southwest.

[e]NW: northwest.

[f]Main analysis using multiple imputations with predictive mean matching, excluding the 29 participants with no follow-up data.

[g]Main analysis but now with all participants including the 29 participants without follow-up data.

[h]QALY: quality-adjusted life year.

Discussion

Principal Findings and Comparison With Prior Work

This study was the first to evaluate the cost effectiveness and cost utility of an internet-based relapse prevention program for recurrent MDD. The results of this study suggest that M-CT added to TAU is not cost effective compared to TAU alone over a 24-month period.

The results revealed that the total costs in both conditions were dominated by losses in productivity, which is consistent with Krol et al [44] and Biesheuvel-Leliefeld et al (unpublished data, 2018). Krol et al [44] found that choices made regarding the inclusion or exclusion of indirect costs influence cost-effectiveness studies to a great extent. As MDD is a disorder with extensive losses in worker productivity [4-7,44], we believe that the societal perspective, in which both direct and indirect costs are included, is justified. Our main analysis showed that participants using M-CT had slightly more depression-free days (5.6 days) and better QALYs (0.004) compared to TAU but achieved at higher costs (€1008 [US $1143]). The probability that M-CT was cost effective in terms of depression-free days and QALYs was dependent on the willingness-to-pay for an additional health gain. For depression-free days, a substantial investment had to be made before reaching an acceptable probability that M-CT was deemed cost effective. For QALYs, a substantial investment had to be made but the probability that M-CT was cost effective did not rise above 40%. We conclude that adding M-CT to TAU is not a cost-effective strategy compared to TAU alone. The results of the sensitivity analyses were partly inconsistent with the main analysis and will be further discussed.

The results contrast with our expectations based on the positive short-term results of M-CT [45] and with the promising findings of Holländare [18] regarding the internet-based relapse prevention therapy in partially remitted individuals. Moreover, the results contrast with the long-term effectiveness of face-to-face PCT that is found by Bockting et al [27-29,46] and De Jonge et al (unpublished data, 2018) and with PCT administered as bibliotherapy with therapist support [47]. As reported elsewhere [20], treatment adherence in this study was comparable to other studies (68.2% [90/132] finished at least 5 lessons) [15] and, therefore, we believe this did not explain the results. Our short-term clinical results demonstrated a positive effect of M-CT on residual depressive symptoms during the first months compared to TAU [45], whereas our long-term results showed no protective effects of M-CT on cumulative relapse/recurrence rate, number of depressive relapses/recurrences, and depressive symptoms after 24 months [20]. Therefore, we hypothesize that M-CT might generate better effects at lower costs during the first months and hence might be more cost effective compared to TAU but that the costs and effects become less favorable for M-CT in the long run. This hypothesis is consistent with the meta-analytic review of So et

al [48] that found short-term effects of internet-based treatments for MDD but no effects beyond 6 months posttreatment. However, a recent review on the long-term effects of internet-based guided cognitive behavioral therapy showed enduring effects for several disorders [49]. We hypothesize that more therapist support and/or booster sessions are needed for M-CT to become more cost effective in the long term. In this study, minimal therapist support was enlisted by the participants (ie, a mean total therapist time of only 17.3 minutes [range 0-70] per participant). Several systematic reviews and meta-analyses showed that internet-based interventions are especially effective when provided with therapist support [13,15,16]. Therapist guidance also seems an important factor in economic evaluations of internet-based interventions. A recent systematic review on economic evaluations of internet- and mobile-based interventions for the treatment and prevention of depression concluded that the internet-based interventions that were likely to be cost effective were all guided interventions whereas the unguided interventions were not likely to be cost effective or the results were ambiguous [50]. However, it is important to note that for most of these guided interventions, a considerable investment was needed in order to reach an acceptable probability that the intervention was deemed cost effective. In their systematic review on the cost effectiveness of internet-based interventions for a wide range of mental health disorders, Donker et al [51] highlighted that the most robust evidence in terms of cost effectiveness was found for guided interventions. A recent individual-participant data meta-analysis on the cost effectiveness of guided internet-based interventions for depression based on 5 studies concluded that considerable investments had to be made for an acceptable probability that the intervention would be cost effective compared to controls in terms of treatment response and depressive symptoms and that for QALYs, this probability was low at the widely used willingness-to-pay-threshold [52]. Differences between the systematic reviews and meta-analysis might be caused by differences in methodology, control group, and/or differences in when an intervention is perceived cost effective. For example, in the systematic review of Paganini et al [50], studies examining only direct health care costs were also included and 4 out of the 7 studies classified as cost effective used a wait list control group as comparator, whereas in the meta-analysis of Kolovos et al [52], only studies that also took into account productivity losses were included and only one study included a wait list control group.

Thus, more information is needed under which circumstances internet-based treatments are effective and cost effective regarding short-term and long-term outcomes. Based on their systematic review, Erbe et al [53] concluded that combining the strength of both face-to-face and internet-based interventions might be a promising direction, although more studies are needed. In addition, besides examining the (cost) effectiveness of specific internet-based interventions, future studies should focus more on the implementation in clinical practice, taking into account specific barriers (eg, preferences of individuals and professionals) [54,55]. Furthermore, more information is needed under which circumstances face-to-face or other forms of PCT are cost effective. To date, one study examined the cost effectiveness of PCT administered as guided bibliotherapy for remitted recurrently depressed individuals (Biesheuvel-Leliefeld et al, unpublished data, 2018) and another study examined the cost effectiveness of PCT in remitted recurrently depressed individuals that had received acute-phase cognitive therapy (De Jonge et al, unpublished data, 2018). Both studies concluded that investments had to be made for an acceptable probability that the intervention would be cost effective in terms of recurrences. For QALYs, substantial investments had to be made but probabilities that the intervention would be cost effective remained low.

Limitations

Some limitations of this study are important to acknowledge. First, cost data were collected with a self-report questionnaire approximately every 3 months during the 24-month follow-up and therefore a substantial amount of participants missed one or more measurement. To deal with missing data, multiple imputations, which is a recommended strategy to handle missing data in cost-effectiveness studies performed alongside RCTs [56,57], were used in our main analysis. We can assume the data were at least partly missing at random since baseline characteristics predicted whether the data were missing. Nevertheless, the missing completely at random assumption cannot be proved, and it is possible data were missing not at random because drop-out could be related to depressive symptom severity. Because of the amount of missing data, we did not want to rely on a single imputation technique and therefore performed several sensitivity analyses that each handled missing data in a different way. The main analysis, the analysis including all participants, and the analysis incorporating participants for whom at least 50% of the data were available (the latter only regarding QALYs) showed similar results. The analysis including participants with at least 50% data (regarding depression-free days) and complete cases showed higher costs and worse outcomes for M-CT compared to TAU. Multiple imputations are preferred over a complete case analysis because of the potential selection bias that might occur due to missing values. The results of the complete cases and cases with at least 50% of the data do suggest a possible selection bias in drop-out, which is also suggested when inspecting a baseline table displaying only the complete cases and cases with 50% of the data. Altogether, we regard our main analysis as primary. Second, the data were obtained in the Netherlands, and generalizability into other countries with other treatment settings is questionable. Third, the cost data and data for the cost-utility analysis were based on retrospective self-report questionnaires which may have affected the reliability. The TiC-P has shown to be a reliable and valid questionnaire for collecting cost data [58]. However, the EQ-5D might be subjected to a possible ceiling effect when estimating changes in QALYs for this group of remitted individuals. Moreover, the EQ-5D refers to the current health state and therefore does not capture all relapses/recurrences during the 24 months of the study.

Conclusions

Although the effectiveness of internet-based therapy for depression is currently established, only a few studies examined the cost effectiveness of these interventions [17,50-52,59,60]. We conclude that adding M-CT to TAU is not an effective or

cost-effective strategy to prevent relapse and recurrence. Future studies should examine the long-term effectiveness of internet-based interventions and the optimal dosage of guidance by therapists. MDD is highly recurrent [2] and one of the leading causes of disability [1,3]. Therefore, it is important that future studies continue to examine highly accessible, scalable, and (potentially) cost-effective interventions to treat depression including interventions that prevent relapse and recurrence. These studies are needed to inform decisions in mental health care. Since treatment effects can manifest differently over time [48], it is important that these cost-effectiveness studies on face-to-face and internet-based interventions include long follow-up periods.

Acknowledgments

We are grateful to all participants, recruitment sites, therapists, doctoral students, and volunteers who contributed to this study. The Netherlands Organization for Health Research and Development (Department of Disease Management and Chronic Illnesses, grant number 300020014) funded this RCT. The funder did not have a role in the study design, the data collection procedure, in analyzing or interpreting the data, or in publication of the results. This article has partly been written by CLHB as part of a fellowship at the Netherlands Institute for Advanced Study in the Humanities and Social Science Study.

Authors' Contributions

NSK was involved in the data collection process, interpretation of the data, and drafting and critically revising the manuscript. CLHB was involved in the design and development of M-CT, data collection, interpretation of the data, and critically revising the manuscript. BW was involved in the analysis and interpretation of the data and in drafting and critically revising the content of the manuscript. GDK was involved in piloting M-CT, recruitment, data collection, interpretation of the data, and critically revising the manuscript. EvV was involved in the development of M-CT and supervision of therapists and critically revising the content of the manuscript. HB, HR, PC, JD, FS, and CvdH were involved in the design of the study and critically revising the manuscript. All authors have significantly contributed to the manuscript and have approved the final version.

Conflicts of Interest

CLHB and EvV developed M-CT, which was integrated in the platform of the Trimbos Institute in collaboration with FS. No other disclosures are reported.

Multimedia Appendix 1

CONSORT‐EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 4MB - jmir_v20i11e10437_app1.pdf]

Multimedia Appendix 2

Consolidated Health Economic Evaluation Reporting Standards statement.

[PDF File (Adobe PDF File), 1MB - jmir_v20i11e10437_app2.pdf]

Multimedia Appendix 3

Consolidated Standards of Reporting Trials flow diagram over the 24-month follow-up.

[PNG File, 293KB - jmir_v20i11e10437_app3.png]

Multimedia Appendix 4

Cumulative costs (€) inside and outside the health care sector over 24 months.

[PDF File (Adobe PDF File), 31KB - jmir_v20i11e10437_app4.pdf]

References

1. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017 Sep 16;390(10100):1211-1259 [FREE Full text] [doi: 10.1016/S0140-6736(17)32154-2] [Medline: 28919117]
2. Richards D. Prevalence and clinical course of depression: a review. Clin Psychol Rev 2011 Nov;31(7):1117-1125. [doi: 10.1016/j.cpr.2011.07.004] [Medline: 21820991]

3. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. Lancet 2013 Nov 9;382(9904):1575-1586. [doi: 10.1016/S0140-6736(13)61611-6] [Medline: 23993280]
4. Greenberg PE, Fournier A, Sisitsky T, Pike CT, Kessler RC. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). J Clin Psychiatry 2015 Feb;76(2):155-162 [FREE Full text] [doi: 10.4088/JCP.14m09298] [Medline: 25742202]
5. Gustavsson A, Svensson M, Jacobi F, Allgulander C, Alonso J, Beghi E, et al. Cost of disorders of the brain in Europe 2010. Eur Neuropsychopharmacol 2011 Oct;21(10):718-779. [doi: 10.1016/j.euroneuro.2011.08.008] [Medline: 21924589]
6. Luppa M, Heinrich S, Angermeyer MC, König H, Riedel-Heller SG. Cost-of-illness studies of depression: a systematic review. J Affect Disord 2007 Feb;98(1-2):29-43. [doi: 10.1016/j.jad.2006.07.017] [Medline: 16952399]
7. Olesen J, Gustavsson A, Svensson M, Wittchen H, Jönsson B, CDBE2010 study group, European Brain Council. The economic cost of brain disorders in Europe. Eur J Neurol 2012 Jan;19(1):155-162. [doi: 10.1111/j.1468-1331.2011.03590.x] [Medline: 22175760]
8. Practice Guidelines for the Treatment of Patients with Major Depressive Disorder, Third Edition.: American Psychiatric Association; 2010. URL: https://psychiatryonline.org/guidelines [accessed 2017-12-04] [WebCite Cache ID 6vSO7AKMR]
9. Demyttenaere K, Bruffaerts R, Posada-Villa J, Gasquet I, Kovess V, Lepine JP, et al. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. JAMA 2004 Jun 2;291(21):2581-2590. [doi: 10.1001/jama.291.21.2581] [Medline: 15173149]
10. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bull World Health Organ 2004 Nov;82(11):858-866 [FREE Full text] [Medline: 15640922]
11. Boerema A, Kleiboer A, Beekman A, van Zoonen K, Dijkshoorn H, Cuijpers P. Determinants of help-seeking behavior in depression: a cross-sectional study. BMC Psychiatry 2016 Mar 23;16(78):78-79 [FREE Full text] [doi: 10.1186/s12888-016-0790-0] [Medline: 27009062]
12. Olfson M, Blanco C, Marcus SC. Treatment of adult depression in the United States. JAMA Intern Med 2016 Oct 01;176(10):1482-1491. [doi: 10.1001/jamainternmed.2016.5057] [Medline: 27571438]
13. 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]
14. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 01;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
15. 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]
16. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
17. Ahern E, Kinsella S, Semkovska M. Clinical efficacy and economic evaluation of online cognitive behavioral therapy for major depressive disorder: a systematic review and meta-analysis. Expert Rev Pharmacoecon Outcomes Res 2018 Feb;18(1):25-41. [doi: 10.1080/14737167.2018.1407245] [Medline: 29145746]
18. Holländare F, Anthony SA, Randestad M, Tillfors M, Carlbring P, Andersson G, et al. Two-year outcome of internet-based relapse prevention for partially remitted depression. Behav Res Ther 2013 Nov;51(11):719-722. [doi: 10.1016/j.brat.2013.08.002] [Medline: 24021360]
19. Buntrock C, Berking M, Smit F, Lehr D, Nobis S, Riper H, et al. Preventing depression in adults with subthreshold depression: health-economic evaluation alongside a pragmatic randomized controlled trial of a web-based intervention. J Med Internet Res 2017 Dec 04;19(1):e5 [FREE Full text] [doi: 10.2196/jmir.6587] [Medline: 28052841]
20. Klein NS, Kok GD, Burger H, van Valen E, Riper H, Cuijpers P, et al. No sustainable effects of an internet-based relapse prevention program over 24 months in recurrent depression: primary outcomes of a randomized controlled trial. Psychother Psychosom 2018;87(1):55-57. [doi: 10.1159/000485039] [Medline: 29306953]
21. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
22. Bockting CLH, Kok GD, van der Kamp L, Smit F, van Valen E, Schoevers R, et al. Disrupting the rhythm of depression using Mobile Cognitive Therapy for recurrent depression: randomized controlled trial design and protocol. BMC Psychiatry 2011 Jan 14;11(12):1-9 [FREE Full text] [doi: 10.1186/1471-244X-11-12] [Medline: 21235774]
23. Spitzer RL, Williams JB, Gibbon M, First MB. The Structured Clinical Interview for DSM-III-R (SCID). I: history, rationale, and description. Arch Gen Psychiatry 1992 Aug;49(8):624-629. [Medline: 1637252]
24. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960 Feb;23:56-62 [FREE Full text] [Medline: 14399272]

25. Hardeveld F, Spijker J, De Graaf R, Nolen WA, Beekman ATF. Prevalence and predictors of recurrence of major depressive disorder in the adult population. Acta Psychiatr Scand 2010 Sep;122(3):184-191. [doi: 10.1111/j.1600-0447.2009.01519.x] [Medline: 20003092]
26. Bockting C. Preventive Cognitive Therapy for Recurrent Depression. Houten: Bohn Stafleu van Loghum; 2009.
27. Bockting CLH, Schene AH, Spinhoven P, Koeter MWJ, Wouters LF, Huyser J, et al. Preventing relapse/recurrence in recurrent depression with cognitive therapy: a randomized controlled trial. J Consult Clin Psychol 2005 Aug;73(4):647-657. [doi: 10.1037/0022-006X.73.4.647] [Medline: 16173852]
28. Bockting CLH, Smid NH, Koeter MWJ, Spinhoven P, Beck AT, Schene AH. Enduring effects of Preventive Cognitive Therapy in adults remitted from recurrent depression: a 10 year follow-up of a randomized controlled trial. J Affect Disord 2015 Oct 01;185:188-194. [doi: 10.1016/j.jad.2015.06.048] [Medline: 26188380]
29. Bockting CLH, Spinhoven P, Wouters LF, Koeter MWJ, Schene AH, DELTA Study Group. Long-term effects of preventive cognitive therapy in recurrent depression: a 5.5-year follow-up study. J Clin Psychiatry 2009 Dec;70(12):1621-1628. [doi: 10.4088/JCP.08m04784blu] [Medline: 20141705]
30. Bockting CLH, van Valen E. Ingredients of Mobile Preventive Cognitive Therapy for Recurrent Depression. Groningen: University of Groningen; 2009.
31. Zorginstituut Nederland. Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. Diemen: Zorginstituut Nederland; 2015.
32. Koopmanschap MA, Rutten FF, van Ineveld BM, van Roijen L. The friction cost method for measuring indirect costs of disease. J Health Econ 1995 Jun;14(2):171-189. [Medline: 10154656]
33. Hakkaart-van Roijen L. Trimbos/iMTA Questionnaire for Costs Associated with Psychiatric Illness (TIC-P). Rotterdam: Erasmus University; 2002.
34. Zorginstituut Nederland. Methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg. Geactualiseerde versie 2015. Diemen: Zorginstituut Nederland; 2015.
35. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001 Jul;33(5):337-343. [Medline: 11491192]
36. Lamers LM, McDonnell J, Stalmeier PFM, Krabbe PFM, Busschbach JJV. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. Health Econ 2006 Oct;15(10):1121-1132. [doi: 10.1002/hec.1124] [Medline: 16786549]
37. White IR, Carlin JB. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. Stat Med 2010 Dec 10;29(28):2920-2931. [doi: 10.1002/sim.3944] [Medline: 20842622]
38. Gosh D, Vogt A. Outliers: An Evaluation of Methodologies. San Diego: American Statistical Association; 2012.
39. Thomas JW, Ward K. Economic profiling of physician specialists: use of outlier treatment and episode attribution rules. Inquiry 2006;43(3):271-282. [doi: 10.5034/inquiryjrnl_43.3.271] [Medline: 17176969]
40. Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddart G. Methods for the Economic Evaluation of Health Care Programmes, Third Edition. Oxford: Oxford University Press; 2005.
41. Efron B, Tibshirani R. An Introduction to the Bootstrap. New York: Chapman & Hall/CRC; 1994.
42. Black WC. The CE plane: a graphic representation of cost-effectiveness. Med Decis Making 1990;10(3):212-214. [doi: 10.1177/0272989X9001000308] [Medline: 2115096]
43. Fenwick E, O'Brien BJ, Briggs A. Cost-effectiveness acceptability curves—facts, fallacies and frequently asked questions. Health Econ 2004 May;13(5):405-415. [doi: 10.1002/hec.903] [Medline: 15127421]
44. Krol M, Papenburg J, Koopmanschap M, Brouwer W. Do productivity costs matter? The impact of including productivity costs on the incremental costs of interventions targeted at depressive disorders. Pharmacoeconomics 2011 Jul;29(7):601-619. [doi: 10.2165/11539970-000000000-00000] [Medline: 21545189]
45. Kok G, Burger H, Riper H, Cuijpers P, Dekker J, van Marwijk H, et al. The three-month effect of mobile internet-based cognitive therapy on the course of depressive symptoms in remitted recurrently depressed patients: results of a randomized controlled trial. Psychother Psychosom 2015 Feb 21;84(2):90-99. [doi: 10.1159/000369469] [Medline: 25721915]
46. Bockting CLH, Klein NS, Elgersma HJ, van Rijsbergen GD, Slofstra C, Ormel J, et al. Effectiveness of preventive cognitive therapy while tapering antidepressants versus maintenance antidepressant treatment versus their combination in prevention of depressive relapse or recurrence (DRD study): a three-group, multicentre, randomised controlled trial. Lancet Psychiatry 2018 May;5(5):401-410. [doi: 10.1016/S2215-0366(18)30100-7] [Medline: 29625762]
47. Biesheuvel-Leliefeld KEM, Kok GD, Bockting CLH, Cuijpers P, Hollon SD, van Marwijk HWJ, et al. Effectiveness of psychological interventions in preventing recurrence of depressive disorder: meta-analysis and meta-regression. J Affect Disord 2015 Mar 15;174:400-410. [doi: 10.1016/j.jad.2014.12.016] [Medline: 25553400]
48. So M, Yamaguchi S, Hashimoto S, Sado M, Furukawa TA, McCrone P. Is computerised CBT really helpful for adult depression? A meta-analytic re-evaluation of CCBT for adult depression in terms of clinical implementation and methodological validity. BMC Psychiatry 2013;13(1):1-14 [FREE Full text] [doi: 10.1186/1471-244X-13-113] [Medline: 23587347]
49. Andersson G, Rozental A, Shafran R, Carlbring P. Long-term effects of internet-supported cognitive behaviour therapy. Expert Rev Neurother 2018 Jan;18(1):21-28. [doi: 10.1080/14737175.2018.1400381] [Medline: 29094622]

50. Paganini S, Teigelkötter W, Buntrock C, Baumeister H. Economic evaluations of internet- and mobile-based interventions for the treatment and prevention of depression: a systematic review. J Affect Disord 2018 Jan 01;225:733-755. [doi: 10.1016/j.jad.2017.07.018] [Medline: 28922737]
51. Donker T, Blankers M, Hedman E, Ljótsson B, Petrie K, Christensen H. Economic evaluations of Internet interventions for mental health: a systematic review. Psychol Med 2015 Dec;45(16):3357-3376. [doi: 10.1017/S0033291715001427] [Medline: 26235445]
52. Kolovos S, van Dongen JM, Riper H, Buntrock C, Cuijpers P, Ebert DD, et al. Cost effectiveness of guided Internet-based interventions for depression in comparison with control conditions: an individual-participant data meta-analysis. Depress Anxiety 2018 Jan 12;35(3):209-219. [doi: 10.1002/da.22714] [Medline: 29329486]
53. Erbe D, Eichert H, Riper H, Ebert DD. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. J Med Internet Res 2017 Sep 15;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]
54. Drozd F, Vaskinn L, Bergsund HB, Haga SM, Slinning K, Bjørkli CA. The implementation of internet interventions for depression: a scoping review. J Med Internet Res 2016 Sep 08;18(9):e236 [FREE Full text] [doi: 10.2196/jmir.5670] [Medline: 27608548]
55. Rost T, Stein J, Löbner M, Kersting A, Luck-Sikorski C, Riedel-Heller SG. User acceptance of computerized cognitive behavioral therapy for depression: systematic review. J Med Internet Res 2017 Sep 13;19(9):e309. [doi: 10.2196/jmir.7662] [Medline: 28903893]
56. Manca A, Palmer S. Handling missing data in patient-level cost-effectiveness analysis alongside randomised clinical trials. Appl Health Econ Health Policy 2005;4(2):65-75. [Medline: 16162026]
57. Burton A, Billingham LJ, Bryan S. Cost-effectiveness in clinical trials: using multiple imputation to deal with incomplete cost data. Clin Trials 2007;4(2):154-161. [doi: 10.1177/1740774507076914] [Medline: 17456514]
58. Bouwmans C, De Jong K, Timman R, Zijlstra-Vlasveld M, Van der Feltz-Cornelis C, Tan Swan S, et al. Feasibility, reliability and validity of a questionnaire on healthcare consumption and productivity loss in patients with a psychiatric disorder (TiC-P). BMC Health Serv Res 2013;13:217 [FREE Full text] [doi: 10.1186/1472-6963-13-217] [Medline: 23768141]
59. Kaltenthaler E, Brazier J, De Nigris E, Tumur I, Ferriter M, Beverley C, et al. Computerised cognitive behaviour therapy for depression and anxiety update: a systematic review and economic evaluation. Health Technol Assess 2006 Sep;10(33):1-168 [FREE Full text] [Medline: 16959169]
60. Arnberg FK, Linton SJ, Hultcrantz M, Heintz E, Jonsson U. Internet-delivered psychological treatments for mood and anxiety disorders: a systematic review of their efficacy, safety, and cost-effectiveness. PLoS One 2014;9(5):e98118 [FREE Full text] [doi: 10.1371/journal.pone.0098118] [Medline: 24844847]

Abbreviations

CEAC: cost-effectiveness acceptability curve
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition
EQ-5D-3L: three level version of the EuroQol Five Dimensional Questionnaire.
HRSD: Hamilton Rating Scale for Depression
ICER: incremental cost-effectiveness ratio
M-CT: mobile cognitive therapy
MDD: major depressive disorder
MDE: major depressive episode
PCT: preventive cognitive therapy
QALY: quality-adjusted life year
RCT: randomized controlled trial
SCID: Structured Clinical Interview for DSM Disorders
TAU: treatment as usual
TiC-P: Trimbos and Institute for Medical Technology Assessment Questionnaire on Costs Associated with Psychiatric Illness

Edited by G Eysenbach; submitted 23.03.18; peer-reviewed by Y Lee; comments to author 05.09.18; accepted 24.09.18; published 26.11.18.

Please cite as:
Klein NS, Bockting CLH, Wijnen B, Kok GD, van Valen E, Riper H, Cuijpers P, Dekker J, van der Heiden C, Burger H, Smit F
Economic Evaluation of an Internet-Based Preventive Cognitive Therapy With Minimal Therapist Support for Recurrent Depression: Randomized Controlled Trial
J Med Internet Res 2018;20(11):e10437
URL: http://www.jmir.org/2018/11/e10437/
doi:10.2196/10437
PMID:30478021

Original Paper

Mining Open Payments Data: Analysis of Industry Payments to Thoracic Surgeons From 2014-2016

Xu Na[1], MSc; Haihong Guo[1], MSc; Yu Zhang[1], BMS; Liu Shen[1], MSc; Sizhu Wu[1], PhD; Jiao Li[1], PhD

Institute of Medical Information / Medical Library, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

Corresponding Author:
Jiao Li, PhD
Institute of Medical Information / Medical Library
Chinese Academy of Medical Sciences & Peking Union Medical College
3rd Yabao Road, Chaoyang District
Beijing,
China
Phone: 86 01052328740
Email: li.jiao@imicams.ac.cn

Abstract

Background: The financial relationship between physicians and industries has become a hotly debated issue globally. The Physician Payments Sunshine Act of the US Affordable Care Act (2010) promoted transparency of the transactions between industries and physicians by making remuneration data publicly accessible in the Open Payments Program database. Meanwhile, according to the World Health Organization, the majority of all noncommunicable disease deaths were caused by cardiovascular disease.

Objective: This study aimed to investigate the distribution of non-research and non-ownership payments made to thoracic surgeons, to explore the regularity of financial relationships between industries and thoracic surgeons.

Methods: Annual statistical data were obtained from the Open Payments Program general payment dataset from 2014-2016. We characterized the distribution of annual payments with single payment transactions greater than US $10,000, quantified the major expense categories (eg, Compensation, Consulting Fees, Travel and Lodging), and identified the 30 highest-paying industries. Moreover, we drew out the financial relations between industries to thoracic surgeons using chord diagram visualization.

Results: The three highest categories with single payments greater than US $10,000 were Royalty or License, Compensation, and Consulting Fees. Payments related to Royalty or License transferred from only 5.38% of industries to 0.75% of surgeons with the highest median (US $13,753, $11,992, and $10,614 respectively) in 3-year period. In contrast, payments related to Food and Beverage transferred from 93.50% of industries to 98.48% of surgeons with the lowest median (US $28, $27, and $27). The top 30 highest-paying industries made up approximately 90% of the total payments (US $21,036,972, $23,304,996, and $28,116,336). Furthermore, just under 9% of surgeons received approximately 80% of the total payments in each of the 3 years. Specifically, the 100 highest cumulative payments, accounting for 52.69% of the total, transferred from 27 (6.05%) pharmaceutical industries to 86 (1.89%) thoracic surgeons from 2014-2016; 7 surgeons received payments greater than US $1,000,000; 12 surgeons received payments greater than US $400,000. The majority (90%) of these surgeons received tremendous value from only one industry.

Conclusions: There exists a great discrepancy in the distribution of payments by categories. Royalty or License Fees, Compensation, and Consulting Fees are the primary transferring channels of single large payments. The massive transfer from industries to surgeons has a strong “apical dominance” and excludability. Further research should focus on discovering the fundamental driving factors for the strong concentration of certain medical devices and how these payments will affect the industry itself.

(J Med Internet Res 2018;20(11):e11655) doi:10.2196/11655

KEYWORDS

open payments data; pharmaceutical industry; thoracic surgeons; transfer of value

Introduction

Background

The term transfer of value means a direct or indirect transfer of value, whether in cash, in kind or otherwise, in connection with the development or sale of medicine [1]. Potential conflicts of interest arising from the transfer of value between the pharmaceutical industries and physicians could significantly affect clinical care, research findings, and physician decision-making. The risks and benefits of physician-industry financial relationships have long been hotly debated [2-4]. A recent investigation found a pattern of after-the-fact compensation by industry to those advising the US government on drug approvals [5].

Open health care–related data have been widely collected and analyzed [6,7]. The Physician Payments Sunshine Act of the US Patient Protection and Affordable Care Act (2010) was enacted in response to growing public interest and scrutiny regarding the financial relationship between physicians and the pharmaceutical and product industries [8,9]. The act mandates that drug and device manufacturers report individual payments of greater than US $10, or US $100 in aggregate annually, received by physicians and teaching hospitals. Physicians include doctors of Medicine, Osteopathy, Dentistry, Dental Surgery, Podiatry, Optometry, and Chiropractic Medicine who are legally authorized to practice. Industry reporting of financial remuneration for travel, gifts, and services rendered is now mandated by the US Centers for Medicare and Medicaid Services (CMS), and the resulting data are made publicly available through the Open Payments Program (OPP) database [10]. General payment records in the OPP provide the total value of general payments or other transfers of value to a particular recipient for a particular date. Each record includes identifying information for the physicians and teaching hospitals in the United States, identifying information for the applicable manufacturer, and applicable group purchasing organizations who made the payment, the total amount of payment, date of payment, nature of payment, associated drug or biological, etc [11].

According to the World Health Organization's 2018 annual report [12], the majority of all noncommunicable disease deaths were caused by cardiovascular disease, accounting for 44% of 41 million deaths. The American Heart Association 2018 update on heart disease and stroke statistics has disclosed that approximately 92.1 million American adults are living with cardiovascular disease or the after-effects of stroke. Direct and indirect costs of cardiovascular diseases and stroke are estimated to total more than US $329.7 billion [13].

Study Aims

Research has been conducted to quantify industry payments to different specialties by analyzing the OPP database [14-17]. However, compared with previous studies, we focused on analyzing the payment characteristics during a 3-year period and explored the regularity of the financial relationship between industries and thoracic surgeons.

Methods

Data Sources

We accessed the OPP general payment dataset, which is publicly accessible, from 2014-2016. We excluded the 2013 dataset due to incompleteness and inconsistencies in the OPP records, which may have led to deviation in reflecting the real situation. Our study focused on the non-research and non-ownership payments received by doctors of Thoracic Surgery. Therefore, we excluded payments for current or prospective ownership or investment interest and limited the physician specialty studied to Thoracic Surgery (Cardiothoracic Vascular Surgery). Payments valued at US $0 were also excluded. There are a few industries with the same name and different ID numbers in the OPP database; however, almost all annual cumulative payment amounts for the different IDs were either less than US $1000, or the payment amounts were no more than US $10, and the majority were transferred to the same surgeon or were associated with the same drug or biological agent. Therefore, we counted the number of industries with distinct names. Our final cohort included 197,592 payments from 446 industries to 4552 surgeons in 3 years.

Payment Categories

The CMS has defined 16 categories for the nature of payment. The six major expense categories involved in this study included the following: Consulting Fees; Compensation for services other than consulting, including serving as faculty or as a speaker at a venue other than a continuing education program (abbreviated as "Compensation"); Travel and Lodging; Food and Beverage; Royalty or License; and Education (see Multimedia Appendix 1). These six categories were selected because they account for greater than 95% of total payments in the database.

Statistical Analysis

All analyses were performed with R 3.4.1. Due to the significant skew of the data, the results are presented as median payments with interquartile ranges (IQR), and a log transformation was performed on payments prior to graphing boxplots to enable visual clarity. Descriptive statistics were calculated to analyze the distribution of annual payments with a single payment greater than US $10,000, payments by major expense categories, and payments by the 30 highest-paying industries. A chord diagram was used to show the distribution of the 100 highest cumulative payments from pharmaceutical industries to thoracic surgeons in the 3-year period.

Results

Distribution of Annual Payments

There were 61,963 payments totaling US $21,036,972 transferred from 299 industries to 3667 thoracic surgeons in 2014; 64,558 payments totaling US $23,304,996 transferred from 282 industries to 3613 surgeons in 2015; and 71,071 payments totaling US $28,116,336 transferred from 283 industries to 3717 surgeons in 2016. The registered number for Thoracic Surgery (Cardiothoracic Vascular Surgery) in the National Provider Identifier (NPI) Database was 5614, excluding

medical groups [18]. The NPI number is a unique 10-digit number issued by the CMS to health care providers in the United States. Based on the NPI, 81.08% of the registered surgeons received a transfer of value in the 3 years; furthermore, two-thirds of the surgeons received a payment each year. The distribution of payments to thoracic surgeons significantly skewed toward smaller payments over the 3 years. The median dollar amount (IQR) of the annual payments were 44 (17-125), 42 (17-123), and 44 (17-126). Furthermore, the proportion of maximum payment increased dramatically from 5.58% (US $1,173,913), to 10.95% (US $2,551,007), to 17.78% (US $5,000,000) of the total. The result illustrated that the skewed distribution kept expanding overall from 2014-2016.

Distribution by Six Major Expense Categories

Six major expense categories accounted for 95.34%, 96.91%, and 95.53% of all payments in 2014, 2015, and 2016, respectively. The most common category associated with payments to thoracic surgeons was Compensation, accounting for 22.84% (US $4,804,701) and 32.18% (US $7,499,516) of all payments in 2014 and 2015, respectively. However, the most common payment changed to Royalty or License (US $7,846,886), which received 27.91% of the total in 2016. The second largest share of payments was Consulting Fees, accounting for 21.95%, 21.23%, and 20.91%, followed by Travel and Lodging 19.84%, 17.81%, and 17.54% (see Figure 1 and Multimedia Appendix 2). In addition, the proportion of Consulting Fees, Travel and Lodging, and Food and Beverage decreased gradually, whereas payments in the Education category increased by comparison.

The payment category with the highest median was Royalty or License. The median dollar payment (IQR) for this category and the percentage of surgeons receiving them were US $13,753 (4122-38,260) to 0.68%, US $11,992 (1132-43,054) to 0.55%, and US $10,614 (766-50,000) to 0.70%. In contrast, the payment category with the lowest median was Food and Beverage, in which the median payments in dollars (IQR) were US $28 (15-80) to 97.14%, US $27 (15-79) to 97.92%, and US $27 (14-77) to 98.33% (see Figures 2 and 3). In addition, the median of the Compensation and Education categories increased annually, while the median of the Royalty or License, and Consulting Fees categories decreased continuously.

Single payments greater than US $10,000 accounted for 35.19%, 39.44%, and 41.16% of the total during the 3 years, which transferred from 14.05% of the industries to 2.35% of the surgeons; from 13.83% of the industries to 2.13% of the surgeons; and from 11.66% to 2.02% of the surgeons, respectively (Table 1). The proportion of single payments greater than US $10,000 increased substantially; in contrast, the percentage of surgeons and industries decreased gradually. Royalty or License, Compensation, and Consulting Fees took up the three highest proportions in the 3 years. Moreover, 97.40%, 96.92%, and 98.95% of payments in the Royalty or License category were greater than US $10,000 from 2014-2016; 45.96%, 63.89%, and 30.29% of payments in the Compensation category were greater than US $10,000; and 36.11%, 35.21%, and 35.91% of payments in the Consulting Fees category were greater than US $10,000. Additionally, 8.62%, 8.36%, and 8.77% of surgeons with annual cumulative payment amounts greater than US $10,000 received 78.42% (US $16,496,272), 80.63% (US $18,790,319), and 81.82% (US $23,004,994), respectively.

Figure 1. Distribution of pharmaceutical industries payments to thoracic surgeons by major categories, 2014-2016.

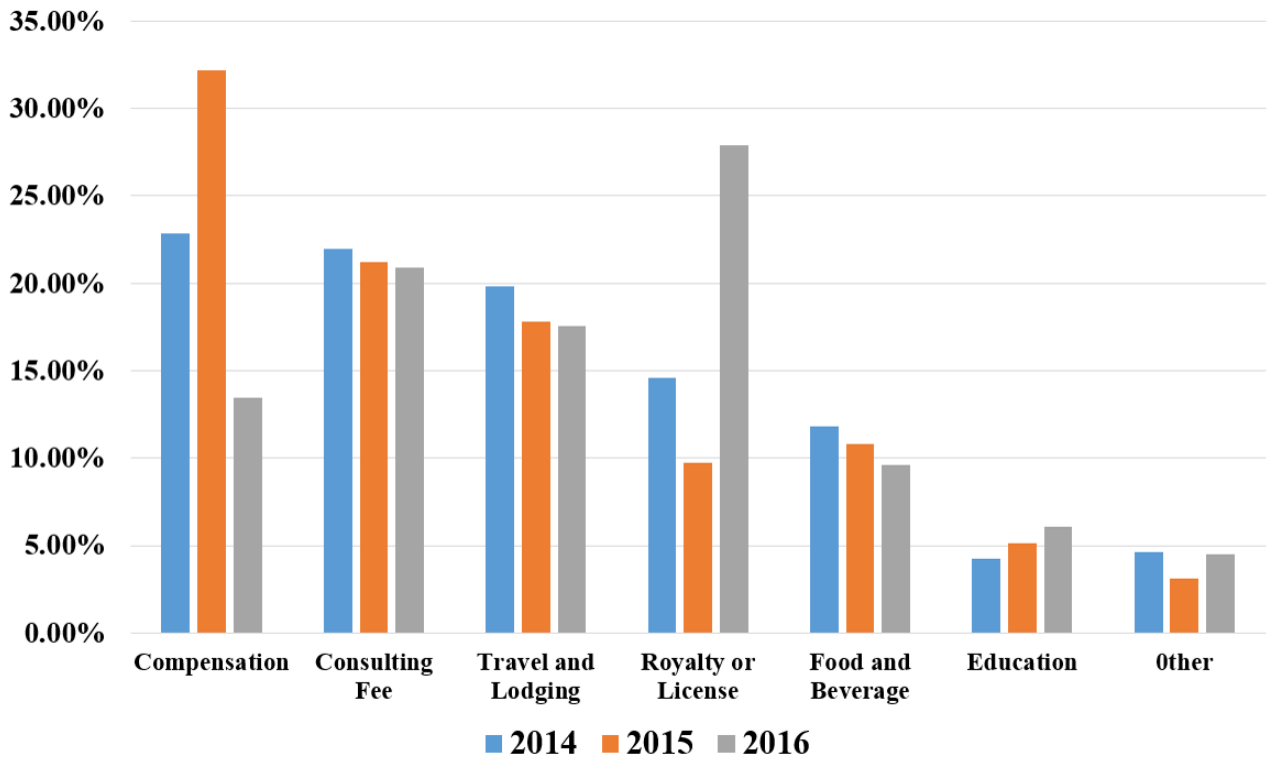


Figure 2. Boxplot of pharmaceutical industries payments to thoracic surgeons by six major categories, 2014-2016 (log transformation performed on payments prior to graphing boxplots for visual clarity).

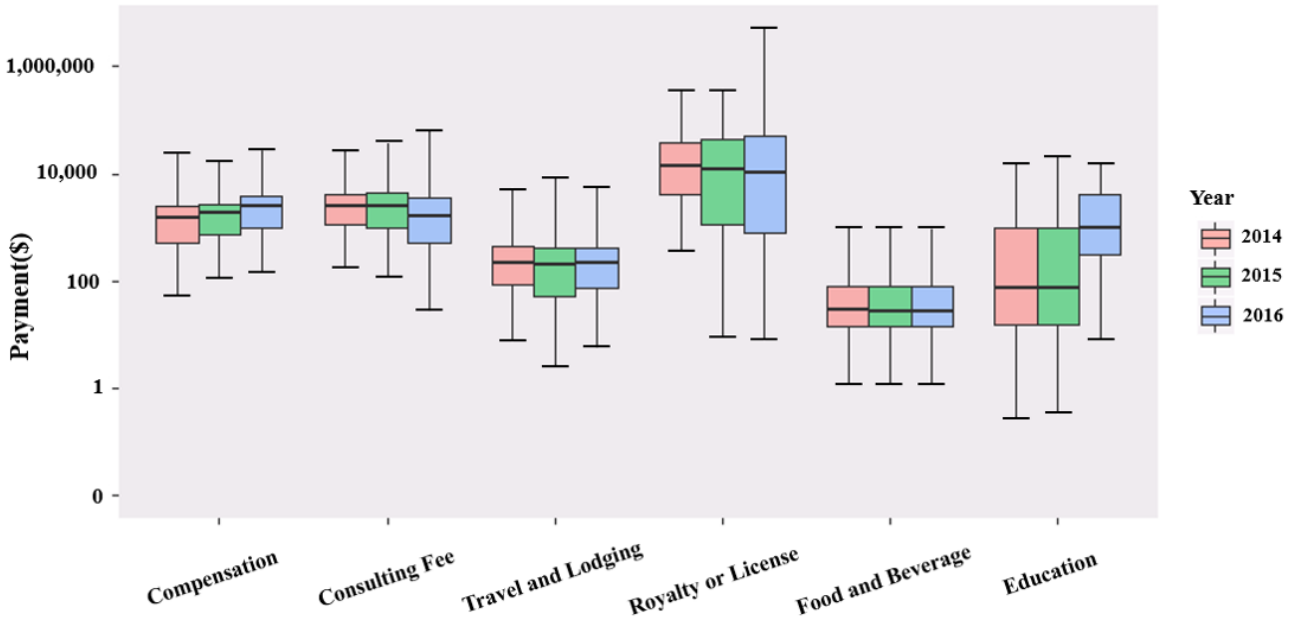
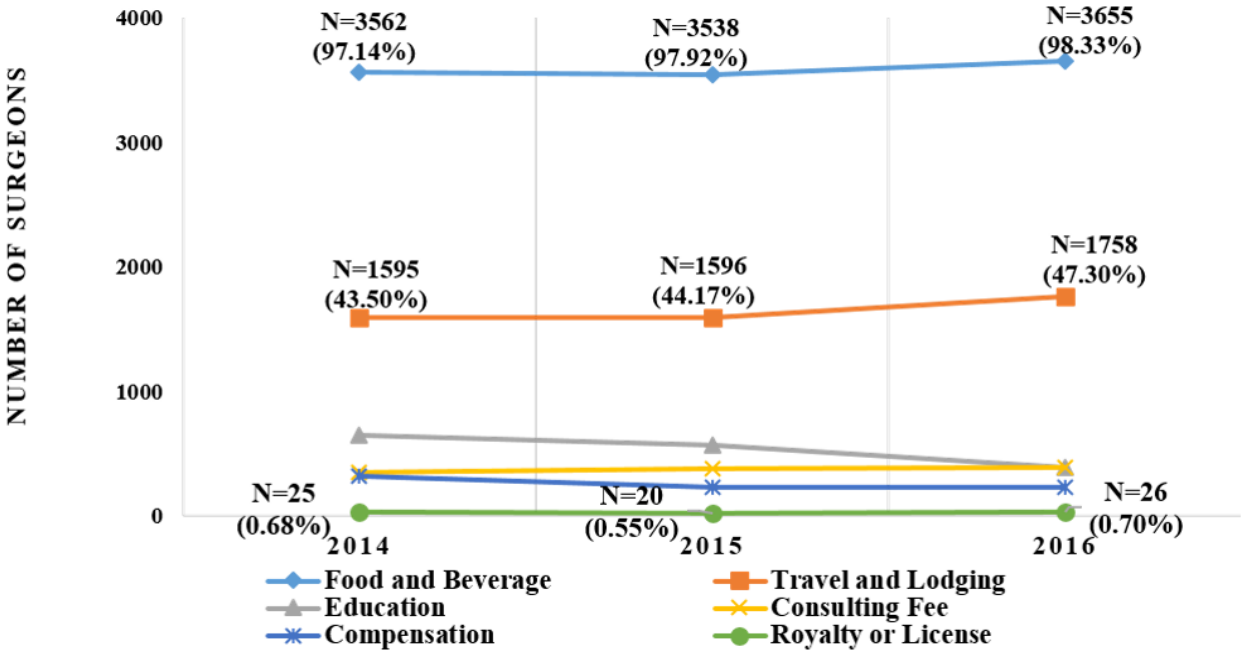


Figure 3. Distribution of pharmaceutical industries payments to thoracic surgeons by number of surgeons, 2014-2016.



Distribution by 30 Highest-Paying Industries and 100 Highest Cumulative Payments

The 30 highest-paying industries made up the vast majority (89.08%) of the total transferred to 3201 (87.29%) surgeons through 50,656 (81.75%) payments, 91.20% of the total transferred to 3127 surgeons through 52,010 payments, and 94.31% of the total transferred to 3372 surgeons through 60,975 payments during the 3 years (see Figure 4).

The 100 highest cumulative payments accounted for 52.69% (US $38,179,324) of the total transferred from 6.05% of the pharmaceutical industries to 1.89% of the thoracic surgeons in 3 years. The five highest-paying industries were Medtronic Vascular, Inc. (41.00% of the 100 highest cumulative payments), AtriCure, Inc. (9.85%), Baxter (6.36%), Intuitive Surgical, Inc. (6.14%), and Abiomed (5.77%). Moreover, 7 surgeons received payments greater than US $1,000,000; 12 surgeons received payments greater than US $400,000; and 68 surgeons received payments greater than US $80,000. Only 8 surgeons received a transfer of value from more than one industry (see Figure 5).

Table 1. Distribution of payments with single payments greater than US $10,000 by major categories, 2014-2016[a].

Year and Category	Royalty or License	Compensation	Consulting Fee	Travel and Lodging	Education	Other	Total
2014							
Surgeons, n	16	17	37	21	2	N/A[b]	86 (2.35)[c]
Industries, n	10	9	25	11	2	N/A	42 (14.05)[c]
Payments, US $ (%)	2,992,270 (97.40)	2,208,319 (45.96)	1,667,119 (36.11)	162,183 (3.89)	27,000 (3.00)	345,129 (N/A)	7,402,020 (35.19)
Payments, n	51	20	62	14	2	26	175
2015							
Surgeons, n	13	19	33	12	3	N/A	77 (2.13)[c]
Industries, n	9	8	18	8	2	N/A	39 (13.83)[c]
Payments, US $ (%)	2,197,409 (96.92)	4,791,506 (63.89)	1,741,916 (35.21)	214,015 (5.16)	60,500 (5.04)	185,200 (N/A)	9,190,546 (39.44)
Payments, n	41	36	72	17	4	15	190
2016							
Surgeons, n	16	21	32	17	1	N/A	75 (2.02)[c]
Industries, n	10	3	20	6	1	N/A	33 (11.66)[c]
Payments, US $ (%)	7,764,333 (98.95)	1,146,097 (30.29)	2,111,663 (35.91)	351,156 (7.12)	15,000 (0.87)	184,093 (N/A)	11,572,342 (41.16)
Payments, n	41	52	53	26	1	14	187

[a]The percentage of payments for major categories all mean the proportion of payments with a single payment greater than $10,000 of the total payments for the category (see Multimedia Appendix 2).

[b]N/A: not applicable.

[c]Value is reported as n (%).

Figure 4. Distribution of 30 highest-paying pharmaceutical industries, 2014-2016.

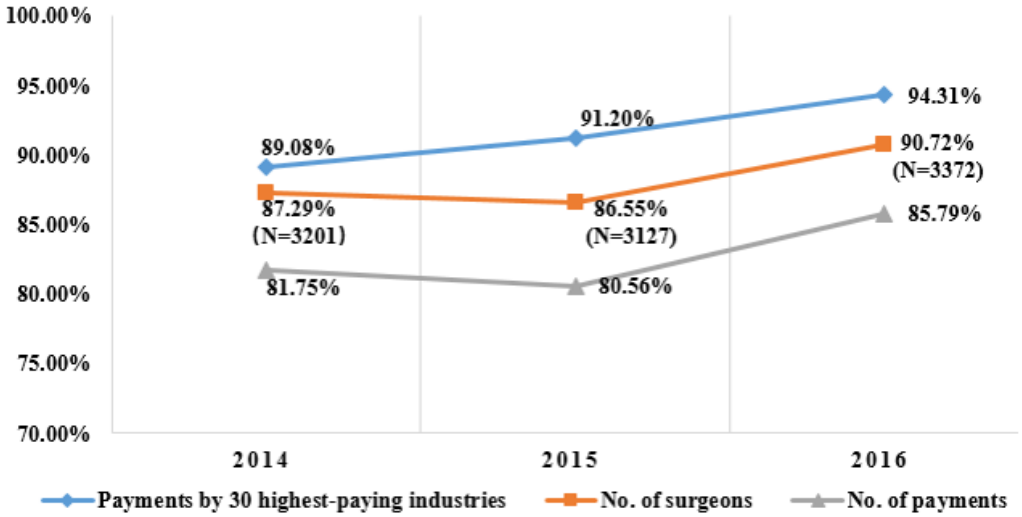
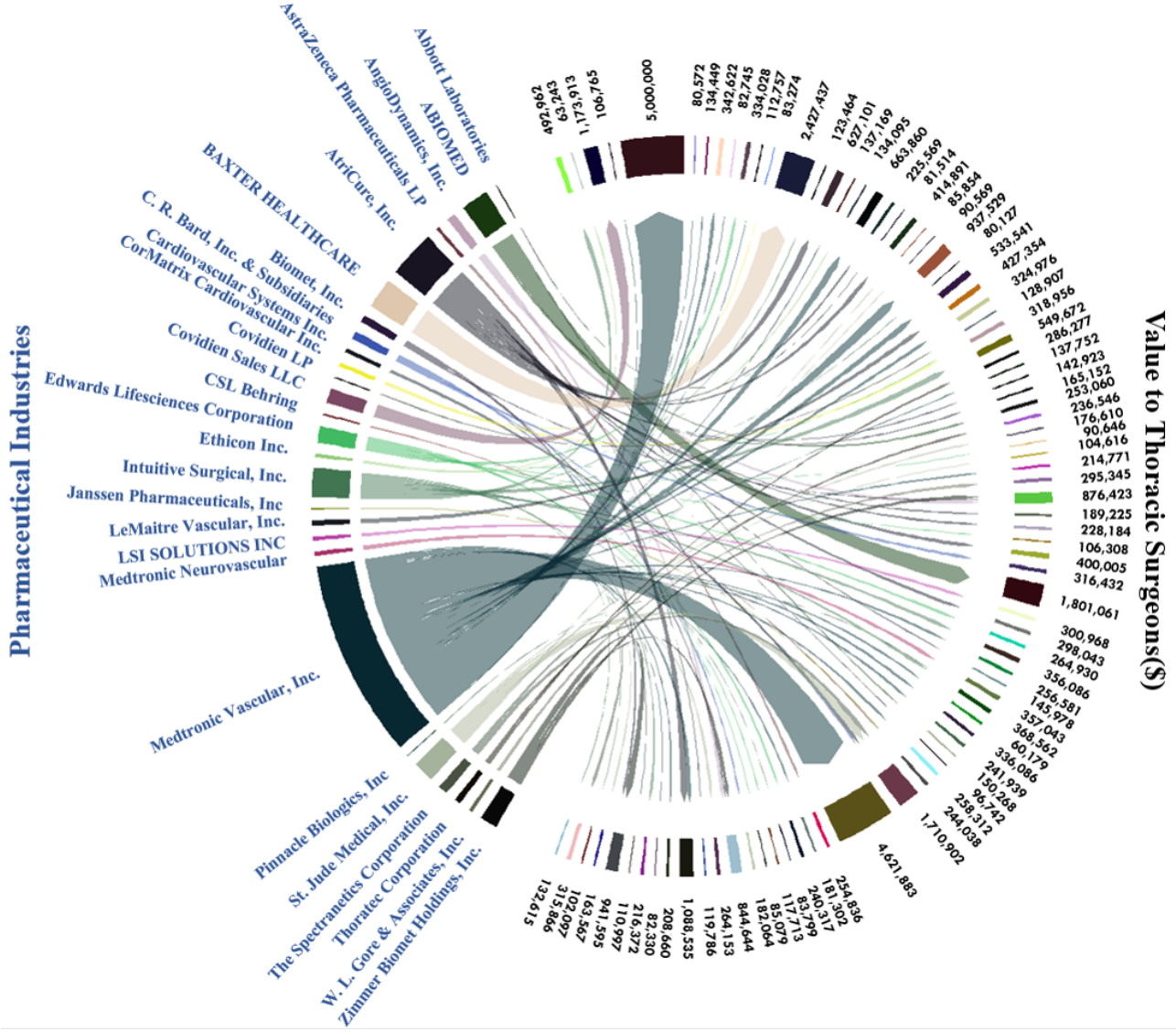


Figure 5. Financial relationships of the 100 highest cumulative payments from pharmaceutical industries (N=27) to thoracic surgeons (N=86) in 3 years.



Discussion

Principal Findings

The number of surgeons and industries was stable, while the total annual payments and the number of payments grew steadily from 2014-2016. Our study showed that just under 9% of surgeons with annual cumulative payment amounts greater than US $10,000 received approximately four-fifths of the annual payments in 3 consecutive years, indicating that the distribution of payments to thoracic surgeons appears to have a remarkable skewness. The result was quite consistent with those from related studies [19,20]. We also found that the industries are inclined to transfer tremendous value to surgeons in a single payment (great than US $10,000). This trend has gradually expanded, with the rapid growth of the proportion of payments of single payments greater than US $10,000 and three-quarters of payments less than US $130 in the 3 years. There were significant differences in the categories of the industry payments (Table 2). The six most common payments accounted for more than 95% of all payments during the 3 years. This finding indicated that the unevenly distributed payments are highly concentrated in certain categories. The extremely unbalanced payment distribution by categories in Thoracic Surgery is quite similar to those of other specialties, although there exists variation in different specialties [21-24].

By comparison, we realized that the three highest-paying categories with single payments greater than US $10,000 were Royalty or License, Compensation, and Consulting Fees. This finding revealed that these categories are the most primary transferring channels of major values from industries to surgeons. Moreover, the distribution of these payments showed a remarkable skewness. In our study, Royalty or License category was transferred from the lowest percentage (5.38%) of industries to 0.75% of the surgeons, of which single payments greater than US $10,000 accounted for 98.24% in the 3-year period and were based on physicians' intellectual property. Similarly, approximately half of the Compensation category and one-third of the Consulting Fees category were transferred to less than 1% of the surgeons, with single payments greater than US $10,000, requiring expertise on a medical product or treatment as well as surgeon participation as faculty or a speaker for noncontinuing education.

Table 2. Payment characteristics by major categories in the 3-year period[a].

Payment category	Payments, %	Industry coverage, %	Surgeon coverage, %	Single payment >US $10,000, %	Professional skills requirements[b]
Royalty or license	18.20	5.38	0.75	98.24	Physician's intellectual property
Consulting fee	21.31	30.04	14.94	35.75	Advice on medical product or treatment
Compensation	22.20	13.90	10.96	50.63	Speaking, training, and noncontinuing education
Travel and lodging	18.30	39.01	57.62	5.49	None
Food and beverage	10.64	93.50	98.48	None	None
Education	5.26	17.94	27.39	2.69	Imparting or acquiring of particular knowledge or skills

[a]The percentage of payments, industry coverage, surgeon coverage, and single payment >US $10,000 all refer to the cumulative percentage in the 3-year period.

[b]Professional skills requirements are based on the Centers for Medicare and Medicaid Services' definitions of six major expense categories.

One plausible reason for these results is that the monopoly of the patent market leads to a tiny proportion of industries that occupy vast market shares; hence, more emphasis is being placed on technological innovation and development to maintain superiority. To do this quickly, industries may seek out leading authorities on thoracic surgeons as cooperative partners, who, due to their deeper scientific and medical experience, are more likely to make patents of inventions in Thoracic Surgery. Industries could establish a long-term partnership, obtain the technology license, and sell their highly profitable patented products on the market by paying large-value Royalty or License, Compensation, and Consulting Fees to these surgeons. In another aspect, related research warned that enormous payments from a few major industries to certain surgeons who advised the US Food and Drug Administration on the approval for the industries' new drugs fit a pattern of what might be called pay-later conflicts of interest [5].

Characteristics of Food and Beverage payments are in sharp contrast to those of Royalty or License. In fact, 10.64% of payments in this category, with 75% of payments less than US $80, transferred to more than 98% of the surgeons from over 93% of the industries in 3 consecutive years, which was consistent with the findings from multiple related studies [25-27]. These results imply that the transfer of value from industries to surgeons in this category was much more widespread in comparison to others. Additionally, this payment has few connections with surgeons' clinical or professional skills. The outcome might be explained by the fact that patients may have a more negative view on payments of food in related studies [28,29]. Similarly, nearly one-fifth of the payments from 39.01% of the industries transferred to 57.62% of the surgeons through the Travel and Lodging category, with 75% of payments less than US $450 in 3 consecutive years; this category is also unrelated to professional skills. It is highly noteworthy that approximately 30% of the total transfer of value was widely transferred to surgeons with no requirement of professional knowledge or skills. Related research found that the receipt of industry-sponsored meals was associated with an increased rate of prescribing the brand-name medication that was being promoted [30]. Pharmaceutical industries provide hundreds of millions of dollars to physicians for food and beverages with the expectations of good returns [31]. Therefore, we are deeply convinced that these sizable payments have quite a widespread impact on industries' product promotion to surgeons.

In another aspect, it should be noted that total payments in the Education category grew remarkably. Furthermore, the number of industries and surgeons reduced significantly over the 3 years, indicating that payment tendency in the Education category was inclined to a highly centralized model. The major reason for this result is that there were substantial imbalances of payments among industries, with the highest-paying industry accounting for 67.94% (US $610,789) in 2014, 77.96% (US $936,205) in 2015, and 94.65% (US $1,622,728) in 2016. Our research found that the largely uneven distribution of industry payments also exists in other expense categories. Therefore, we can confirm that the payment growth is driven mainly by the few highest-paying industries.

One of the most interesting findings in our study is that the proportion of payments from the 30 highest-paying industries was approximately nine out of ten over 3 years, consistently accounting for only one-tenth of total industries. It is noteworthy that the payment amount is highly correlated with the size of the industries. The majority of 30 highest-paying industries were in the list of the world's top 20 leading medical device and diagnostic companies in 2016 based on Evaluate MedTech's annual report [32]. The vast majority of these payments only transferred to less than 9% of surgeons, accordingly. Moreover, 7 surgeons (0.15%) received more than US $1,000,000 in 3-year period. A large share of all payments was skewed toward a small fraction of top earners; this pattern has also been observed in other specialties [23,33,34]. Based on the above results, we tentatively suggest that there exists the famous "apical dominance" in the transfer of value. This effect was described as the control exerted by the terminal bud over the outgrowth of lateral buds in plant physiology, which is also a widespread phenomenon in economics. Due to the extremely uneven development, tremendous transfers are highly concentrated on the dominant industries and leading surgeons.

Additionally, the 100 highest cumulative payments, accounting for more than half of the total, transferred from a tiny percentage of industries to a minority of surgeons. Over 90% of these surgeons received hundreds of thousands of dollars from only one specific industry during the 3 years. These results strongly indicate that leading surgeons have a long-term exclusive partnership with dominant industries, particularly for large payment amounts. On the one hand, Royalty or License could be transferring to an individual industry due to its monopoly and exclusivity. On the other hand, key partners might understand the industries' commercial secrets through consultations and other forms of cooperation. Therefore, industries do not wish for them to cooperate with competitors in the current and fierce business competition.

Limitations

There are several limitations to this study. First, it should be noted that this study has examined only non-research and non-ownership payments, which may lead to deviations in reflecting the real situation. Second, the database relies on accurate reports from manufacturers. There may be inaccurate attribution of payments into categories, to individuals, or to affiliated institutions by reporting companies in the database. The CMS redacted more than 40% of all reported payment records in the first year. Finally, our research mainly focuses on the payment characteristics of six major expense categories, which may not completely reflect all data features.

Conclusions

There exists a great discrepancy in the distribution of payments by categories. Royalty or License, Compensation, and Consulting Fees are the primary transferring channels of single large payments. The massive transfer from industries to surgeons has a strong "apical dominance" and excludability. Furthermore, our study provides evidence that payments by the 30 highest-payment manufacturers were specifically targeting certain medical devices during the 3 years. Further research should focus on discovering the fundamental driving factors for the strong concentration of certain medical devices and how these payments will affect the industry itself.

Acknowledgments

This work was supported by the National Population and Health Scientific Data Sharing Program of China, the Knowledge Centre for Engineering Sciences and Technology (Medical Centre), the Key Laboratory of Medical Information Intelligent Technology of Chinese Academy of Medical Sciences, and the Key Laboratory of Knowledge Technology for Medical Integrative Publishing.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The CMS definitions of six major expense categories.

[PDF File (Adobe PDF File), 12KB - jmir_v20i11e11655_app1.pdf]

Multimedia Appendix 2

Annual payments distribution of payments from industries to surgeons, 2014-2016.

[PDF File (Adobe PDF File), 251KB - jmir_v20i11e11655_app2.pdf]

References

1. Association of the British Pharmaceutical Industry (ABPI). Scope of the Code and Definition of Certain Terms in the ABPI code of practice for the pharmaceutical industry. URL: http://www.abpi.org.uk/ethics/ethical-responsibility/disclosure-uk/about-disclosure-uk/definitions [accessed 2018-10-07] [WebCite Cache ID 72xvjYwF6]
2. Rosenbaum L. Conflicts of interest: part 1: Reconnecting the dots--reinterpreting industry-physician relations. N Engl J Med 2015 May 07;372(19):1860-1864. [doi: 10.1056/NEJMms1502493] [Medline: 25946288]
3. Steinbrook R, Kassirer JP, Angell M. Justifying conflicts of interest in medical journals: a very bad idea. BMJ 2015 Jun 02;350:h2942. [doi: 10.1136/bmj.h2942] [Medline: 26036926]
4. Pelliccia F, Coats AJS, Pani L, Gaudio C, Rosano G. Transparency in medical research: Time for a paradigm shift. Int J Cardiol 2015;186:259-260. [doi: 10.1016/j.ijcard.2015.03.219] [Medline: 25828129]
5. Piller C. Hidden conflicts? Science 2018 Jul 06;361(6397):16-20. [doi: 10.1126/science.361.6397.16] [Medline: 29976808]
6. Stadler JG, Donlon K, Siewert JD, Franken T, Lewis NE. Improving the Efficiency and Ease of Healthcare Analysis Through Use of Data Visualization Dashboards. Big Data 2016 Dec;4(2):129-135. [doi: 10.1089/big.2015.0059] [Medline: 27441717]
7. Vie LL, Scheier LM, Lester PB, Ho TE, Labarthe DR, Seligman MEP. The U.S. Army Person-Event Data Environment: A Military-Civilian Big Data Enterprise. Big Data 2015 Dec;3(2):67-79. [doi: 10.1089/big.2014.0055] [Medline: 27447431]
8. Arkinson J, Holbrook A, Wiercioch W. Public perceptions of physician - pharmaceutical industry interactions: a systematic review. Healthc Policy 2010 May;5(4):69-89 [FREE Full text] [Medline: 21532771]

9. Agrawal S, Brown D. The Physician Payments Sunshine Act--Two Years of the Open Payments Program. N Engl J Med 2016 Mar 10;374(10):906-909. [doi: 10.1056/NEJMp1509103] [Medline: 26962725]
10. Centers for Medicare and Medicaid Services. Fact Sheet For Physicians. Open payments (Physician Payments Sunshine Act) URL: https://www.regeneron.com/sites/default/files/Sunshine-Physician-fact-sheet.pdf [accessed 2018-10-07] [WebCite Cache ID 72xzahzU4]
11. Centers for Medicare and Medicaid Services. Open Payments Public Use Files: Methodology Overview & Data Dictionary. URL: https://www.cms.gov/OpenPayments/Downloads/OpenPaymentsDataDictionary.pdf [accessed 2018-10-07] [WebCite Cache ID 72y0thp2R]
12. World Health Organization. World Health Statistics 2018. URL: http://apps.who.int/iris/bitstream/handle/10665/272596/9789241565585-eng.pdf [accessed 2018-11-17] [WebCite Cache ID 73ybt3Ikj]
13. Heart Disease and Stroke Statistics-2018 Update: A Report From the American Heart Association. 2018 Mar 20. URL: https://www.ahajournals.org/doi/pdf/10.1161/CIR.0000000000000558 [accessed 2018-10-07] [WebCite Cache ID 72y27jZuO]
14. Parreco J, Donath E, Kozol R, Faber C. Comparing industry compensation of cardiothoracic surgeons and interventional cardiologists. J Surg Res 2017 Dec;208:51-59. [doi: 10.1016/j.jss.2016.09.022] [Medline: 27993217]
15. Lichter PR. Implications of the Sunshine Act--revelations, loopholes, and impact. Ophthalmology 2015 Apr;122(4):653-655. [doi: 10.1016/j.ophtha.2014.12.029] [Medline: 25797085]
16. Frishman WH. Reflections on the Physicians Payment Sunshine Act. Am J Med 2016 Jan;129(1):3-4. [doi: 10.1016/j.amjmed.2015.08.001] [Medline: 26291906]
17. Santhakumar S, Adashi EY. The Physician Payment Sunshine Act: testing the value of transparency. JAMA 2015 Jan 06;313(1):23-24. [doi: 10.1001/jama.2014.15472] [Medline: 25473969]
18. National Provider Identifier Database. NPI Lookup Thoracic Surgery (Cardiothoracic Vascular Surgery) URL: https://npidb.org/doctors/allopathic_osteopathic_physicians/thoracic-surgery-cardiothoracic-vascular-surgery_208g00000x/ [accessed 2018-10-07] [WebCite Cache ID 72y2GtoYS]
19. Bandari J, Turner RM, Jacobs BL, Davies BJ. Urology Payments from Industry in the Sunshine Act. Urol Pract 2016 Sep;3(5):332-337 [FREE Full text] [doi: 10.1016/j.urpr.2015.12.002] [Medline: 27722187]
20. Karas DJ, Bandari J, Browning DN, Jacobs BL, Davies BJ. Payments to Pediatricians in the Sunshine Act. Clin Pediatr (Phila) 2017 Jul;56(8):723-728. [doi: 10.1177/0009922816670981] [Medline: 27663966]
21. Feng H, Wu P, Leger M. Exploring the Industry-Dermatologist Financial Relationship: Insight From the Open Payment Data. JAMA Dermatol 2016 Dec 01;152(12):1307-1313. [doi: 10.1001/jamadermatol.2016.3037] [Medline: 27706478]
22. Lopez J, Ahmed R, Bae S, Hicks CW, El Dafrawy M, Osgood GM, et al. A New Culture of Transparency: Industry Payments to Orthopedic Surgeons. Orthopedics 2016 Nov 01;39(6):e1058-e1062. [doi: 10.3928/01477447-20160719-04] [Medline: 27459138]
23. Rathi VK, Samuel AM, Mehra S. Industry ties in otolaryngology: initial insights from the physician payment sunshine act. Otolaryngol Head Neck Surg 2015 Jun;152(6):993-999. [doi: 10.1177/0194599815573718] [Medline: 25779469]
24. Fleischman W, Ross JS, Melnick ER, Newman DH, Venkatesh AK. Financial Ties Between Emergency Physicians and Industry: Insights From Open Payments Data. Ann Emerg Med 2016 Dec;68(2):153-158.e4. [doi: 10.1016/j.annemergmed.2016.01.014] [Medline: 26973175]
25. Chang JS. The Physician Payments Sunshine Act: data evaluation regarding payments to ophthalmologists. Ophthalmology 2015 Apr;122(4):656-661 [FREE Full text] [doi: 10.1016/j.ophtha.2014.11.003] [Medline: 25578254]
26. Iyer S, Derman P, Sandhu HS. Orthopaedics and the Physician Payments Sunshine Act: An Examination of Payments to U.S. Orthopaedic Surgeons in the Open Payments Database. J Bone Joint Surg Am 2016 Mar 02;98(5):e18. [doi: 10.2106/JBJS.O.00343] [Medline: 26935466]
27. Tierney NM, Saenz C, McHale M, Ward K, Plaxe S. Industry Payments to Obstetrician-Gynecologists: An Analysis of 2014 Open Payments Data. Obstet Gynecol 2016 Feb;127(2):376-382. [doi: 10.1097/AOG.0000000000001270] [Medline: 26942368]
28. Perry JE, Cox D, Cox AD. Trust and transparency: patient perceptions of physicians' financial relationships with pharmaceutical companies. J Law Med Ethics 2014;42(4):475-491. [doi: 10.1111/jlme.12169] [Medline: 25565614]
29. Licurse A, Barber E, Joffe S, Gross C. The impact of disclosing financial ties in research and clinical care: a systematic review. Arch Intern Med 2010 Apr 26;170(8):675-682. [doi: 10.1001/archinternmed.2010.39] [Medline: 20421551]
30. DeJong C, Aguilar T, Tseng C, Lin GA, Boscardin WJ, Dudley RA. Pharmaceutical Industry-Sponsored Meals and Physician Prescribing Patterns for Medicare Beneficiaries. JAMA Intern Med 2016 Dec 01;176(8):1114-1122. [doi: 10.1001/jamainternmed.2016.2765] [Medline: 27322350]
31. Steinbrook R. Physicians, Industry Payments for Food and Beverages, and Drug Prescribing. JAMA 2017 May 02;317(17):1753-1754. [doi: 10.1001/jama.2017.2477] [Medline: 28464155]
32. EvaluateMedTech. World Preview 2017, Outlook to 2022 URL: http://info.evaluategroup.com/rs/607-YGS-364/images/MTWP2017.pdf [accessed 2018-10-07] [WebCite Cache ID 72y2vNeKy]
33. Samuel AM, Webb ML, Lukasiewicz AM, Bohl DD, Basques BA, Russo GS, et al. Orthopaedic Surgeons Receive the Most Industry Payments to Physicians but Large Disparities are Seen in Sunshine Act Data. Clin Orthop Relat Res 2015 Oct;473(10):3297-3306 [FREE Full text] [doi: 10.1007/s11999-015-4413-8] [Medline: 26088767]

34. Parikh K, Fleischman W, Agrawal S. Industry Relationships With Pediatricians: Findings From the Open Payments Sunshine Act. Pediatrics 2016 Dec;137(6):- [FREE Full text] [doi: 10.1542/peds.2015-4440] [Medline: 27244857]

Abbreviations

CMS: Centers for Medicare and Medicaid Services
IQR: interquartile ranges
NPI: National Provider Identifier
OPP: Open Payments Program

Edited by G Eysenbach; submitted 22.07.18; peer-reviewed by B Davies, M Cosentino; comments to author 16.08.18; revised version received 07.10.18; accepted 25.10.18; published 30.11.18.

Please cite as:
Na X, Guo H, Zhang Y, Shen L, Wu S, Li J
Mining Open Payments Data: Analysis of Industry Payments to Thoracic Surgeons From 2014-2016
J Med Internet Res 2018;20(11):e11655
URL: http://www.jmir.org/2018/11/e11655/
doi:10.2196/11655
PMID:30504119

Original Paper

Relations Between the Use of Electronic Health and the Use of General Practitioner and Somatic Specialist Visits in Patients With Type 1 Diabetes: Cross-Sectional Study

Anne Helen Hansen[1,2], BSW, MD, PhD; Jan Broz[3], MD; Tor Claudi[4], MD; Eirik Årsand[5,6], BSE, PhD

[1]Centre for Quality Improvement and Development, University Hospital of North Norway, Tromsø, Norway

[2]Department of Community Medicine, Faculty of Health Sciences, UiT - The Arctic University of Norway, Tromsø, Norway

[3]Department of Internal Medicine, Second Faculty of Medicine, Charles University, Prague, Czech Republic

[4]Medical Department, Nordland Hospital, Bodø, Norway

[5]Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway

[6]Department of Clinical Medicine, UiT - The Arctic University of Norway, Tromsø, Norway

Corresponding Author:
Anne Helen Hansen, BSW, MD, PhD
Centre for Quality Improvement and Development
University Hospital of North Norway
PO Box 35
Tromsø, 9038
Norway
Phone: 47 91619655
Email: anne.helen.hansen@unn.no

Abstract

Background: The prevalence of diabetes and the use of electronic health (eHealth) are increasing. People with diabetes need frequent monitoring and follow-up of health parameters, and eHealth services can be of great value. However, little is known about the association between the use of eHealth and provider-based health care services among people with diabetes.

Objective: The objective of this study was to investigate the use of 4 different eHealth platforms (apps, search engines, video services, and social media sites) and associations with the use of provider-based health care visits among people diagnosed with type 1 diabetes mellitus (T1DM).

Methods: We used email survey data collected from 1250 members of the Norwegian Diabetes Association (aged 18 to 89 years) in 2018. Eligible for analyses were the 523 respondents with T1DM. Using descriptive statistics, we estimated the use of eHealth and the use of general practitioners (GPs) and somatic specialist outpatient services. By logistic regressions, we studied the associations between the use of these provider-based health services and the use of eHealth, adjusted for gender, age, education, and self-rated health.

Results: Of the sample of 523 people with T1DM, 90.7% (441/486) had visited a GP once or more, and 61.0% (289/474) had visited specialist services during the previous year. Internet search engines (such as Google) were used for health purposes sometimes or often by 84.0% (431/513), apps by 55.4% (285/514), social media (such as Facebook) by 45.2% (232/513), and video services (such as YouTube) by 23.3% (118/506). Participants aged from 18 to 39 years used all forms of eHealth more than people aged 40 years and older, with the exception of social media. The use of search engines was positively associated with the use of somatic specialist services (odds ratio 2.43, 95% CI 1.33-4.45). GP visits were not associated with any kind of eHealth use.

Conclusions: eHealth services are now widely used for health support and health information by people with T1DM, primarily in the form of search engines but often in the form of apps and social media as well. We found a positive association between the use of search engines and specialist visits and that people with T1DM are frequent users of eHealth, GPs, and specialist services. We found no evidence that eHealth reduces the use of provider-based health care; these services seem to be additional rather than alternative. Future research should focus on how health care services can meet and adapt to the high prevalence of eHealth use. Our results also indicate that many patients with T1DM do not visit specialist clinics once a year as recommended. This raises questions about collaboration in health care services and needs to be followed up in future research.

(J Med Internet Res 2018;20(11):e11322) doi:10.2196/11322

KEYWORDS

eHealth; internet; health care utilization; general practitioners; specialist; cross-sectional study; diabetes mellitus, type 1; Norway

Introduction

Background

Internet-based health information, sensors, apps, and other solutions for self-management, as well as new treatment strategies have developed rapidly in recent years, becoming an important support for both patients and health services. Of particular interest in this regard are patients with chronic diseases, such as diabetes, who are in need of frequent monitoring and follow-up of health parameters.

Increasing Prevalence of Diabetes Mellitus

The prevalence of diabetes is increasing worldwide. Estimates of 415 million cases in 2015 (age group 20-79 years) are expected to rise to 642 million in 2040 [1]. Global prevalence in adults is estimated at 8.8% [1] and the Norwegian prevalence at 4.7% [2]. Around 245,000 persons are diagnosed with diabetes in Norway, of whom around 28,000 have type 1 diabetes mellitus (T1DM) [2]. Costs attributable to diabetes represent around 1.4% of the total Norwegian expenditure on health care [3]. Diabetes is a considerable burden on patients in terms of morbidity and mortality [4]. Most patients do not reach the combined national treatment targets for prevention of complications [5-7].

Increasing Use of Electronic Health Services

The World Health Organization states that “eHealth is the use of information and communication technologies (ICT) for health” [8]. The use of electronic health (eHealth) has increased over the past decades. Back in 2005, 44% of the general population of 7 European countries reported using the internet for health purposes [9,10], increasing to 52.2% by 2007 [11]. Consistent with European trends, in Poland, 66.7% used the internet for health purposes in 2012 [12]. Around 75% to 80% of internet users in the United States and Europe conduct health-related searches [9,13]. Most Norwegian households (97%) had internet access in 2015 [9-11,14], and 78% of the population aged 15 years and older have reported using the internet for health purposes [15]. In the Czech Republic, more than 25% of insulin-treated patients visited a professional diabetes internet portal in the period between 2009 and 2013 [16]. However, eHealth use among people with T1DM in Norway has yet to be explored.

Unclear Relations Between the Use of Electronic Health and Provider-Based Health Services

Andreassen et al found that the use of eHealth in a general population was positively associated with general practitioner (GP) visits (yes or no) [10], whereas others have reported no or inverse associations with the *frequency* of regular provider visits [17,18]. A German study found that frequent users of health services were 73% more likely to seek health information on the internet compared with nonusers [12]. Research on the associations between the use of eHealth and provider-based health care is scarce, both in general populations and for populations with specific diseases [19,20].

Norwegian Health Care Services

The Norwegian health care system is based on universal insurance. Primary health care is run by the municipalities. All residents are provided a regular GP according to the patient list system. Specialist outpatient services are operated by regional and local health enterprises owned by the national government, consisting of public and private somatic and psychiatric specialist services. Access to specialist care is usually achieved by referral from the regular GP (the gatekeeper role). However, persons with T1DM are recommended to make at least one annual visit to specialist health services [21] and are most often invited directly for annual checks. GP and specialist visits for adults have a small co-payment, with a total maximum limit of 2258 Norwegian Kroner (around US $280) within a year (2018).

Planning for Future Electronic Health and Provider-Based Health Care Services

The use of eHealth is an area of continuous and rapid development, which varies between regions, countries, diagnostic groups, health care services, and health care systems. Hence, research from different settings is important to achieve an overall epidemiological view. A comprehensive understanding of the influence of eHealth on health care utilization in patients with T1DM is thus important for patients, health care providers, administrators, policy makers, and society to enable evidence-based planning for future eHealth and provider-based health care services.

Aim

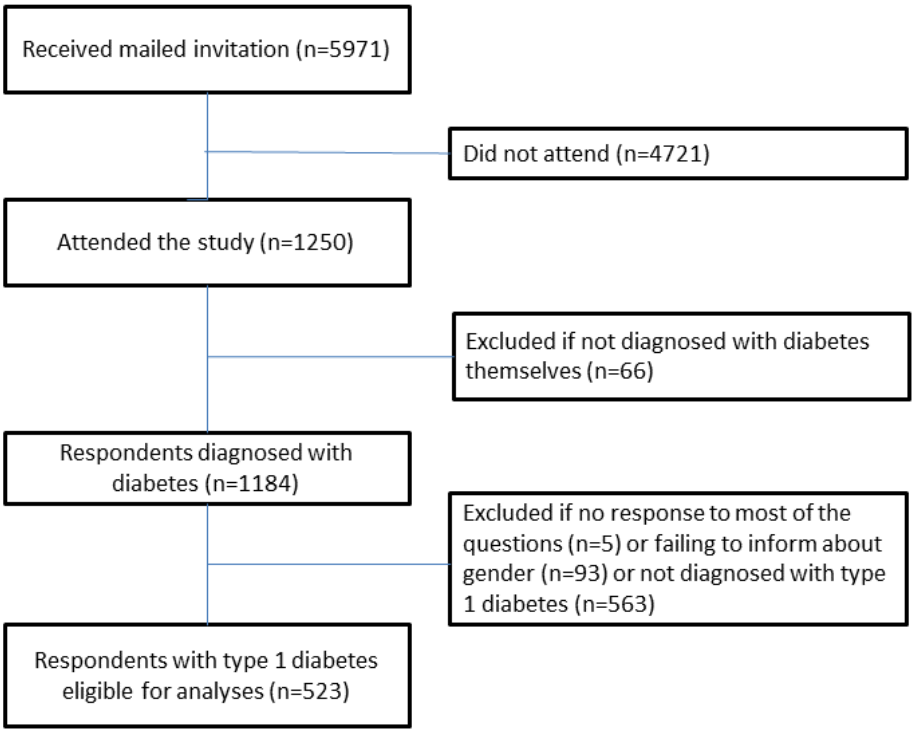
The aim of this study was to investigate which eHealth services are used among people with T1DM and whether the use of eHealth is associated with the use of primary and specialist health care services. Specifically, we tested whether the use of apps, search engines (such as Google), video services (such as YouTube), and social media (such as Facebook) was associated with the use of GPs and somatic outpatient specialist services.

Methods

Data

For this cross-sectional study, we used email survey data obtained in January and February 2018 from members of the Norwegian Diabetes Association. On December 31, 2017, the organization had 33,908 members—53% women and 47% men. Around 30% of the members have T1DM [22]. The Norwegian Centre for Research Data (NSD) Web survey distributed the invitations to a randomly selected sample of 5971 individuals (about 18% of all members).

Figure 1. Flowchart of study population.



Initially, as described in our protocol paper [23], we planned to use data from the seventh Tromsø study conducted in 2015-2016, where we participated in the planning. Comprehensive population-based health surveys have been conducted in Tromsø since 1974. In the seventh study, all residents aged 40 years and older (around 33,000 persons) were invited, and questions about the use of eHealth were included for the first time. However, after data collection, the Tromsø study was not able to give us access. Consequently, we had to change our data collection plans. We developed a tailored questionnaire based on the specific objectives of our study [23], using relevant questions from other published surveys on health care utilization and health information seeking [24,25].

Information about the study purpose and what participation would entail was distributed together with the invitation. The questionnaire (Multimedia Appendix 1) included questions about demographic and socioeconomic characteristics, health status including diabetes-specific questions about duration, severity and treatment, and use of and experiences with eHealth and health care services. Before data collection, the questionnaire was reviewed and tested several times by 2 persons diagnosed with diabetes and by experts from our research group (EÅ and AHH). Nonrespondents were given 1 reminder submitted by email 15 days after the first request.

Participants

It was not possible for the same respondent to fill in the questionnaire more than once. Starting from 1250 participants, we first excluded those who did not suffer from diabetes themselves (n=66). This group consisted of 61 family members, 4 health personnel (2 overlapping), and 3 others. We also excluded participants who failed to respond to most of the questions (n=5) and those who did not give information about gender (n=93). Finally, as we had decided to investigate T1DM in this part of the study, participants with type 2 diabetes mellitus (T2DM) and other diabetes types were excluded. The sample finally consisted of 523 respondents (Figure 1).

Variables

The dependent variables were use of GPs and outpatient specialist services during the previous 12 months. Specialist services use refers to any somatic specialist clinic visit, regardless of clinical issue (not only endocrinologists/diabetologists). For GP services, 2 dichotomous outcome variables were applied, 1 for use or no use and 1 for less frequent use (0 to 2 visits) or more frequent use (3 visits or more). The distinction between more and less frequent use was the 50th percentile, a cutoff point that has been used in previous research [26]. For outpatient specialist visits, the 50th percentile was set at 1, making a frequency variable redundant.

Respondents were asked about their use of eHealth in the same period. eHealth was subdivided into apps for mobile phone or tablet computer, search engines (such as Google), social media (such as Facebook), and video services (such as YouTube). eHealth variables were dichotomized by merging the original 4 answering options into "never or once" and "sometimes or often."

The use of eHealth (apps, search engines, social media, and video services) was the key independent variable. Adjustment independent variables were gender, age, education, and self-rated health. We grouped age in 20-year intervals. The 4 education categories were labeled as follows: low (primary/part of secondary school), middle (completed secondary school), high (college/university<4 years), and highest (college/university 4 years or more). Response options for self-rated health were excellent, good, fair, bad, and very bad. The bad and very bad categories were merged because of low numbers in the very bad category (4 respondents).

Analyses

Data were analyzed by means of descriptive statistics and logistic regressions. Correlations were tested with Spearman as well as Pearson correlation coefficients. We constructed 1 multivariable regression model for each of the dependent variables. The independent variables (apps, search engines, social media, video services, gender, age, education, and self-rated health) were introduced collectively into the models.

Due to a relatively low response rate, we compared respondents who did not respond initially but eventually consented with early respondents, assuming that late respondents were more similar to nonrespondents [27]. This was done by descriptive statistics (stratification), and by subsequently introducing the response time variable into the regression models.

We used 95% CIs throughout the study. All analyses were accomplished using Stata, version 14.2.

Ethics

This project has been presented to the Regional Committee for Medical and Health Research Ethics (REK), which found that an application was not required according to the Norwegian Health Research Act (2015/1779/REK nord). The study has been approved by the data protection officer (Personvernombudet) at the University Hospital of North-Norway (ref 2017/6579). The data bureau NSD received no other information about the participants than the email addresses.

Results

Participation

In total, 1250 persons aged 18 to 89 years participated, constituting a minimum response rate of 20.93% (1250/5971) (Figure 1). However, we experienced more than 400 bounce backs from servers unable to deliver the invitation. Consequently, the real response rate is assumed higher. Eligible for analysis in our study were the 523 persons who reported having T1DM (Figure 1).

Sample Characteristics

Mean age was 47.0 years, 48.9 years for men and 45.3 years for women. Median age was 48 years. Mean disease duration was 23.2 years (median 22 years). Most participants (75.7%) had suffered from diabetes for 10 years or more. The largest groups consisted of women (281/523, 53.7%), persons aged 40-59 years (223/523, 42.6%), married/cohabitants (338/380, 89.0%), full-time or part-time employed persons (308/481, 64.0%), persons with high education (152/480, 31.7%), high household income (238/467, 51.0%), good self-rated health (269/521, 51.6%), and good self-rated regulation of diabetes (292/520, 56.2%). Among the late respondents (those who responded after the reminder), older people (60 years and older) represented a larger proportion (57/180, 31.7% compared with 65/343, 19.0% among the early respondents Table 1).

The Use of Electronic Health and Provider-Based Health Care Services

During the previous year, 90.7% (441/486) visited a GP once or more, and 61.0% (289/474) visited somatic outpatient services (Table 2). Overall, 87.0% (447/514) used eHealth in one or more forms. Search engines were used sometimes or often by 84.0% (431/513), apps were used by 55.4% (285/514), social media were used by 45.2% (232/513), and video services were used by 23.3% (118/506) of participants (Table 2). People aged 40 years and older used all health care services more and all forms of eHealth less than younger people, with the exception of social media (Table 2).

Relations Between the Use of Provider-Based Health Care Services and Electronic Health

GP visits (yes or no, and frequency) were not associated with any type of eHealth use. The use of somatic outpatient services was positively associated with the use of search engines (odds ratio [OR] 2.43, CI 1.33-4.45; Table 3). The use of apps, social media, or video services was not associated with the use of health care services (Table 3).

The probability of 1 or more GP visits during the previous year increased with poorer health. People in good and fair self-rated health were more likely to visit the GP compared with those in excellent health (OR 2.20, CI 1.04-4.68 and OR 6.73, CI 1.78-25.49, respectively). All the 42 individuals reporting bad/very bad health had visited their GP during the previous year. As there were 0 observations in the nonvisiting group; OR was not calculated (Table 3).

The probability of more frequent GP visits also increased with poorer health and was almost 7 times higher in the bad/very bad health group compared with those in excellent health (OR 6.77, CI 2.63-17.45).

People in bad/very bad health were more than twice as likely to visit specialist services compared with those in excellent health (OR 2.51, CI 1.07-5.88).

Table 1. Percentage of sample characteristics.

Variables	Total sample, n	Early respondents, n	Late respondents, n	18-39 years, n	40-59 years, n	60 years and older, n
Gender	523	343	180	178	223	122
Female	53.7	52.2	56.7	59.5	52.0	48.4
Male	46.3	47.8	43.3	40.5	48.0	51.6
Age (years)	523	343	180	178	223	122
18-39	34.0	37.0	28.3	N/A[a]	N/A	N/A
40-59	42.6	44.0	40.0	N/A	N/A	N/A
60 and above	23.4	19.0	31.7	N/A	N/A	N/A
Marital status	380	254	126	129	163	88
Single	11.0	13.8	5.6	27.1	4.3	0.0
Married/cohabitant	89.0	86.2	94.4	72.9	95.7	100.0
Main daily activity	481	320	161	155	210	116
Working[b]	64.0	67.5	57.1	66.4	81.9	28.5
Pensioner old age	13.5	10.6	19.3	0.0	0.0	56.0
Pensioner disability	11.0	9.4	14.3	5.8	14.3	12.1
Pupil/student	7.3	8.1	5.6	22.6	0.0	0.0
Other	4.2	4.4	3.7	5.2	3.8	3.4
Education[c]	480	319	161	155	210	115
Low	8.1	6.3	11.8	5.2	7.6	13.0
Middle	29.0	27.9	31.1	32.9	26.7	27.8
High	31.7	33.8	27.3	29.7	32.9	32.2
Highest	31.2	32.0	29.8	32.2	32.8	27.0
Household income[d]	467	309	158	153	207	107
Low	14.1	13.3	15.8	24.2	7.7	12.2
Middle	34.9	33.3	38.0	37.3	26.1	48.6
High	51.0	53.4	46.2	38.5	66.2	32.2
Duration of diabetes	522	343	179	178	222	122
<10 years	24.3	23.3	26.2	36.0	17.6	19.7
10-19 years	20.5	20.7	20.1	28.6	17.6	13.9
20-29 years	19.4	20.7	16.8	27.5	17.6	10.7
30 years and above	35.8	35.3	36.9	7.9	47.2	55.7
Self-rated regulation of diabetes	520	341	179	178	222	120
Excellent	19.4	17.0	24.0	19.1	17.6	23.3
Good	56.2	57.2	54.2	50.0	57.7	62.5
Fair	19.8	20.5	18.4	23.6	20.7	12.5
Bad/very bad	4.6	5.3	3.4	7.3	4.0	1.7
Self-rated health	521	342	179	178	222	121
Excellent	17.9	15.5	22.4	14.6	19.4	19.8
Good	51.6	53.2	48.6	59.6	47.3	47.9
Fair	21.7	21.9	21.2	15.7	23.4	27.3
Bad/very bad	8.8	9.4	7.8	10.1	9.9	5.0

[a]N/A: not applicable.

[b]Full-time or part-time.

[c]Low (primary/part of secondary school), middle (completed secondary school), high (college/university <4 years), and highest (college/university 4 years or more).

[d]Low (Norwegian Kroner [NOK] <350,000), middle (NOK 351,000-750,000), high (NOK >750,000).

Table 2. Proportion using provider-based health care services and 4 types of electronic health (eHealth) services during the previous 12 months.

Variables	Total T1DM[a] sample, n (%)	Early respondents, n (%)	Late respondents, n (%)	18-39 years, n (%)	40-59 years, n (%)	60 years and above, n (%)
Use of health care services once or more						
GP[b]	441 (90.7)	293 (90.4)	148 (91.2)	132 (84.1)	198 (93.8)	111 (94.1)
Somatic outpatient specialist	289 (61.0)	182 (58.2)	107 (66.5)	82 (53.3)	140 (67.3)	67 (59.8)
Use of eHealth sometimes or often						
Apps	285 (55.4)	194 (57.6)	91 (51.4)	108 (62.4)	121 (55.3)	56 (45.9)
Search engines	431 (84.0)	293 (86.9)	138 (78.4)	159 (91.9)	184 (84.4)	88 (72.1)
Social media	232 (45.2)	154 (45.8)	78 (44.1)	78 (45.1)	108 (49.5)	46 (37.7)
Video services	118 (23.3)	80 (24.1)	38 (21.8)	45 (26.5)	51 (23.5)	22 (18.5)

[a]T1DM: type 1 diabetes mellitus.

[b]GP: general practitioner.

Among people aged 60 years and older, the probability of visiting a GP was more than 5 times higher than in the youngest (18 to 39 years) group (OR 5.25, CI 1.82-15.14). The probability of frequent visits was twice as high among people aged 60 years and older (OR 1.98, CI 1.15-3.40). The probability of visiting specialist services was significantly higher among people aged from 40 to 59 years, compared with the youngest group (OR 1.97, CI 1.25-3.11), but not significantly higher among those aged 60 years and older (Table 3).

The group with high education was more likely to visit somatic outpatient services compared with the low education group (OR 2.97, CI 1.36-6.51). GP visits were not associated with educational level. Gender was not associated with the use of health care services (Table 3). Goodness of fit was tested by Hosmer/Lemeshow goodness-of-fit test. All tests showed nonsignificant *P* values, indicating acceptable fit for all models.

All findings regarding GP visits persisted after introducing the response time variable into the regression models. The probability of using somatic outpatient services was increased among the late respondents compared with the early respondents (OR 1.84, CI 1.18-2.86). The positive association between the use of somatic outpatient services was increased among those in fair health (OR 2.04, CI 1.07-3.91) and among those in bad/very bad health (OR 2.74, CI 1.16-6.45). Otherwise, the introduction of the response time variable did not alter the results.

There were no strong correlations (defined as Spearman rho >.5) between the independent variables in any of the models. A similar result was found using Pearson correlation test.

Table 3. Probability of using general practitioners (GPs) and somatic outpatient services during the previous year in a population with diabetes type 1 (multivariable logistic regressions).

Variables	GP visits (yes/no; n=432), OR[a](95% CI)	GP visits (0 to 2 vs 3 visits or more; n=474), OR (95% CI)	Somatic outpatient visits (yes/no; n=468), OR (95% CI)
Apps[b]	1.29 (0.62-2.70)	1.21 (0.77-1.88)	0.65 (0.41-1.02)
Search engines[b]	0.92 (0.33-2.52)	1.10 (0.60-2.02)	*2.43*[c] (*1.33-4.45*)
Social media[b]	1.33 (0.59-3.01)	1.38 (0.88-2.17)	1.15 (0.73-1.82)
Video services[b]	1.73 (0.64-4.70)	0.94 (0.56-1.56)	1.06 (0.64-1.76)
Gender[d]	0.64 (0.32-1.29)	0.77 (0.51-1.15)	0.99 (0.66-1.49)
Age, in years			
18-39[e]	1.00	1.00	1.00
40-59	*3.50* (*1.63-7.52*)	*1.74* (*1.10-2.75*)	*1.97* (*1.25-3.11*)
60 and above	*5.25* (*1.82-15.14*)	*1.98* (*1.15-3.40*)	1.63 (0.96-2.79)
Education[f]			
Low[e]	1.00	1.00	1.00
Middle	2.48 (0.66-9.26)	1.03 (0.46-2.33)	1.69 (0.78-3.66)
High	2.66 (0.70-10.11)	1.17 (0.52-2.64)	*2.97* (*1.36-6.51*)
Highest	1.56 (0.44-5.59)	0.70 (0.31-1.58)	2.16 (0.98-4.74)
Self-rated health[f]			
Excellent[f]	1.00	1.00	1.00
Good	*2.20* (*1.04-4.68*)	1.35 (0.80-2.28)	1.43 (0.84-2.42)
Fair	*6.73* (*1.78-25.49*)	*4.15* (*2.15-8.01*)	1.86 (0.98-3.52)
Bad/very bad[g]	N/A[h]	*6.77* (*2.63-17.45*)	*2.51* (*1.07-5.88*)

[a]OR: odds ratio.

[b]Apps/search engines/social media/video services in 2 groups: 1=never or once, 2=sometimes or often.

[c]Statistically significant findings are marked in italics.

[d]Gender: 1=women, 2=men.

[e]Reference groups.

[f]Education: low (primary/part of secondary school), middle (high school), high (college/university <4 years), highest (college/university 4 years or more).

[g]Odds ratio was not calculated for general practitioner visits (yes/no) because of zero (0) observations in the nonvisiting group.

[h]N/A: not applicable.

Discussion

Principal Findings

We found that 90.7% (441/486) of study participants visited a GP once or more during the previous year, and 61.0% (289/474) participants visited somatic outpatient services. Search engines were used sometimes or often by 84.0% (431/513), apps by 55.4% (285/514), social media by 45.2% (232/513), and video services by 23.3% (118/506) of the participants. Participants aged 18 to 39 years used the investigated forms of eHealth more than people aged 40 years and above, with the exception of social media. GP visits were not associated with the use of eHealth, whereas visits to specialist services were positively associated with the use of search engines. Poorer self-rated health and higher age were associated with increased use of GPs and specialist services, whereas higher education was associated with increased use of specialist services. Gender, social media use, and video services use were not associated with the use of health services.

Low Specialist and High General Practitioner Visit Rates

We were surprised that only 61.0% (289/474) reported 1 or more visits to any somatic specialist service during the previous 12 months. This is a higher rate than reported for the general population [28] but still remarkably low as at least one annual checkup visit is recommended for people diagnosed with T1DM [21]. In Salten, Norway, around 80% of insulin-treated patients reported visits to specialist services regarding their diabetes during 2014 (unpublished data from the Rosa4 study, communicated by TCL). We have not found other studies

reporting specialist checkup rates for patients with T1DM in Norway; however, it has been suggested that many older patients are monitored by their GP [7]. Our finding that people aged 60 years and older less likely visited specialists compared with people aged 40 to 59 years (Table 3) supports this view. However, it is important to note that low specialist visit rates apply to all age groups, and in particular to younger ages (18 to 39 years, Table 2).

On the other hand, a GP visit rate of 90.7% (441/486) is high compared with around 80% in the general population [25,28]. Patients with T1DM might see their GP for a variety of health problems that are connected, directly or indirectly, to their diabetes. A few GPs are specialized in diabetes care and might partly provide a substitute for specialist clinic checkup. This might explain some of the low specialist visit rate, but it is unlikely to explain all of it. The notion that people with T1DM are followed up by annual checkups in specialist services thus needs to be questioned or at least nuanced. It may be problematic if the GP and the specialist both believe that the other is performing the checkup of their patients, with the risk of dropouts. Our finding raises questions about the collaboration between GPs, specialist health services, and patients and needs to be followed up in further research.

Extensive Use of Electronic Health

Our finding that people with T1DM are heavy users of all 4 forms of eHealth is not surprising. Among people aged 40 years and older, the different types of eHealth were used from around twice (search engines) to 5, 6, or 7 times (apps, video services, and social media) more than reported for the general population in the Tromsø Study [25]. However, it should be noted that our data were collected 2 to 3 years later than the Tromsø Study. Considering the rapid development in this field, some of the differences might be due to changing of trends over time. Nevertheless, reports are quite consistent that people with chronic conditions or poorer self-rated health are more likely to use eHealth than the general population [29-33]. This conforms with the illness behavior model [34], indicating that people in poor health are more likely to seek Web-based disease-related information, where an obvious prerequisite is access [35]. Concerns about one's own disease or poor health will naturally lead to demand for relevant information. This extensive use may reinforce the notion that eHealth and provider-based health care are additional rather than alternative services at present, possibly interacting in a reciprocal way [33,36] and that this applies at least as much to people with T1DM as to the general population.

The Age Divide

This study confirms that younger people use the internet for health purposes more than older people, particularly apps and search engines. Previous research is consistent regarding this age divide for general populations, elderly populations, and populations with chronic disease [10,11,15,17,33,37-41]. Findings by Tarver et al may indicate that age differences among internet users decreased in the period from 2003 to 2013, although this finding was not statistically significant [41]. This possible trend may amplify when cohorts exposed to digital technology from childhood get older, and the present inverse association between age and eHealth use might not be sustained to the same extent in the future. Elderly people are a rapidly growing age group in Europe and a fast-growing group of eHealth users [12,42].

Positive Association Between Specialist Visits and the Use of Search Engines

GP visits were not associated with any kind of eHealth use, whereas specialist visits were positively associated with the use of search engines. Back in 2008, Lee suggested that internet use for health information increased contact with health professionals [36]. In line with this, 2 Asian studies recently found that internet use was significantly associated with more outpatient clinic visits [29,43]. Medlock's study of elderly people in the Netherlands, however, reported that use of health professionals was not associated with internet use [24]. None of these studies can be directly compared with this study, as study methodology, health care systems, and cultures differ substantially, and the other studies were not performed in disease-specific populations. The specific finding of an association between specialist visits and the use of search engines among people with T1DM might point to a need for additional information concerning specialist visits, which may be greater than the need connected to GP visits. We know that eHealth might be used *before* the visit to seek information or to decide about the need to see a doctor and *after* the visit for additional information [31,33,44]. The GP-patient relationship may have a longer duration, and GP visits may be more frequent than specialist visits. Thus, a closer relationship with continuity combined with room for questions and discussions may develop [45]. In Norwegian specialist care, the patient will not necessarily see the same physician from 1 visit to the next, and the internet may be of great value as a source of supplemental information.

Self-Rated Health, Education, and Late Respondents

People in poorer self-rated health used the surveyed provider-based health services significantly more than people in better health. This adds to solid documentation in previous research, both for disease-specific and general populations [26,29].

The probability of visiting somatic specialist services was higher among those with higher education, even if health is usually worse in lower socioeconomic groups. Our finding is consistent with previous findings for the general Norwegian population [26].

The increased probability of using somatic outpatient services among the late respondents in our study is most likely because of the higher age in this group and the consequences related to increased health care needs. Otherwise, our investigation of early versus late respondents did not alter the results.

Limitations

The main limitation of this study is the low participation rate, one of the indicators of study representativeness [46]. However, response rate must not be confused with response quality [27]. More important is the assessment of the possible influence of nonparticipation on exposure, outcome, or the relation of interest

[47]. In our study, older people dominated among the late respondents compared with the early respondents. Assuming that late respondents are more similar to nonrespondents, younger individuals may be overrepresented in our study.

The distribution of the questionnaire to people with email addresses excluded those who do not use the internet or do not have an email address. The distribution of functioning email addresses might have been skewed, for instance, toward younger members. However, as 97% of Norwegian households have internet access, 90% of Norwegians use the internet every day, and around 91% use email [42], we do not think this has affected our results significantly. Nevertheless, some of those invited might not use email regularly, and some may use mobile phones more than computers. In both cases, there might be barriers to filling out a large questionnaire. In Norway, younger people use email less and mobile phones more than middle-aged and older people [42]. In addition, participation in surveys is generally lower among younger people [48]. These factors might contribute to balance a possible overrepresentation of younger people in this study, and consequently add to its generalizability.

It is well known that women, healthier persons, higher socioeconomic groups, and middle-aged people are more likely to participate in surveys [47,48]. This suggests that women, people aged from 40 to 80 years, people in better health, and higher socioeconomic groups might be overrepresented in our study, thus tending to level out a possible skewness in the opposite directions.

Furthermore, we presume that people with interest in eHealth might be overrepresented, as interest in the topic studied has shown to increase responses [49]. If this is the case, our rates of eHealth use might be higher than the true rates. However, this point applies to most other eHealth studies as well.

In questionnaire data, there is always a potential for recall bias, particularly regarding minor events and distant past, usually leading to underreporting [50]. In addition, the validity of self-reported data of health care utilization may be questioned, although agreement between self-reported and registered health care use is generally high [51]. The cross-sectional study design implies that no causal relationships can be established. Furthermore, we cannot exclude the possibility of unmeasured confounders of the reported associations.

Overall, we conclude that younger individuals might be overrepresented in this study. It is not possible to judge the magnitude of a possible bias, as different factors might pull the tendency in different directions or level each other out. The low response rate is in itself not an indication of low representativeness, as nonresponse bias may be a problem even if response rates are high [52]. Moreover, our results seem reasonable and not contradictory to prior research where such research is available. We suggest that bias poses a limited threat to our study's validity. Nevertheless, generalization must be approached with caution.

Strengths

One strength of this study is the focus on an area that has been scarcely investigated. Another strength is the fact that we were able to design a more detailed questionnaire specifically tailored to people with diabetes. This enabled us to distinguish between T1DM and T2DM. Further, we were able to recruit participants from all of Norway, not only from Tromsø municipality, as planned in the protocol using data from the Tromsø Study. This study included individuals from 18 years of age, whereas participants in the Tromsø Study were 40 years and older. Moreover, we were able to analyze data shortly after they were collected, which we consider of great importance in the rapidly developing field of eHealth. Finally, yet importantly, the collection of data in cooperation with the Norwegian Diabetes Association enabled us to develop excellent user participation with a large and important group of health care users.

Future Plans

We plan to extend this study by investigating how eHealth use and use of other health care services, such as emergency departments and hospitalizations, might be related. Furthermore, we will make efforts to contribute to a deeper understanding of possible causal relationships regarding the use of eHealth and provider-based health care services. In further studies, this will be applied to populations with T2DM as well as T1DM.

Conclusions

We found that the eHealth services are widely used for health information by people with T1DM, primarily in the form of search engines, but often in the form of apps and social media as well. Our study suggests a positive association between the use of search engines and specialist visits and that people with T1DM are frequent users of eHealth, GPs, and specialist services. We found no evidence that eHealth reduces the use of provider-based health care, suggesting that these services are additional rather than alternative in today's health care. For future research, it would be interesting to investigate how eHealth services may replace some of today's face-to-face consultations. Moreover, our results indicate that many patients with T1DM do not visit specialist clinics once a year as recommended. This raises questions about the collaboration between GPs, specialist services, and patients and needs to be followed up in future research.

Acknowledgments

The authors would like to thank the Norwegian Diabetes Association for their cooperation in the conduct and performance of this study. Without their generous contribution, this study could not have been realized. The authors also thank the Northern Norway Health Authorities who provided funding for this research (DIAcare, project number HST1306-16) and approved the changes regarding data collection that had to be made to complete the study.

Authors' Contributions

All authors contributed to the design and conduct of the study. AHH drafted the manuscript. All authors contributed with improvements and critical revisions and approved the final version for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questionnaire.

[PDF File (Adobe PDF File), 33KB - jmir_v20i11e11322_app1.pdf]

References

1. Ogurtsova K, da Rocha Fernandes JD, Huang Y, Linnenkamp U, Guariguata L, Cho NH, et al. IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. Diabetes Res Clin Pract 2017 Jun;128:40-50. [doi: 10.1016/j.diabres.2017.03.024] [Medline: 28437734]
2. Norwegian Institute of Public Health. Public Health in Norway 2018 URL: https://www.fhi.no/globalassets/dokumenterfiler/rapporter/2018/helsetilstanden-i-norge-2018.pdf [accessed 2018-06-16] [WebCite Cache ID 70DeT070T]
3. Solli O, Jenssen T, Kristiansen IS. Diabetes: cost of illness in Norway. BMC Endocr Disord 2010 Sep 20;10:15 [FREE Full text] [doi: 10.1186/1472-6823-10-15] [Medline: 20854689]
4. Sandahl K, Nielsen LB, Svensson J, Johannesen J, Pociot F, Mortensen HB, et al. Increased mortality in a Danish cohort of young people with type 1 diabetes mellitus followed for 24 years. Diabet Med 2017 Dec;34(3):380-386. [doi: 10.1111/dme.13124] [Medline: 27027777]
5. Jenssen TG, Tonstad S, Claudi T, Midthjell K, Cooper J. The gap between guidelines and practice in the treatment of type 2 diabetes A nationwide survey in Norway. Diabetes Res Clin Pract 2008 May;80(2):314-320. [doi: 10.1016/j.diabres.2007.12.025] [Medline: 18279994]
6. Mouland G. [Diabetes in general practice--were treatment goals reached?]. Tidsskr Nor Laegeforen 2014 Jan 28;134(2):168-172 [FREE Full text] [doi: 10.4045/tidsskr.13.0375] [Medline: 24477150]
7. Cooper JG, Claudi T, Thordarson HB, Løvaas KF, Carlsen S, Sandberg S, et al. Treatment of type 1 diabetes in the specialist health service--data from the Norwegian Diabetes Register for Adults. Tidsskr Nor Laegeforen 2013 Nov 12;133(21):2257-2262 [FREE Full text] [doi: 10.4045/tidsskr.13.0153] [Medline: 24226332]
8. World Health Organization. http://www.who.int/ehealth/en/. eHealth URL: http://www.who.int/ehealth/en/ [accessed 2018-06-16] [WebCite Cache ID 70DfcHSuE]
9. Wangberg SC, Andreassen HK, Prokosch HU, Santana SM, Sørensen T, Chronaki CE. Relations between internet use, socio-economic status (SES), social support and subjective health. Health Promot Int 2008 Mar;23(1):70-77. [doi: 10.1093/heapro/dam039] [Medline: 18083686]
10. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. BMC Public Health 2007 Apr 10;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]
11. Kummervold PE, Chronaki CE, Lausen B, Prokosch HU, Rasmussen J, Santana S, et al. eHealth trends in Europe 2005-2007: a population-based survey. J Med Internet Res 2008 Nov 17;10(4):e42 [FREE Full text] [doi: 10.2196/jmir.1023] [Medline: 19017584]
12. Bujnowska-Fedak MM. Trends in the use of the Internet for health purposes in Poland. BMC Public Health 2015 Feb 27;15:194 [FREE Full text] [doi: 10.1186/s12889-015-1473-3] [Medline: 25886280]
13. Fox S, Duggan M. Pew Internet & American Life Project. Washington DC: Pew Research Center; 2013. Health Online URL: http://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2018-06-16] [WebCite Cache ID 70Dh4ruWf]
14. Nölke L, Mensing M, Krämer A, Hornberg C. Sociodemographic and health-(care-)related characteristics of online health information seekers: a cross-sectional German study. BMC Public Health 2015 Jan 29;15:31 [FREE Full text] [doi: 10.1186/s12889-015-1423-0] [Medline: 25631456]
15. Sørensen T, Andreassen HK, Wangberg SC. Nasjonalt senter for e-helseforskning. 2014. Prosjektrapport e-helse i Norge 2013 URL: https://ehealthresearch.no/files/documents/Prosjektrapporter/NST-rapport_2014-02_e-helse_i_Norge_2013.pdf [accessed 2018-06-16] [WebCite Cache ID 70DhrW5OT]
16. Broz J, Brabec M, Brozova K, Cibulkova I, Zdarska D, Hartmann D. Professional internet information source used as educational resource for patients with insulin-treated diabetes in the Czech Republic: a 5-year analysis of operations. Wiener klinische Wochenschrift 2016;128:153 [FREE Full text] [doi: 10.1007/s00508-015-0946-9]
17. Koch-Weser S, Bradshaw YS, Gualtieri L, Gallagher SS. The Internet as a health information source: findings from the 2007 Health Information National Trends Survey and implications for health communication. J Health Commun 2010;15 Suppl 3:279-293. [doi: 10.1080/10810730.2010.522700] [Medline: 21154099]

18. AlGhamdi KM, Moussa NA. Internet use by the public to search for health-related information. Int J Med Inform 2012 Jun;81(6):363-373. [doi: 10.1016/j.ijmedinf.2011.12.004] [Medline: 22217800]
19. Catan G, Espanha R, Mendes RV, Toren O, Chinitz D. Health information technology implementation - impacts and policy considerations: a comparison between Israel and Portugal. Isr J Health Policy Res 2015;4:41 [FREE Full text] [doi: 10.1186/s13584-015-0040-9] [Medline: 26269740]
20. Cline RJ, Haynes KM. Consumer health information seeking on the Internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692. [Medline: 11780707]
21. Norwegian Directorate of Health. National Professional Guideline for Diabetes: Follow-up, investigation and organization of diabetes care URL: https://helsedirektoratet.no/retningslinjer/diabetes/seksjon?Tittel=oppfolging-utredning-og-organisering-7125 [accessed 2018-06-16] [WebCite Cache ID 70DjGRLu1]
22. The Norwegian Diabetes Association. Årsberetning 2017 URL: https://issuu.com/knutjarle/docs/a_rsberetning_2017_ [accessed 2018-06-16] [WebCite Cache ID 70DkKIFTA]
23. Hansen AH, Bradway M, Broz J, Claudi T, Henriksen O, Wangberg SC, et al. The use of eHealth and provider-based health services by patients with diabetes mellitus: protocol for a cross-sectional study. JMIR Res Protoc 2016 Oct 31;5(4):e207 [FREE Full text] [doi: 10.2196/resprot.6529] [Medline: 27799136]
24. 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 Jan 08;17(1):e10 [FREE Full text] [doi: 10.2196/jmir.3749] [Medline: 25574815]
25. The Tromsø Study. URL: http://tromsoundersokelsen.uit.no/tromso/ [accessed 2018-06-16] [WebCite Cache ID 70Dkah0hU]
26. Hansen AH, Halvorsen PA, Ringberg U, Førde OH. Socio-economic inequalities in health care utilisation in Norway: a population based cross-sectional survey. BMC Health Serv Res 2012 Sep 25;12:336 [FREE Full text] [doi: 10.1186/1472-6963-12-336] [Medline: 23006844]
27. Johnson TP, Wislar JS. Response rates and nonresponse errors in surveys. J Am Med Assoc 2012 May 02;307(17):1805-1806. [doi: 10.1001/jama.2012.3532] [Medline: 22550194]
28. Hansen AH, Halvorsen PA, Forde OH. The ecology of medical care in Norway: wide use of general practitioners may not necessarily keep patients out of hospitals. J Public Health Res 2012 Jun 15;1(2):177-183 [FREE Full text] [doi: 10.4081/jphr.2012.e28] [Medline: 25180941]
29. Hsieh RW, Chen L, Chen TF, Liang JC, Lin TB, Chen YY, et al. The association between internet use and ambulatory care-seeking behaviors in Taiwan: a cross-sectional study. J Med Internet Res 2016 Dec 07;18(12):e319 [FREE Full text] [doi: 10.2196/jmir.5498] [Medline: 27927606]
30. 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]
31. Rice RE. Influences, usage, and outcomes of Internet health information searching: multivariate results from the Pew surveys. Int J Med Inform 2006 Jan;75(1):8-28. [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]
32. Fox S. After Dr Google: peer-to-peer health care. Pediatrics 2013 Jun;131 Suppl 4:S224-S225. [doi: 10.1542/peds.2012-3786K] [Medline: 23729765]
33. Ayers SL, Kronenfeld JJ. Chronic illness and health-seeking information on the Internet. Health (London) 2007 Jul;11(3):327-347. [doi: 10.1177/1363459307077547] [Medline: 17606698]
34. Li J, Theng YL, Foo S. Predictors of online health information seeking behavior: changes between 2002 and 2012. Health Informatics J 2016 Dec;22(4):804-814. [doi: 10.1177/1460458215595851] [Medline: 26261218]
35. Fox S, Purcell K. Pew Internet & American Life Project. Washington DC: Pew Research Center; 2010. Chronic disease and the internet URL: http://www.pewinternet.org/2010/03/24/chronic-disease-and-the-internet/ [accessed 2018-06-16] [WebCite Cache ID 70Dljfgyl]
36. Lee CJ. Does the internet displace health professionals? J Health Commun 2008;13(5):450-464. [doi: 10.1080/10810730802198839] [Medline: 18661387]
37. Bauer AM, Rue T, Keppel GA, Cole AM, Baldwin LM, Katon W. Use of mobile health (mHealth) tools by primary care patients in the WWAMI region Practice and Research Network (WPRN). J Am Board Fam Med 2014;27(6):780-788 [FREE Full text] [doi: 10.3122/jabfm.2014.06.140108] [Medline: 25381075]
38. Campbell RJ, Nolfi DA. Teaching elderly adults to use the Internet to access health care information: before-after study. J Med Internet Res 2005 Jun 30;7(2):e19 [FREE Full text] [doi: 10.2196/jmir.7.2.e19] [Medline: 15998610]
39. Satterlund MJ, McCaul KD, Sandgren AK. Information gathering over time by breast cancer patients. J Med Internet Res 2003;5(3):e15 [FREE Full text] [doi: 10.2196/jmir.5.3.e15] [Medline: 14517106]
40. Takahashi Y, Ohura T, Ishizaki T, Okamoto S, Miki K, Naito M, et al. Internet use for health-related information via personal computers and cell phones in Japan: a cross-sectional population-based survey. J Med Internet Res 2011 Dec 14;13(4):e110 [FREE Full text] [doi: 10.2196/jmir.1796] [Medline: 22169526]
41. Tarver WL, Menser T, Hesse BW, Johnson TJ, Beckjord E, Ford EW, et al. Growth dynamics of patient-provider internet communication: trend analysis using the Health Information National Trends Survey (2003 to 2013). J Med Internet Res 2018 Mar 29;20(3):e109 [FREE Full text] [doi: 10.2196/jmir.7851] [Medline: 29599107]

42. Norwegian Media Barometer 2017. Oslo-Kongsvinger: Statistics Norway; 2018. URL: https://www.ssb.no/en/kultur-og-fritid/statistikker/medie/aar/2018-04-19 [accessed 2018-06-16] [WebCite Cache ID 70DmN6Jxv]
43. Nguyen HT, Nakamura K, Seino K, Vo VT. Association between a wider availability of health information and health care utilization in Vietnam: cross-sectional study. J Med Internet Res 2017 Dec 18;19(12):e405 [FREE Full text] [doi: 10.2196/jmir.8328] [Medline: 29254911]
44. McMullan M. Patients using the internet to obtain health information: how this affects the patient-health professional relationship. Patient Educ Couns 2006 Oct;63(1-2):24-28. [doi: 10.1016/j.pec.2005.10.006] [Medline: 16406474]
45. Hansen AH, Halvorsen PA, Aaraas IJ, Førde OH. Continuity of GP care is related to reduced specialist healthcare use: a cross-sectional survey. Br J Gen Pract 2013 Jul;63(612):482-489 [FREE Full text] [doi: 10.3399/bjgp13X669202] [Medline: 23834885]
46. Tolonen H, Koponen P, Borodulin K, Männistö S, Peltonen M, Vartiainen E. Differences in participation rates between urban and rural areas are diminishing in Finland. Scand J Public Health 2018 Jan 01:1403494817748737. [doi: 10.1177/1403494817748737] [Medline: 29307266]
47. Galea S, Tracy M. Participation rates in epidemiologic studies. Ann Epidemiol 2007 Sep;17(9):643-653. [doi: 10.1016/j.annepidem.2007.03.013] [Medline: 17553702]
48. Langhammer A, Krokstad S, Romundstad P, Heggland J, Holmen J. The HUNT study: participation is associated with survival and depends on socioeconomic status, diseases and symptoms. BMC Med Res Methodol 2012 Sep 14;12:143 [FREE Full text] [doi: 10.1186/1471-2288-12-143] [Medline: 22978749]
49. Kristoffersen AE, Fønnebø V, Norheim AJ. Do cancer patients with a poor prognosis use complementary and alternative medicine more often than others? J Altern Complement Med 2009 Jan;15(1):35-40. [doi: 10.1089/acm.2008.0262] [Medline: 19769474]
50. Roberts RO, Bergstralh EJ, Schmidt L, Jacobsen SJ. Comparison of self-reported and medical record health care utilization measures. J Clin Epidemiol 1996 Sep;49(9):989-995. [Medline: 8780606]
51. Reijneveld SA, Stronks K. The validity of self-reported use of health care across socioeconomic strata: a comparison of survey and registration data. Int J Epidemiol 2001 Dec;30(6):1407-1414. [Medline: 11821355]
52. Halbesleben JR, Whitman MV. Evaluating survey quality in health services research: a decision framework for assessing nonresponse bias. Health Serv Res 2013 Jun;48(3):913-930 [FREE Full text] [doi: 10.1111/1475-6773.12002] [Medline: 23046097]

Abbreviations

eHealth: electronic health
GP: general practitioner
NSD: Norwegian Centre for Research Data
OR: odds ratio
REK: Regional Committee for Medical and Health Research Ethics
T1DM: type 1 diabetes mellitus
T2DM: type 2 diabetes mellitus

Edited by G Eysenbach; submitted 19.06.18; peer-reviewed by W Tarver, YY Chen; comments to author 19.07.18; revised version received 11.09.18; accepted 29.09.18; published 07.11.18.

Please cite as:
Hansen AH, Broz J, Claudi T, Årsand E
Relations Between the Use of Electronic Health and the Use of General Practitioner and Somatic Specialist Visits in Patients With Type 1 Diabetes: Cross-Sectional Study
J Med Internet Res 2018;20(11):e11322
URL: http://www.jmir.org/2018/11/e11322/
doi:10.2196/11322
PMID:30404766

Original Paper

A Practical Do-It-Yourself Recruitment Framework for Concurrent eHealth Clinical Trials: Simple Architecture (Part 1)

Hannah L Palac[1,2], MS; Nameyeh Alam[2], MPH; Susan M Kaiser[2], MPH; Jody D Ciolino[3], PhD; Emily G Lattie[4], PhD; David C Mohr[2], PhD

[1]AbbVie Inc, North Chicago, IL, United States

[2]Center for Behavioral Intervention Technologies, Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[3]Division of Biostatistics, Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[4]Center for Behavioral Intervention Technologies, Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

Corresponding Author:
Emily G Lattie, PhD
Center for Behavioral Intervention Technologies
Department of Medical Social Sciences
Northwestern University Feinberg School of Medicine
10th Floor
750 North Lake Shore Drive
Chicago, IL, 60611
United States
Phone: 1 3125033741
Email: emily.lattie@northwestern.edu

Abstract

Background: The ability to identify, screen, and enroll potential research participants in an efficient and timely manner is crucial to the success of clinical trials. In the age of the internet, researchers can be confronted with large numbers of people contacting the program, overwhelming study staff and frustrating potential participants.

Objective: This paper describes a "do-it-yourself" recruitment support framework (DIY-RSF) that uses tools readily available in many academic research settings to support remote participant recruitment, prescreening, enrollment, and management across multiple concurrent eHealth clinical trials.

Methods: This work was conducted in an academic research center focused on developing and evaluating behavioral intervention technologies. A needs assessment consisting of unstructured individual and group interviews was conducted to identify barriers to recruitment and important features for the new system.

Results: We describe a practical and adaptable recruitment management architecture that used readily available software, such as REDCap (Research Electronic Data Capture) and standard statistical software (eg, SAS, R), to create an automated recruitment framework that supported prescreening potential participants, consent to join a research registry, triaging for management of multiple trials, capture of eligibility information for each phase of a recruitment pipeline, and staff management tools including monitoring of participant flow and task assignment/reassignment features. The DIY-RSF was launched in July 2015. As of July 2017, the DIY-RSF has supported the successful recruitment efforts for eight trials, producing 14,557 participant records in the referral tracking database and 5337 participants in the center research registry. The DIY-RSF has allowed for more efficient use of staff time and more rapid processing of potential applicants.

Conclusions: Using tools already supported at many academic institutions, we describe the architecture and utilization of an adaptable referral management framework to support recruitment for multiple concurrent clinical trials. The DIY-RSF can serve as a guide for leveraging common technologies to improve clinical trial recruitment procedures.

(J Med Internet Res 2018;20(11):e11049) doi:10.2196/11049

KEYWORDS

eHealth; mHealth; online recruitment; REDCap; referral management

Introduction

The ability to identify, screen, and enroll potential research participants in an efficient and timely manner is crucial to the success of clinical trials. However, researchers continually emphasize insufficient accrual and subsequent challenges [1-3], with an estimated 31% of trials failing to meet original recruitment targets and 53% of trials requiring additional time to meet study recruitment goals [4,5]. Failure to reach recruitment targets may result in a loss of statistical power and increase trial duration and costs [6,7]. Furthermore, even when studies are completed on time, a large part of the costs are associated with participant recruitment [8].

In recent years, clinical trial recruitment support systems (CTRSS) have aimed at improving the efficiency and effectiveness of clinical trial recruitment. These CTRSS use electronic patient data, typically from electronic health records or data warehouse services, to assess patient eligibility for one or more trials. The system then alerts a physician or patient of study eligibility or provides a list of potential trial participants to a study investigator [9]. Similarly, Web-based recruitment strategies have become increasingly popular, especially for electronic and mobile health (eHealth/mHealth) studies [10-12]. A 2011 systematic review concluded that prescreening is the most effective part of the recruitment process addressed by CTRSS [13]. Prescreening optimizes recruitment procedures by reserving staff time for interactions with participants who are likely to be enrolled [14]. Dividing recruitment into a series of different examinations of increasing complexity has also been shown to be successful in recruitment for large trials [15,16], although few studies describe their prescreening methods and subsequent record keeping. There are few, if any, published descriptions on optimization of these processes for eHealth clinical trials. Wide-reaching and cost-effective recruitment strategies are essential for trial success. There remains a need for systems that can function across strategies and capture data necessary to provide insight into the efficiency of different types of recruitment methods to enable investigators to appropriately allocate recruitment resources.

This paper (the first of a two-part series) describes the process of acquiring design requirements and the architecture for a practical “do-it-yourself” recruitment support framework (DIY-RSF) using technologies readily available at many academic institutions. The framework was developed to support recruitment efforts for multiple, concurrent institutional review board-approved clinical trials conducted by the Center for Behavioral Intervention Technologies (CBITs). This paper characterizes the important features of this framework and discusses the uptake and use at a single academic research center. The second part to this paper by Lattie and colleagues describes the use of the DIY-RSF to identify the most cost-effective and time-efficient recruitment strategies for eHealth clinical trials and provide a guide for recruitment decision making [17].

Methods

Setting

Located within the Northwestern University Feinberg School of Medicine, CBITs is an academic research center focused on developing and evaluating behavioral intervention technologies, including Web, mobile, and sensor technologies that help people make positive behavior changes to support physical and emotional health. It is supported by an interdisciplinary team of research faculty, software engineers, and research support staff. The trials at CBITs generally focus on evaluating the use and efficacy of these digital mental health interventions and tools. At the time of the needs assessment, CBITs had approximately one dozen research protocols underway, with studies requiring anywhere from five to 100 enrolled participants each month. Participant recruitment methods included digital strategies (eg, social media, Craigslist), print advertisements (eg, flyers, posters on public transportation), research registries, clinics, and traditional media strategies (eg, press releases), which directed prospective participants to the center website and/or a prescreening Web survey via various links. The second part of this paper outlines detailed descriptions of the center’s recruitment strategies [17].

Needs Assessment

To inform the DIY-RSF, a needs assessment was conducted. A research program manager conducted a series of individual, unstructured interviews with investigators and research personnel to identify current barriers to recruitment and determine important features for the new system. Informal focus groups were also conducted during standing meetings. Interviewees were encouraged to speak freely regarding pain points with current processes and to generate a “wish list” for a new system. The research program manager then generated a prioritized list of requirements for a recruitment management framework, met with a data manager to review these requirements, and designed a general structure and workflow to support the identified needs and pain points. Some suggestions, such as interactive visualizations to show the lifecycle of a participant from the time of referral to the center through enrollment, were deemed impractical due to technology constraints. There was also a request to develop a comprehensive trial management toolkit; however, due to the magnitude of this task, it was decided early on to focus on recruitment as this was the area of greatest need in the center. System features were chosen and prioritized by determining which components of recruitment were experiencing the greatest deficiencies and the overall feasibility of implementation given the technologies available to us. Follow-up meetings were scheduled as needed to further discuss and revise the proposed DIY-RSF.

Results

Needs Assessment Findings

The results of the needs assessment revealed a number of challenges, inefficiencies, and wishes for a new system. Challenges included concurrent, ongoing trials with competition for participants across studies with similar goals and entry

criteria, the addition of new studies over time, and a multidisciplinary research team with shared, yet unique, roles and responsibilities. Examples of overlapping entry criteria for trials included access to technology (personal computer and/or mobile phone with internet access) and at least moderate symptoms of depression and/or anxiety as measured by the Patient Health Questionnaire-9 (PHQ-9) [18] and Generalized Anxiety Disorder-7 (GAD-7) [19]. Inefficiencies included multiple disconnected systems for tracking referrals, insufficient staffing to manage time-intensive contact with the growing number of ineligible individuals, and inability for management to track recruiter caseloads in real time. Requirements for the new system are listed in Textbox 1. In developing the DIY-RSF to meet these requirements, we aimed to optimize the design to minimize participant burden and staff time.

Recruitment Support Framework

Multiple software, including REDCap (Research Electronic Data Capture); SAS software, version 9.4 (SAS Institute Inc, Cary, NC, USA); and the R programming language, version 3.4.3, supported the development and implementation of the DIY-RSF. Microsoft Excel was used to create a recruitment dashboard; however, the DIY-RSF framework does not require Excel, nor does Excel itelf support any other framework components. We deployed two different REDCap projects: (1) a prescreening Web survey and registry, and (2) a referral tracking database (Figure 1). REDCap is a Web-based platform used for collecting and managing research data in noncommercial settings [20]. Records housed in REDCap are stored to meet an institution's local security standards. We chose the REDCap platform to create this new framework because the center already used it for traditional, study-specific data collection, our institution housed its own platform with accompanying support staff, and there were no additional technology costs to use it. We employed two separate REDCap projects to allow for new, prospective participants to take the Web survey via a public survey link and also to allow for the capture of legacy records for those who did not have prescreening data. We used SAS software and the R programming language to export data from the REDCap projects using REDCap's application programming interface (API), perform data manipulations to make eligibility determinations, and transfer data between REDCap projects. The REDCap API is an interface that allows external programs, such as statistical software, to remotely connect to the REDCap platform and import and export data automatically via a programming script. Both SAS and R were used simply due to programmer preferences.

We modified the preexisting multistage recruitment pipeline within the center to include the prescreening Web survey, a consent form and eligibility questionnaire, a final eligibility interview, and enrollment into the trial (Figure 2). Study staff conducted a brief follow-up phone call to confirm interest and verify contact information after the prescreening Web survey for select studies. The recruitment pipeline employed a scaffold approach to screening, where the length and complexity of questions increased over time and each key point in the pipeline involved an eligibility decision. Traditional REDCap databases for housing study-specific data also played a role in the DIY-RSF via data transfer between projects. The Northwestern University Institutional Review Board (IRB) approved all registry and recruitment strategies, including all recruitment methods and materials, prescreening survey questions, informed consent language, and data handling practices.

REDCap Project #1: Prescreening Web Survey and Registry

The Prescreening Web Survey and Registry project is a single-survey REDCap project that also captures consent and responses to questions pertaining to a research registry. We designed this project to meet the requirements of self-referral, prescreening Web survey (requirement #1), permission of data capture on recruitment sources (requirement #3), and integration of a participant research registry (requirement #4). Prospective participants who wish to be screened for at least one CBITs clinical trial or join the center registry click the public survey link hosted on the center webpage or posted via other recruitment sources, including website and print advertisements.

First, prospective participants review a brief IRB-approved description of each study with active recruitment. After specifying study preferences, additional IRB-approved study information and/or waiver of informed consent are displayed. Participants imply consent to proceed to prescreening by continuing with the survey. Respondents then answer a series of questions regarding referral source, preliminary inclusion/exclusion criteria for studies, and general registry questions. In some instances, the referral source was prefilled in the survey by appending appropriate URL parameters to the survey link. This allowed participants to bypass the question on referral source and mitigate self-report biases. The research team made the decision to avoid asking for certain identifying or sensitive information, such as detailed mental health history, until after participants provided consent pertaining to specific study protocols.

Textbox 1. System requirements of the "do-it-yourself" recruitment support framework (DIY-RSF).

1. Self-referral Web screening survey accessible via computers and mobile devices for participants
2. Ability to make automated eligibility determinations for multiple studies based on survey responses and route participants to appropriate studies
3. Ability to capture data on recruitment sources and enrollment outcomes
4. An integrated participant research registry
5. A centralized database for tracking prospective participants throughout the enrollment pipeline
6. Staff management tools that support real-time monitoring of recruiter caseloads and assignment/reassignment of cases

Figure 1. REDCap framework. API: application programming interface.

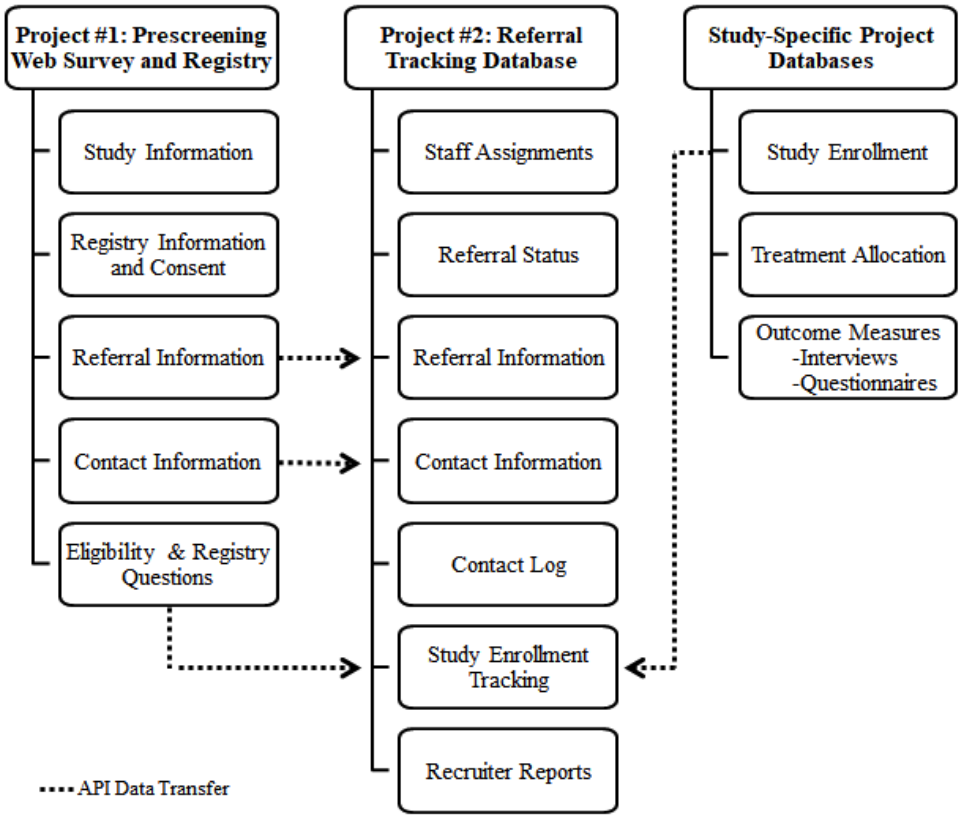
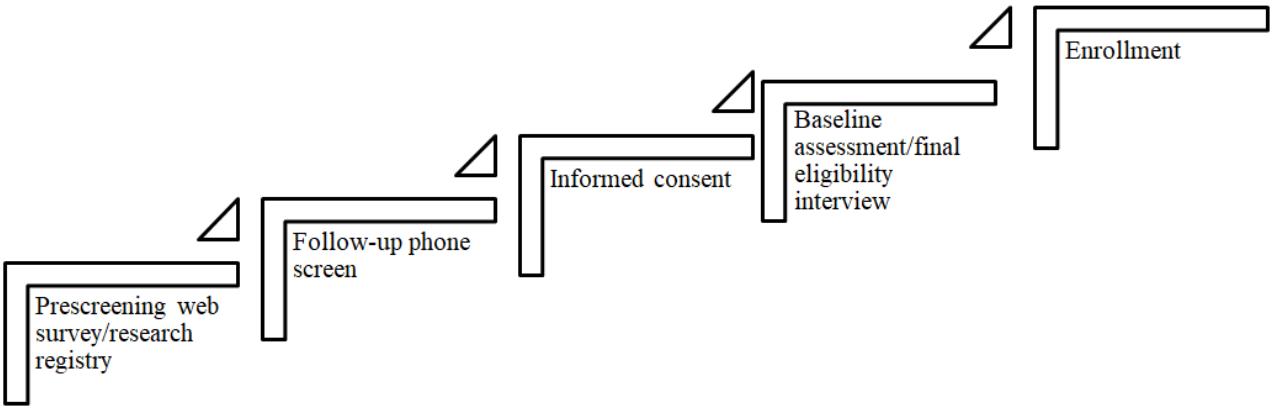


Figure 2. Multistage recruitment pipeline.



Prospective participants end the survey by providing contact information and preferred methods for outreach. When a participant clicks "Submit" on the last page of the survey, responses are saved in REDCap and may be accessed by the research team. This project accomplishes the need of having an efficient prescreening process that allows prospective participants to take the Web survey at any time from an internet browser or mobile device and consent to join the center research registry at the same point of contact.

REDCap Project #2: Referral Tracking Database

We designed the referral tracking database to serve as the central repository for data pertaining to participant recruitment sources and screening and enrollment outcomes for all trials within the center (requirement #5) and provide management with tools to support real-time monitoring of recruiter caseloads and assignment/reassignment of cases (requirement #6). The referral tracking database contains two data collection instruments: (1) referral info and (2) tracking. These data collection instruments provide an organized structure for staff to access participant data for recruitment as a whole and by study. The project is set up longitudinally with one "referral info" event and a "tracking" event for every study utilizing the DIY-RSF for recruitment. The referral tracking project serves as the "hub" for center staff to access participant records, recruiter caseloads, and basic reports for center management and staff. The referral info form contains recruiter assignments, participant referral status, detailed recruitment source information, contact information, contact preferences, and a log of all outreach attempts. The tracking form contains study-specific information concerning screening, consent, and enrollment dates, as applicable. Ultimately, this format contains eligibility determinations for

each step in the recruitment pipeline and keeps a record of study enrollment across all center trials.

New referrals are batch processed daily by importing data from the prescreening survey project and reading it into SAS software via the REDCap API, which addresses the requirement to provide automated eligibility determinations for multiple studies based on survey responses and route participants to appropriate studies (requirement #2). Specifically, the API allows the REDCap and the statistical software to programmatically perform automated activities between the two systems, including reading data in directly from the referral tracking database into the SAS software without needing to manually log-in to REDCap using the user interface. A SAS program then processes survey responses to subset surveys for completed and new Web surveys since the last batch processing, merge referral tracking records by email and first and last name, and assess participant eligibility for each study based on responses to survey questions. Participants are routed to a specific study based on eligibility, study choice preferences, and center recruitment targets. The SAS program was written from scratch by the research team and includes a series of DATA steps, IF and WHERE expressions, and sorting procedures. The program code is updated as needed as studies open and close, and further upon modifications to eligibility criteria or study routing preferences. The program then generates an import template that contains a list of new referrals with their eligibility status, study to route to (if eligible), and recruitment source information. The template is formatted for the referral tracking project, which can be remotely imported to the project via the REDCap API or uploaded manually using the data import tool from within the REDCap platform.

Recruiter Reports

A key component of the referral tracking project is the recruiter reports, which utilize the REDCap Data Exports, Reports, and Stats application. These are a series of individualized reports for each recruiter based on a prospective participant's current referral status and can also be used to monitor recruiter caseloads and assign or reassign cases (requirement #6). Report filters are used to determine which records should appear on specific reports. Center workflow instructs recruiters to check reports on a daily basis and complete a series of action items, which may differ depending on a participant's referral status. For example, the referral status of "Action required: eligible Web screener" would trigger a recruiter to attempt to contact a participant to complete the next phase of the recruitment pipeline (eg, follow-up phone screen or informed consent). A referral status of "In progress: scheduling phone screen" would advise recruiters to follow up with participants, adhering to center operating procedures regarding frequency and methods of outreach. When recruiters make appropriate updates to the participant record in the database per the action items, such as finalizing a referral status, participants will no longer appear on the reports per filter specifications. Examples of final referral status categories include "ineligible Web screener," "could not be contacted," or "screened." The workflows for each stakeholder in the DIY-RSF are shown in Figures 3 and 4.

Recruitment Dashboard

We built an Excel dashboard to provide an at-a-glance view of key performance indicators pertaining to recruitment, such as cumulative counts of referrals and enrollment by recruitment source that can be filtered for a specified date range or trial. On a weekly basis, the REDCap API reads data into R, makes minor aggregations from data housed in the referral tracking database, and outputs an Excel template that updates the source data for the dashboard. Center staff involved in recruitment review the dashboard to maintain awareness of recruitment processes and allow for open dialog regarding reasons for lags and appropriate action, such as making adjustments to various recruitment strategies [21]. Future directions for the recruitment dashboard include improving efficiency and user interface with an R Shiny Web app [22].

Implementation

We first launched the CBITs DIY-RSF on July 27, 2015. Prior to the launch, the center's Data Core led a training session for all staff involved in participant recruitment. The training session utilized live demonstrations and PowerPoint, and staff received an instruction guide for continued reference after the training session. The primary cost of framework development and implementation involved staff effort. There were no outside technology costs other than those already supported at the institution and within the center (eg, SAS licenses). Development and management of the DIY-RSF were supported by a master's-level statistician and clinical data manager and overseen by a research program manager. No new personnel were hired to support the framework and all duties pertaining to the development and management of the effort were worked into existing day-to-day operations.

Data Transfer

Following the initial launch, recruitment personnel were encouraged to provide feedback and suggestions for improvement about the new processes. Research assistants identified the need for manual duplicate data entry across multiple databases as a key inefficiency. In addition to compiling new records from the prescreening Web survey project and transferring to the referral tracking project, we created programming scripts using the REDCap API to transfer data between study-specific databases and the tracking form in the referral tracking database. This allowed for central access of certain information entered in the study-specific databases, such as consent and eligibility interview outcomes, from the referral tracking database (Figure 1), and it allowed center research assistants to avoid double data entry. Thus, this mitigated staff burden, improved efficiency, and reduced opportunity for human error and inconsistencies.

Figure 3. Framework structure and workflow for the participant. CBITs: Center for Behavioral Intervention Technologies.

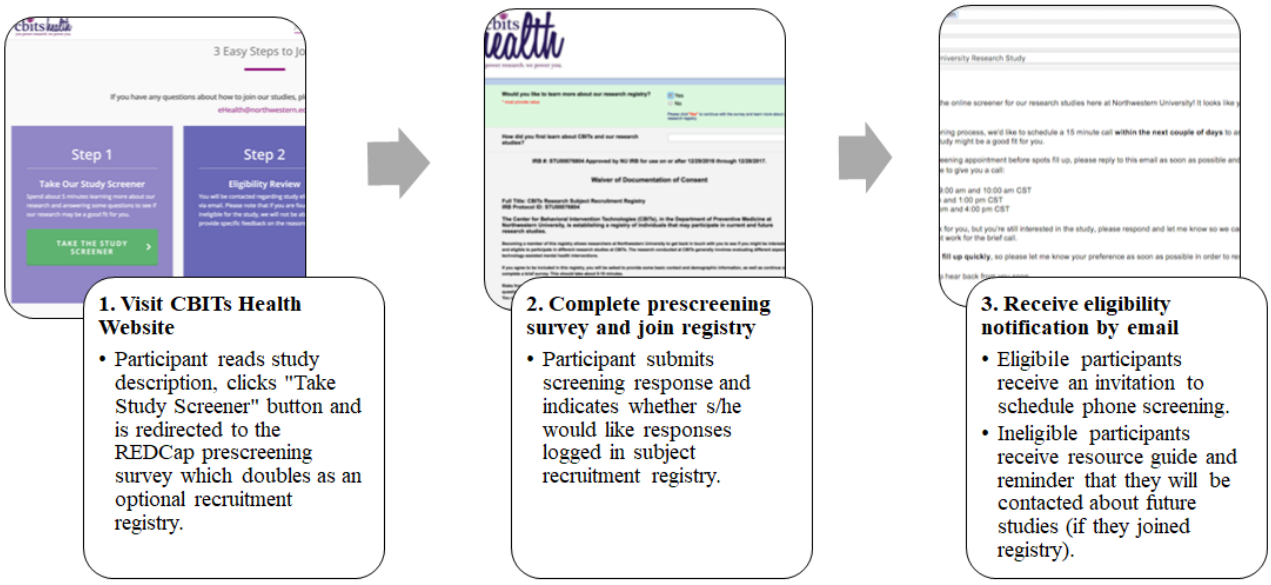
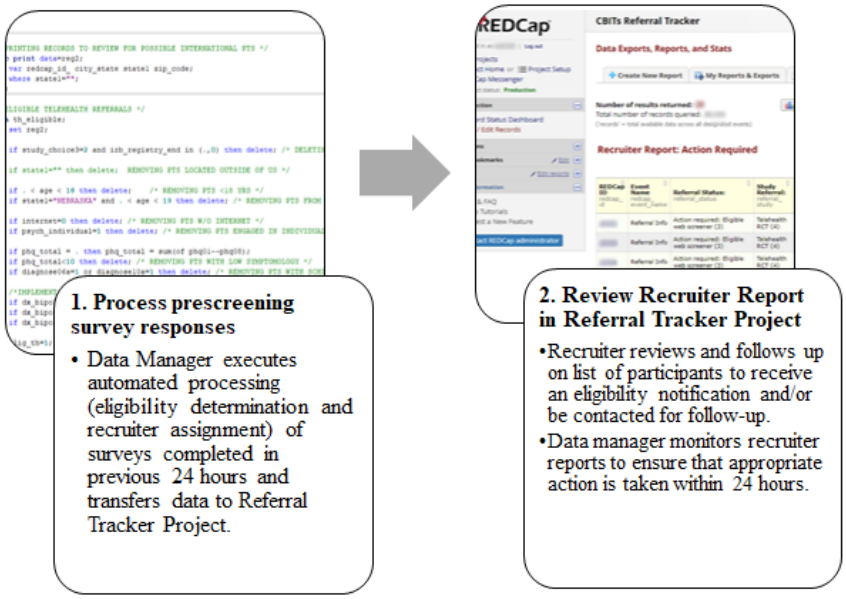


Figure 4. Framework structure and workflow for the data manager and recruiter.



Uptake

In the 2 years following the launch, the center has used the system to recruit participants for 8 research projects, including the successful completion of four NIH-funded projects, which recruited across the life span, from a project targeting youth ages 14 to 19 years, to another project targeting adults 65 years and older [23,24]. As of July 26, 2017, a total of 6525 prospective participants had completed the prescreening survey and there are currently 14,557 records entered into the referral tracking database. Of the 6525 participants who have completed the prescreening survey, 5337 (81.54%) consented to participate in the registry and agreed to be contacted at a later date for future center studies. The majority of participants enrolled reported hearing about the center via digital recruitment strategies (eg, Instagram, Reddit) and research registries (eg, ResearchMatch) [17]. Of the 14,557 records in the referral tracking database, 770 (5.29%) have been enrolled in a clinical trial. Figure 5 displays participant accrual summaries at each phase of the recruitment pipeline. Tables 1 and 2 summarize characteristics of subjects who consented to the research registry and subjects enrolled in at least one trial.

There are 453 fields, including legacy fields, across both REDCap projects. This highlights the breadth of data the framework captures. Figure 6 presents the number of referrals and enrollments in the 2-year timeframe following the launch of the DIY-RSF. In January 2017, we received a large increase in referrals due to a press release and subsequent news coverage in CBITs research. Valleys primarily correspond with holidays, such as our institution's winter recess.

Figure 5. Flowchart of participant accrual outcomes at each phase of the recruitment pipeline. Note: denominator remains 14,557 referrals throughout flow diagram.

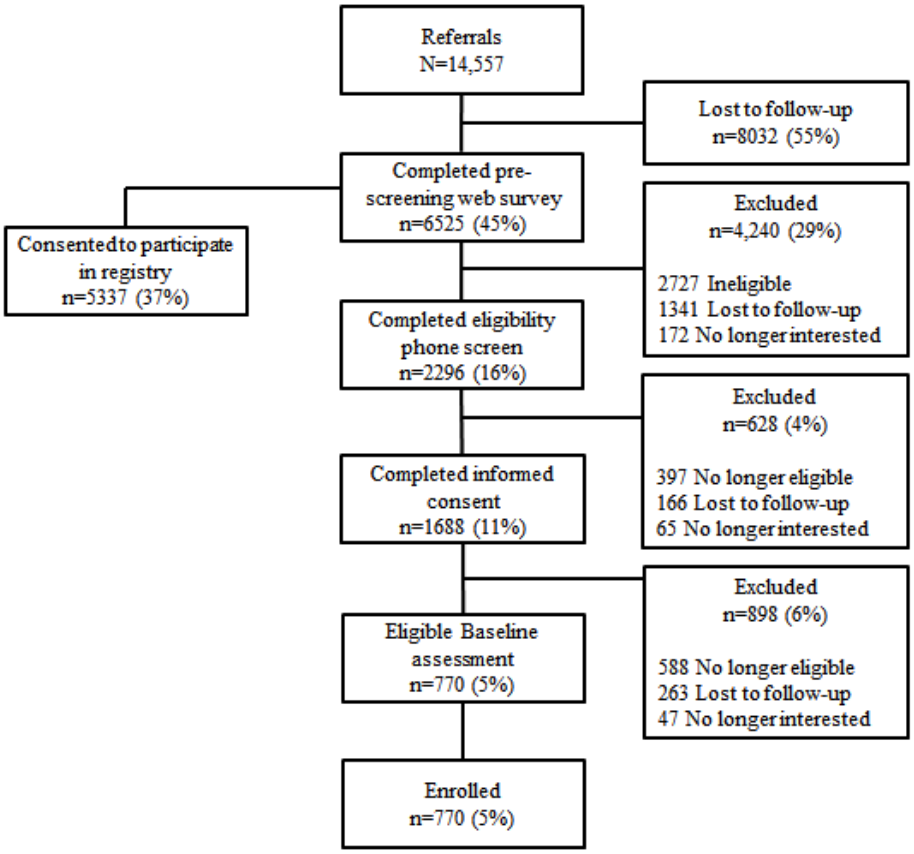


Table 1. Characteristics of participants in the Center for Behavioral Intervention Technologies recruitment registry and enrolled in at least one clinical trial. Percentages were calculated using nonmissing observations.

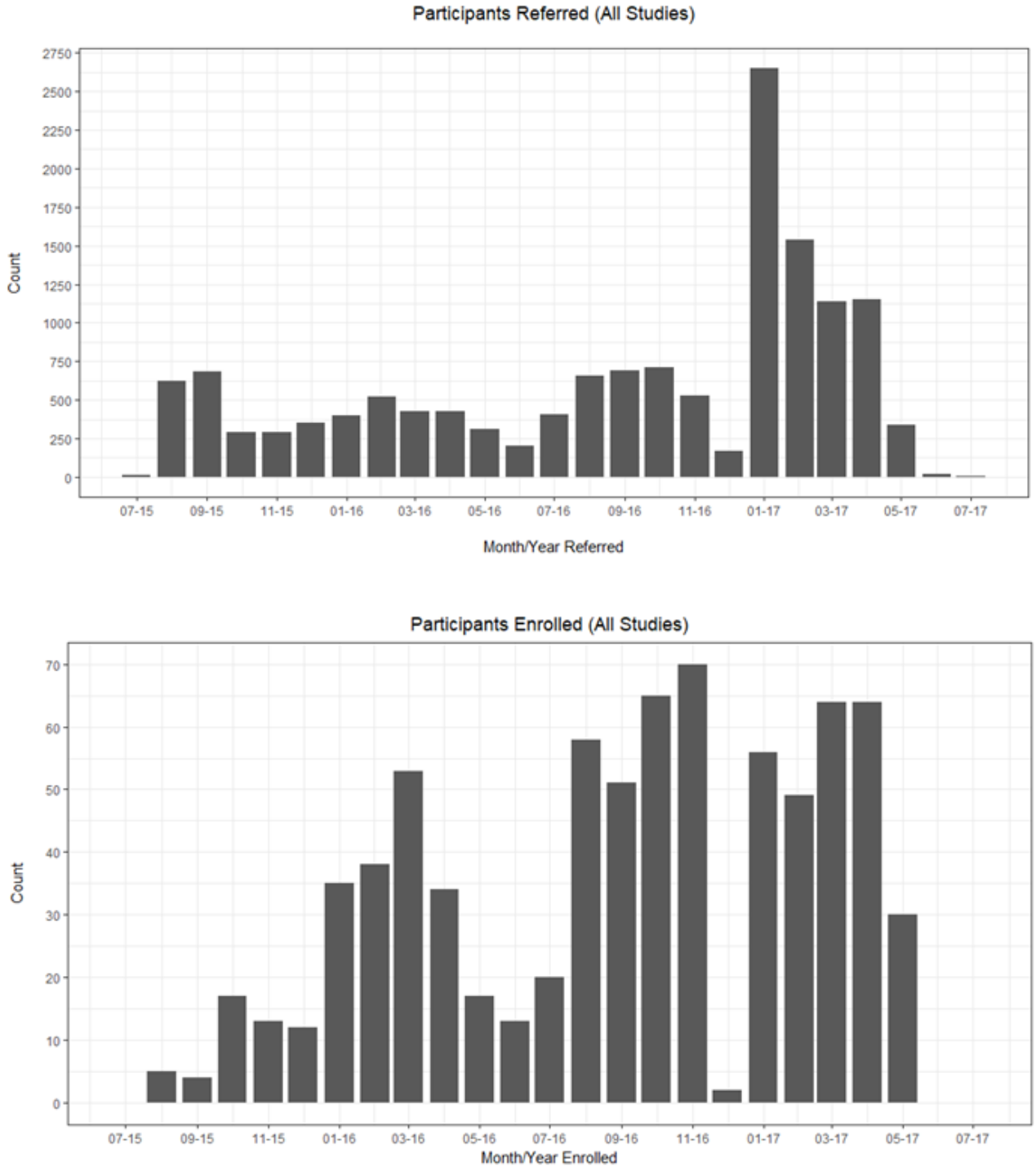
Variable	Registry participants (N=5337), n (%)	Enrolled participants (n=770), n (%)
Gender	**N=5333**	**N=770**
Female	4062 (76.2)	584 (75.8)
Male	1217 (22.8)	183 (23.8)
Other	54 (1)	3 (0.4)
Mobile device	**N=5334**	**N=769**
Yes	5141 (96.4)	741 (96.4)
No	193 (3.6)	28 (3.6)
Mobile device type	**N=5136**	**N=695**
Android	3512 (68.4)	556 (80)
iPhone	1569 (30.5)	138 (19.9)
Windows	27 (0.5)	0 (0)
Other	28 (0.5)	1 (0.1)
Access to internet	**N=5141**	**N=770**
Yes	5090 (99.0)	770 (100)
No	51 (1.0)	0 (0)

Table 2. Characteristics of participants in the Center for Behavioral Intervention Technologies recruitment registry and enrolled in at least one clinical trial.

Variable	Registry (N=5337)		Enrolled (n=770)	
	Participants, n (%)	Mean (SD)	Participants, n (%)	Mean (SD)
Age (years)	5307 (99.44)	35.2 (13.7)	770 (100.00)	38.3 (15.6)
Patient Health Questionnaire-8[a]	5337 (100.00)	14.3 (5.4)	729 (94.68)	14.8 (4.5)
Generalized Anxiety Disorder-7	4548 (85.22)	12.2 (5.4)	578 (75.06)	12.7 (4.4)

[a]Final question omitted from Patient Health Questionnaire-9 during prescreening stage.

Figure 6. Participant referrals and enrollments over the 2 years following launch of the "do-it-yourself" recruitment support framework.



Discussion

This manuscript explains the process of developing a "do-it-yourself" recruitment support framework (DIY-RSF) that streamlines and improves participant accrual for a single center in an academic setting conducting concurrent similar prospective studies. The DIY-RSF was successfully implemented in our center to support recruitment and referral management efforts. To accomplish this, we conducted a needs assessment, which identified a number of challenges and inefficiencies pertaining to current recruitment workflows. This needs assessment guided the development of our DIY-RSF. The motivation behind the implementation of the framework was to create a practical, flexible system that would mitigate current inefficiencies and support a range of trials with both overlapping and unique entry criteria and a multidisciplinary research team with varying roles and responsibilities. Since participant recruitment in clinical trials continues to be time- and cost-intensive, we used tools already supported at many academic institutions. As described previously, REDCap seemed the most logical option given the needs assessment and the availability of the platform: the local institution supported it already and made it freely available to the parties involved, which is not uncommon for those institutions housing a REDCap platform; further, REDCap is easily accessible to a wide range of skill sets with a minimal learning curve and provides an API to permit external software, such as SAS and R, to programmatically export and import data across projects.

The DIY-RSF supported automated prescreening of a large number of prospective participants and rapid identification of ineligible participants, preserving staff time for participants who have a high probability of enrolling in a trial. Other advantages include the ability to balance recruiter caseloads using features that support monitoring and assignment or reassignment of cases to other staff. The capture of objective data pertaining to study eligibility at each stage of the recruitment pipeline allowed for ongoing adjustments and optimization of recruitment strategies to improve efficiency and cost-effectiveness (see Lattie et al [17] for a more detailed description). In addition to staff-facing advantages, participant-facing advantages include the ability to take the survey from a browser or mobile device at any time and it is an easy way to send contact information. Thus, the DIY-RSF we describe in this paper is ideal for efficiently processing large numbers of prospective participants, while providing a streamlined process for participants, tools for managing staff, and data to refine recruitment processes.

The problems facing CBITs are representative of the problems facing many clinical research groups and centers as recruitment increasingly uses online modalities as the point of first contact. To date, the majority of large clinical trials within the center have been online in nature to evaluate eHealth and mHealth interventions, which did not require in-person visits with the study team. The center receives a large volume of community-based referrals with a large proportion of these referrals being ineligible or becoming lost to follow-up. "Lost to follow-up" refers to prospective participants who at one point were actively engaging in the enrollment pipeline, but were unable to be reached to move forward in the process at a point of follow-up in the recruitment process. The DIY-RSF permits a brief prescreening Web survey and automated eligibility determinations as the first step of engagement, which substantially reduces staff time spent on initial outreach and telephone screening procedures. This framework is particularly optimized for eHealth/mHealth trials that do not require in-person visits; however, many of these features may also be useful for other research groups that conduct clinic-based trials, as the internet is increasingly the medium through which people search for information, communicate, and contact research groups.

As with any new tool, the center continues to make updates to the DIY-RSF to support changing workflow of the center's recruitment procedures, new trials, and incorporate feedback from stakeholders. Support for ongoing management and updates to the framework and associated programs are currently supported by a master's-level data manager. We also note the center allocates a significant amount of staffing time and resources to recruitment and participant outreach, and this framework is only a single component of broader recruitment efforts within the center.

A limitation of this study is that data on referrals to the center prior to the implementation of the framework were not routinely collected and therefore prerecruitment and postrecruitment metrics are not available for comparison. The use of existing tools that were not specifically designed for recruitment processing also resulted in some limitations. First, our DIY-RSF's capacity for data clean-up, such as reconciliation of duplicate records, was not optimal. Although the statistical program used to process referrals evaluates for duplicate records based on first and last name and email address, some instances require manual removal (eg, when survey respondents enter nicknames and secondary email addresses that the program failed to identify). Second, REDCap lacks complex visualization capabilities making visual representations of the enrollment pipeline require external software, although the API may simplify this issue. Lastly, this framework requires a strong understanding of the REDCap platform and some programming expertise for API utilization. Although the research team used SAS and R for API utilization in this framework, other languages, such as Python, PHP, Java, or Ruby, may also be used. Support is required for the ongoing management and updates to the framework and associated programs. However, we feel this time is more than offset by the more efficient use of staff time in recruitment efforts.

Using tools already supported at many academic institutions, we developed and implemented a practical referral and recruitment framework in the context of multiple concurrent trials. This framework has shown to be a valuable tool for the management and acquisition of community-based research participants in a single academic research center, particularly in regards to efficiently prescreening large volumes of participants for multiple concurrent trials. As REDCap is already supported at over 2000 institutions in over 100 countries and statistical software programs are ubiquitous in research settings, it is our hope that other research centers will be able to leverage

common technologies to adapt and implement similar frameworks to support clinical trial recruitment efforts.

Acknowledgments

This research was supported by the United States National Institute of Mental Health grants MH095753 and MH100482 to DCM. EGL is supported by a research grant K08 MH112878 from the National Institute of Mental Health. The authors wish to thank the volunteers who have participated in research through the Center for Behavioral Technologies (CBITs). REDCap is supported at the Feinberg School of Medicine by the Northwestern University Clinical and Translational Science (NUCATS) Institute. Research reported in this publication was supported, in part, by the National Institutes of Health National Center for Advancing Translational Sciences, Grant Number UL1TR001422. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

DCM has accepted consulting fees from Optum Behavioral Health, and has an ownership interest in Actualize Therapy. EGL and SMK have accepted consulting fees from Actualize Therapy. HLP is currently employed by AbbVie, Inc. Contributions to the recruitment framework described in this manuscript were made while she was employed by Northwestern University. None of the other authors have any conflicts of interest to declare.

References

1. Madan A, Tracy S, Reid R, Henry A. Recruitment difficulties in obstetric trials: a case study and review. Aust N Z J Obstet Gynaecol 2014 Dec;54(6):546-552. [doi: 10.1111/ajo.12233] [Medline: 25350684]
2. Donovan JL, Paramasivan S, de Salis I, Toerien M. Clear obstacles and hidden challenges: understanding recruiter perspectives in six pragmatic randomised controlled trials. Trials 2014;15:5 [FREE Full text] [doi: 10.1186/1745-6215-15-5] [Medline: 24393291]
3. Ridda I, MacIntyre CR, Lindley RI, Tan TC. Difficulties in recruiting older people in clinical trials: an examination of barriers and solutions. Vaccine 2010 Jan 22;28(4):901-906. [doi: 10.1016/j.vaccine.2009.10.081] [Medline: 19944149]
4. McDonald AM, Knight RC, Campbell MK, Entwistle VA, Grant AM, Cook JA, et al. What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. Trials 2006;7:9 [FREE Full text] [doi: 10.1186/1745-6215-7-9] [Medline: 16603070]
5. Treweek S, Lockhart P, Pitkethly M, Cook JA, Kjeldstrøm M, Johansen M, et al. Methods to improve recruitment to randomised controlled trials: Cochrane systematic review and meta-analysis. BMJ Open 2013;3(2):e1 [FREE Full text] [doi: 10.1136/bmjopen-2012-002360] [Medline: 23396504]
6. Watson JM, Torgerson DJ. Increasing recruitment to randomised trials: a review of randomised controlled trials. BMC Med Res Methodol 2006 Jul 19;6:34 [FREE Full text] [doi: 10.1186/1471-2288-6-34] [Medline: 16854229]
7. Kadam RA, Borde SU, Madas SA, Salvi SS, Limaye SS. Challenges in recruitment and retention of clinical trial subjects. Perspect Clin Res 2016;7(3):137-143 [FREE Full text] [doi: 10.4103/2229-3485.184820] [Medline: 27453831]
8. Huynh L, Johns B, Liu S, Vedula SS, Li T, Puhan MA. Cost-effectiveness of health research study participant recruitment strategies: a systematic review. Clin Trials 2014 Oct;11(5):576-583. [doi: 10.1177/1740774514540371] [Medline: 24980278]
9. Köpcke F, Prokosch H. Employing computers for the recruitment into clinical trials: a comprehensive systematic review. J Med Internet Res 2014 Jul 01;16(7):e161 [FREE Full text] [doi: 10.2196/jmir.3446] [Medline: 24985568]
10. Lane TS, Armin J, Gordon JS. Online recruitment methods for web-based and mobile health studies: a review of the literature. J Med Internet Res 2015;17(7):e183 [FREE Full text] [doi: 10.2196/jmir.4359] [Medline: 26202991]
11. Graham AL, Milner P, Saul JE, Pfaff L. Online advertising as a public health and recruitment tool: comparison of different media campaigns to increase demand for smoking cessation interventions. J Med Internet Res 2008;10(5):e50 [FREE Full text] [doi: 10.2196/jmir.1001] [Medline: 19073542]
12. Laws RA, Litterbach EV, Denney-Wilson EA, Russell CG, Taki S, Ong K, et al. A comparison of recruitment methods for an mHealth intervention targeting mothers: lessons from the Growing Healthy Program. J Med Internet Res 2016 Sep 15;18(9):e248 [FREE Full text] [doi: 10.2196/jmir.5691] [Medline: 27634633]
13. Cuggia M, Besana P, Glasspool D. Comparing semi-automatic systems for recruitment of patients to clinical trials. Int J Med Inform 2011 Jun;80(6):371-388. [doi: 10.1016/j.ijmedinf.2011.02.003] [Medline: 21459664]
14. Rose GL, Ferraro TA, Skelly JM, Badger GJ, MacLean CD, Fazzino TL, et al. Feasibility of automated pre-screening for lifestyle and behavioral health risk factors in primary care. BMC Fam Pract 2015 Oct 23;16:150 [FREE Full text] [doi: 10.1186/s12875-015-0368-9] [Medline: 26497902]
15. Torgerson JS, Arlinger K, Käppi M, Sjöström L. Principles for enhanced recruitment of subjects in a large clinical trial. The XENDOS (XENical in the prevention of Diabetes in Obese Subjects) study experience. Control Clin Trials 2001 Oct;22(5):515-525. [Medline: 11578785]

16. Stewart TM, Bhapkar M, Das S, Galan K, Martin CK, McAdams L, CALERIE Study Group. Comprehensive Assessment of Long-term Effects of Reducing Intake of Energy Phase 2 (CALERIE Phase 2) screening and recruitment: methods and results. Contemp Clin Trials 2013 Jan;34(1):10-20 [FREE Full text] [doi: 10.1016/j.cct.2012.08.011] [Medline: 22981898]
17. Lattie E, Kaiser S, Alam N, Tomasino K, Sargent E, Rubanovich C, et al. A practical, do-it-yourself recruitment support framework for concurrent eHealth clinical trials: part two. J Med Internet Res 2018:e1 (forthcoming). [doi: 10.2196/11050]
18. 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]
19. 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]
20. 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]
21. Friedman L, Furberg C, DeMets D. Fundamentals of Clinical Trials. New York: Springer; 2010.
22. Chang W, Cheng J, Allaire J, Xie Y, McPherson J. The Comprehensive R Archive Network. shiny: web application framework for R URL: https://cran.r-project.org/web/packages/shiny/index.html [accessed 2018-09-27] [WebCite Cache ID 72kSdq2hh]
23. Tomasino KN, Lattie EG, Ho J, Palac HL, Kaiser SM, Mohr DC. Harnessing peer support in an online intervention for older adults with depression. Am J Geriatr Psychiatry 2017 Oct;25(10):1109-1119. [doi: 10.1016/j.jagp.2017.04.015] [Medline: 28571785]
24. Lattie EG, Ho J, Sargent E, Tomasino KN, Smith JD, Brown CH, et al. Teens engaged in collaborative health: the feasibility and acceptability of an online skill-building intervention for adolescents at risk for depression. Internet Interv 2017 Jun;8:15-26. [doi: 10.1016/j.invent.2017.02.003] [Medline: 28584734]

Abbreviations

API: application programming interface
CBITs: Center for Behavioral Intervention Technologies
CTRSS: clinical trial recruitment support system
DIY-RSF: “do-it-yourself” recruitment support framework
IRB: institutional review board
REDCap: Research Electronic Data Capture

Edited by G Eysenbach; submitted 14.05.18; peer-reviewed by T Lane, M Musker; comments to author 20.06.18; revised version received 02.07.18; accepted 05.07.18; published 01.11.18.

Please cite as:
Palac HL, Alam N, Kaiser SM, Ciolino JD, Lattie EG, Mohr DC
A Practical Do-It-Yourself Recruitment Framework for Concurrent eHealth Clinical Trials: Simple Architecture (Part 1)
J Med Internet Res 2018;20(11):e11049
URL: https://www.jmir.org/2018/11/e11049/
doi:10.2196/11049
PMID:30389650

Original Paper

A Practical Do-It-Yourself Recruitment Framework for Concurrent eHealth Clinical Trials: Identification of Efficient and Cost-Effective Methods for Decision Making (Part 2)

Emily G Lattie[1], PhD; Susan M Kaiser[2], MPH; Nameyeh Alam[2], MPH; Kathryn N Tomasino[2], PhD; Elizabeth Sargent[2], BA; Caryn Kseniya Rubanovich[2], MS; Hannah L Palac[3], MPH; David C Mohr[2], PhD

[1]Center for Behavioral Intervention Technologies, Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[2]Center for Behavioral Intervention Technologies, Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[3]AbbVie Inc, North Chicago, IL, United States

Corresponding Author:
Emily G Lattie, PhD
Center for Behavioral Intervention Technologies
Department of Medical Social Sciences
Northwestern University Feinberg School of Medicine
10th Floor
750 North Lake Shore Drive
Chicago, IL, 60611
United States
Phone: 1 3125033741
Email: emily.lattie@northwestern.edu

Abstract

Background: The ability to successfully recruit participants for electronic health (eHealth) clinical trials is largely dependent on the use of efficient and effective recruitment strategies. Determining which types of recruitment strategies to use presents a challenge for many researchers.

Objective: The aim of this study was to present an analysis of the time-efficiency and cost-effectiveness of recruitment strategies for eHealth clinical trials, and it describes a framework for cost-effective trial recruitment.

Methods: Participants were recruited for one of 5 eHealth trials of interventions for common mental health conditions. A multipronged recruitment approach was used, including digital (eg, social media and Craigslist), research registry-based, print (eg, flyers and posters on public transportation), clinic-based (eg, a general internal medicine clinic within an academic medical center and a large nonprofit health care organization), a market research recruitment firm, and traditional media strategies (eg, newspaper and television coverage in response to press releases). The time costs and fees for each recruitment method were calculated, and the participant yield on recruitment costs was calculated by dividing the number of enrolled participants by the total cost for each method.

Results: A total of 777 participants were enrolled across all trials. Digital recruitment strategies yielded the largest number of participants across the 5 clinical trials and represented 34.0% (264/777) of the total enrolled participants. Registry-based recruitment strategies were in second place by enrolling 28.0% (217/777) of the total enrolled participants across trials. Research registry-based recruitment had a relatively high conversion rate from potential participants who contacted our center for being screened to be enrolled, and it was also the most cost-effective for enrolling participants in this set of clinical trials with a total cost per person enrolled at US $8.99.

Conclusions: On the basis of these results, a framework is proposed for participant recruitment. To make decisions on initiating and maintaining different types of recruitment strategies, the resources available and requirements of the research study (or studies) need to be carefully examined.

(J Med Internet Res 2018;20(11):e11050) doi:10.2196/11050

KEYWORDS

eHealth; mHealth; mental health; recruitment

Introduction

Background

Recruiting participants into electronic health (eHealth) intervention efficacy trials has long been a challenge [1,2]. Although internet access has become increasingly widespread and the digital divide has narrowed in recent years [3], difficulties remain in reaching individuals who are both representative of the target population and interested in taking part in these trials [4,5]. There are ever increasing ways of recruiting, from older, more traditional methods such as mailing or public print advertising, to newer methods such as social media, and resources such as registries and marketing firms, and each method comes with a set of costs and benefits.

In recent years, difficulties associated with developing and testing new eHealth programs under traditional research grant timelines have been identified [6-8]. Given the focus of the National Institute of Mental Health on information technologies for social and behavioral health [9] and the increase in health researchers who are now capitalizing upon the widespread adoption of personal technologies in attempts to expand the reach and accessibility of behavioral interventions, it is increasingly important that researchers choose efficient recruitment strategies to maximize their research funds and timelines and hit recruitment targets to allow for robust evaluation of program quality, efficacy, and effectiveness. Time and costs required to design and program technologies, as well as the unanticipated, albeit inevitable development problems, often squeeze out time and resources intended for trial recruitment. Although past reviews [10,11] have highlighted the value of using Facebook and other social media methods for health research recruitment, there have been few studies reporting on the efficiency of these recruitment methods relative to other recruitment methods for health intervention research. Thus, there remains a need to examine the costs and benefits of multiple methods of recruitment to identify those methods that are likely to be efficient and cost-effective.

Objectives

The Center for Behavioral Intervention Technologies (CBITs) at Northwestern University recently completed enrollment for 5 simultaneous clinical trials of eHealth interventions for common mental health conditions (ie, depression and anxiety). To support this enrollment effort, CBITs developed a clinical trial recruitment support system [12] and a set of recruitment methods that were flexible to the target populations required for each of the trials. This paper presents descriptive information regarding the recruitment strategies employed by CBITs during a nationwide recruitment for eHealth clinical trials, the efficiency of these strategies in producing referred and enrolled participants, and the estimated cost of using these strategies. Given the diverse set of responsibilities needed to successfully employ these strategies, we provide a description of the roles and relevant expertise of our research study staff.

The aim of this paper was to propose a decision framework for cost-effective trial recruitment. To support this aim, we describe a set of procedures that were used to recruit and enroll participants across 5 trials using a *do-it-yourself* (DIY) recruitment support framework described in the companion paper [12]. We then analyze cost-effectiveness of the recruitment strategies used. Finally, these data, along with lessons learned, are used to propose a framework for recruitment decision making.

Methods

Study Descriptions

During the recruitment period reported on in this paper, we conducted 3 trials for adults older than 18 years and 2 trials for targeted age groups (ie, high school students, adults aged 65 years and older), all of which evaluated eHealth interventions for the treatment or prevention of common mental health conditions (ie, depression, anxiety) and included a national recruitment strategy. The companion paper by Palac et al [12] also includes a trial that was conducted exclusively in the Chicago area.

The trials for adults older than 18 years are described below:

Stepped Care Randomized Controlled Trial

The Stepped Care randomized controlled trial (RCT) recruited adults older than 18 years who were currently experiencing a depressive episode. Through random assignment, the study compared up to 20 weeks of (1) a telephone-administered cognitive behavioral therapy (T-CBT) and (2) a stepped care intervention that initiated treatment with a coached internet CBT program called ThinkFeelDo, stepping those participants who did not show improvement up to T-CBT (outcome paper currently under review). Follow-up assessments were administered by phone and Web-based questionnaire up to 2 times during the 20-week treatment period and at 3 and 6 months post treatment.

IntelliCare Field Trial

The IntelliCare Field Trial evaluated a suite of 13 Android apps with adults older than 18 years with symptoms of anxiety, depression or both [13]. Of these, 12 apps provided different clinical therapy skills for treating anxiety and/or depression, and 1 app, named the IntelliCare Hub, served as a central place to manage the other apps. All participants used the apps for 8 weeks and were provided with access to a coach via SMS text messaging (short message service, SMS). Participants completed Web-based questionnaires assessing symptom change and provided user feedback about the apps at 4 and 8 weeks into the study.

IntelliCare Randomized Controlled Trial

This RCT continued the evaluation of the IntelliCare platform using a 2×2 factorial design in which participants were randomized to receive (1) coaching or no coaching and (2) automatic weekly recommendations versus no automated

recommendations [14]. Participants were asked to use the apps for 8 weeks, completing 2 Web-based questionnaires during the active app use study period, and again 3 and 6 months after the end of the 8-week active app use period (primary outcome paper is currently in preparation).

The 2 trials for targeted age groups, which both utilized a group social networking component, are described below:

ProjectTECH Field Trial

ProjectTECH tested an online and Web-app based group intervention for the prevention of teenage depression and substance use disorders [15]. Youth in the age group of 14 to 19 years were placed into peer groups and provided an adapted, responsive version of ThinkFeelDo that was available on phones with age-appropriate content and was embedded in an activity feed that supported communication among group members. The peer groups were facilitated by either a clinical psychologist or a high school student peer guide. Participants were asked to use the Web platform for 8 weeks. They were sent online questionnaires at 4, 8, and 12 weeks after beginning the study.

MoodTech Field Trial

MoodTech adapted the ThinkFeelDo program for the treatment of depression among adults older than 65 years [16]. All users had the support of the same clinical psychologist to coach them on how to use the website. Participants were assigned to 1 of 3 groups and either had access to a version of the website they could use independently, a version of the website that included peer support features and as well as an online space where they could interact with a small group of their peers or they were assigned to a wait list control group. Both versions of treatment were 8 weeks long. Follow-up assessments were administered by phone and online questionnaires at 4, 8, and 12 weeks after starting to use the website. Participants placed on the wait list before using the website; they completed 2 additional assessments during the waiting period and then had access to the independent version of the site.

Recruitment Strategies

Participants were recruited for these trials using a multipronged approach, including digital (eg, social media and Craigslist), research registry-based, print (eg, flyers and posters on public transportation), clinic-based (eg, a general internal medicine clinic within an academic medical center and a large nonprofit health care organization), a market research recruitment firm, and traditional media strategies (eg, newspaper and television coverage in response to press releases). Participants self-reported their recruitment source on an initial online screening survey, and recruitment source was clarified when contact was made with study staff.

For each recruitment strategy, the research team prepared verbiage (and in some cases images) with target populations in mind. All verbiage and images were approved by the institutional review board (IRB) before use and all online advertisements directed to full study information on a separate website and/or the online study prescreen survey. An example of these recruitment advertisements can be seen in Multimedia Appendix 1.

For social media ads including Facebook, Instagram, and Twitter, a bank of IRB-approved verbiage and images was used to create brief teaser ads. This allowed staff members posting ads to mix and match the headlines, body text, and associated photos of advertisements, thus avoiding repetitive content. Messages were targeted toward the group of interest for each study (eg, an older adult sample, a teen sample, and a general adult sample) and on criteria including but not limited to geographic location, gender, and those who had indicated interest in any number of relevant keywords such as “depression” and “anxiety.” We typically ran social media ads for a month at a time but were more interested in how the ads were performing while adhering to budgetary constraints. Once an ad set started running, we would let it run for several days to 1 week before checking in the progress. Ads that were not performing well were turned off, and resources were reallocated to better performing ads. We typically used the standard delivery type (which spreads out the ads throughout a selected campaign schedule); however, when we were trying to recruit quickly, we would turn on the accelerated delivery type.

For research registry-based recruitment effort, invitations were crafted for various research registries with verbiage directed toward individuals interested in helping further research knowledge. With some registries such as ResearchMatch, the research team could target invitations based on age, race/ethnicity, previous diagnoses reported, and area of residence in the United States.

For print-based recruitment, flyers and posters were designed by research staff members and printed through companies that had partnerships with the research team’s university. Flyers were placed in various businesses with community boards such as coffee shops and on various university boards in common areas and in medical office waiting rooms with medical staff permission. Research staff utilized their university’s partnership with the Chicago Transit Authority to purchase flyer space on buses and trains at a discounted rate. Train and bus routes that were generally busy and ran close to the university were chosen for advertising.

For clinic-based recruitment, research staff partnered with physicians to refer patients by providing information about the studies. For 1 study, the research team partnered with a large nonprofit health care organization that orchestrated referrals from their clinics.

For market research recruitment firm-based recruitment, the research team worked with a research recruitment firm that was able to recruit interested volunteers from across the United States. Email invitations were first sent internally by the research recruitment firm to participant panelists. The email contained a link to a study screener that was adapted for and hosted on the market research firm’s site. Research staff had to develop unique recruitment verbiage and a separate online screener for individuals from the market research firm.

For traditional media strategy-based recruitment, the research team’s university media relations department typically wrote and released an article about results of studies previously conducted within our research center and provided contact information (eg, email, website, and phone) for those interested

in enrolling in an ongoing study. The research team would then receive calls or emails from interested participants. Once an online prescreening system was implemented, interested volunteers were directed to the research website and online prescreening link.

Staff Roles and Expertise

Recruitment efforts were conducted under the leadership of an MPH-level research manager (SMK) with experience in community mental health and clinical trial management. This individual managed a team of research staff for the clinical trials unit (CTU), which was composed of bachelor's and master's level staff. A total of 23 individuals supported study recruitment over the recruitment period, including 3 staff members from Northwestern Clinical and Translational Sciences Institute who specialized in study screening and were brought on when our team exceeded capacity to manage recruitment and clinical interviews. At the peak of recruitment, there was a core team of 10 CTU staff members supporting recruitment efforts. Most of the research staff members had primary roles as clinical interviewers or technology support specialists on these trials and managed specific recruitment strategies as a smaller component of their work week. For a 6-month period, a digital marketing manager worked with the CTU on developing a robust social media strategy focused on Instagram, Facebook, and Twitter.

The recruitment strategies mentioned above directed participants to a centralized online screening survey to be prescreened for the center's actively recruiting clinical trials. The online survey was used to automate initial eligibility decisions, eliminating individuals who would be ineligible for all studies, and allowing research assistants more time to interact with potentially eligible participants and confirm eligibility via a brief phone screener.

A master's level data manager, experienced in programming language for data wrangling, managed the back-end automation and routing of potential participants from various recruitment sources through this centralized online screening survey. For the back-end automation, programming code was written to automatically screen participants for entry into the center's active clinical trials and route to a study based on specific study eligibility, participants' preferred study choice, and the center's recruitment targets for each active study. Code was updated as new recruitment sources were added and center recruitment targets changed. New referrals were processed daily and based on the number of eligible participants received. The data manager notified team members to increase or decrease recruitment efforts, particularly on digital strategies such as social media ads and research registry pulls. This is described in further detail in the study by Palac and colleagues [12] and was a critical and cost-efficient contributor to the success of the recruitment strategies described in this paper. As the data component of the framework described by Palac et al [12] was based on technologies already supported by our university (and commonly found at other universities), there were no additional technology costs to maintain this support system.

Recruitment Process

Potential participants could contact the center via email, telephone, our Web screening survey, or from an in-app interest form. The IntelliCare apps were publicly available on the Google Play Store [17], and people who had already downloaded an IntelliCare app could complete a form within the app that indicated their interest in participating in relevant research projects. These potential participants are labeled as "contacted" throughout this paper. Then, all potential participants went through a brief screening measure and, if initially eligible, were phone-screened by a research assistant. These potential participants are labeled as "screened" throughout this paper. Finally, eligible potential participants who passed the 2-stage screening and enrolled in 1 of the 5 clinical trials described above are labeled as "enrolled" throughout this paper.

Data Analysis

Descriptive statistics were computed to characterize (1) the number of potential participants labeled as "contacted," (2) the number of those potential participants labeled as "screened," and (3) the number of participants from each recruitment source labeled as "enrolled." To highlight differences between the intervention trials included in this study, the demographics of participants and the number of enrolled participants from each recruitment source by trial were also computed. Per-participant costs were calculated for each recruitment method based on a ratio of participant yield to expenditures. The time costs of each recruitment method were calculated based on objective review of study records (eg, meeting minutes) and through estimates made in consultation with study staff regarding the time study staff members spent on the launch and maintenance of each research strategy while it was being utilized. Time estimates were then converted to time costs by multiplying hours spent by the relevant hourly wage (eg, US $17.50 for research assistant time and US $30.17 for research manager time). Fees for each of the recruitment methods were calculated based on billing records. Then, the participant yield on recruitment costs was calculated by dividing the number of enrolled participants by the total cost for each method. This analytic method allows for the identification of methods that were particularly cost efficient and time efficient for recruiting eligible participants, while providing transparency into the inner and outer system fees associated with each set of recruitment methods. Results from these analyses were then used to outline a framework for recruitment decision making in the Discussion section.

Results

As shown in Table 1, there was considerable variability in the staff skills and time required to establish and maintain the recruitment strategy.

Table 1. Summary of recruitment strategies

Recruitment strategy	Digital	Registry	Print	Clinic	Firm	Media
Topperforming sites	Instagram, Reddit, Craigslist	ResearchMatch	Ads on Chicago Transit Authority bus and train lines	Health partners, Group Health	Focus Pointe Global	Unable to be determined
Techniques	Social media marketing, content marketing, direct email, eNewsletters, app advertising, study description on website, blog posts	Email (direct and through Web portal) to registry participants	Approximately 800 study-specific banner ads were placed on 2 of Chicago's busiest train lines and on 18 bus routes.	Invitation mailed via United States Postal Service, email invitation sent via electronic medical record portal, phone call from research assistant	Email invitation through firm	Planned press release, reprints
Target population	US general public (adults and adolescents)	Registry participants (adults)	Chicago general public (adults)	Individuals engaged in care systems (adults)	Market research firm panelists (adults)	US general public (adults)
Staff skills required for startup/management	Social media marketing, analytics, design, public relations (crisis response), REDCap [18]	Human subjects recruitment	Design	Project management, relationship management, stakeholder management, database management, human subjects recruitment	Project management, database management, clinical trials recruitment	Public relations, journalism
Management effort	Daily management	Weekly management	Monthly management	Weekly management	Weekly management	As needed (but labor intensive during initial media blitz)
Resource considerations	Nearly infinite in terms of reaching new potential participants	Finite number of registry participants	Cost prohibitive. University discount made it possible to advertise broadly	Finite number of patients	Finite number of participants	Nearly infinite in terms of reaching new potential participants

Table 2. Potential participants by recruitment source.

Recruitment strategy	Digital	Registry	Print	Clinic	Firm	Media	Other	Unknown	Unknown (IntelliCare) app/ Web form	Total
Raw numbers (n)										
Contacted	3318	2030	789	3261	290	297	33	472	6727	17,217
Screened	895	627	308	266	138	144	9	33	86	2506
Enrolled	271	225	89	75	55	49	3	6	4	777
Outcomes (%)										
Percent of patients contacted that were screened	26.97	30.89	39.04	8.16	47.59	48.48	27.27	6.99	1.28	—[a]
Percent of paitents contacted that were enrolled	8.17	11.08	11.28	2.30	18.97	16.5	9.09	1.27	0.06	—
Percent of patients screened that were enrolled	30.28	35.89	28.9	28.2	39.86	34.03	33.33	18.18	4.65	—

[a]Not applicable.

A total of 17,217 potential participants contacted the recruitment site, 2506 completed screening, and 777 were enrolled across the studies. Table 2 displays that the number of potential participants from each recruitment source who had contact with our research center during the trial enrollment period varied greatly. The largest portion of potential participants came from an unknown source (ie, the recruitment source was missing from their record, usually because of the participant's failure to respond to that query) and had contacted the research center through the IntelliCare in-app interest form (labeled IC app/Web form in Table 2). This means that these potential participants had already downloaded an IntelliCare app, but we do not know

how they first learned about the IntelliCare suite of apps. Among potential participants from a known source, the majority came from digital recruitment strategies (eg, social media and Craigslist), followed by clinic-based recruitment (eg, a general internal medicine clinic in an academic medical center and a large nonprofit health care organization), research registries (eg, ResearchMatch website), print-based advertising (eg, flyers and posters on public transportation), and media (eg, news stories prompted by press releases from our research center that included information about ongoing trial recruitment). The smallest portion of participants were recruited from "other" sources, which included recruitment sources such as being referred to our center by another research lab and learning about our center through public events.

The Outcomes section of Table 2 shows that the potential participants who contacted our center and failed to identify how they arrived at the site (both the "unknown" and the "unknown IC app/Web form") had extremely low rates of screening completion (<7% for the general unknown category and <2% for those who contacted us through the IntelliCare in-app interest form), while those who identified how they arrived at the site had substantially better rates of Web-screening completion, ranging from 8% for clinic-based recruitment to 48% for media-based recruitment and market research firms, with digital recruitment strategies yielding 27%. The strategies that yielded the highest rates of conversion from contact to screening were the use of a market research recruitment firm (48%) and the use of research registries (31%), both of which target individuals who are likely to be interested in research participation. Overall, digital recruitment strategies yielded the largest number of participants across the 5 clinical trials, with nearly 35% of the total enrolled participants coming in from digital recruitment strategies. Registry-based recruitment strategies were in second place by enrolling nearly 29% of the total enrolled participants across trials.

To highlight differences in the use and success of recruitment strategies for the different targeted trials, Table 3 presents the number of participants enrolled by each recruitment source by trial.

As seen in Table 4, enrolled participants were similar to the overall demographic of United States and largely representative of individuals seeking mental health treatment in the United States in that there was overrepresentation of women and non-Hispanic white individuals.

Table 5 displays the fees and time costs per person screened, and cost per person enrolled varied considerably across recruitment strategies. During the recruitment period for the 5 clinical trials included in this paper, a total of US $144,537.67 were spent on recruitment fees, and there was a total estimated time cost of US $19,834.59 for a combined total of US $164,372.26. The fees, which included those fees that were paid to enact and maintain the recruitment strategies, ranged from US $1 per person enrolled for research registry-based recruitment to US $1,218.33 per person enrolled for clinic-based recruitment. The time costs, or research staff hourly wages required to implement and maintain the recruitment strategies, ranged from US $8.99 per person enrolled for research registry-based recruitment to US $75.01 per person enrolled for clinic based recruitment.

Research registry-based recruitment had particularly low fees (eg, many registries were free to post in, and nominal fees amounted to US $150 total) and had an associated moderate time cost. As research registry-based recruitment had a relatively high conversion rate from potential participants who contacted our center to be screened to be enrolled, registries have presented as the most cost-effective method for enrolling participants in this set of clinical trials, with a total cost per person enrolled at US $8.99. However, these registries are typically a finite resource. As recruitment progressed, the research team exhausted the supply of registry participants such that the registries were not accumulating new potentially eligible participants at a rate that kept up with recruitment needs.

Table 3. Participants enrolled by recruitment source.

Name of trial	Stepped Care RCT, n (%)	IntelliCare RCT, n (%)	IntelliCare Field Trial, n (%)	ProjectTech Field Trial, n (%)	MoodTech Field Trial, n (%)
Digital	111 (35.6)	103 (34.2)	25 (23.8)	30 (76.9)	8 (17.0)
Registry	99 (31.7)	72 (23.9)	17 (16.2)	0 (0)	35 (74.5)
Print	35 (11.2)	41 (13.6)	10 (9.5)	0 (0)	3 (6.4)
Clinic	35 (11.2)	0 (0)	39 (37.1)	0 (0)	1 (2.1)
Firm	0 (0)	55 (18.3)	0 (0)	0 (0)	0 (0)
Media	17 (5.4)	25 (8.3)	7 (6.7)	0 (0)	0 (0)
Other	13 (4.2)	5 (1.7)	3 (2.9)	9 (23.1)	0 (0)
Unknown/in-app referral	2 (0.6)	0 (0)	4 (3.8)	0 (0)	0 (0)

Table 4. Participant demographics by trial.

Demographics	Stepped Care RCT[a] (N=312)	IntelliCare RCT (N=301)	IntelliCare Field Trial (N=105)	ProjectTECH Field Trial (N=39)	MoodTech Field Trial (N=47)
Age in years, mean (SD)	37.7 (14.2)	36.5 (11.8)	38.9 (14.1)	16.23 (0.99)	69.6 (4.1)
Gender, n (%)					
Female	229 (73.4)	228 (75.7)	80 (76.2)	29 (74)	32 (68)
Male	81 (26.0)	71 (23.6)	25 (23.8)	9 (23)	15 (31)
Other	2 (0.6)	2 (0.7)	0 (0)	1 (3)	0 (0)
Race, n (%)					
American Indian or Alaska Native	0 (0)	0 (0)	1(1)	0 (0)	0 (0)
White	275 (88.1)	237 (78.7)	88 (83.8)	24 (62)	41 (87)
African American	21 (6.7)	29 (9.6)	8 (7.6)	3 (8)	2 (4)
Asian	14 (4.5)	10 (3.3)	6 (5.7)	4 (10)	0 (0)
More than one race	8 (2.6)	18 (6.0)	1 (1)	4 (10)	3 (6)
Unknown/declined to report	0 (0)	7 (2.3)	1 (1)	4 (10)	1 (2)
Ethnicity, n (%)					
Hispanic or Latino	32 (10.3)	30 (9.9)	5 (4.8)	10 (26)	1 (2)
Not Hispanic or Latino	275 (88.1)	268 (89.0)	99 (94.3)	29 (74)	46 (98)
Hispanic or Latino—unknown or not reported	5 (1.6)	3 (1.0)	1 (1)	0 (0)	0 (0)

[a]RCT: randomized controlled trial.

Table 5. Fees and time costs for recruitment strategies (in USD).

Recruitment strategy	Digital	Registry	Print	Clinic	Firm	Media	Total
Fees	$11,726.01	$150.00	$9318.66	$91,375.00	$31,968.00	$0.00	$144,537.67
Fees per person screened	$13.10	$0.24	$30.26	$343.52	$231.65	$0.00	
Fees per person enrolled	$43.27	$1	$104.70	$1218.33	$581.24	$0.00	
Time cost	$8601.25	$1872.50	$761.53	$5767.52	$1896.52	$935.27	$19,834.59
Time cost per person screened	$9.61	$2.99	$2.47	$21.68	$13.74	$6.49	
Time cost per person enrolled	$31.74	$8.32	$8.56	$76.90	$34.48	$19.09	
Total cost	$20,327.26	$2022.50	$10,080.19	$97,142.52	$33,864.52	$935.27	$164,372.26
Total cost per person screened	$22.71	$3.23	$32.73	$365.20	$245.40	$6.49	
Total cost per person enrolled	$75.01	$8.99	$113.26	$1295.23	$615.72	$19.09	

Discussion

Principal Findings

Results from this set of 5 eHealth intervention trials focused on common mental health problems (ie, depression and anxiety) indicate that use of digital recruitment strategies (eg, Facebook, Instagram, and Craigslist) and research registry-based recruitment strategies (eg, ResearchMatch) were the most fruitful, time-efficient, and cost-effective methods for recruiting a nationwide sample of participants who were largely representative of the populations of interest. These results add to the literature on clinical trial recruitment methods and the benefits of technology-enabled recruitment strategies. Findings are partially consistent with systematic review results recently reported by Whitaker et al [10] on the topic of using Facebook to recruit participants for health research purposes. Whitaker et al [10] found growing evidence that, when compared with traditional recruitment methods (eg, print, radio, and email), Facebook recruitment had multiple benefits including lower costs and shorter recruitment periods. However, that review only included 1 study focused on mental health and did not examine the utility of other digital recruitment methods such as Instagram and Craigslist. These results also partially support findings of a scoping review by Topolovec-Vranic and Natarajan [11] in which digital recruitment strategies (eg, Facebook and Craigslist) were compared with other recruitment strategies for medical research study recruitment. Of the 30 studies included in their review, 12 studies found that digital strategies were more effective than other methods, and an additional 3 studies

found that digital strategies were equally effective as another recruitment strategy. However, only 10 of the 30 studies were on behavioral interventions, and none of them were on eHealth interventions for common mental health problems. Although these studies provide support for the use of digital strategies for medical/health-related study recruitment, they do not reflect the unique nature of recruiting participants with common mental health problems for eHealth interventions. Thus, the results presented in this paper contribute to the broader literature by honing in on this population for eHealth intervention research and by examining additional recruitment strategies (eg, Instagram and ResearchMatch).

Results of analyses, combined with research staff experiences, have been used to develop a framework for recruitment strategy decision making for eHealth interventions depicted in the questions to guide strategic decision making presented in Table 6 and the matrix of recruitment strategy benefits presented in Table 7. In Table 7, we have highlighted the recruitment strategies that offer primary benefits of low fees, a high degree of control over the number and flow of referrals being directed to research staff, access to large numbers of people, access to targeted populations (eg, with specific clinical diagnoses and with specific demographic profiles), and 2 benefits associated with easier management/maintenance of the recruitment strategy (ie, a lack of specialized skills needed and a relatively low burden/time effort for study staff).

Using a variety of recruitment strategies is recommended, and the tools presented in Tables 6 and 7 are intended to help researchers determine the best subset of strategies to use for a particular study or set of studies. To efficiently manage multiple strategies, we recommend implementing a recruitment support framework as described by Palac et al [12], which is structured around an online screening survey and a central tracking database overseen by a data manager. To make decisions on initiating and maintaining different types of recruitment strategies, careful examination of the resources available (ie, budget, staff, relationships, and discounts) and requirements of the research study (ie, target recruitment number, target participant flow/timeline, and entry criteria) is essential. However, before reviewing the Table 6 question set and Table 7 matrix to determine one's optimal recruitment strategies, one should conduct a literature review to determine if there are relevant studies that suggest what the outcomes or conversion rates for screening to enrollment could be for one's target population using recruitment strategies that may already be under consideration. Early identification of conversion rate estimates for screening to enrollment will help the research team make appropriate time-cost and fee-related investments from the beginning of a trial. If there are no estimates available, then researchers will need to experiment with their selected set of recruitment strategies to fine-tune their approach.

Throughout the question set in Table 6, one is prompted to consider the existing resources and requirements for a specific study. These resources include available funds (ie, the budget), staff expertise, staff effort, existing relationships, and access to discounts. As our research center was concurrently recruiting for multiple clinical trials, we were afforded some flexibility using recruitment funds to test multiple recruitment strategies and to start and stop the use of those strategies as needed.

Table 6. Questions to guide strategic decision making for recruitment.

Topics	Questions
Resources	
Budget	Do you have a budget for paid advertising? Do you have a budget to support staff to manage the strategy?
Staff expertise	Can you recruit or train staff to learn skills required to set up/manage this strategy?
Staff effort	Do you have staff who will be available to establish/manage this strategy?
Relationships	Do you have relationships to establish this strategy?
Discounts	Do you have or can you make connections to reduce the overall cost of this strategy?
Requirements	
Target (N)	How many people do you need to recruit overall (<100, >100)?
Flow/timeline	How quickly do you need to enroll subjects (months, years)? Do you have enough time to experiment?
Entry criteria	How stringent are your entry criteria (ie, how targeted do you need to be with your advertising?)

Table 7. Matrix of recruitment strategy benefits.

Benefits	Digital	Registry	Print	Clinic	Firm	Media
Low fees		✓				✓
High degree of control (can control number and flow of referrals)	✓	✓			✓	
Broad reach (access large numbers of people)	✓				✓	✓
Access to a targeted population	✓			✓		
No specialized skills required for maintenance/management		✓	✓			✓
Low effort required for maintenance/management		✓	✓			

Furthermore, most research staff members had a primary role as a clinical interviewer or as a technology support specialist on these trials and managed specific recruitment strategies as a smaller component of their work week. As research staff employed in a primary capacity for clinical interviewing typically had times during the workday in which no interviews were taking place, there was bandwidth to develop specialized skills and to manage more time-intensive recruitment strategies. Thus, the capacity for recruitment strategies requiring specialized skills (such as digital, clinic, and firm-based strategies) and higher levels of effort for management (such as those needed to maintain digital, clinic, firm, and media-based strategies) was already built into the structure of the research team. As seen in Table 6, study requirements include the target sample size, the target flow/timeline of participants getting screened and enrolled in the study, and the study's entry criteria, which can all be assessed to determine which recruitment strategies are most likely to be fruitful. Studies requiring a large sample size will need to utilize strategies capable of tapping into large numbers of potential participants, and for studies that have a limited timeline for recruitment, it will be important to pick a few recruitment strategies and monitor their success closely so that the research team can adjust the strategies as needed. Studies with stringent entry criteria need to be more targeted in their advertising (relative to studies that are recruiting a general adult sample), and this can increase the fees associated with certain types of recruitment (eg, online advertisements) and increase the time necessary to develop and design appropriate recruitment advertisements.

As identified in our results, the cost-effectiveness and time-efficiency of the recruitment strategies employed varied significantly, with digital and registry-based recruitment strategies demonstrating the greatest degree of cost-effectiveness and time-efficiency. This was likely because of the ability of our research team to control the number and flow of referrals using these 2 strategies, and thus, we were able to get large numbers of potentially eligible participants into our studies in a relatively efficient manner. However, many of the costs presented in this paper are dependent on multiple factors and thus can be estimated differently based on resources available in different research settings. For example, the expertise that staff members already possess (eg, social media expertise) can contribute to certain recruitment efforts (eg, digital strategies) in ways that reduce the need for hiring outside consultants or contractors. Alternately, a lack of these types of internal expertise would not preclude a research team from undertaking these types of recruitment strategies but could increase the costs of engaging in these strategies, as it may be a less efficient use of a staff member's time. Similarly, the existing state of relationships with clinics and health care systems can dramatically impact the time costs and fees associated with clinic-based recruitment. Building new relationships takes significant time, and strong existing relationships may come with reduced fees within certain clinics and health care systems.

Furthermore, recruitment-associated fees can vary depending on existing institutional relationships and access to support such as discounts. For example, the price that our research center paid for recruitment advertisements on public transportation was at a reduced cost because of an arrangement previously established by our Northwestern University's Clinical and Translational Sciences Institute with the public transportation service. Recruitment-associated fees can also vary depending on changing advertising fee structures and the popularity of keywords used in the advertisements [19,11]. One recent systematic review on the cost of recruiting for research studies using Facebook found that researchers paid between US $1.36 and US $110 per completing participant [20]. Although the majority of studies (80%) included in this review were cross-sectional surveys, and, thus, those ad clicks were more likely to convert to active study participation compared with intervention studies that last several weeks to months, findings by Thornton et al [20] demonstrate the broad range of fees that can be applied to use of a single digital recruitment strategy.

For research studies with a limited staff that are targeting fee-related cost-efficiency, reliance on registry-based and media-based strategies as primary recruitment efforts could prove to be both realistic and successful to hit recruitment targets, provided that the research registries utilized include a feasible number of potential participants (see Table 7). Print strategies may also be considered for these cases if the research team is able to locate low-cost print outlets that are likely to reach their target population. The use of digital recruitment strategies (eg, Facebook, Twitter, and Instagram) can also be feasible for studies with limited staff if the study team contains at least one individual with a firm understanding of digital marketing, or if there is support for a study team member to develop this expertise. The use of these strategies requires initial management decisions (eg, reliance on paid ad campaigns vs time developing more robust but free Web presence) but can be designed to require less staff time than was used by our group while still allowing researchers to draw from a very large number of potential participants and exert a high degree of control over the flow of potential participants from targeted populations.

Our personnel cost estimates are pulled from a private university in a large Midwestern city and may not accurately reflect pay rates in other areas of the United States or in other cities around the world conducting similar research. Indeed, clinical research costs are largely driven by personnel costs, and these costs can be substantially lower or higher in other locations where similar research could feasibly take place [21]. Some researchers may struggle with personnel-related decisions because of financial costs, and we note that having an experienced research manager can be more costly upfront but has the potential to save money over time because of skill at managing other research staff time and at negotiating relationships with new recruitment partners. This was particularly important during the set of trials used in this paper, as our experienced research manager was key in negotiating and navigating relationships and keeping recruitment targets on track to ensure that money was being well spent. This tracking system is further described in our companion piece by Palac et al [12].

We found that digital and research registry-based recruitment strategies brought in a faster flow of participants than other strategies examined and that this can be particularly useful for studies with a limited recruitment timeline. This is partially

consistent with past review papers on using social media for research recruitment [10,11]. Although digital strategies can be designed to tap into a growing audience through slight shifts in targeting, strategies such as clinics and research registries may limit recruitment efforts as they tend to have a relatively fixed number of potential participants. Not surprisingly, the recruitment strategies used and their relative success varied by target population. Although recruitment using digital and research-registry based strategies were similarly successful in our studies of general adult samples, some differences were noted in our studies focused on specific age groups, as seen in Table 3. In our ProjectTECH study of high school students [15], the vast majority of participants were recruited through Instagram, a social media platform that was particularly popular with teenagers during the recruitment period. In the MoodTech study for older adults [16], recruitment via digital platforms was less successful, and the vast majority of participants were recruited through the ResearchMatch registry.

To our knowledge, the time-efficiency and cost-effectiveness of research registry-based recruitment for eHealth interventions has not previously been reported upon and compared with other methods of recruitment such as digital strategies and more traditional strategies such as clinic-based recruitment and print advertisements. Results of this study suggest that, as the most cost-effective method of recruitment that also yielded a high percentage of eligible participants, researchers should strongly consider strategies such as the ResearchMatch registry to identify individuals who are likely to be interested and eligible for their eHealth intervention studies. The use of research registries appears to be far more efficient and inexpensive compared with print advertisements, recruitment firms, and clinic-based strategies. However, given that research registries are typically drawing from a finite group of potential participants, the use of supplementary recruitment strategies is valuable.

Research Considerations

An issue that emerges here is the denominator problem, as previously discussed by Mohr et al [22]. The denominator problem notes that most eHealth interventions recruit from very large pools of potential participants, and thus those individuals who choose to participate in an eHealth program are likely uniquely motivated. Although this paper focuses on recruitment for early efficacy trials of eHealth programs, we note that the time-efficient and cost-effective recruitment strategies discussed in this paper may further contribute with regard to testing eHealth interventions on the select group of individuals in the general population who are likely to engage and benefit from these interventions. A broader use of recruitment strategies produces the possibility of a wider range of participants, but this does not necessarily solve the denominator problem. As the goal of eHealth program development is to ultimately have the potential for larger scale implementation and public health benefits, an exclusive focus on recruiting for efficacy trials is likely to have a detrimental impact on the potential for developing programs to be successfully implemented. Thus, researchers may be wise to consider at least preliminary assessment of implementation factors in early evaluations of eHealth interventions following the guidelines for type 1 hybrid trials described by Curran et al [23].

Although findings indicate that clinic-based recruitment strategies were expensive and inefficient in this set of trials, we do not conclude that researchers avoid partnerships with clinical care settings when evaluating eHealth interventions for common mental health problems. Rather, the data presented here demonstrate that digital and research-registry recruitment strategies are efficient and relatively inexpensive for enrolling participants in these types of studies. For researchers focused on bringing their eHealth programs into clinical practice settings, the additional time and effort needed to enroll participants from a clinical practice setting is vital and will come with valuable insights into barriers and facililators to larger scale program implementation. To maximize time-efficiency and cost-effectiveness, the strategies described in this paper should be used in tandem with clinical trial recruitment support systems focused on prescreening referral (as described by Palac et al [12]).

Limitations

The examination of recruitment strategy efficiency and cost-effectiveness and the resulting decision-making framework presented in Tables 6 and 7 are not without limitations. This was based on a limited number of clinical trials of eHealth interventions for common mental health conditions. However, the recruitment principles listed within this paper are likely generalizable to clinical trials focusing other types of digital behavior change and health interventions. Furthermore, the time spent on various recruitment efforts was not closely tracked during these trials, and thus the time costs of many strategies were estimated through a combination of objective review of study records (eg, meeting minutes) and through estimates made in consultation with study staff regarding the time study staff members spent on the launch and maintenance of each research strategy while it was being utilized.

Another limitation of this study is the large percentage of potential participants who came from unknown sources and after completing an in-app research interest form, did not proceed with screening. Media-based recruitment, in which press releases from our research center included information about ongoing trial recruitment, initially appear to be a relatively low-cost recruitment strategy. Yet, many of these “unknown” participants contacted our research center following periods of media coverage, and although we can hypothesize that a sizeable portion of these individuals learned about our trials and downloaded 1 or more IntelliCare apps through media coverage, we cannot substantiate this hypothesis. Although it is clear that media coverage generated a small stream of referrals who went on to complete screening and enroll in the study, the influx of potential participants (many of which are labeled as being from unknown sources) contacting our research staff following press releases required a fairly high management effort by study staff.

This decision-making framework is less relevant if it is important for an intervention to be tested within a specific clinic. In those cases, the recruitment strategies will have to be focused within the clinic, and recruitment timelines and budgets will have to be established to account for a potentially slow

recruitment speed/low recruitment yield and to account for what could be substantial time costs and fees associated with establishing the clinic relationship, navigating acceptable referral methods, and advertising to clinic patients (eg, mailing study advertisments to all potentially eligible patients within clinic can be high cost and low yield, while personnel time required to conduct daily chart review and identify potentially eligible participants to approach may be more fruitful). These barriers to quick recruitment in clinic settings have been well documented previously and must be planned around [24,25]. For a review of best practices in study site selection and recommendations to plan efficient recruitment efforts in these clinical contexts, see Huang et al's' recent paper on the Clinical Trials Transformation Initiative [26]. More commonly, eHealth interventions are being tested for efficacy and can draw from a broader pool of potential participants. In these cases, the framework can be used to evaluate the resources available and the requirements (ie, main aims and constraints) of the study.

Despite these weaknesses, in tandem with the system described by Palac et al [12], this is the first framework for designing and monitoring recruitment efforts for eHealth clinical trials. This framework can be used by fellow researchers to make recruitment decisions at the outset of an eHealth clinical trial to target a set of efficient and cost-effective recruitment efforts and can be used as recruitment needs and priorities may shift over the course of a clinical trial.

Conclusions

In our study, digital and research registry-based recruitment strategies are more efficient and cost-effective for engaging potential participants in trials evaluating eHealth interventions aimed at common mental health problems (ie, depression and anxiety) when compared with traditional recruitment strategies such as print-based advertisements and recruitment from within clinical care systems. These results also demonstrate how a DIY recruitment framework can be used to track recruitment success and cost-effectiveness and support recruitment strategy decision making. These methods, along with the topics proposed in the recruitment strategy framework, should be considered by researchers when designing their recruitment strategies, with specific focus on the overarching aims of the study (eg, getting participants in quickly to test an intervention, compared with focusing on how an intervention would fit into a specific clinical care setting).

Acknowledgments

The intervention studies included in this paper were supported by research grants from the National Institute of Mental Health (P20 MH090318; R01 MH095753; R01 MH100482) to DCM, and EGL is supported by a research grant K08 MH112878 from the National Institute of Mental Health. Recruitment methods cited in this paper included use of ResearchMatch, a national health volunteer registry supported by the National Institutes of Health as part of the Clinical Translational Science Award program that is funded by grants UL1TR000445, U24TR00157 9, and 5 U24 TR001579-02 and the Aging Research Registry, a database of approximately older people in the Chicagoland area who have expressed a willingness to participate in research studies on the provision of care to aging patients and that was created and is supported by Northwestern University's Buehler Center on Aging, Health, and Society. The authors wish to thank the volunteers who have participated in research through the Center for Behavioral Intervention Technologies.

Conflicts of Interest

DCM has accepted consulting fees from Optum Behavioral Health and has an ownership interest in Actualize Therapy. EGL has accepted consulting fees from Actualize Therapy. None of the other authors have any conflicts of interest to declare. HLP is currently employed by AbbVie, Inc. Contributions to the recruitment framework described in this manuscript were made while she was employed by Northwestern University.

Multimedia Appendix 1

Examples of recruitment advertisements used for various recruitment sites.

[PDF File (Adobe PDF File), 209KB - jmir_v20i11e11050_app1.pdf]

References

1. Thompson D, Canada A, Bhatt R, Davis J, Plesko L, Baranowski T, et al. eHealth recruitment challenges. Eval Program Plann 2006 Nov;29(4):433-440. [doi: 10.1016/j.evalprogplan.2006.08.004] [Medline: 17950873]
2. Ahern DK, Patrick K, Phalen JM, Neiley JD. An introduction to methodological challenges in the evaluation of eHealth research: perspectives from the Health e-Technologies Initiative. Eval Program Plann 2006 Nov;29(4):386-389. [doi: 10.1016/j.evalprogplan.2006.08.008] [Medline: 17950866]
3. Smith A. Pew Research. 2017. Record shares of Americans now own smartphones, have home broadband URL: http://www.pewresearch.org/fact-tank/2017/01/12/evolution-of-technology/ [accessed 2018-10-25] [WebCite Cache ID 73R5FYVjo]
4. 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 Dec 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]

5. Nelson LA, Zamora-Kapoor A. Challenges in conducting mHealth research with underserved populations: lessons learned. J Telemed Telecare 2016 Oct;22(7):436-440. [doi: 10.1177/1357633X15609853] [Medline: 26468214]
6. Muessig KE, Nekkanti M, Bauermeister J, Bull S, Hightow-Weidman LB. A systematic review of recent smartphone, Internet and Web 2.0 interventions to address the HIV continuum of care. Curr HIV/AIDS Rep 2015 Mar;12(1):173-190 [FREE Full text] [doi: 10.1007/s11904-014-0239-3] [Medline: 25626718]
7. Wu YP, Steele RG, Connelly MA, Palermo TM, Ritterband LM. Commentary: pediatric eHealth interventions: common challenges during development, implementation, and dissemination. J Pediatr Psychol 2014 Jul;39(6):612-623 [FREE Full text] [doi: 10.1093/jpepsy/jsu022] [Medline: 24816766]
8. Pagliari C. Design and evaluation in eHealth: challenges and implications for an interdisciplinary field. J Med Internet Res 2007 May 27;9(2):e15 [FREE Full text] [doi: 10.2196/jmir.9.2.e15] [Medline: 17537718]
9. National Institute of Mental Health. Opportunities and Challenges of Developing Information Technologies on Behavioral and Social Science Clinical Research. 2017. URL: https://tinyurl.com/y7noqvqz [WebCite Cache ID 73VLp5ngO]
10. 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]
11. Topolovec-Vranic J, Natarajan K. The use of social media in recruitment for medical research studies: a scoping review. J Med Internet Res 2016 Dec 07;18(11):e286 [FREE Full text] [doi: 10.2196/jmir.5698] [Medline: 27821383]
12. Palac HL, Alam N, Kaiser SM, Ciolino JD, Lattie EG, Mohr DC. A Practical Do-It-Yourself Recruitment Framework for Concurrent eHealth Clinical Trials: Simple Architecture (Part 1). J Med Internet Res 2018 Nov 01;20(11):e11049 [FREE Full text] [doi: 10.2196/11049] [Medline: 30389650]
13. Mohr D, Tomasino KN, Lattie EG, Palac HL, Kwasny MJ, Weingardt K, et al. IntelliCare: an eclectic, skills-based app suite for the treatment of depression and anxiety. J Med Internet Res 2017 Dec 05;19(1):e10 [FREE Full text] [doi: 10.2196/jmir.6645] [Medline: 28057609]
14. Cheung K, Ling W, Karr CJ, Weingardt K, Schueller SM, Mohr DC. Evaluation of a recommender app for apps for the treatment of depression and anxiety: an analysis of longitudinal user engagement. J Am Med Inform Assoc 2018 Aug 01;25(8):955-962. [doi: 10.1093/jamia/ocy023] [Medline: 29659857]
15. Lattie EG, Ho J, Sargent E, Tomasino KN, Smith JD, Brown CH, et al. Teens engaged in collaborative health: the feasibility and acceptability of an online skill-building intervention for adolescents at risk for depression. Internet Interv 2017 Jun;8:15-26 [FREE Full text] [doi: 10.1016/j.invent.2017.02.003] [Medline: 28584734]
16. Tomasino KN, Lattie EG, Ho J, Palac HL, Kaiser SM, Mohr DC. Harnessing peer support in an online intervention for older adults with depression. Am J Geriatr Psychiatry 2017 Oct;25(10):1109-1119 [FREE Full text] [doi: 10.1016/j.jagp.2017.04.015] [Medline: 28571785]
17. Lattie E, Schueller SM, Sargent E, Stiles-Shields C, Tomasino KN, Corden ME, et al. Uptake and usage of IntelliCare: a publicly available suite of mental health and well-being apps. Internet Interv 2016 May;4(2):152-158 [FREE Full text] [doi: 10.1016/j.invent.2016.06.003] [Medline: 27398319]
18. 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]
19. Ramo DE, Prochaska JJ. Broad reach and targeted recruitment using Facebook for an online survey of young adult substance use. J Med Internet Res 2012 Feb 23;14(1):e28 [FREE Full text] [doi: 10.2196/jmir.1878] [Medline: 22360969]
20. Thornton L, Batterham PJ, Fassnacht DB, Kay-Lambkin F, Calear AL, Hunt S. Recruiting for health, medical or psychosocial research using Facebook: systematic review. Internet Interv 2016 May;4:72-81 [FREE Full text] [doi: 10.1016/j.invent.2016.02.001] [Medline: 30135792]
21. Glickman SW, McHutchison JG, Peterson ED, Cairns CB, Harrington RA, Califf RM, et al. Ethical and scientific implications of the globalization of clinical research. N Engl J Med 2009 Feb 19;360(8):816-823. [doi: 10.1056/NEJMsb0803929] [Medline: 19228627]
22. Mohr DC, Lyon AR, Lattie EG, Reddy M, Schueller SM. Accelerating digital mental health research from early design and creation to successful implementation and sustainment. J Med Internet Res 2017 May 10;19(5):e153 [FREE Full text] [doi: 10.2196/jmir.7725] [Medline: 28490417]
23. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012 Mar;50(3):217-226 [FREE Full text] [doi: 10.1097/MLR.0b013e3182408812] [Medline: 22310560]
24. Peters-Lawrence MH, Bell MC, Hsu LL, Osunkwo I, Seaman P, Blackwood M, Sickle Cell Disease Clinical Research Network (SCDCRN). Clinical trial implementation and recruitment: lessons learned from the early closure of a randomized clinical trial. Contemp Clin Trials 2012 Mar;33(2):291-297 [FREE Full text] [doi: 10.1016/j.cct.2011.11.018] [Medline: 22155024]
25. Sumi E, Teramukai S, Yamamoto K, Satoh M, Yamanaka K, Yokode M. The correlation between the number of eligible patients in routine clinical practice and the low recruitment level in clinical trials: a retrospective study using electronic medical records. Trials 2013 Dec 11;14:426 [FREE Full text] [doi: 10.1186/1745-6215-14-426] [Medline: 24326039]

26. Huang G, Bull J, Johnston McKee K, Mahon E, Harper B, Roberts JN. Clinical trials recruitment planning: a proposed framework from the Clinical Trials Transformation Initiative. Contemp Clin Trials 2018 Mar;66:74-79. [doi: 10.1016/j.cct.2018.01.003]

Abbreviations

CBITs: Center for Behavioral Intervention Technologies
CTU: clinical trials unit
DIY: do-it-yourself
eHealth: electronic health
IRB: institutional review board
RCT: randomized controlled trial
T-CBT: telephone-administered cognitive behavioral therapy

Edited by G Eysenbach; submitted 14.05.18; peer-reviewed by C Botella, A Poli, K Reuter; comments to author 19.07.18; revised version received 03.09.18; accepted 10.09.18; published 29.11.18.

Please cite as:
Lattie EG, Kaiser SM, Alam N, Tomasino KN, Sargent E, Rubanovich CK, Palac HL, Mohr DC
A Practical Do-It-Yourself Recruitment Framework for Concurrent eHealth Clinical Trials: Identification of Efficient and Cost-Effective Methods for Decision Making (Part 2)
J Med Internet Res 2018;20(11):e11050
URL: https://www.jmir.org/2018/11/e11050/
doi:10.2196/11050
PMID:30497997

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org